

In Search of New Insights into Teacher-Learner Interactions:
The Potential of Students' (Non)Attention-related Behavior
During Instruction and Its Measurement

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
M.A. Patricia Charlotte Goldberg

Tübingen

2020

1. Betreuer:	Prof. Dr. Kathleen Stürmer
2. Betreuer:	Prof. Dr. Ulrich Trautwein
Tag der mündlichen Prüfung:	22.10.2020
Dekan:	Prof. Dr. Josef Schmid
1. Gutachter:	Prof. Dr. Kathleen Stürmer
2. Gutachter:	Prof. Dr. Benjamin Nagengast

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisors Prof. Dr. Kathleen Stürmer and Prof. Dr. Ulrich Trautwein for providing me with the opportunity to benefit from their expertise and their foresighted guidance throughout the last three years. I have learned a lot and I am grateful for any feedback and inspiration I received from you. A special thanks goes to Kathleen: I highly appreciate your patience when talking me through argumentation structures over and over again and your continuous support in all situations!

Furthermore, I would like to thank all my colleagues from the Hector Research Institute of Education Sciences and Psychology for their incredible support and encouragement over the past years. Particular thanks goes to Dr. Wolfgang Wagner for his methodological advice and guidance. Furthermore, I would like to thank my further co-authors for their support and cooperation in the formation of the studies: Jakob Schwerter, Prof. Dr. Tina Seidel, Ömer Sümer, Prof. Dr. Richard Göllner, Prof. Dr. Peter Gerjets, and Prof. Dr. Enkelejda Kasneci. Moreover, I express gratitude to Prof. Dr. Benjamin Nagengast and Prof. Dr. Kathleen Stürmer for assessing this thesis.

Additionally, I would like to thank the LEAD Graduate School & Research Network and the Tübingen School of Education for providing me a great network and supporting structures that allowed me to benefit from an interdisciplinary exchange with other researchers and people from the field. I also would like to thank my Hiwi students, especially Julian Lohmann and Selina Bohler, for their untiring efforts in the behavioral ratings.

Moreover, I would like to thank Moritz Fleischmann for sharing my ups and downs and for giving me a different perspective in various situations. I am happy we agreed to call each other friends. Big thanks go to Xian Cheng for being my constant during the whole PhD time and for listening to all the small and big things that moved me. Further, I would like to thank Lisa Hasenbein and Dr. Till Fütterer for always having an open ear and valuable advice on different occasions, as well as Dr. Laura Braun and Molly Hammer for all their pep talks.

Particular thanks goes to Isabel Faller for always being one of my greatest supporters and never tired of helping me move forward. Furthermore, I owe great thanks to Arianna Tolone for taking care of my physical health and Carlo Forst-Gill for doing everything he could to keep me sane. Finally, I would like to thank my family: my mother and my sister for giving me the emotional support I needed and my father, for inspiring me to think outside the box ever since.

Many thanks to all of you!

ABSTRACT

Attention plays a central role in teacher-learner interactions. Students' attention is considered a main goal of teaching (Diederich & Tenorth, 1997), as it is an important prerequisite for students' learning. The definitions of attention are manifold and range from those that focus on cognitive dispositions (i.e., students being attentive) to those that focus on active behaviors (i.e., students paying attention). However, theoretical considerations and empirical findings have supported the idea that there is a close connection between observable behaviors and covert cognitive processes (e.g., Chi & Wylie, 2014; Fredricks et al., 2004; Olney et al., 2015). Therefore, behavior that is associated with students' attention can help explain the processes involved in teacher-learner interactions.

When investigating the processes involved in teacher-learner interactions, it is necessary to integrate the level of individual students with the level of interactions between teachers and learners. Whereas psychological theories about information processing tend to focus on the level of individual students, the level of interactions between teachers and learners in the instructional context is usually the focus of educational models. Students' behavior within teacher-learner interactions can indicate whether students are paying attention and also provide hints about the intensity of their attention, as certain learning activities require more attentional resources than others do. Teachers need to use these attention-related behaviors as a reference point to remain aware of what is going on in the classroom and to adjust their teaching accordingly (Kounin, 1970; Stürmer & Seidel, 2015). The degree to which teachers succeed in guiding their students' attention has been acknowledged in fundamental aspects of teaching quality, making students' (non)attention-related behavior a valuable indicator for classroom management and cognitive activation. However, existing observational instruments are not yet suitable for properly investigating the processes involved in teacher-learner interactions on the basis of students' (non)attention-related behavior. Either the temporal resolution is too rough to account for the situation-specific effects of the instructional context or the category system does not provide enough information about the quality of the respective student behaviors. These shortcomings have prevented research from making use of the potential of students' (non)attention-related behavior for investigating the processes involved in teacher-learner interactions more closely.

The objective of the present dissertation was to provide new insights into teacher-learner interactions by evaluating (a) the adequate measurement and (b) the potential of students'

(non)attention-related behavior during instruction for offering such new insights. Therefore, the present dissertation elaborated on the adequate assessment of students' (non)attention-related behavior by presenting and validating a new observation instrument (Study 1) that provides behavioral indicators that enable a continuous assessment of students' (non)attention-related behavior (Continuous (non)Attention-related Behavior Indicators – *CABI*). Furthermore, the present dissertation evaluated the potential of students' (non)attention-related behavior by using the *CABI* to investigate the mechanisms underlying students' behavior on the individual level (Study 2) as well as on the level of interactions between teachers and learners in instruction (Study 3).

Study 1 (*Attentive or Not? Toward a Machine-Learning Approach for Assessing Students' Visible Engagement in Classroom Instruction*) tested the validity of the new observation instrument in an instructional setting to strengthen the close connection between observable behavior and internal cognitive processes. Results supported the *CABI*'s construct validity, which was tested by using students' self-reported cognitive engagement, situational interest, and involvement. Results also supported the *CABI*'s predictive validity, examined via students' performance on a subsequent knowledge test on the session's topic. In addition, Study 1 provided a proof of concept for a machine-vision-based approach for assessing visible indicators of students' (non)attention-related behavior. The automated approach was based on visual parameters such as head pose, gaze direction, and facial expressions.

Study 2 (*Why Do Students Exhibit Different Attention-Related Behavior During Instruction? Investigating the Effects of Individual Prerequisites, Class Membership, and Classroom Activities*) focused on determinants of students' (non)attention-related behavior on the individual level. Given that situational aspects of students' (non)attention-related behavior during classroom instruction have not yet been studied intensively, Study 2 explored how students' individual prerequisites, class membership, and classroom activities – as determinants to consider in addition to teachers' quality of instruction – can explain differences in variability within individual students' (non)attention-related behavior across a lesson as well as differences between students. This study made use of the intensive longitudinal data structure that resulted from the continuous annotation of behavior with the *CABI*. Using dynamic structural equation modeling, students' (non)attention-related behavior was primarily determined by factors that were specific to single classrooms, but within the same classroom, students' (non)attention-related behavior appeared to be affected by their self-concept.

Focusing on effects on the interaction level, Study 3 (*How does learners' behavior attract preservice teachers' attention during teaching?*) investigated the relationship of preservice teachers' attentional focus and students' (non)attention-related behavior in standardized teaching situations. Teachers need to notice and identify relevant cues in students' behavior to make reasoned decisions about their practices; however, novice teachers in particular have trouble distributing their attentional focus evenly across all students while teaching. To investigate the possible determinants more closely, Study 3 examined how students' (non)attention-related behavior (annotated with the CABI) guided preservice teachers' attentional focus (operationalized via eye tracking) while teaching. The results demonstrated that when inexperienced teachers were faced with the demands of interacting with students while conveying learning material, they were more likely to focus on students who engaged in behavior that supported instruction, such as active participation.

The potential of students' (non)attention-related behavior during instruction for offering new insights into teacher-learner interactions and the measurement of this behavior are discussed. The results of this dissertation have practical implications, especially with regard to teacher training. Moreover, they provide implications for future research, for example, on teachers' professional vision and the automated assessment of students' attentional processes via behavior.

ZUSAMMENFASSUNG

Aufmerksamkeit spielt eine zentrale Rolle bei der Interaktion zwischen Lehrenden und Lernenden. Die Aufmerksamkeit von Schülerinnen und Schülern wird als ein Hauptziel des Unterrichts betrachtet (Diederich & Tenorth, 1997), da sie eine wichtige Voraussetzung für erfolgreiches Lernen ist. Die Definitionen von Aufmerksamkeit sind vielfältig und reichen von solchen, die sich auf kognitive Dispositionen fokussieren, bis zu solchen, die sich auf aktive Verhaltensweisen konzentrieren. Theoretische Überlegungen und empirische Befunde betonen jedoch den engen Zusammenhang zwischen beobachtbaren Verhaltensweisen und verdeckten kognitiven Prozessen (z.B. Chi & Wylie, 2014; Fredricks et al., 2004; Olney et al., 2015). Daher kann Verhalten, das mit der Aufmerksamkeit der Schülerinnen und Schüler assoziiert ist, dazu beitragen Prozesse zu erklären, welche den Interaktionen zwischen Lehrenden und Lernenden zugrunde liegen.

Bei der Untersuchung der Interaktionsprozesse von Lehrenden und Lernenden, ist es notwendig, die Ebene der einzelnen Schülerinnen und Schüler mit der Ebene der Interaktionen zwischen Lehrenden und Lernenden zu verbinden. Während sich psychologische Theorien zur Informationsverarbeitung eher auf die Ebene der einzelnen Schülerinnen und Schüler konzentrieren, steht in Modellen der empirischen Bildungsforschung in der Regel die Ebene der Interaktionen zwischen Lehrenden und Lernenden im Unterrichtskontext im Mittelpunkt. Das Verhalten der Schülerinnen und Schüler innerhalb von Lehrenden-Lernenden-Interaktionen kann sowohl Hinweise darauf geben, ob die Schülerinnen und Schüler aufmerksam sind, als auch Hinweise auf die Intensität ihrer Aufmerksamkeit, da bestimmte Lernaktivitäten mehr Aufmerksamkeitsressourcen erfordern als andere. Lehrende müssen diese aufmerksamkeitsbezogenen Verhaltensweisen kontinuierlich als Bezugspunkt nutzen, um sich über die Vorgänge im Klassenzimmer bewusst zu bleiben, und ihren Unterricht entsprechend anzupassen (Kounin, 1970; Stürmer & Seidel, 2015). Der Grad, zu dem es den Lehrern gelingt, die Aufmerksamkeit ihrer Schülerinnen und Schüler zu lenken, wurde in grundlegenden Aspekten der Unterrichtsqualität berücksichtigt, so dass das (nicht)aufmerksamkeitsbezogene Verhalten der Schülerinnen und Schüler zu einem wertvollen Indikator für die strukturierte Klassenführung und die kognitive Aktivierung wird. Bereits existierende Beobachtungsinstrumente sind jedoch noch nicht geeignet, Prozesse der Lehrenden-Lernenden-Interaktionen auf der Grundlage des (nicht)aufmerksamkeitsbezogenen Verhaltens der Schülerinnen und Schüler angemessen zu untersuchen. Entweder ist die zeitliche Auflösung zu

grob, um die situationsspezifischen Effekte des Unterrichtskontextes zu berücksichtigen, oder das Kategoriensystem liefert nicht genügend Informationen über die Qualität des jeweiligen Verhaltens der Schülerinnen und Schüler. Diese Defizite hinderten die bisherige Forschung daran, das Potenzial des (nicht)aufmerksamkeitsbezogenen Verhaltens der Schülerinnen und Schüler für eine genauere Untersuchung der Prozesse in Lehrenden-Lernenden-Interaktionen zu nutzen.

Ziel der vorliegenden Dissertation war es, neue Einblicke in die Interaktion zwischen Lehrenden und Lernenden zu gewinnen, indem (a) die adäquate Messung des (nicht)aufmerksamkeitsbezogenen Verhaltens der Schülerinnen und Schüler während des Unterrichts und (b) das Potenzial aus diesem Verhalten neuen Einsichten abzuleiten evaluiert wurde. In der vorliegenden Dissertation wurde die adäquate Bewertung des (nicht)aufmerksamkeitsbezogenen Verhaltens der Schülerinnen und Schüler evaluiert, indem ein neues Beobachtungsinstrument vorgestellt und validiert wurde (Studie 1), das Verhaltensindikatoren liefert, die eine kontinuierliche Bewertung des (nicht)aufmerksamkeitsbezogenen Verhaltens von Lernenden ermöglichen (Continuous (non)Attention-related Behavior Indicators - *CABI*). Darüber hinaus evaluierte die vorliegende Dissertation das Potenzial des (nicht)aufmerksamkeitsbezogenen Verhaltens von Schülerinnen und Schülern, indem sie das *CABI* nutzte, um die dem Verhalten der Schülerinnen und Schüler zugrundeliegenden Mechanismen auf der individuellen Ebene (Studie 2) sowie auf der Ebene der Interaktionen zwischen Lehrenden und Lernenden im Unterricht (Studie 3) zu untersuchen.

Studie 1 (*Attentive or Not? Toward a Machine-Learning Approach for Assessing Students' Visible Engagement in Classroom Instruction*) testete die Validität des neu entwickelten Beobachtungsinstruments in einer Instruktionssituation, um die enge Verbindung zwischen beobachtbarem Verhalten und internen kognitiven Prozessen zu stärken. Die Ergebnisse untermauerten die Konstruktvalidität des Beobachtungsinstruments, die anhand des selbstberichteten kognitiven Engagements, situativen Interesses und Beteiligung der Studierenden getestet wurde. Die Ergebnisse untermauerten zudem die prädiktive Validität des Beobachtungsinstruments, die anhand der Leistung der Studierenden bei einem anschließenden Wissenstest zum Thema der Sitzung untersucht wurde. Darüber hinaus lieferte Studie 1 einen Machbarkeitsnachweis für einen auf maschinellem Lernen basierenden Ansatz zur Bewertung sichtbarer Indikatoren für das (nicht)aufmerksamkeitsbezogene Verhalten von Lernenden. Der automatisierte Ansatz basierte auf visuellen Parametern wie Kopfhaltung, Blickrichtung und Gesichtsausdrücken.

Studie 2 (*Why Do Students Exhibit Different Attention-Related Behavior During Instruction? Investigating the Effects of Individual Prerequisites, Class Membership, and Classroom Activities*) konzentrierte sich auf Determinanten des (nicht)aufmerksamkeitsbezogenen Verhaltens von Schülerinnen und Schülern auf individueller Ebene. Da situative Aspekte des (nicht)aufmerksamkeitsbezogenen Verhaltens von Schülerinnen und Schülern während des Unterrichts noch nicht intensiv berücksichtigt wurden, untersuchte Studie 2, wie die individuellen Voraussetzungen, die Klassenzugehörigkeit und die Klassenaktivitäten der Schülerinnen und Schüler - als Determinanten, die zusätzlich zur Unterrichtsqualität der Lehrenden zu berücksichtigen sind - Unterschiede in der Variabilität des (nicht)aufmerksamkeitsbezogenen Verhaltens einzelner Schülerinnen und Schüler innerhalb einer Unterrichtsstunde sowie Unterschiede zwischen den Schülerinnen und Schülern erklären können. Studie 2 machte sich die intensive Längsschnittdatenstruktur zunutze, die sich aus der kontinuierlichen Beurteilung des Verhaltens mit dem CABI ergab. Mit Hilfe dynamischer Strukturgleichungsmodelle wurde gezeigt, dass das (nicht)aufmerksamkeitsbezogene Verhalten der Schülerinnen und Schüler in erster Linie durch Faktoren bestimmt wurde, die spezifisch für einzelne Klassen waren. Innerhalb derselben Klasse schien das (nicht)aufmerksamkeitsbezogene Verhalten der Schülerinnen und Schüler allerdings primär durch ihr Selbstkonzept beeinflusst zu werden.

Mit dem Schwerpunkt auf Auswirkungen auf der Interaktionsebene zwischen Lehrenden und Lernenden untersuchte Studie 3 (*How does learners' behavior attract preservice teachers' attention during teaching?*) die Beziehung zwischen dem Aufmerksamkeitsfokus von angehenden Lehrkräften und dem (nicht)aufmerksamkeitsbezogenen Verhalten von Lernenden in standardisierten Unterrichtssituationen. Lehrende müssen relevante Hinweise im Verhalten von Lernenden wahrnehmen und identifizieren, um begründete Entscheidungen über ihre Unterrichtsmethoden und Lehrtechniken treffen zu können. Allerdings haben insbesondere angehende Lehrpersonen Schwierigkeiten, ihren Aufmerksamkeitsfokus während des Unterrichts gleichmäßig auf alle Lernenden zu verteilen. Um die möglichen Determinanten genauer zu untersuchen, wurde in Studie 3 untersucht, wie das (nicht)aufmerksamkeitsbezogene Verhalten von Lernenden (beurteilt mit dem CABI) den Aufmerksamkeitsfokus von angehenden Lehrpersonen (operationalisiert durch Eye Tracking) während des Unterrichts lenkte. Die Ergebnisse zeigten, dass unerfahrene Lehrpersonen, wenn sie mit den Anforderungen der Interaktion mit den Lernenden bei der Vermittlung von Lernmaterial konfrontiert waren, sich primär auf Lernende

konzentrierten, die ein den Unterricht förderliches Verhalten zeigten, wie beispielsweise aktive Teilnahme.

Das Potenzial des (nicht)aufmerksamkeitsbezogenen Verhaltens von Schülerinnen und Schülern während des Unterrichts für neue Einblicke in die Lehrenden-Lernenden-Interaktionen und die Messung dieses Verhaltens wird entsprechend diskutiert. Die Ergebnisse dieser Dissertation haben praktische Implikationen, insbesondere im Hinblick auf die Lehrkräfteausbildung. Darüber hinaus liefern sie Implikationen für die zukünftige Forschung, beispielsweise im Bereich der professionellen Unterrichtswahrnehmung von Lehrenden und der automatisierten verhaltensbasierten Erfassung von Aufmerksamkeitsprozessen von Schülerinnen und Schülern.

CONTENTS

1	INTRODUCTION AND THEORETICAL BACKGROUND.....	13
1.1	Students' Attention on the Individual Level.....	16
1.1.1	Students' Attention: Defining the Construct of Interest	16
1.1.2	How is Attention Related to Observable Behavior?	20
1.1.3	Attention as a Precondition for Learning.....	23
1.1.4	Factors That Influence Students' (Non)Attention-related Behavior	25
1.2	Students' (Non)Attention-related Behavior on the Interaction Level	27
1.2.1	Students' (Non)Attention-related Behavior as an Indicator of Teaching Quality.....	27
1.2.2	Students' (Non)Attention-related Behavior as a Point of Reference for Teachers	30
1.2.3	How to Structure the Mechanisms Underlying Students' (Non)Attention-related Behavior in Teacher-Learner Interactions?	32
1.3	Measuring Students' (Non)Attention-related Behavior	36
1.3.1	Approaches for Measuring Students' Attention During Teacher-Learner Interactions	36
1.3.2	Observational Approaches to Measure Students' Attention- and Engagement-related Behavior During Instruction.....	38
1.3.3	Considerations for New Approaches to Measure Students' (Non)Attention-related Behavior During Instruction.....	44
2	RESEARCH QUESTIONS.....	47
3	STUDY 1.....	53
4	STUDY 2.....	89
5	STUDY 3.....	125
6	GENERAL DISCUSSION AND CONCLUSION.....	167
6.1	Discussion of the Results.....	169
6.1.1	Measurement of Students' (Non)Attention-related Behavior during Teacher-Learner Interactions .	169
6.1.2	New Insights Offered by Students' (Non)Attention-related Behavior during Teacher-Learner Interactions	170
6.2	Strengths and Limitations	173
6.3	General Implications and Future Directions	175
6.3.1	Implications for Practice.....	175
6.3.2	Implications for Research and Future Directions	177
6.4	Conclusion	180
7	REFERENCES.....	183
8	APPENDIX.....	201

1

Introduction and Theoretical Background

1 Introduction and Theoretical Background

Attention please! Whenever we want to say or explain something and we want to be sure that our interlocutor understands what we are talking about, we request their attention. This is because we intuitively know that attention is central to the adequate processing of information. Consequently, it is no surprise that attention plays a central role in teacher-learner interactions as well. Students need to pay attention to their teachers to understand what the teachers are talking about and to learn something. In turn, when students fail to pay attention, they risk missing important information that might be crucial for them to adequately process the learning material.

The term attention covers a broad range of definitions, and depending on the domain, such definitions may be more likely to focus on cognitive dispositions (i.e., being attentive) than on active behaviors (i.e., paying attention; Brünken & Seufert, 2006). Nevertheless, all of them consider attention to be an important determinant of individual students' learning. Psychological models of information processing have emphasized the central role that attention exhibits during learning (e.g., Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Craik & Lockhart, 1972), and empirical research from education science has supported the strong connection between behaviors that are associated with attention and learning outcomes (e.g., Karweit & Slavin, 1982; Lei et al., 2018; Stipek, 2002). Consequently, students' attention is considered a main goal of teaching and learning (Diederich & Tenorth, 1997). In turn, teachers need to use students' behaviors that are related to attention as a reference point in order to remain aware of what's going on in the classroom and to adapt their teaching accordingly (Kounin, 1970; Stürmer & Seidel, 2015). For example, when the teacher introduces a new topic and explains the Pythagorean Theorem, the students need to pay attention in order to understand that in a right triangle, the sum of the square areas over the legs is equal to the square area over the hypotenuse. However, students can only process a certain amount of information at a time, so they need to ignore the nice weather outside or the person sitting next to them who wants to chat about the latest gossip, and they must find a way to focus on the teacher instead. However, this might be quite difficult, so teachers need to provide a learning environment that reduces distractions and helps students remain on task. Furthermore, to avoid superficial learning, teachers need to choose the right level of challenge in their tasks in order to encourage students to pay attention with the desired intensity. This enterprise requires teachers to detect whether and how intensely their students are paying attention so that they can adapt their overall teaching. For example, teachers need to identify students who are chatting about the weekend

instead of the new math topic and encourage them to turn their attention toward the material that is being taught, or they must notice which students are struggling to understand Euclidean geometry and provide additional explanations for them.

When investigating the processes involved in teacher-learner interactions, two different levels need to be considered: students' individual level and the level of interaction in the instructional context. The individual level is concerned with students' psychology and internal processes of successful learning. The interaction level considers the surrounding factors that determine the educational environment. Whereas psychological theories focus on the individual level, the interaction level in the instructional context is the focus of educational models. Therefore, to understand the processes involved in teacher-learner interactions, these two perspectives need to be integrated. This is also true for students' attentional processes within teacher-learner interactions. On the one hand, attentional processes play a central role in students' learning success on the individual level. On the other hand, attentional processes provide important information about interactions between teachers and learners as such processes are a reference point from which teachers evaluate their choice of instructional methods and techniques and can serve as an indication of how well teachers have designed their instruction.

However, research on the mechanisms underlying students' attention-related behavior in the interactions between teachers and learners is still scarce. This gap might be due to the situation-specific and dynamic nature of teacher-learner interactions. Approaches for measuring students' behavior commonly use aggregated values that average out situational effects, thus making it impossible to have deeper insights into the processes behind teacher-learner interactions. Additionally, it is important to consider the entire spectrum of students' behavior during instruction, as content-related as well as distracted or disruptive behavior provide valuable information (a) about students' internal cognitive processes on the individual level and (b) as a reference point for teachers and the quality of instruction on the level of the interaction. The present dissertation offers a novel approach for measuring students' behavior. On the one hand, this approach allows students' behavior to be captured continuously throughout instruction. On the other hand, it covers the entire (non)attention-related behavioral spectrum that students can exhibit during teacher-learner interactions. The present dissertation aims to provide new insights into teacher-learner interactions by evaluating the adequate measurement and the potential of students' (non)attention-related behavior during instruction for offering such new insights.

Therefore, the theoretical foundation will be presented in the introductory chapter of this dissertation. This chapter is split into three parts: First, the importance of students' (non)attention-related behavior with regard to students' individual level and learning is outlined in Chapter 1.1. Subsequently, Chapter 1.2 describes the relevance of students' (non)attention-related behavior on the interaction level. The last chapter of the theoretical background is dedicated to the measurement of students' (non)attention-related behavior (Chapter 1.3). Chapter 2 introduces the research questions that guide the three empirical studies that were conducted for this dissertation and are presented in Chapters 3 to 5. The first study investigated the validity of a novel approach for measuring students' (non)attention-related behavior continuously throughout the instructional process. The second study investigated the potential of students' (non)attention-related behavior to offer further insights into the underlying mechanisms on the individual level, whereas the third study elaborated on new insights provided by students' (non)attention-related behavior on the interaction level during instruction. The final chapter (Chapter 6) discusses the findings of the three empirical studies with regard to the research questions and examines their strengths and limitations. The dissertation closes with implications for research, practice, and future directions.

1.1 Students' Attention on the Individual Level

To provide common ground in the use of terminology, I first provide a definition of attention as I will use it in the present dissertation and elaborate on the relation between students' behavior and their cognitive processes. Further, to point out the relevance of students' attention, I emphasize the importance of attention for students' learning and describe the factors that influence individual students' behavior in the context of teacher-learner interactions.

1.1.1 Students' Attention: Defining the Construct of Interest

“Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought.”

(p. 403, James, 1890)

As James (1890) already pointed out, everybody has a natural understanding of what attention might comprise. However, we have trouble precisely defining attention because it is used in everyday language, and researchers from various domains have conducted extensive research on it during the past 70 years, resulting in many different definitions. The semantic range of attention includes current dispositions (i.e., being attentive) as well as active behaviors

(i.e., paying attention; Brünken & Seufert, 2006). For example, if a psychologist is talking about attention, he or she is most likely referring to a selection mechanism that directs the cognitive processing system to focus on a subset of accessible information to optimize performance because cognitive capacity is limited (Cohen, 2014; McDowd, 2007). Then again, a neuroscientist would view attention as a collection of interrelated mechanisms and processes that involve almost every area of the human brain rather than a unitary concept (Corbetta & Shulman, 2002; Posner & Rothbart, 2007). And an education scientist would focus on whether the students in a classroom are *paying attention* to their teacher and the learning material, thus allocating a rather behavioral interpretation. In the following, I present some of the most prominent considerations and terms that are associated with attention, especially in research on education, as this dissertation follows a cross-disciplinary understanding of attention.

Theoretical considerations that have emerged from research in psychology and cognitive science can be differentiated into models that focus on the selectivity aspect of attention and those that focus on aspects of capacity. Since the 1950s, models on the early and late selection of attention have emerged. Such models assume some kind of bottleneck with a filtering device prior to it. However, the two perspectives differ in where this filter is located. Two of the most influential models on early selection are the filter theory by Broadbent (1958) and the attenuator model by Treisman (1964). According to Broadbent (1958), the stream of information is reduced by the selection/filtering of relevant information based on sensory properties at an early processing stage, whereas the identification and categorization of these stimuli occur later. Treisman (1964) on the other hand assumed that neglected information is not blocked entirely but is only attenuated and can be analyzed semantically when it is very important. By contrast, according to Deutsch and Deutsch (1963), information gets selected in a later stage of processing: All incoming information is analyzed the same way, and selection is based on semantic properties when entering consciousness or memory. Therefore, which pieces of information get processed does not depend on perceptual restrictions (see filter theory), but the situational context defines whether the information is relevant (Goldhammer & Moosbrugger, 2006). The debate about where the selection mechanism is located has yet to be entirely resolved, but Lavie's (2005) load theory provides a framework from which to determine the level of processing of unattended stimuli. Lavie's (2005) load theory assumes that a person's capacity to process the unattended information depends on how difficult the person finds it to process the attended information (Chun et al., 2011). If the primary task is easy, spare attentional resources will spill over to distractors, suggesting late selection (Lavie, 1995). By

contrast, if the primary task is difficult, all attentional resources will be devoted to the primary task, and the distractors will not be processed as well, indicating early selection (Lavie, 1995). Lavie (2005) concluded that a high perceptual load and cognitive control are crucial for focusing and actively maintaining attention. With regard to educational settings, theoretical considerations that focus on the selectivity aspects of attention emphasize that it is crucial to reduce competing streams of information as much as possible so that relevant information can be processed reliably.

Models that focus on the capacity of attention can be differentiated into models of unspecific and specific capacity. The model of central capacity by Kahneman (1973) assumes that attentional resources are limited and can only be allocated to a restricted number of tasks at the same time. Depending on the difficulty of one task, some proportions of attentional resources can be simultaneously allocated to another task. However, when engaging in a highly demanding task that requires a lot of effort, the pool of available attentional resources may already be exhausted. By contrast, Allport (1980) proposed that attention is based on the sensory properties of several specialized processing modules (e.g., one module for auditory information, one for visual information). Each of these models is limited in its capacity so that simultaneous tasks that require the same modularity compete for the available capacity (Goldhammer & Moosbrugger, 2006). Similarly, in their economics theory of scarce resources, Navon and Gopher (1979) suggested a model of multiple resources but with the difference that when a certain resource is limited due to task demands, processing can be supported by other resources but with less efficiency. However, in both models, the number and kind of modules remained unspecified (Goldhammer & Moosbrugger, 2006). Wickens' (2002) multiple resource model differentiated four dimensions to classify tasks with different manifestations: processing stage (perception/cognition vs. response), perceptual modality (visual vs. auditory), processing code (visual vs. spatial), and visual channel (focal vs. ambient). Wickens assumed that the more the simultaneous tasks resemble the manifestations along these dimensions, the more they interfere with each other. Altogether, theoretical models on the limited capacity of attention have suggested that it is important to guide students' attention to save their cognitive resources (see also Mayer, 2002; Sweller et al., 1998).

Other perspectives that include findings from neuroscientific and neuropsychological perspectives have stated that attention does not constitute a unitary concept and that multidimensional aspects of attention should be considered (Goldhammer & Moosbrugger, 2006). Multicomponent models have systematized important components of attention (see

Table 1) and have related them to distinct neuronal networks (Posner & Boies, 1971; Posner & Rafal, 1987; Posner & Raichle, 1994; Van Zomeren & Brouwer, 1994). With regard to classroom instruction and teacher-learner interactions, the ability to sustain attention over a longer period is particularly important, as classroom instruction usually lasts for at least 45 min, which are filled with learning-relevant information that students need to pay attention to (Chun et al., 2011). Additionally, students' selective attention to the learning material is essential (Janssen et al., 2014), as students need to select the relevant pieces out of the ongoing stream of information provided by the teacher. Furthermore, students need to engage in processes of (covert) attention switching or divide their attention during teacher-learner interactions, for example, when they are distracted by peers or other situation-specific stimuli (e.g., a loud noise outside).

Table 1

Dimensions and Components of Attention with Descriptions Structured according to Van Zomeren and Brouwer (1994) and Posner and Raichle (1994; in parentheses)

Dimension	Component	Description
Intensity (alertness and vigilance)	Alertness	Regulation of physical and mental responsiveness
	Vigilance/sustained attention	Sustained selective attention under monotonous conditions
Selectivity (executive attention)	Selective/focused attention	Selective processing of one source of information
	Attention switching	Alternating between more than one source of information
	Divided attention	Processing of more than one source of information
Spatial attention (orienting)	Visual-spatial attention	Covert attention shifting without eye movements

As the variety of theoretical models demonstrates, there is no unambiguous definition of attention, nor can attention be described as a unitary construct. Moreover, researchers in education science often use the term attention interchangeably with or they do not clearly explain its delimitations from the notions of engagement, effort, time on task, or concentration¹ but mostly with a focus on observable aspects of student behavior. In the present dissertation, I interpret attention as a construct that is based on a combination of definitions used by education scientists and psychologists. Speaking from an educational perspective, attention is a central prerequisite for successful student learning, and as such, it is associated with certain kinds of behavior. When students pay attention, they increase the probability that they will actually remember what the teacher is telling them and thus have the opportunity to learn the respective material. With regard to psychological considerations, attention is a situation-specific mechanism that filters the ongoing stream of information and can vary in its intensity (Cohen, 2014). For example, a student might listen passively to their teacher until the teacher poses a question. In this moment, the student may begin to pay attention more closely, think about the question, try to find an answer, and raise their hand to participate in the classroom discussion. The student may focus more on the instructional situation than before, thus paying attention more intensely and providing the foundation for successful learning. The interpretation of attention in the present dissertation is therefore based on the assumption that students' behavior in teacher-learner interactions can indicate (a) whether or not students are paying attention and (b) how intensely they are focusing their attention on the learning material. This means that a student's attention plays a particular role, as it constitutes the transition from observable behavior to covert cognitive processes.

1.1.2 How is Attention Related to Observable Behavior?

The intermediate position between overt behavior and covert cognitive processes makes the observation of attention quite challenging. It has been shown that some parts of cognitive processes are observable from the outside and can be seen in students' behavior. For example, visual orientation toward a stimulus (i.e., overt behavior) improves internal processing efficiency (i.e., covert cognitive processes; Posner, 1988). Likewise, students will better

¹ As the term concentration is more common in German compared with English-language literature, I will not expand on the distinction between attention and concentration further. To avoid confusion, however, I would like to point out that concentration is defined as one dimension of attention and refers to the deliberate decision to invest mental effort (also referred to as effortful awareness). See Moran, A. (2012). Concentration: Attention and performance. In S. M. Murphy (Ed.), *The Oxford handbook of sport and performance psychology* (pp. 117-130). Oxford University Press.

understand what the teacher is trying to explain when they actually look at the teacher. In this case, students' behavior can signal whether students are paying attention and consequently whether they are engaging in learning-related cognitive processes.

Attention is specified as part of the behavioral component of student engagement (Fredricks et al., 2004). Student engagement is a multidimensional construct that is commonly defined along emotional, cognitive, and behavioral dimensions. Different from other researchers who identified only two or one additional component of engagement, in an extensive literature review, Fredricks and McColskey (2012) argued that the empirical and theoretical basis is strongest for this three-dimensional approach. The emotional dimension comprises affective reactions in classroom situations, such as interest, curiosity, or boredom (Connell, 1990; Fredricks et al., 2004). Investment in learning and making an effort to comprehend information are in turn part of cognitive engagement (Fredricks et al., 2004; Pintrich & De Groot, 1990). Whereas these dimensions are considered to be rather internal processes, behavioral engagement is observable. Next to attention, overt participation, positive conduct, and the absence of disruptive behavior are classified as part of the behavioral dimension (Fredricks et al., 2004). Despite their different interpretations, most definitions of engagement agree on the high levels of relations between the different dimensions of constructs. Therefore, whether or not someone agrees that attention should be defined as purely behavioral, overt student behavior can provide visible indicators of ongoing cognitive processes.

However, not only does students' behavior indicate the absence or presence of attention, but it can further provide hints about the quality of ongoing cognitive processes. In their ICAP framework, Chi and Wylie (2014) proposed that different observable learning activities can signal different covert learning processes. They differentiated cognitive processes into four levels on the basis of students' overt behavior: interactive, constructive, active, and passive (ICAP). According to their hypothesis, students engage more with the learning material, which results in greater learning gains as they move from passive to active to constructive to interactive learning activities. For example, when listening to a lecture, students exhibit passive behavior by orienting toward the instruction without doing anything else. Active behavior includes, for example, copying the steps involved in a solution, whereas students who display constructive behavior also ask questions. Finally, defending and arguing for a position or answering comprehension questions is categorized as interactive behavior (Chi & Wylie, 2014). On the basis of their theory of mind wandering, Olney et al. (2015) argued that this learning task structure accounts for enhanced student learning because, within the ICAP structure,

students' attention gets enhanced. Their ICAP-A hypothesis asserts that students' attention improves as they move from passive to interactive learning activities because opportunities for proactive (in terms of sequential actions) and reactive control (in terms of monitoring) increase as they move from passive to interactive learning activities (Olney et al., 2015).

As certain overt behaviors of students can provide hints about whether and how intensely they are paying attention, other behaviors can signal the opposite. Students can either not use learning opportunities by exhibiting passive behavior (e.g., looking out the window), or they can further engage in interactive, unrelated activities (e.g., talking to the student in the next seat). It is important to distinguish between different kinds of unrelated activities because, equivalent to the ICAP framework, different manifestations of unrelated (i.e., off-task) behavior can indicate different cognitive processes as well. From passive to interactive off-task behavior, attention to the respective activity is enhanced. Following Lavie's (2005) theory, the demands of the primary activity determine how well the competing information is processed. Transferred to students' behavior, this interdependency means that the more students are engaged in interactive off-task activities, the less likely they will be to follow what the teacher is teaching. For example, when students play on their phones, this activity might happen rather incidentally so that competing (i.e., instruction-related) information can still be processed to some extent. When students chat with their friends about weekend plans, focusing on the conversation will be more demanding for students, resulting in having even fewer resources available to process the instructional material. In both cases, they would not be paying attention to the material as the primary activity. By contrast, the more students are focused on what the teacher is teaching, and the more they engage in the respective learning activities, the fewer attentional resources they have available to put toward distracting stimuli.

Even though there is a strong empirical and theoretical connection between overt behavior and covert cognitive processes, I want to point out that I do not assume that students' behavior will always approximate the underlying proportions of cognitive attention or that it will do so without bias. There is always the possibility that students will appear to be disengaged but will still be paying close attention to the learning material or that students who appear attentive because they are looking at the teacher are engaging in mind wandering (see spatial attention). It would be presumptuous to claim that inferences about cognitive processes are possible with absolute certainty by just looking at students' behavior. However, in accordance with Chi and Wylie (2014), I would argue that elaborate cognitive processes become more probable the more overt behavior moves from passive to interactive manifestations.

Additionally, higher order cognitive processes, such as are required in constructive and interactive learning activities, require many attentional resources. When learning activities require many attentional resources, less attentional capacity is available for other tasks, making it more likely that students are actually engaging in the respective cognitive processes. In turn, interactive off-task activities also require more attentional resources compared with passive off-task behaviors. Therefore, it is less likely that students are paying attention to what the teacher is teaching the more distracted (i.e., interactive) their behavior gets. Consequently, to increase the probability of making exact estimations of cognitive processes on the basis of behavioral activities, the entire behavioral spectrum that students can exhibit during teacher-learner interactions should be considered. To emphasize the behavioral interpretation of attention and to avoid misinterpretations, I will subsequently refer to *attention-related* or *(non)attention-related behavior*.

1.1.3 Attention as a Precondition for Learning

Attention is an important prerequisite for learning (Brünken & Seufert, 2006). Learning and knowledge construction are commonly used synonymously and are a main concern of educational institutions. In looking at the most prevalent psychological theories, theoretical perspectives on knowledge construction have changed over time (Zoelch et al., 2019). Today, a rather pluralistic perspective of learning is prominent, which views learning as knowledge construction involving situational aspects through active information processing (Zoelch et al., 2019). Theories on information processing define learning as a set of cognitive processes through which we acquire information and store it in memory (Winne, 2001). Memory is considered to be an active and dynamic information processing system that encodes, stores, and retrieves information (Zimbardo et al., 2008). Due to extensive research in past decades, several theoretical models have become important in cognitive psychology. Some of the most famous models, such as the Multi Store Model by Atkinson and Shiffrin (1968) or the Working Memory Model by Baddeley and Hitch (1974) consider attention to be a control mechanism that limits the amount of information that can be processed further. By contrast, according to the theoretical considerations of Cowan (2001), information that is in the focus of attention is processed and stored better. In their Levels of Processing framework, Craik and Lockhart (1972) proposed that deeper (i.e., more elaborate) analysis requires more attentional resources. Comparably, in the Adaptive Control of Thought theory by Anderson and Funke (2013), it is assumed that controlled processes require extensive attentional resources as opposed to automatic processes.

Even though the underlying structures of theoretical models of information processing differ, all of them view attention as a central component that enables successful learning. The central role of attention for learning has further been supported by empirical evidence in education science. This field of research is more focused on the behavioral aspects of attention, which can be reflected by student engagement, active participation, or on-task behavior. Although existing research has different names for the construct under investigation, many results have supported the strong connection between behavior that is associated with attention and learning outcomes. For example, early research found that the time students spent engaged in a task predicted their achievement (Fisher et al., 1981; Karweit & Slavin, 1982) and that students' on-task behavior was correlated with their learning outcomes (Helmke & Renkl, 1992). More recent research has further demonstrated that active participation in classroom discussions contributes to explaining learning success (Pauli et al., 2008). Stipek (2002) found that student engagement predicted student learning (performance on an achievement test), which was later supported by Lei et al.'s (2018) meta-analysis, indicating that students' engagement is positively correlated with their academic achievement. Furthermore, existing research demonstrated that a lack of attention leads to reduced student learning (performance on deep reasoning questions; D'Mello, 2016). When students cannot sustain their attentional focus or engage in mind-wandering (see Smallwood & Schooler, 2006), they will demonstrate different behaviors that result in superficial understanding compared with deep learning (D'Mello, 2016). Finn et al. (1995) showed that inattentive and disruptive behaviors were associated with decreased academic performance. Further support for the strong connection of (non)attention-related behavior and the ability to learn has come from the broad research field on attention-deficit/hyperactivity disorder (ADHD). ADHD affects students' performance levels in school due to inattention and hyperactivity (American Psychiatric Association, 2013). Affected children have trouble controlling their behavior and sustaining their attention. This results in superficial and erroneous task processing, which in turn prohibits deep learning. Attentional/behavioral difficulties have been shown to be associated with poorer academic performance (Aronen et al., 2005).

Therefore, various approaches underline that it is important for students to pay attention so that they can learn new information successfully. However, there are multiple factors that influence students' (non)attention-related behavior during teacher-learner interactions.

1.1.4 Factors That Influence Students' (Non)Attention-related Behavior

Students show different kinds of (non)attention-related behavior during instruction. These differences can be attributed to the classroom environment but also to variations in students' individual cognitive prerequisites (Vygotsky, 1978). From an education psychology perspective, instruction is regarded as offering a structure that learners must use actively (Baumert et al., 2002; Brühwiler & Blatchford, 2011; Helmke, 2009; Seidel, 2014). Teachers design learning environments that are supposed to provide optimal conditions for all students (Seidel & Reiss, 2014). The extent to which students make use of these opportunities depends on their individual characteristics and the surrounding educational context (Seidel, 2014). Individual factors as well as context-dependent factors thus influence which activities students engage in and whether and how intensely they pay attention during teacher-learner interactions. Individual factors can comprise person-specific cognitive prerequisites as well as motivational-affective characteristics (e.g., emotions, motivation, or distractibility; Jurik et al., 2013; Rollett, 2001; Sacher, 1995; Turner & Patrick, 2004). Situational and context-dependent factors include, for example, the quality of instruction, teacher-student relationships, number of learning opportunities, teachers' choice of practices, or the quality of the learning material (Brühwiler & Blatchford, 2011; Helmke, 2009; Seidel, 2014).

Kelly (2007) found that the frequency of students' participation (i.e., asking and answering questions) during instruction varied substantially between different classes but also within one classroom. This variation was independent from teachers' dialogic instruction or teaching quality, indicating that differences in students' participation and engagement are partly caused by individual factors. He supported this assumption 1 year later when he demonstrated that cognitive prerequisites, such as prior abilities, affect students' number of verbal contributions (Kelly, 2008). Other research found that motivational-affective processes, such as students' task values (interest, perceived importance, and perceived utility) and self-concept of ability, predicted students' self-reported attention and degree of participation in classroom experiments (Lau & Roeser, 2002). Students who were more confident about their competencies in mathematics, for example, participated more often in classroom discussions than their insecure peers (Böheim et al., 2020; Pauli & Lipowsky, 2007). Also the combination of students' cognitive prerequisites and individual motivational-affective characteristics was found to explain how students participated and engaged in classroom instruction (Jurik et al., 2013; Turner & Patrick, 2004). For example, distinct student profiles that are based on cognitive (prior knowledge and cognitive abilities) and motivational-affective (subject-related self-

concept and interest) prerequisites predicted verbal participation in teacher-student interactions (Jurik et al., 2013).

Context-dependent factors can also affect students' (non)attention-related behavior. Factors from the physical instructional environment can influence whether students are able to exhibit and maintain attention-related behavior. For example, if the acoustics of a room are suboptimal and students are distracted by background noises, it is more difficult for students to pay attention to the material that is being taught (Kamps & Oberdörster, 2002). Furthermore, if the instructional method (e.g., direct instruction) does not match the seating arrangement (e.g., group tables), students might fail to direct their attention to the center of instruction (Imhof, 2004). Also, determinants of classroom processes, such as the quality of instruction, teacher-student relationships, number of learning opportunities, teachers' choice of practices, or the quality of the learning material, can influence how students behave during instruction (Brühwiler & Blatchford, 2011; Helmke, 2009; Seidel, 2014). For example, teachers create certain classroom environments through their choice of practices (e.g., activities such as classroom discussions or individual seatwork), which can support activity-related behavior but also set up occasions for misbehavior (Beyda et al., 2002). Students were shown to exhibit more on-task behavior in settings that provided many opportunities for interactions with peers with minimal interruptions from the teacher compared with teacher-centered settings or during individual seatwork (Beyda et al., 2002). Additionally, Friedman et al. (1988) found higher student engagement during teacher-directed instruction compared with individual seatwork. Social structures as well as teachers' socialization of participation influence how students verbally participate during instruction (Clarke et al., 2016). Helmke and Renkl (1992) discovered that 56% of individual differences in attention behavior could be attributed to differences between school classes. They inferred that determinants, such as class composition or classroom management, might substantially affect individual students' behavior.

Furthermore, the multifaceted relationship between internal and external determinants of individual students' (non)attention-related behavior has been acknowledged in psychological diagnoses of attention in schools. Imhof (2004) underlined the idea that the interaction between personal and situational variables needs to be assessed to reliably determine students' behavior. For example, class climate can influence students' motivation, and whether a student pays attention depends on their motivation to engage with the learning material (Anderman & Anderman, 1999). As students' (non)attention-related behavior plays a particular role in

teacher-learner interactions, I elaborate more on its relation to the quality of instruction in the next chapter.

1.2 Students' (Non)Attention-related Behavior on the Interaction Level

In the previous chapter, I described the central role of attentional processes on the individual level for students' learning and factors that can affect students' behavior. I am now going to focus on the role of students' attention and especially students' (non)attention-related behavior on the level of interactions between teachers and learners in the instructional context, as it plays a crucial role for the teacher and the teaching process in different ways. On the one hand, considering the (non)attention-related behavior of all students in a classroom can serve as an indication of the overall level of teaching quality. On the other hand, teachers need to rely on individual students' (non)attention-related behavior as a point of reference from which to infer students' internal states and to react to the students' needs appropriately. In this chapter, I therefore first describe the relationship between students' (non)attention-related behavior and teaching quality. Second, I point out how students' behavior is relevant to teachers' professional perceptions of the classroom (i.e., their professional vision). Finally, I present a conceptual framework that can help systematize the mechanisms underlying students' (non)attention-related behavior in teacher-learner interactions.

1.2.1 Students' (Non)Attention-related Behavior as an Indicator of Teaching Quality

Diederich and Tenorth (1997) identified students' attentiveness, motivation, and understanding as constitutive goals of teaching and learning. The degree to which teachers succeed in achieving these goals may be described by their teaching quality (Klieme & Rakoczy, 2003; Praetorius et al., 2018). Within the scope of research on teaching effectiveness, different approaches that have identified different dimensions of teaching quality have emerged (e.g., Lipowsky et al., 2009; Pianta & Hamre, 2009; Praetorius et al., 2018; Reyes et al., 2012). Even though these approaches are based on rather independent lines of research, they largely agree on three dimensions that describe the quality of teaching and teacher-learner interactions: student support/emotional support, classroom management/classroom organization, and cognitive activation/instructional support (Pianta & Hamre, 2009; Praetorius et al., 2018).

The dimension of student support (or emotional support) refers to the teacher's promotion of positive interactions to create a supportive classroom environment (Pianta et al.,

2012). This includes sensitivity to individual student needs and the facilitation of students' motivation (Hamre et al., 2013; Klieme, 2018). Classroom management (or classroom organization) comprises a teacher's ability to establish classroom discipline by preventing disruptions and providing a clear set of rules and expectations (Hamre et al., 2013; Klieme, 2018). By managing the instructional time efficiently, teachers should maximize the opportunities to learn and the time students spend on task (Hamre et al., 2013). According to cognitive activation (or instructional support), instruction should be oriented toward understanding and higher order thinking (Pianta et al., 2012; Praetorius et al., 2017), which can be accomplished by engaging students in challenging activities that activate students' prior knowledge and stimulate their deductive thinking (Praetorius et al., 2017).

These quality dimensions classify the opportunities that are provided by the teacher during instruction and are supposed to have effects on students' outcomes. According to Klieme and Rakoczy (2008), student support affects students' motivation through their experience of autonomy, competence, and social relatedness, whereas classroom management and cognitive activation determine the available time on task as well as the processing depth and through this affect students' knowledge and understanding. Students have to actively use the available time on task by focusing on the instructional material (i.e., by being on-task) and by allocating the required amount of attentional resources to the task (i.e., by engaging in the necessary depth of processing). Therefore, the dimensions of classroom management and cognitive activation are particularly important for students' (non)attention-related behavior, as they provide the structures that directly aim to guide students' attention and behavior. However, the dimension of student support can also affect students' (non)attention-related behavior rather indirectly via students' motivation.

As already outlined above, attention determines the success of knowledge construction (Brünken & Seufert, 2006; Chapter 1.1.3). However, attentional resources are limited, and it is important to guide students' attention in teacher-learner interactions during instruction to ensure that students allocate their attention appropriately and engage with the learning material in the desired way (Mayer, 2002; Sweller et al., 1998). Teachers need to manage the classroom in such a way that students find themselves in learning environments that offer the opportunity to focus and pay attention to the learning material. Good classroom management aims to provide optimal conditions for students to become attentive (Praetorius et al., 2018). In turn, a lack of disruptive behavior, for example, can indicate that a teacher is managing their classroom successfully. Whereas classroom management should make sure that students pay attention,

teachers can affect the intensity with which students pay attention by choosing cognitively activating tasks. Cognitive activation is aimed at deep-level thinking and can be accomplished by choosing constructive and interactive learning activities (see Chi & Wylie, 2014; Chapter 1.1.2). Therefore, teachers can guide students' attention-related behavior through cognitive activation in their choice of learning activities, thus providing the conditions necessary for knowledge construction (Praetorius et al., 2018). At the same time, students' behavior can indicate their level of cognitive activation, for example, when students make an argument to support their position in a discussion.

Stipek (2002) demonstrated in the area of mathematics that quality aspects of instruction in terms of analysis, depth of knowledge, problem solving, discourse, and locus of authority were associated with students' engagement. Even though Stipek (2002) did not claim to measure cognitive activation with her choice of quality aspects, the respective conceptualizations refer to independent and higher order student thinking and can thus be considered to cover the cognitive-activation dimension. Furthermore, van de Grift et al. (2017) found that intense and activating teaching was related to student engagement in Dutch and South Korean schools, supporting the importance of cognitive activation for guiding the intensity with which students allocate their attentional resources toward the learning material.

Previous research demonstrated that whether or not teachers made clear statements about their expectations (in terms of classroom management) was associated with higher rates of students' on-task behavior (Beyda et al., 2002). Additionally, studies have shown that teachers' participation in workshops on classroom (behavior) management significantly reduced students' reported levels of disruptive and off-task behavior (Maini, 2011) and increased levels of student engagement (Piwowar et al., 2013). These results support the impact of classroom management for ensuring that students are on task and are paying attention. Current approaches view classroom management as teachers' proactive and preventative controlling effect and emphasize teachers' ability to remain aware of what is going on in the classroom (withitness; Kounin, 1970) as this ability is associated with student work involvement. Maintaining a functional overview is necessary to provide sufficient learning time, engage all students in active learning processes, and elicit their cooperation to create a learning environment that enables all students to engage in relevant cognitive processes (Emmer & Stough, 2001).

1.2.2 Students' (Non)Attention-related Behavior as a Point of Reference for Teachers

In order to avoid disruptions, teachers need to notice cues in students' behavior that indicate off-task behavior as well as cues that indicate problems in understanding to ensure that students can engage in the desired learning activities. However, in order to maintain a functional overview, an adequate perception of the given instructional situation is required to act in a professional way (Bromme, 1992; Seidel et al., 2010).

To ensure effective teaching, teachers' professional knowledge is crucial (Darling-Hammond & Bransford, 2005; Seidel & Shavelson, 2007). According to Shulman's framework (1986) on professional knowledge of teachers, three distinct categories are apparent: content knowledge (i.e., knowledge about principles, concepts, procedures, and meta-knowledge about the subject matter), generic pedagogical knowledge (i.e., knowledge about the nature of learning), and pedagogical content knowledge (i.e., knowledge about how to convey domain-specific topics and possible misconceptions by students). In addition to the type and amount of knowledge, the elaborated and coherent organization of knowledge structures is also essential (Borko & Livingston, 1989; Krauss et al., 2008). Knowledge is organized in so called curriculum scripts, which allow teachers to detect relevant patterns in the classroom and to make reasoned decisions about their instruction (Putnam, 1987). Teachers use their curriculum scripts to understand classroom situations by noticing and interpreting relevant cues that are critical for successful teaching and learning (Lachner et al., 2016). The ability of professionals to notice and interpret features that are relevant to their work is described by a concept called professional vision (Goodwin, 1994; Sherin, 2007).

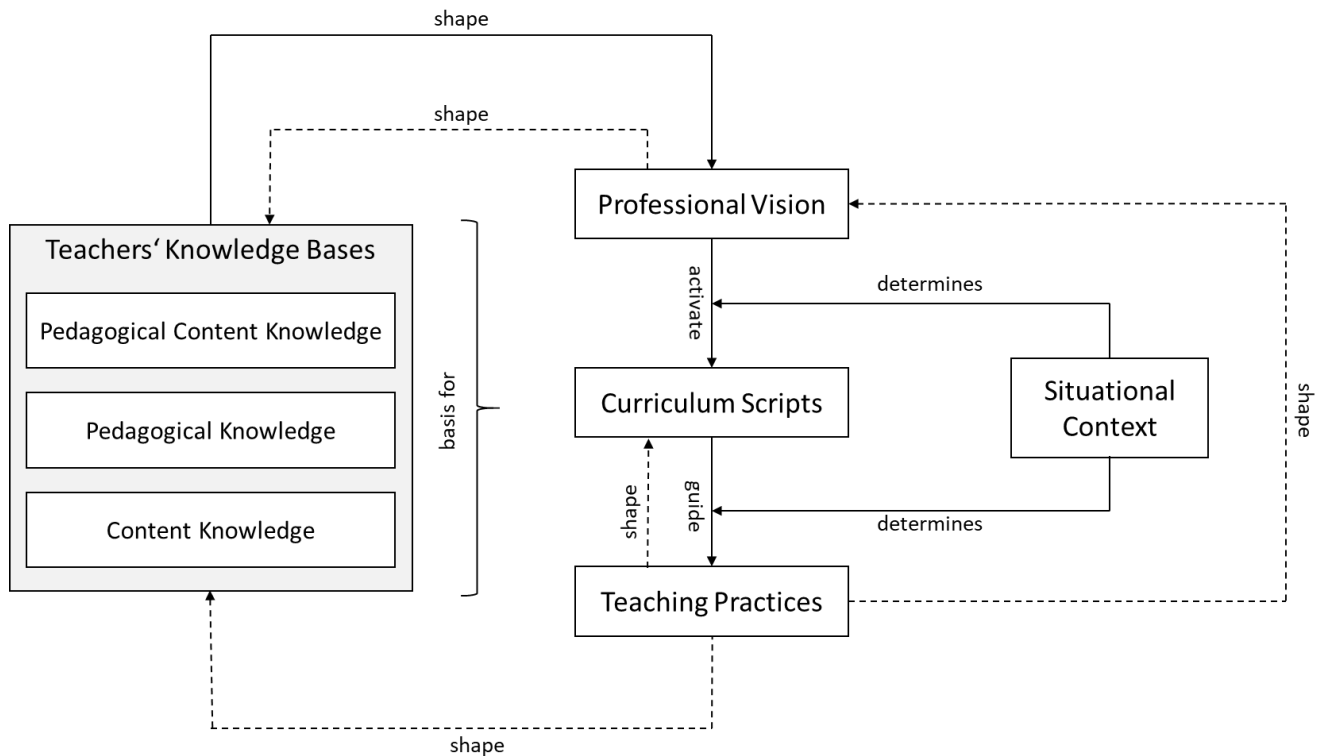
The concept of professional vision was transferred to teaching practice by Sherin et al. (2011) who defined teachers' professional vision as the ability to identify important events for students' learning and to interpret them meaningfully within the classroom context. Teachers' professional vision can thus be viewed as an indicator of knowledge representations that aid the preparation of effective teaching actions (Kersting et al., 2012; Sherin, 2007). It involves knowledge-based processes of attentional control and information processing (van Es & Sherin, 2008) and consists of two components: (a) noticing, which consists of the identification of relevant events that are important for teaching and learning in the classroom, and (b) reasoning, which consists of knowledge-based processing and a reasoned approach to events that are noticed in the classroom (Seidel & Stürmer, 2014). Due to the combination of perception, interpretation, and decision-making, professional vision can be considered a situation-specific

skill (see Blömeke et al., 2015) that requires practical experience. To appropriately respond to disruptions in terms of classroom management, for example, teachers need to detect respective classroom events (e.g., two students talking to each other), identify them as relevant (e.g., the students' conversation is not related to the learning material), and derive reasoned decisions for their subsequent actions (e.g., intervene and prompt students to pay attention). Professional vision thus requires generic pedagogical knowledge about principles of teaching and learning (Grossman & McDonald, 2008; Shulman, 1987) to classify observed information according to its relevance.

Teachers' professional vision is determined by the situational context (Lachner et al., 2016). Figure 1 illustrates how the situational context influences the way professional vision affects a teacher's choice of teaching practices via curriculum scripts that highlight the situation-specificity of the underlying mechanisms (Korthagen, 2010; Lachner et al., 2016). These teaching practices in turn shape what kind of information is relevant in the classroom and needs to be noticed and identified by the teacher. For example, depending on the current situation, if two students are talking to each other, this might indicate that the students are confused during classroom discussions when focusing on the teacher would be desirable. Conversely, conversations can constitute on-task behavior during practice phases, whereas passive behavior (e.g., looking out the window) would indicate a lack of understanding or interest. When teachers can notice and identify a lack of attention-related behavior in students, teachers can act and adapt their teaching methods accordingly or encourage their students to actively engage with the learning material actively.

Figure 1

Model of Teachers' Cognitions in Relation to Teaching Practices and Aspects of the Situational Context



Note. Adapted by permission from Lachner, A., Jarodzka, H. & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science* 44, 197–203. <https://doi.org/10.1007/s11251-016-9376-y>. Copyright 2016 by Springer Nature.

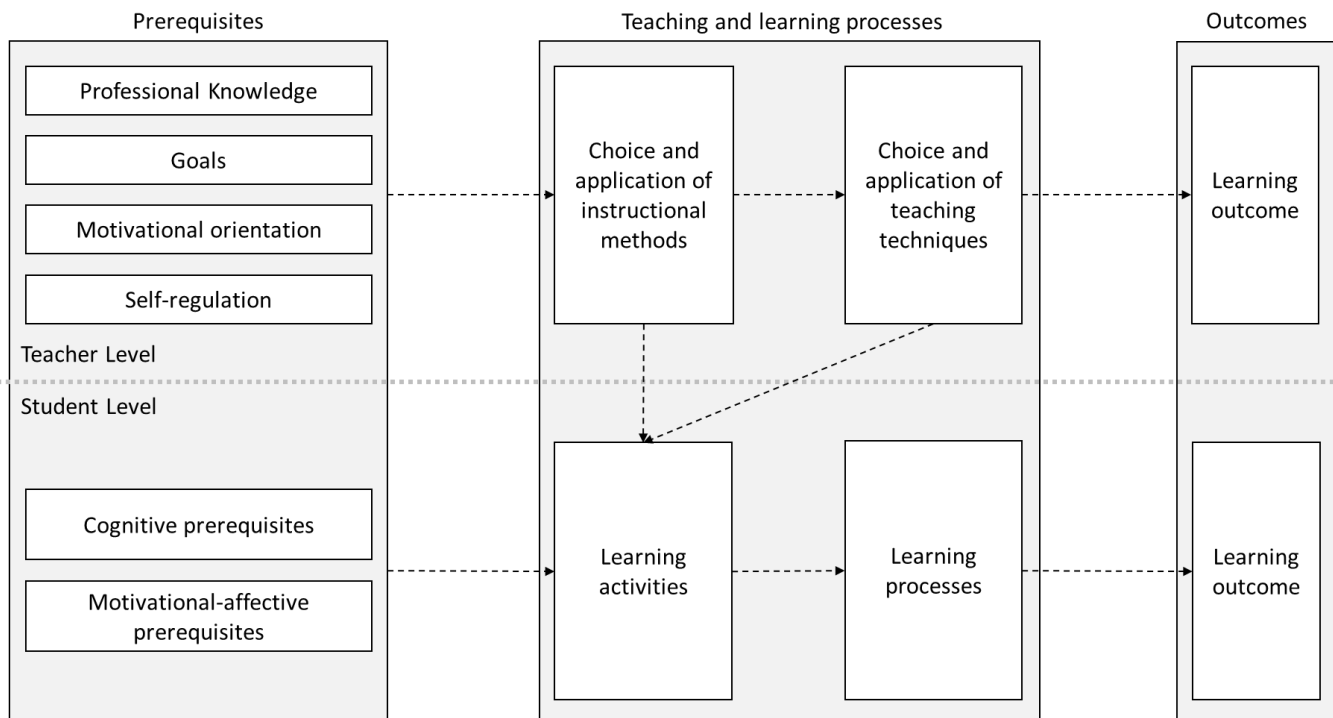
1.2.3 How to Structure the Mechanisms Underlying Students' (Non)Attention-related Behavior in Teacher-Learner Interactions?

A central goal that teachers have is to ensure that students are paying attention to the learning material with the desired level of intensity, as students need to pay attention to build up their respective knowledge structures (Brünken & Seufert, 2006). To design high-quality instruction, teachers need to rely on students' (non)attention-related behavior to identify possible starting points for improvement. For example, teachers need to identify disruptive student behavior to figure out how to manage their classrooms. Then again, they need to observe students closely to estimate whether a given task encourages them to engage in the desired cognitive processes or whether the teachers need to adapt their degree of cognitive activation. In turn, classroom management and cognitive activation qualify the extent to which teachers succeed in realizing this goal with students' (non)attention-related behavior serving as an

indicator. Therefore, students' (non)attention-related behavior needs to be considered on the individual level of students as well as on the interaction level during classroom instruction. The integration of the two levels can help researchers analyze and investigate determinants of teacher-learner interactions during instructional processes.

Kollar and Fischer (2019) developed a conceptual framework for analyzing instructional teaching and learning processes while taking a two-level structure into consideration (see Figure 2). Their framework was based on the supply-use model (Angebot-Nutzungs-Modell; Helmke, 2009) as it distinguishes between individual prerequisites, teaching and learning processes, and outcomes on the teacher and student levels, respectively. Taking Chi and Wylie's (2014) ICAP framework into consideration, Kollar and Fischer (2019) differentiated between overt learning activities and covert learning processes on the student level. They proposed that the teacher's choice and application of instructional methods and teaching techniques affects students' learning activities and through this indirectly influences students' learning processes. This framework was developed to investigate determinants of successful teaching and learning processes. It allows researchers to analyze processes and relationships across the different levels and further acknowledges the connection of overt behavior and covert processes. However, it can be adapted so that it becomes more suitable for structuring the mechanisms underlying students' (non)attention-related behavior during teacher-learner interactions.

As the term *interaction* already denotes, students and teachers affect each other's behavior in a reciprocal manner. For example, teachers can make controversial statements to encourage students to engage in elaborate cognitive processing and participate in classroom discussions. Otherwise, students can pose questions or show a lack of understanding through their facial expressions, prompting the teacher to reflect on their teaching methods and to derive ways to optimize their teaching. Further, in terms of classroom management, students' disruptive behavior would force the teacher to take action, for example, by engaging in eye contact or admonishing the students to make them stop the undesirable behavior. Consequently, to structure the mechanisms underlying students' (non)attention-related behavior during teacher-learner interactions, it is necessary to consider (a) students' individual level and (b) the interaction level during classroom instruction with mutual influences between teachers and learners. Additionally, situation-specific and dynamic components have to be implemented as teacher-learner-interaction processes are characterized by their simultaneity, multidimensionality, and immediacy (Doyle, 1977).

Figure 2*Conceptual Framework for the Analysis and Promotion of Teaching and Learning Processes in Instruction*

Note. Adapted by permission from *Psychologie für den Lehrberuf* (p. 335), by D. Urhahne, M. Dresel, and F. Fischer, 2019. Springer. Copyright 2019 by Springer Deutschland.

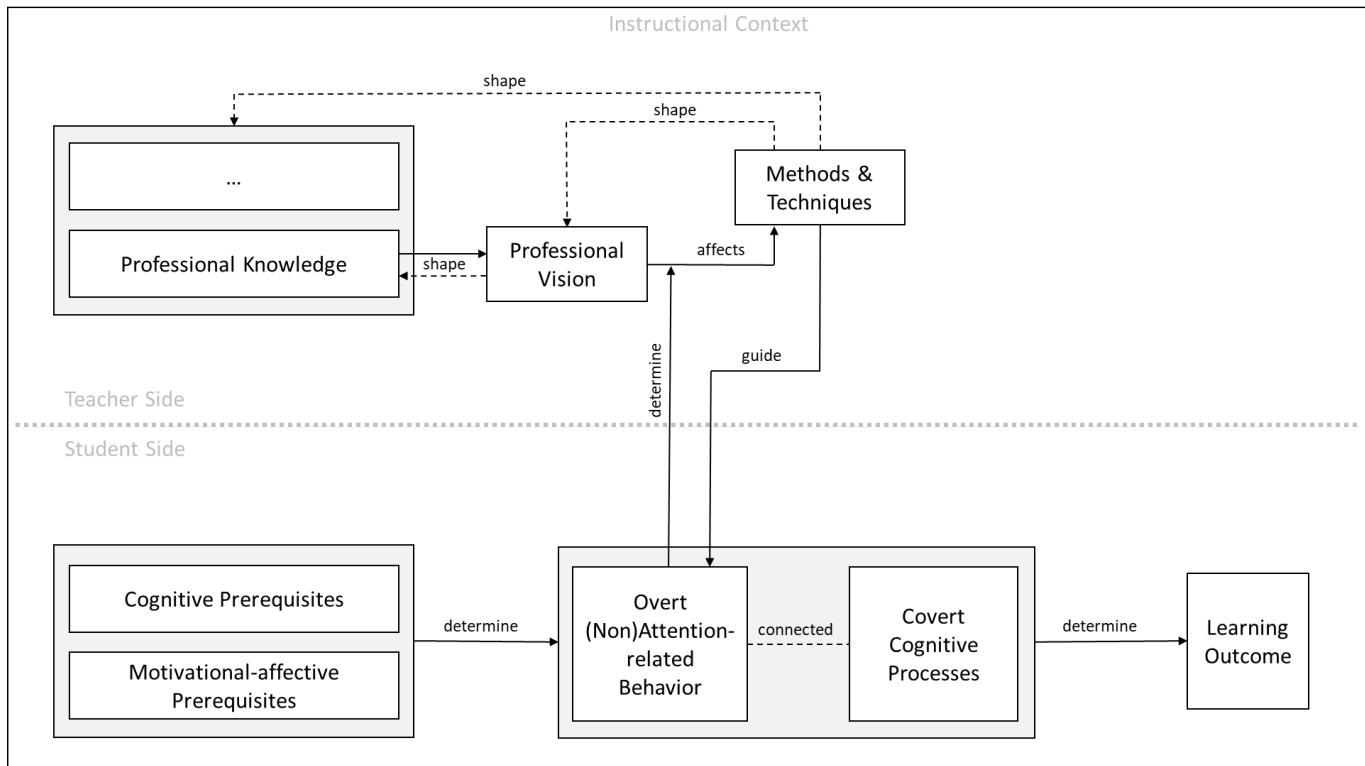
To account for the reciprocal influence and situation-specific components of teacher-learner interactions and their momentum, I used the broad structure of the framework by Kollar and Fischer (2019) as the foundation because the authors took into consideration the relations between students' overt learning activities and covert learning processes and provided a structure that allowed for the implementation of mutual effects between teachers and learners. However, I modified the framework in three ways. First, I changed the labeling of the student and teacher *levels* into the student and teacher *sides* to avoid creating confusion with the terminology that referred to individual and interaction levels as used throughout this dissertation. Second, as teachers must constantly monitor students' behavior, I added teachers' professional vision as a dynamic component on the teacher side. According to Lachner et al. (2016), the situational context determines the way professional vision affects the methods and techniques chosen by the teacher (via the curriculum scripts). In teacher-learner interactions, the situational context includes the students and the ways in which they behave. Therefore, on the basis of how well teachers notice and reason about students' (non)attention-related behavior, they would choose their methods and techniques in such a way that encourages the

appropriate set of cognitive processes. In turn, the choice of methods and techniques would guide students' behavior. For example, by providing feedback during a classroom discussion, teachers can encourage students to pay close attention and to engage in deep-level thinking. Third, I supplemented students' learning activities in Kollar and Fischer's (2019) original framework by students' (non)-attention-related behavior in general. In teacher-learner interactions, the teacher needs to consider the entire spectrum of students' behavior to decide whether students are paying attention and whether they are engaging in the necessary cognitive processes. Additionally, the type of off-task behavior can indicate the extent to which students allocate their attentional resources to non-content-related activities, providing more information for when teachers may attempt to intervene in potential learning outcomes. Thus, students' overall behavior constitutes the dynamic component on the student side. The resulting framework is presented in Figure 3.

This framework provides a way (a) to structure the factors that determine students' (non)attention-related behavior and (b) to investigate how students' (non)attention-related behavior affects teachers' actions during teacher-learner interactions. Therefore, this framework can be used to evaluate the potential of students' (non)attention-related behavior to provide new insights into teacher-learner-interaction processes.

Figure 3

Conceptual Framework for Systematizing the Mechanisms Underlying Students' (Non)Attention-related Behavior in Teacher-Learner Interactions



1.3 Measuring Students' (Non)Attention-related Behavior

1.3.1 Approaches for Measuring Students' Attention During Teacher-Learner Interactions

As outlined in Chapter 1.1, attention has been the focus of research in different domains with different theoretical considerations and approaches. As a consequence, attention and its effects can be measured through brain imaging, electrophysiology, self-reports, and overt behaviors (Chun et al., 2011). However, not all of these measurement methods are suitable in the scope of education research in classroom settings.

Brain imaging techniques, such as electroencephalography (EEG) have already been widely used in laboratory settings, but recent studies have also transferred EEG into real classroom settings. Ko et al. (2017), for example, had students perform a sustained attention task and identified EEG patterns that were linked to visual attention during classroom activities. In another study, Babiker et al. (2019) used EEG to detect situational interest² in students during classroom instruction. However, Ko et al.'s (2017) study results were restricted to visual

² Babiker et al. (2019) argued that situational interest can evoke attention in students.

attention and thus cannot provide information about whether students actively processed the learning material. Additionally, Babiker et al. (2019) used aggregated measures of situational interest to identify EEG patterns, requiring that more research be put toward online measurement to investigate the process of instruction. Even though EEG works well in laboratory settings and is being used more frequently in real classrooms, it is not yet suitable for delivering information that can be used to fully investigate effects of teaching methods and the instructional process.

Another measurement approach uses peoples' electrophysiology, such as their electrodermal activity (EDA) by detecting the change in electrical properties of the skin during mental exertion (Yoshida et al., 2014). EDA is considered a useful indicator of attention, as attention-grabbing and attentionally demanding tasks induce increased EDA responses (Edelberg, 1972; Rushby et al., 2007). Besides the rather noisy nature of EDA data due to movement artefacts and distortions that come from environmental factors (e.g., high temperatures), another issue is that even if it is possible to detect whether a person is attentive, the focus of their attention remains unclear. For example, a student's EDA signals can indicate that they are highly attentive, but instead of listening closely to the teacher, the student may be focused on the latest text message on their phone.

When it comes to internal cognitive processes, it appears obvious that the best way to get the respective information is by asking people themselves. Self-reports are widely used because they are practical, relatively cheap, and easy to administer to a large sample (Christenson et al., 2012). However, self-reports can be biased, for example, due to problems in retrospective recall. This issue can be circumvented by using experience sampling, by which participants carry electronic devices that signal them to fill out a self-report at a predefined point in time (Hektner et al., 2007). Nevertheless, this method also has its disadvantages, as the success of experience sampling strongly depends on participants' compliance and diligence and might not always be given in students. Furthermore, using experience sampling within teacher-learner interactions can disrupt the natural flow of instruction and/or the attention that students put toward the instructional material. To cause as little disruption as possible, the questionnaires administered in experience sampling studies have to be as short as possible, even though this comes with the risk that they might not adequately cover the construct under investigation (see Fredricks & McColskey, 2012).

A special case of overt behaviors are eye movements. Eye movements are largely guided by selective attentional processes (Findlay & Gilchrist, 2003; Hayhoe & Ballard, 2005) and

thus provide a great deal of potential for research in education science. Mobile eye-tracking technology has opened up new possibilities for collecting eye movements in natural settings, which is important as people's eye movements in laboratory settings differ from those in the real world (Foulsham et al., 2011). Besides this advantage, the use of eye trackers with children easily results in erroneous analyses as gaze calibration can deteriorate if people touch the eye-tracking glasses too often. Additionally, equipping an entire class with mobile eye-tracking technology is expensive and thus hardly feasible.

Using external behavioral observation has a long tradition in psychological (see Foster et al., 1988; Tryon, 1998) and education research (e.g., Jackson & Hudgins, 1965). Observational measurements are systematic approaches that are applied to detect and interpret certain behaviors (Girard & Cohn, 2016). They can capture the development of behavior over time and in this way support our understanding of its antecedents and consequences as well as its contribution to dynamic processes (Bakeman & Quera, 2011). Systematic observations involving time-sampling (i.e., fixed intervals with one event per interval) or event-sampling (i.e., all occurrences and durations of events during a certain period) have been used to investigate students' attention during classroom instruction for about six decades (see Chapter 1.3.2). Compared with self-reports, systematic behavioral observations do not underlie biased response tendencies, such as self-presentation or social desirability (Stone et al., 2000), and can be administered without instrumental effort (as opposed to psychophysiological measures, such as EEG or EDA, as well as eye-tracking; Helmke & Renkl, 1992). Even though behavioral observations are restricted from covering the overt parts of attention (Büttner & Schmidt-Atzert, 2004; Cobb & Hops, 1973; Helmke & Renkl, 1992), students' behavior can provide valuable indicators of the underlying cognitive processes (see Chapter 1.1.2). Hence, systematic behavioral observations constitute a suitable approach for investigating effects of specific teaching methods and instructional techniques, as well as the instructional process during classroom lessons.

1.3.2 Observational Approaches to Measure Students' Attention- and Engagement-related Behavior During Instruction

In recent decades, various observational approaches for measuring students' attention- and engagement-related behavior in educational settings have evolved. In the following, I will review the most prominent lines of research that focus on students' behavior during instruction. However, I will not consider instruments that have been developed to identify students with ADHD, as those instruments are designed for a particular target group and thus might not be

suitable for determining the behavior of students in general. A chronological overview of the approaches is provided in Table 2.

One of the first studies to use direct observation in classrooms to investigate the effects of students' attention was conducted in the late 1960s. Lahaderne (1968) used a modified version of the Jackson-Hudgins Observation Schedule (Jackson & Hudgins, 1965) to determine whether students were attentive during instruction. On the basis of Lahaderne's (1968) study, Cobb (1972) asked external observers to classify students' overt classroom behaviors to predict their academic achievement. By using 14 categories of student behaviors, he was able to demonstrate that interactive compared with passive content-related behaviors had stronger connections with achievement (Cobb, 1972). These findings motivated Samuels and Turnure (1974) to replicate Lahaderne's (1968) approach and found that task-relevant behavior was related to achievement in addition to gender-specific differences in classroom attentiveness. Marliave et al. (1977) proposed that differences between students that could not be explained by individual background characteristics could be accounted for by the time students spent actively engaged with the learning material. They developed an instrument that measured students' engagement-behavior patterns in addition to information about teachers and their instructional activities. Also Karweit and Slavin (1981) were interested in the effects of time on learning achievement. They developed a system to categorize the available instruction time and assessed engagement time on two subdimensions. Within the scope of a larger project, Carta et al. (1988) developed an observation system that focused on students' engagement in addition to ecological and teacher information and was later also provided as software (Greenwood et al., 1994). Using distinct aspects of Marliave et al. (1977) and Carta et al. (1988) and comparable to Cobb (1972), Friedman et al. (1988) defined 13 categories of attention-related activities in classrooms to investigate differences between students with and without learning disabilities.

Whereas this line of development originated in North America, Ehrhardt et al. (1981) developed the first German behavioral observation system for students that also relied on time-sampling and direct observations in classrooms. It was later refined by Helmke and Renkl (1992), who not only differentiated between on- and off-task behaviors but also passive and active manifestations. About 20 years later, Hommel (2012) further modified this instrument and applied it to video analyses as opposed to direct classroom observations. Helmke and Renkl (1992) as well as Hommel (2012) further coded the instructional context as additional information.

Table 2

Chronological Overview of Already Existing Observational Approaches to Measure the Attention- and Engagement-related Behavior of Individual Students in Instructional Settings that were not Developed to Measure Symptoms of Attention Deficit/Hyperactivity Disorder (ADHD)

Author(s)	Scale	Sampling rate	Modality	Indicators
Lahaderne (1968) – modified version of Jackson-Hudgins Observation Schedule (Jackson & Hudgins, 1965)	<ul style="list-style-type: none"> ▪ Clearly attentive ▪ Clearly inattentive ▪ Uncertain ▪ Not observable 	n/a	Direct observation	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Activity
Cobb (1972)	<ul style="list-style-type: none"> ▪ Attention ▪ Talk to teacher (positive) ▪ Talk to peer (positive) ▪ Volunteers ▪ Initiation to teacher ▪ Compliance ▪ Self-stimulation ▪ Out of chair ▪ Play ▪ Talk to teacher (negative) ▪ Talk to peer (negative) ▪ Noncompliance ▪ Looking around ▪ Not attending 	Time sampling 10-s intervals	Direct observation	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Activity ▪ (Conversation) content ▪ Raising hand
Marliave et al. (1977)	<ul style="list-style-type: none"> ▪ Engaged – written response ▪ Engaged – oral response ▪ Engaged – covert response ▪ Engaged – engaged directions 	Time sampling 3-6-min intervals	Direct observation	

	<ul style="list-style-type: none"> ▪ Not engaged – interim activity ▪ Not engaged – waiting for help ▪ Not engaged – off task 				
Ehrhardt et al. (1981)	<ul style="list-style-type: none"> ▪ On-Task ▪ Off-Task 	Time sampling 10-s intervals	Direct observation	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Body posture ▪ Activity 	
Karweit and Slavin (1981)	<ul style="list-style-type: none"> ▪ On-Task ▪ Raised Hand ▪ Off-Task (ok) ▪ Off-Task (not ok) ▪ No task opportunity 	Time sampling 30-s intervals	Direct observation		
Carta et al. (1988) – MS-CISSAR	<ul style="list-style-type: none"> ▪ Positive Engagement: Academic Responses ▪ Neutral Engagement: Task Management ▪ Negative Engagement: Inappropriate/Competing Behavior 	Time sampling 20-s intervals	Direct observation	Activity	
Friedman et al. (1988) – CSTOCS	<ul style="list-style-type: none"> ▪ Writing ▪ Game ▪ Reading aloud ▪ Reading silently ▪ Talking appropriately ▪ Answering question ▪ Asking question ▪ Raising hand ▪ Looking ▪ Looking for materials ▪ Disruptive ▪ Self-stimulating ▪ Gazing 	Time sampling 30-s intervals	Direct observation	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Activity ▪ Task-orientation 	

Author(s)	Scale	Sampling rate	Modality	Indicators
Helmke and Renkl (1992) – MAI	<ul style="list-style-type: none"> ▪ On-Task active/self-initiated ▪ On-Task reactive/externally-initiated ▪ On-Task passive ▪ Off-Task not interacting ▪ Off-Task interacting/disruptive ▪ No task 	Time sampling 5-s intervals	Direct observation	
O'Malley et al. (2003) – STROBE	<ul style="list-style-type: none"> ▪ Talk ▪ Listen ▪ Read ▪ Organize ▪ Write ▪ Other 	Time sampling 5-min intervals 4 students in parallel	Direct observation	
Baker et al. (2004) – BROMP	<ul style="list-style-type: none"> ▪ On-Task ▪ On-Task conversation ▪ Off-Task conversation ▪ Off-Task solitary behavior ▪ Inactivity ▪ Gaming the system 	Time sampling 20-s intervals	Direct observation	Wok with tutoring system on computer
Hommel (2012) – ModAI	<ul style="list-style-type: none"> ▪ On-Task active ▪ On-Task passive ▪ Other task ▪ Off-Task passive ▪ Off-Task active 	30-s intervals	Videos	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Body posture ▪ Way of interaction and activity
Whitehill et al. (2014)	<ul style="list-style-type: none"> ▪ Very engaged ▪ Engaged in task ▪ Nominally engaged 	<ul style="list-style-type: none"> ▪ Static frames ▪ 10-s intervals 	Videos/Pictures	

Alimoglu et al. (2014) – IEM	<ul style="list-style-type: none"> ▪ Not engaged at all ▪ Frame unclear/no person ▪ Engaged with noneducational material/browsing a book/notes/whispering to a friend, etc. ▪ Reading or writing something (including following the lecture from a published material or taking notes) ▪ Listening to the instructor or a talking student/looking at slides or board ▪ Talking to the instructor/reading something to entire class or writing something on the board, flipchart, etc. ▪ Talking/discussing with one or a group of students on the subject matter 	Time sampling 20-s intervals	Direct observation	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Activity
Lane and Harris (2015) – BERI	<ul style="list-style-type: none"> ▪ Engaged ▪ Disengaged ▪ Uncertain 	10 students at once, 3-10-s each	Direct observation	<ul style="list-style-type: none"> ▪ Gaze direction ▪ Activity ▪ Facial expressions

The observation of students' attention- and engagement-related behavior was later expanded to go beyond school-based settings and became of interest to other research domains. For example, O'Malley et al. (2003) developed a system for observing students' engagement in the context of medical education, which was based on the work by Greenwood and colleagues (Stanley & Greenwood, 1981). This observation system was later used by Alimoglu et al. (2014) as the foundation of a system for observing students' engagement in different types of classes. Lane and Harris (2015) believed that the contextual coding of O'Malley et al. (2003) was too complicated and developed an observation protocol that allowed researchers to observe the behavior of 10 students at once. Furthermore, within the scope of work from computer science, systems for coding students' behavior have been developed in recent years. Baker et al. (2004) classified students' behavior into six distinct categories while working on an intelligent tutoring software for investigating the effect of off-task activities. The coding system used by Whitehill et al. (2014) was deployed to provide data by human raters, which then served as the ground truth for an automated estimation of students' engagement.

To sum up, a broad variety of behavioral-observation systems exist. Whereas the original focus was on direct observation within classroom instruction, recent approaches have transferred students' behavioral observation to other domains and made use of advances in technology.

1.3.3 Considerations for New Approaches to Measure Students' (Non)Attention-related Behavior During Instruction

Students' (non)attention-related behavior can provide valuable information about teacher-learner interactions as it can indicate the overall level of teaching quality and constitutes a pivotal point of reference for teachers when inferring students' needs, level of understanding, and current cognitive processes (see Chapter 1.2). Knowledge about students' (non)attention-related behavior can thus be used to gain further insights into the determinants of effective teaching and the instructional process. Even though already existing instruments are well-developed, mostly validated, and supported by empirical evidence, their potential has yet to be fully exploited. One avenue for improving existing observational systems is to use approaches from other domains as inspiration and to adapt the available systems accordingly. Another idea is to integrate and expand previous approaches with theoretical frameworks that have yet to be empirically supported.

In general, the decision to use categorical or dimensional measurements depends on the theoretical considerations underlying the construct of interest (Girard & Cohn, 2016). For

example, Dhamija and Boulton (2018) used a dimensional rating system to estimate people's engagement with a web intervention. They coded behavior on a continuous scale ranging from *very disengaged* to *very engaged* sampled on a per second basis. They used joysticks and open-source software for continuous interpersonal behavior annotation in videos (Girard, 2014). Even though this coding system cannot be used to measure students' behavior during instruction, it provides inspiration for further directions in education research. Lizdek et al. (2012) presented this joystick approach and its advantages and possibilities for studying patterns of continuous behavior in the social sciences.

Also, researchers from education science have already discovered the potential of continuous behavioral annotation during teacher-learner interactions (e.g., Pennings & Mainhard, 2016; Pennings et al., 2014). Continuous observations allow researchers to evaluate instructional processes on a different temporal level that goes beyond the opportunities that existing approaches provide. In education science, longitudinal data commonly refer to multiple measurement points that are spread across several years or at least various days. Continuous rating systems of (student) behavior provide intensive longitudinal data that can describe dynamic, situation-specific interpersonal processes and provide information about time-varying determinants and effects (e.g., within one school lesson; Dermody et al., 2017).

Referring to Fredricks et al. (2004) and Peterson et al. (1984), Fredricks and McColskey (2012) noted that individual behavioral observations usually only provide limited information about the quality of people's effort or engagement. Theoretical considerations such as the ICAP framework (Chi & Wylie, 2014) can account for this drawback, as they differentiate between different modes of engagement. Helmke and Renkl (1992) and Hommel (2012) already provided a clear hierarchical structure for the behavioral categories as their distinction included active and passive manifestations. This division of on-task behavior can easily be extended in accordance with the ICAP framework (Chi & Wylie, 2014), resulting in more informative measurements. The finer-grained distinction of on-task behavior into passive, active, constructive, and interactive manifestations can provide hints about the level of cognitive activation of the current task. As already described above (Chapter 1.1.2), however, it is important not only to consider attention-related behavior but also behavior that is associated with inattentiveness in the context of classroom instruction. The quality of alternative activities (i.e., off-task behaviors such as playing on one's phone or talking to peers) during instruction can provide hints about whether students can process instruction-relevant information at least partially (see Lavie, 2005). The failure to focus one's attention on the lesson reduces learning

and results in a superficial understanding (D'Mello, 2016). Additionally, off-task behavior can serve as an indicator for teachers that certain students might not be following what they are teaching and that they need to adapt their teaching method to support these struggling or distracted students. When students are off-task, the teacher might need to reconsider which strategies they are using for classroom management. Therefore, students' off-task behavior is important on the individual level to determine learning (success) as well as on the interaction level as a major indicator of how teachers can adapt their teaching methods to meet students' needs. Consequently, off-task behavior and thus students' failure to adequately focus their attention can be indicators of overall teaching quality, especially classroom management.

New approaches should thus consider the entire behavioral spectrum that students can exhibit during teacher-learner interactions so that estimations can be more precise. Additionally, researchers should make use of the opportunities new technologies provide and deploy continuous ratings that enable the analysis of situation-specific effects.

2

Research Questions

2 Research Questions

The objective of the present dissertation was to provide new insights into teacher-learner interactions by evaluating how to adequately measure students' (non)attention-related behavior during instruction and assess the potential of such behavior to provide these new insights. To this end, perspectives on students' individual level and the level of interactions between teachers and learners need to be integrated. Therefore, I first outlined the importance of students' (non)attention-related behavior with regard to the individual level and learning success (Chapter 1.1) as well as its relevance on the interaction level (Chapter 1.2). On the basis of the theoretical foundations, I derived a conceptual framework for systematizing the mechanisms underlying students' (non)attention-related behavior during teacher-learner interactions (Chapter 1.2.3). Subsequently, I provided an overview of already existing behavioral observation instruments and deduced considerations for new measurement approaches (Chapter 1.3). This resulted in two main research questions that I addressed by conducting three empirical studies:

- 1) How can students' (non)attention-related behavior be measured validly so that the mechanisms underlying the processes involved in teacher-learner interactions can be investigated?
- 2) What new insights can students' (non)attention-related behavior provide during teacher-learner interactions (a) on the individual level and (b) on the interaction level during classroom instruction?

As outlined above, none of the existing observational instruments were suitable for investigating the processes involved in teacher-learner interactions on the basis of students' (non)attention-related behavior. Either the temporal resolution was too rough to account for situation-specific effects of the instructional context, or the category system did not provide enough information about the quality of the respective student behavior. With regard to the first research question and on the basis of the theoretical considerations outlined above, I thus developed a new observation scale that provides behavioral indicators that enable a continuous assessment of students' (non)attention-related behavior (*Continuous (non)Attention-related Behavior Indicators – CABI*). This observation scale expands the approach of Helmke and Renkl (1992) and Hommel (2012) by using the basic structure of on- and off-task behaviors with active and passive manifestations, but it defines a dimensional scale with behavioral indicators that were derived from the entire body of already existing behavioral observation instruments. It was further inspired by Chi and Wylie's (2014) ICAP framework and Lavie's

(2005) load theory of attentional processing to provide more information about the quality of students' (non)attention-related behavior (for more details, see the coding manual in the Appendix). These adjustments were made (a) to improve the estimations about covert cognitive processes while maintaining the greatest possible level of objectivity and (b) to open up new ways to investigate the processes involved in teacher-learner interactions by providing a continuous structure of the data. The CABI is suitable for teacher-centered settings because there is no ambiguity with respect to where the students are supposed to focus their attention. Students are supposed to focus on either the teacher or individual peers during classroom discussions, or the teacher provides clear statements about what the students are supposed to do. In comparison, activities such as group work are characterized by rather unstructured and individual interactions.

In the following, I present three empirical studies that aimed to answer the aforementioned research questions. Figure 4 depicts the placement of the three empirical studies in the conceptual framework to systematize the mechanisms underlying students' (non)attention-related behavior within teacher-learner interactions. Study 1 was concerned with the validation of the new observation scale. Studies 2 and 3 used the CABI to investigate the mechanisms underlying students' (non)attention-related behavior on the individual level (Study 2) as well as on the interaction level within instruction (Study 3), which were only possible due to the continuous nature of the data.

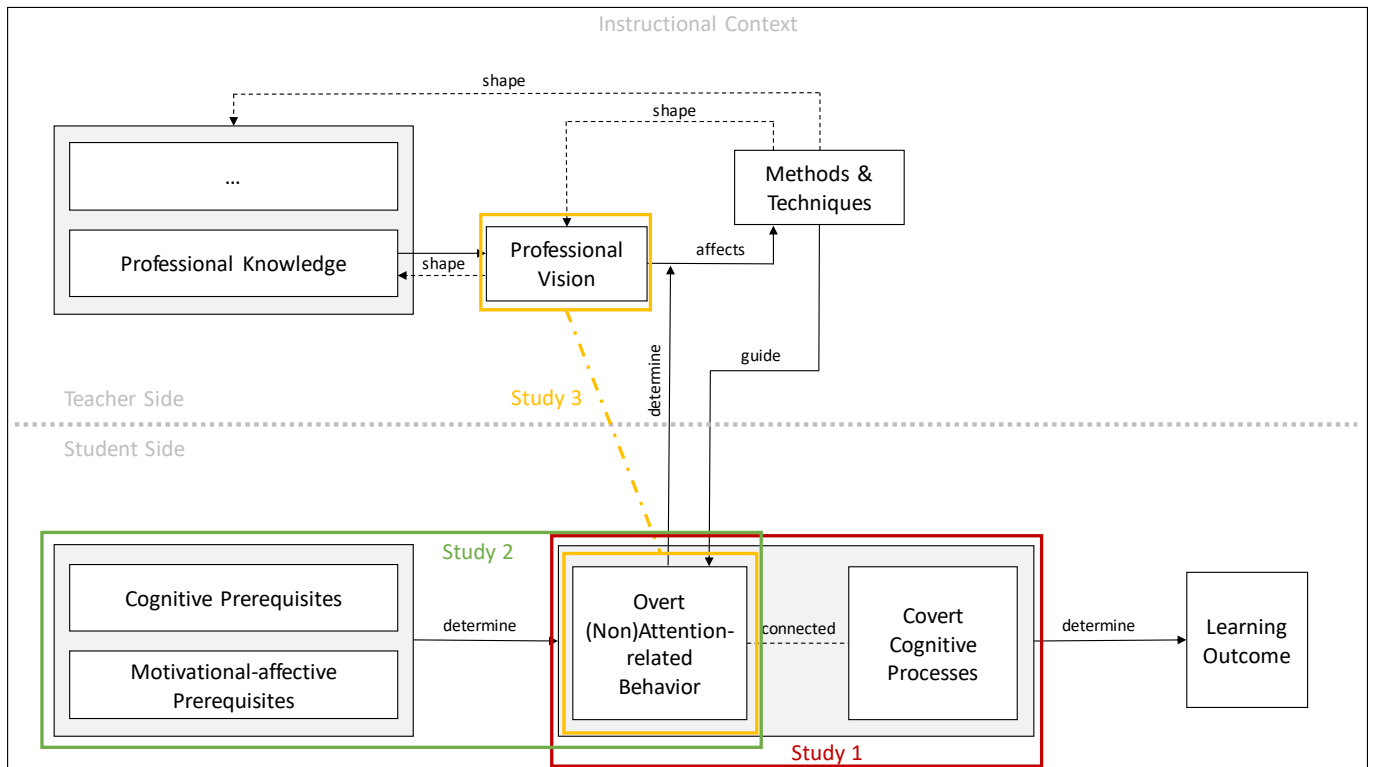
Study 1 (*Attentive or Not? Toward a Machine-Learning Approach for Assessing Students' Visible Engagement in Classroom Instruction*) assessed the validity of the new observation instrument in an instructional setting to strengthen the close connection between observable behavior and internal cognitive processes. Construct validity was tested by using students' self-reported cognitive engagement, situational interest, and involvement. Predictive validity was examined via performance on a subsequent knowledge test on the session topic. In addition, Study 1 aimed to replicate the relation between overt behaviors and covert processes by using a machine-vision-based approach to assess visible indicators of students' (non)attention-related behavior. The automated approach was based on visual parameters, such as head pose, gaze direction, and facial expressions. Therefore, this study contributes to answering Research Question 1 about how to validly measure students' (non)attention-related behavior so that its contribution to the processes involved in teacher-learner interactions can be investigated.

Study 2 (*Why Do Students Exhibit Different Attention-Related Behavior During Instruction? Investigating the Effects of Individual Prerequisites, Class Membership, and Classroom Activities*) focused on determinants of students' (non)attention-related behavior on the individual level. Given that situational aspects of students' (non)attention-related behavior during classroom instruction have not yet been studied intensively, Study 2 explored how students' individual prerequisites, class membership, and classroom activities – as determinants to consider in addition to teachers' quality of instruction – can explain differences in variability within individual students' (non)attention-related behavior across a lesson as well as differences between students. This study made use of the intensive longitudinal data structure that resulted from the use of continuous behavioral annotation with the CABI. Video data were obtained from a larger study by Seidel et al. (2016) in which the teacher-centered parts of eighth-grade introductory mathematics lessons in German high schools were displayed. This study evaluated one way to gain new insights into the contribution of students' (non)attention-related behavior to teacher-learner interactions on the individual student level.

Focusing on effects on the interaction level, Study 3 (*How does learners' behavior attract preservice teachers' attention during teaching?*) investigated the relationship of preservice teachers' attentional focus and students' (non)attention-related behavior in standardized teaching situations. Teachers need to notice and identify relevant cues in students' behavior to make reasoned decisions about their practices; however, novice teachers in particular have trouble distributing their attentional focus evenly across all students while teaching. This mismatch might be due to rather salient student behaviors that catch preservice teachers' attention. To investigate the possible determinants more closely, this study examined how students' behavior guides preservice teachers' attentional focus during teaching. Study 3 was based on video- and eye-tracking data from a previous study conducted by Stürmer et al. (2017) and was supplemented by the continuous annotation of learners' behavior with the CABI. Therefore, Study 3 also aimed to provide new insights into students' (non)attention-related behavior, particularly on the level of interactions between novice teachers and learners.

Figure 4

Placement of the Three Empirical Studies in the Conceptual Framework to Systematize the Mechanisms Underlying Students' (Non)Attention-related Behavior during Teacher-Learner Interactions.



3

Study 1

Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., Kasneci, E., & Trautwein, U. (2019). Attentive or not? Toward a machine-learning approach for assessing students' visible engagement in classroom instruction. *Educational Psychology Review*.

<https://doi.org/10.1007/s10648-019-09514-z>

Abstract

Teachers must be able to monitor students' behavior and identify valid cues in order to draw conclusions about students' actual engagement in learning activities. Teacher training can support (inexperienced) teachers in developing these skills by using videotaped teaching to highlight which indicators should be considered. However, this supposes that (a) valid indicators of students' engagement in learning are known and (b) work with videos is designed as effectively as possible to reduce the effort involved in manual coding procedures and in examining videos. One avenue for addressing these issues is to utilize the technological advances made in recent years in fields such as machine learning to improve the analysis of classroom videos. Assessing students' attention-related processes through visible indicators of (dis)engagement in learning might become more effective if automated analyses can be employed. Thus, in the present study, we validated a new manual rating approach and provided a proof of concept for a machine vision-based approach evaluated on pilot classroom recordings of three lessons with university students. The manual rating system was significantly correlated with self-reported cognitive engagement, involvement, and situational interest and predicted performance on a subsequent knowledge test. The machine vision-based approach, which was based on gaze, head pose, and facial expressions, provided good estimations of the manual ratings. Adding a synchrony feature to the automated analysis improved correlations with the manual ratings as well as the prediction of posttest variables. The discussion focuses on challenges and important next steps in bringing the automated analysis of engagement to the classroom.

Keywords: students' visible engagement, attention-related behavior, machine learning, automated picture analysis, classroom synchronization

Cognitive activation, classroom management, and teacher support are the three central tenants of teaching quality (Klieme, Lipowsky, Rakoczy, & Ratzka, 2006; Praetorius, Klieme, Herbert, & Pinger, 2018). The level of students' (dis)engagement in learning activities can be considered a major indicator of both cognitive activation and classroom management because it signals students' engagement in the deep processing of learning content and reveals the time on task (Carroll, 1963) provided by the teachers for students' learning. To this end, teachers are required to take note of their students' attentional focus and make sure the students are engaging in the desired learning activities. Thus, the ability to monitor students' attention and to keep it at a high level is part of the competencies that novice teachers need to acquire. However, research has indicated that teachers might not always be aware of their students' attentional focus, and this may be particularly true for novice teachers.

In general, beginning teachers have trouble monitoring all students in the classroom evenly and noticing events that are relevant for student learning (Berliner, 2001; Cortina, Miller, McKenzie, & Epstein, 2015; Star & Strickland, 2008; Stürmer, Seidel, Müller, Häusler, & Cortina, 2017). Therefore, teacher training needs to support future teachers in developing the necessary knowledge structures that underlie these abilities (e.g., Lachner, Jarodzka, & Nückles, 2016). Consequently, providing an improved measurement approach for student attention will be beneficial for research and can potentially contribute to teacher training. Research has already demonstrated that both inexperienced and experienced teachers' ability to notice relevant cues in the classroom benefits from observing and reflecting on their own videotaped teaching (Kleinknecht & Gröschner, 2016; Sherin & van Es, 2009). Until now, however, instructors have typically had to watch hours of video material to select the most crucial phases of lessons. Similarly, when it comes to research on teaching effectiveness and the development of teachers' ability to notice relevant cues in classroom instruction (i.e., professional vision skills), researchers typically have to invest considerable resources, especially coding resources, to examine the association between teacher behavior and classroom processes (Erickson, 2007). The required effort further increases when investigating students' attention across an entire lesson and analyzing attention at the group level instead of among individuals. In this vein, attention- and engagement-related behavior during classroom instruction has rarely been studied due to the difficulty of data collection and labeling. However, learners might behave differently in naturalistic settings and show versatile behavior that cannot be found in a lab.

One potentially valuable avenue for addressing these issues is to utilize the technological advances made in recent years in fields such as computer vision and machine learning. Therefore, in an ongoing research project (Trautwein, Gerjets, & Kasneci, 2017), we have been investigating whether and how the automated assessment of students' attention levels can be used as an indicator of their active engagement in learning. This automated assessment can in turn be used to report relevant cues back to the teacher, either simultaneously or by identifying and discussing the most relevant classroom situations (e.g., a situation where students' attention increases or decreases significantly) after a lesson.

In the present study, we present a proof of concept for such a machine vision-based approach by using manual ratings of visible indicators of students' (dis)engagement in learning as a basis for the automated analysis of pilot classroom recordings of three lessons with university students. More specifically, by combining multiple indicators from previous research (i.e., Chi & Wylie, 2014; Helmke & Renkl, 1992; Hommel, 2012), we developed a manual rating instrument to continuously measure students' observable behavior. In addition, we performed an automated analysis of the video recordings to extract features of the students' head pose, gaze direction, and facial expressions using modern computer vision techniques. Using these automatically extracted features, we aimed to estimate manually annotated attention levels for each student. Because we had continuous labeling, this could be done by training a regressor between the visible features and the manual labels. We investigated the predictive power of both the manual and automatic analyses for learning (i.e., performance on a subsequent knowledge test). To account for complexity within classrooms and enrich the automated analysis, we also considered synchronous behavior among neighboring students. In the present article, we report initial empirical evidence on the reliability and validity of our automated assessments and their association with student performance.

Attention in Classroom Instruction

Student attention is a key construct in research on both teaching and learning. However, definitions vary widely and are discussed from multiple perspectives. Here, we focus on describing three lines of research that inspired our research program: cognitive psychology models that describe attention as part of information processing, engagement models in which attention makes up part of a behavioral component, and teaching quality models in which student attention is a crucial factor.

In current models in the psychology of learning, *attention* denotes a filtering mechanism that determines the kind and amount of information that enters working memory (Driver, 2001).

This mechanism is crucial for preventing working memory overload and allows the learner to focus on the right kind of information. Only sensory information that enters working memory is encoded, organized, and linked to already existing knowledge. Thus, attention serves as a selection process for all incoming sensory information as it dictates which pieces of information will be processed further and will get the chance to be learned. Thus, attention determines the success of knowledge construction (Brünken & Seufert, 2006). Engle (2002) further proposed that executive attention, which actively maintains or suppresses current representations in working memory, is part of working memory. Certain instructional situations strongly depend on executive processes such as shifting, inhibition, or updating (Miyake et al., 2000) and thus necessitate top-down attentional control. Although information processing occurs in a covert manner, some aspects of attentional processes are likely to be observed from the outside: for example, visually orienting toward a certain stimulus, which improves processing efficiency (Posner, 1988).

Attention is often mistaken for engagement, even though it constitutes only part of it. *Engagement* is defined as a multidimensional meta-construct and represents one of the key elements for learning and academic success (Fredricks, Blumenfeld, & Paris, 2004). It includes observable behaviors, internal cognitions, and emotions. Covert processes such as investment in learning, the effort expended to comprehend complex information, and information processing form part of cognitive engagement (Fredricks et al., 2004; Pintrich & De Groot, 1990). Emotional engagement in the classroom includes affective reactions such as excitement, boredom, curiosity, and anger (Connell, 1990; Fredricks et al., 2004). Attention is considered a component of behavioral engagement alongside overt participation, positive conduct, and persistence (Connell, 1990; Fredricks et al., 2004). Per definition, cognitive engagement refers to internal processes, whereas only the emotional and behavioral components are manifested in visible cues. Nevertheless, all engagement elements are highly interrelated and do not occur in isolation (Fredricks et al., 2004). Thus, attention plays a crucial role because it may signal certain learning-related processes that should become salient in students' behavior to some extent.

Learners' attention also plays a crucial role in *research on teaching*. Teachers must determine whether their students are attentive by considering visible cues, continually monitoring the course of events in order to manage the classroom successfully (Wolff, Jarodzka, van den Bogert, & Boshuizen, 2016) and providing ambitious learning opportunities. A student's attention or lack thereof (e.g., when distracted or engaging in mind wandering) can

signal whether she or he is on-task or off-task. This in turn can provide hints about instructional quality and the teacher's ability to engage his or her students in the required learning activities. Thus, it is important to help teachers develop the skills needed to monitor and support student attention and engagement and adapt their teaching methods. Consequently, accounting for student attention and more broadly student engagement in teaching is considered crucial for ensuring teaching quality, including classroom management, cognitive activation, and instructional support (Klieme, Schümer, & Knoll, 2001; Pianta & Hamre, 2009).

In sum, the definitions, theoretical backgrounds, and terminology used in various lines of research to describe observable aspects of students' cognitive, affective, or behavioral attention/engagement in learning are diverse, but experts agree on their importance and key role in learning. As teachers must rely on visible cues to judge their students' current attention levels (Büttner & Schmidt-Atzert, 2004; Yamamoto & Imai-Matsumura, 2012), we focused on observable aspects of attention and inferences that were based on visible indicators. In the remainder of the article, we use the term *visible indicators of (dis)engagement in learning* to describe these aspects. These visible indicators are highly likely to be associated with learning, but this assumption needs to be validated.

Previous Approaches for Measuring Visible Indicators of Engagement in Learning

The difficulty in assessing students' engagement-related processes in real-world classroom settings consists of externalizing learners' internal (covert) states through visible overt aspects to the greatest extent possible. In psychology, affective states and cognitive processes such as attentional control are usually determined from physiological signals, such as heart rate, electrodermal activity, eye tracking, or electroencephalography (e.g., Gerjets, Walter, Rosenstiel, Bogdan, & Zander, 2014; Krumpe, Scharinger, Rosenstiel, Gerjets, & Spueller, 2018; Poh, Swenson, & Picard, 2010; Yoshida et al., 2014). Using this kind of psychologically sound measurements makes it possible to detect covert aspects of learning-related processes; however, these measures are hardly feasible in classroom instruction, especially when teachers must be equipped with knowledge about what indicators to look for in students. Furthermore, these approaches are useful for answering very specific research questions. However, they are not sufficient for determining whether students' ongoing processes are actually the most appropriate for the situation. By contrast, overt behavior can provide visible indicators of appropriate learning-related processes in students.

Overt classroom behavior is an important determinant of academic achievement (Lahaderne, 1968; McKinney, Mason, Perkerson, & Clifford, 1975). Although overt behavior

does not always represent a reliable indicator of covert mental processes, previous findings have demonstrated a link between cognitive activity and behavioral activity (Mayer, 2004). Previous studies have analyzed students' behavior and have determined its relation to achievement (Helmke & Renkl, 1992; Hommel, 2012; Karweit & Slavin, 1981; Stipek, 2002). Furthermore, in research on engagement, correlations between student engagement and academic achievement have been found (Lei, Cui, & Zhou, 2018). Other studies have found opposing results (e.g. Pauli & Lipowsky, 2007); however, these studies either relied on self-reports as opposed to observer ratings or only focused on certain facets of engagement-related behavior (e.g., only active on-task behavior).

There have been various attempts to systematically assess visible indicators of engagement in classroom learning. For example, Helmke and Renkl (1992) based their research on an idea by Ehrhardt, Findeisen, Marinello, and Reinartz-Wenzel (1981) and related observable student behavior to internal processes using time-on-task as an indicator of whether a student was paying attention to classroom-related content. Assessing observable content-related behavior is essential to this operationalization of higher order attention. Hommel (2012) modified this approach and applied it to the video-based analysis of instructional situations. Rating behavior as either on- or off-task with varying subcategories demonstrated the interrelation between visual cues and achievement or reduced learning (Baker, Corbett, Koedinger, & Wagner, 2004; Helmke & Renkl, 1992).

However, learners can differ in their learning activities but still be engaged in a certain task. The ICAP framework proposed by Chi and Wylie (2014) distinguishes between passive, active, constructive, and interactive overt behavior, which differ across various cognitive engagement activities. This framework focuses on the amount of cognitive engagement, which can be detected from the way students engage with learning materials and tasks (Chi & Wylie, 2014). This theoretical model provides a promising approach for further expanding the different types of on-task behavior so that variations in student behavior can be accounted for.

In sum, considering learning content has been shown to be useful; however, there is a lack of research involving the continuous analysis of attention or engagement over the course of one or more lessons. A unique feature of the present study is that we aimed to acquire a continuous assessment (i.e., a score for every student in the classroom for every second of instruction time) of students' visible indicators of (dis)engagement in learning. This temporal resolution was crucial in our approach because we aimed to provide comparable data that could be used to train a machine-learning algorithm. To reach this high level of temporal resolution,

we decided to annotate learners' behavior continuously. The free software CARMA (Girard, 2014) enables the continuous interpersonal behavior annotation by using joysticks (see Lizdek, Sadler, Woody, Ethier, & Malet, 2012). However, this new approach limited us in terms of using already existing rating instruments because existing instruments do not allow for a high enough level of temporal resolution. Furthermore, the CARMA software requires annotations on a scale rather than rating the behavior in terms of categories as already existing instruments do. When developing the new instrument, we mainly oriented on the MAI (Helmke & Renkl, 1992; Hommel, 2012). However, we needed to define more fine-grained indicators of student behavior to make annotations along a continuous scale possible. Therefore, we added indicators from various established instruments to extend our rating scale. We assumed that the manual observer annotations would serve only as approximations of the actual cognitive states of the students and that the averaged (i.e., intersubjective) manual annotations would reflect the "true score" of the visible indicators of (dis)engagement in learning better than a single rater could. Subsequent to the ratings, we thus calculated the mean of the raters for every second. The mean values for each second and student were used as the ground truth to train a machine-learning approach.

Using Machine Learning to Assess Visible Indicators of (Dis)Engagement in Learning

Machine learning and computer vision methods have made tremendous progress over the past decade and have been successfully employed in various applications. In the context of teaching, these methods might offer an efficient way to measure student engagement, thereby decreasing the need for human rating efforts. However, any machine-learning method that is aimed at estimating covert engagement-related processes in learning needs to depend on visible indicators such as head pose, gaze direction, facial action unit intensity, or body pose and gestures. State-of-the-art methodologies for the automated assessment of engagement can be divided into two categories: single-person and classroom-based analyses.

In a single-person analysis, facial expressions can provide hints about ongoing cognitive processes and can be analyzed by considering action unit (AU) features. Related studies by Grafsgaard, Wiggins, Boyer, Wiebe, and Lester (2013), Bosch, D'Mello, Baker, et al. (2016), and Bosch, D'Mello, Ocumpaugh, Baker, and Shute (2016) investigated the relations between AU features and several response items and affective states. Even though these studies found that several facial AUs were associated with engagement, they were limited to affective features and did not consider head pose or gaze direction.

In another work, Whitehill, Serpell, Lin, Foster, and Movellan (2014) introduced a facial analysis approach to estimating the level of engagement on the basis of manually rated engagement levels. Although their facial analysis approach was able to predict learning just as accurately as participants' pretest scores could, the correlation between engagement and learning was moderate due to the limited amount of data and the short-term nature of the situations.

In a classroom-based analysis, the focus shifts away from single individuals onto shared features and interactions among participants. In this context, a number of notable contributions (e.g., Raca, 2015; Raca & Dillenbourg, 2013) have utilized various sources of information to understand features of audience behavior, such as the amount of estimated movement and synchronized motions among neighboring students. They found that immediate neighbors had a significant influence on a student's attention, whereas students' motion was not directly connected with reported attention levels (Raca & Dillenbourg, 2013; Raca, Tormey, & Dillenbourg, 2013). Furthermore, Raca, Tormey, and Dillenbourg (2014) analyzed students' reaction time upon presentation of relevant information (sleeper's lag). In addition to estimating head pose, they considered the class period, student's row, how often faces were automatically detected (as a precursor to eye contact), head movement, and the amount of still time (i.e., 5-s periods without head movement) because these features had previously been shown to be good predictors of engagement in learning (Raca, Kidzinski, & Dillenbourg, 2015). Although these results were promising, they were limited to correlational studies of reported attention levels; predictive approaches were not used due to limits in the performance of computer vision methodology.

A recent study estimated human-annotated attention levels by using 3D vision cameras to identify individuals using face and motion recognition without any physical connection to people and solely on the basis of visual features (Zaletelj, 2017; Zaletelj & Košir, 2017). Due to technological limitations associated with 3D vision cameras, the analysis was based on a single row of students rather than the entire classroom. Fujii, Marian, Clark, Okamoto, and Rekimoto (2018) used head-up and head-down states and classroom synchronization in terms of head pose as informative tools that could provide feedback to teachers. However, they did not validate their system using educational measures (pretests, posttests, or observations) and only reported user experiences with three teachers.

In sum, few previous studies have investigated classroom-based attention and engagement beyond the single-person context due to the poor performance of computer vision

approaches for face and body pose recognition in unconstrained settings (e.g., varying illumination, occlusion, motion, challenging poses, low resolution, and long distance). However, recent advances in deep learning technology have resulted in the availability of new methods for the robust extraction of such features from videos. By employing such technology in this study, we aim to bring a fine-scaled analysis of visible indicators to classroom studies and augment individual engagement analysis with another useful feature: classroom synchronization.

Research Questions

The present study is part of an ongoing research project in which researchers from education science, psychology, and computer science are working to create an automatic assessment of students' engagement that could one day be implemented in an interface that can be used for research as well as teacher training purposes. The present study lays the basis for achieving these goals by developing and testing an automated approach to assessing visible indicators of students' (dis)engagement in learning. Such a remote approach requires comparable data (generated by human raters) that can be used as the ground truth in order to train a classifier. However, existing instruments (Helmke & Renkl, 1992; Hommel, 2012) for measuring engagement-related processes in learning (a) require human observers to make a huge number of inferences and (b) require data to be collected in 30-s or 5-min intervals. This is problematic for our context because an automated analysis can only rely on visible indicators, does not consider content-specific information at all, and operates at a more fine-grained temporal resolution. Therefore, we developed a new instrument to annotate student behavior manually by applying a rating method with visible indicators over time. This manual rating served as the starting point from which to train an algorithm by applying methods from machine learning and computer vision.

The present study addressed the following research questions:

- 1) Is the new manual annotation of visible indicators of (dis)engagement in learning related to students' learning processes and outcomes? To validate our instrument, we examined how the manual ratings were correlated with students' self-reported cognitive engagement, involvement, and situational interest. We expected these self-reported learning activities to cover different facets of (dis)engagement in learning, and when combined, we expected them to account for cognitive parts of the construct. Furthermore, we tested whether the scores resulting from the manual annotation would predict students' performance on a knowledge test at the end of an instructional session.

2) Is it possible to adequately replicate the relation to students' learning processes and outcomes by using visible indicators of (dis)engagement in learning based on the machine-learning techniques that estimated the manual ratings? We used gaze, head posture, and facial expressions to estimate the manual ratings. To test the quality of our machine vision-based approach, we examined the associations between the scores generated from the automated approach and the manual ratings and students' self-report data regarding their learning processes, and we used the machine-learning scores to predict achievement on the knowledge test.

3) How does adding synchrony aspects of student behavior affect the automated estimations of the manual ratings? The results of previous studies have indicated that immediate neighbors have a significant influence on a student's engagement (Raca & Dillenbourg, 2013; Raca et al., 2013). As a first step toward including indicators of synchrony in our project, we added students' synchrony with the person sitting next to them as an additional variable to our prediction models, which were based on the automated assessment of student engagement.

Method

The ethics committee from the Leibniz-Institut für Wissensmedien in Tübingen approved our study procedures (approval #2018-017), and all participants gave written consent to be videotaped.

Sample and Procedure

We decided to conduct a study involving university students in order to validate our approach before administering it in school classrooms. A total of $N = 52$ university students (89.5% women, 8.8% men, mean age = 22.33, $SD = 3.66$) at a German university volunteered to take part in the study. The study was conducted during regular university seminar sessions on quantitative data analysis (90 min). A total of three different seminar groups were assessed. The topics of the sessions were either *t tests for independent samples* (Sessions 1 and 2) or *regressions* (Session 3) and ranged from 30 to 45 min. The sessions were videotaped with three cameras (one teacher camera, two cameras filming the students). If students refused to be videotaped, they were either seated outside the scope of the cameras or switched to a parallel seminar. Participants were informed in advance of the study's purpose, procedure, and ethical considerations such as data protection and anonymization. To avoid confounding effects of the teacher, the same person taught all sessions in a teacher-centered manner. Before the session started, students filled out a questionnaire on background variables (age, gender, final high school examination [Abitur] grade, school type) and individual learning prerequisites. After the

session, participants completed a knowledge test on the specific topic of the session and completed another questionnaire about learning activities during the seminar.

Instruments

Individual learning prerequisites. We used established questionnaire measures to assess three individual learning prerequisites: Dispositional interest in the session's topic was captured with four items ($\alpha = .93$) adapted from Gaspard, Häfner, Parrisius, Trautwein, and Nagengast (2017). Self-concept in quantitative data analysis was assessed with five items ($\alpha = .80$; adapted from Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006), and 13 items were used to test for self-control capacity ($\alpha = .83$; Bertrams & Dickhäuser, 2009). Moreover, we administered the short version of the quantitative subscale (Q3) of the cognitive abilities test (Heller & Perleth, 2000). Measuring these learning prerequisites allowed us to control for potential confounding variables in the analyses.

Learning outcomes. The knowledge test consisted of 12 and 11 items that referred to participants' declarative and conceptual knowledge of the session topic, respectively. We z-standardized the knowledge test scores within each group for subsequent analysis.

Self-reported learning activities. After the session, we assessed students' involvement (four items, $\alpha = .61$; Frank, 2014), cognitive engagement (six items, $\alpha = .79$; Rimm-Kaufman, Baroody, Larsen, & Curby, 2015), and situational interest (six items, $\alpha = .89$; Knogler, Harackiewicz, Gegenfurtner, & Lewalter, 2015) during the seminar session (see Table 1).

Table 1

Item wording for learning activities

Construct	Items
Cognitive engagement	<p>I exerted myself as much as possible during the session.</p> <p>I thought about different things during the session.</p> <p>I only paid attention when it was interesting during the session.</p> <p>It was important for me to really understand things during the session.</p> <p>I tried to learn as much as possible during the session.</p> <p>I pondered a lot during the session.</p>
Involvement	<p>During the session...</p> <p>... I strongly concentrated on the situation.</p> <p>... I occasionally forgot that I was taking part in a study.</p> <p>... I was mentally immersed in the situation.</p> <p>... I was fully engaged with the content.</p>
Situational interest	<p>When you think about today's session...</p> <p>... the seminar session aroused your curiosity.</p> <p>... the seminar session attracted your attention.</p> <p>... you were completely concentrated on the seminar session.</p> <p>... the seminar session was entertaining for you.</p> <p>... the seminar session was fun for you.</p> <p>... the seminar session was exciting for you.</p>

Note: Items have been translated

Analysis

Continuous manual annotation. To develop a continuous manual annotation that included potential valid indicators of students' visible (dis)engagement in learning, we used the instruments developed by Helmke and Renkl (1992) and Hommel (2012) as a basis. However, these instruments label behavior in categories and thus cannot be used as a continuous scale.

Therefore, we combined the idea of on-/off-task behavior and active/passive subcategories with existing scales from the engagement literature. Furthermore, we used the theoretical assumptions about students' learning processes and related activities in classrooms pointed out by the ICAP framework (Chi & Wylie, 2014) as an inspiration to define more fine-grained differentiations within the possible behavioral spectrum. The distinction into passive, active, constructive, and interactive behavior allowed us to make subtler distinctions between the different modes of on-task behavior, and this concept could be transferred to off-task behavior (i.e., passive, active, deconstructive, and interactive) as well. By combining different approaches, we could define visible indicators of (dis)engagement in learning on a continuous scale. The resulting scale ranged from -2, indicating interruptive and disturbing off-task behavior, to +2, indicating highly engaged on-task behavior where, for example, learners ask questions and try to explain the content to fellow learners (see Figure 1). When a person could not be seen or was not present in the classroom, the respective time points were coded as missing values in subsequent analyses.

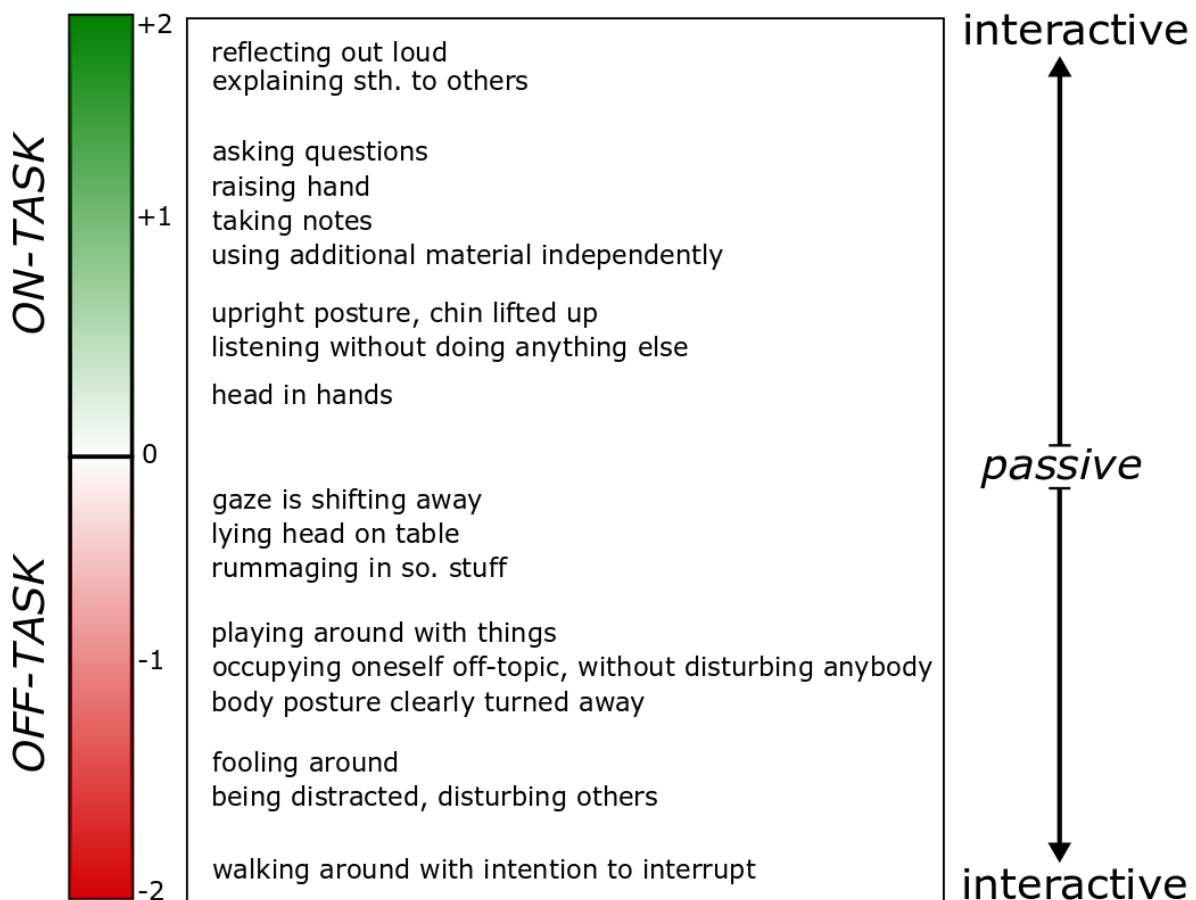


Figure 5 Scale with exemplary behavioural indicators

The behavior of each observed person throughout the instructional session was coded in 1-s steps using the CARMA software (Girard, 2014) and a joystick. A total of six raters annotated the videotaped seminar sessions, and each session was annotated by a total of three raters. The raters consisted of student assistants and one researcher, all of whom were trained carefully before annotating the videos. First, raters were introduced to the conceptual idea of the rating and the rating manual. They were told to concentrate on observable behavior to avoid making inferences and considering information from previous ratings. The raters focused on one student at a time in a random order. Every rater had to code one of two specific sections of the video for training, and the raters had to annotate special students who showed different types of behavior. To ensure that we could use all the video material for our analysis, raters who used Video-Section A for training annotated Video-Section B later and vice versa. The respective video sections used for training purposes were not included in the analysis. Only after their annotations reached an interrater reliability with an expert rating of at least $ICC(2,1) = .60$ were raters allowed to annotate the study material. We report the $ICC(2,1)$ here as an indicator of interrater reliability because our data were coded on a metric scale level, and we had more than two raters per participant. We calculated the $ICC(2,1)$ for every student, indicating the interrater reliability averaged across all time points, whereby values between .60 and .74 indicated good interrater reliability (Hallgren, 2012); the $ICC(2,1)$ for each student was .65 on average (absolute agreement). When the annotations between the raters deviated strongly, critical situations were discussed among the raters and recoded following consensus. The raters were not informed about the students' individual prerequisites, their learning outcomes, or their self-reported learning activities.

Machine-learning approach. In addition to the manual ratings (see previous section), we employed a machine vision approach to estimate (dis)engagement in learning using visible indicators and analyzed the same videos with this approach. More specifically, we first detected the faces in the video (Zhang et al., 2017) and automatically connected the faces detected in the video stream to each student so that we could track their behavior. Faces were aligned, and their representative features extracted automatically based on the OpenFace library (Baltrušaitis, Zadeh, Lim, & Morency, 2018). However, this procedure was not applicable to all students and all frames due to occlusions by peers, laptops, or water bottles. The subsequent analyses were therefore based on a subsample of $N = 30$ students.

In contrast to typical facial analysis tasks such as face recognition, the number of participants in classrooms is limited. We used the following three modalities as feature

representations: *head pose*, *gaze direction*, and *facial expressions* (represented by facial action units). The head pose features consist of the head's location with respect to the camera and the rotation in radians around three axes. Gaze is represented by unit gaze vectors for both eyes and gaze direction in radians in world coordinates. Facial action units (AU) were estimated according to the Facial Action Coding System (FACS; Ekman & Friesen, 1978), for which each AU can be expressed at five intensity levels. More specifically, to estimate the occurrence and intensity of FACS AUs, we used the following 17 AUs: upper face AUs are AU1 (inner brow raiser), AU2 (outer brow raiser), AU4 (brow lowerer), AU5 (upper lid raiser), AU6 (cheek raiser), and AU7 (lid tightener); the lower face AUs are AU9 (nose wrinkler), AU10 (upper lip raiser), AU12 (lip corner puller), AU14 (dimpler), AU15 (lip corner depressor), AU17 (chin raiser), AU20 (lip stretcher), AU23 (lip tightener), AU25 (lips part), AU26 (jaw drops), and AU45 (blink). Given that our videos were recorded at 24 frames per second, and the manual annotations were conducted each second, we used the mean values of these features for time sequences of 24 frames to predict engagement intensities. More specifically, we regressed the engagement intensities using linear Support Vector Regression (Fan, Chang, Hsieh, Wang, & Lin, 2008) in a subject-independent manner. Excluding the subject whose engagement intensity was to be predicted, individual regression models were trained using all other student features and labels. Subsequently, the test subject's engagement during each 1-s period was predicted. Finally, the average estimated engagement intensity during the instructional session was taken as the final descriptor for each participant.

The label space for students' manually annotated engagement was between -2 and +2; however, the distribution of the data was highly imbalanced. Nearly 80% of all of the annotated data ranged from 0.2 to 0.8. Therefore, we had to clip the label values to fit the range of -0.5 and 1.5 and then rescale them to 0 and 1 in our regression models.

In summary, the visible indicators we used could be differentiated into two categories: engagement-related features (i.e., head pose and gaze direction) and emotion-related features (AU intensities). In order to compare their contributions with visible indicators of (dis)engagement in learning, we used them both separately and in combination.

In order to go beyond a single-person analysis, we further integrated an indicator of *synchrony*. Because simultaneous (i.e., synchronous) behavior in a group of students or an entire classroom can have an impact on individual students, in this first step toward an automated approach, we considered the behavior of neighboring students sharing the same desk. First, we measured the cosine similarities between neighboring students' manual ratings ($N =$

52, 26 pairs). Second, we calculated the relation between neighbors' synchrony (cosine similarities) and their mean engagement levels during instruction. Because synchronization is a precursor to engagement, we expected the neighbors to provide valuable information for estimating (dis)engagement in learning. Therefore, in the final step of our analysis, we concatenated the feature vector of each student and his or her neighbor into a single vector and trained the same regression models as for the estimation of each individual student's engagement.

Results

Relation between Continuous Manual Annotation and Student Learning

We tested the validity of our manual rating instrument in two steps. First, we investigated construct validity by correlating the manual ratings with the self-reported learning activities. The manual annotations were significantly correlated with students' self-reported cognitive engagement, situational interest, and involvement ($.49 \leq r < .62$; Table 2).

Additionally, we calculated a multiple linear regression with the three self-reported learning activities as regressors. Together, they explained 42.9% of the variance in the manual ratings. This corresponds to a multiple correlation of $r = .66$. Second, we examined the predictive validity of our new instrument. We inspected the intercorrelations between all variables with the knowledge test (Table 2). The knowledge test scores (the dependent variable in this study) were significantly correlated with the manual ratings, cognitive abilities, and situational interest ($.30 \leq r < .42$). To test for effects of possible confounding variables, we calculated two additional linear regression models in which we added background variables (Model 2) and learning prerequisites (Model 3) into the regression and compared them with the prediction that involved only manual ratings (Table 3). The effect of the manual ratings remained robust and still explained a significant proportion of the variance in the knowledge test results.

Table 2

Correlations between individual characteristics, learning activities, achievement, and manual rating, with confidence intervals in brackets (N = 52)

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Female												
2. Age	-.29* [-.52, -.01]											
3. Abitur-grade	-.10 [-.36, .19]	.31* [.03, .54]										
4. School type	-.09 [-.36, .20]	-.01 [-.29, .27]	-.26 [-.51, .01]									
5. Dispositional Interest	-.12 [-.39, .16]	.04 [-.24, .32]	-.07 [-.34, .21]	.01 [-.27, .29]								
6. Self-concept	.16 [-.12, .42]	-.22 [-.47, .06]	.16 [-.12, .42]	.00 [-.28, .28]	-.62** [-.77, -.41]							
7. Self-control capacity	.09 [-.20, .36]	-.10 [-.37, .18]	-.15 [-.41, .14]	-.14 [-.41, .14]	.28 [-.00, .52]	-.38** [-.60, -.12]						
8. Cognitive abilities	.04 [-.24, .31]	.07 [-.22, .34]	-.39** [-.60, -.12]	.08 [-.21, .35]	.14 [-.15, .40]	-.22 [-.47, .06]	-.02 [-.30, .26]					

9. Manual rating	-.21 [-.46, .08]	.04 [-.24, .32]	.02 [-.26, .30]	-.25 [-.49, .03]	.18 [-.10, .44]	-.21 [-.46, .07]	.20 [-.08, .45]	.01 [-.27, .29]				
10. Cognitive engagement	-.14 [-.40, .14]	-.14 [-.41, .14]	.03 [-.25, .30]	-.11 [-.38, .18]	.30* [.03, .54]	-.26 [-.50, .02]	.31* [.03, .54]	-.12 [-.38, .17]	.60** [.39, .75]			
11. Situational interest	.05 [-.23, .33]	-.27 [-.51, .01]	-.06 [-.33, .22]	-.11 [-.37, .18]	.51** [.27, .69]	-.32* [-.55, -.05]	.11 [-.18, .37]	-.05 [-.32, .23]	.49** [.24, .67]	.60** [.39, .75]		
12. Involvement	-.11 [-.38, .17]	-.06 [-.34, .22]	.14 [-.15, .40]	-.23 [-.47, .06]	.30* [.02, .53]	-.28 [-.52, .00]	.35* [.07, .57]	-.17 [-.42, .12]	.62** [.42, .77]	.76** [.61, .86]	.68** [.50, .81]	
13. Knowledge test	.09 [-.20, .36]	-.07 [-.34, .21]	-.23 [-.48, .05]	-.19 [-.45, .09]	.20 [-.08, .45]	-.03 [-.30, .25]	-.08 [-.35, .21]	.33* [.06, .56]	.30* [.03, .54]	.12 [-.16, .39]	.42** [.16, .62]	.21 [-.07, .46]

Note: To better understand the relations between individual prerequisites, learning activities, and learning outcomes we calculated correlations across all variables. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). Abitur-grade: lower values indicate better results according to the German grading system. * $p < .05$, ** $p < .01$.

Table 3

Prediction of knowledge test results (N = 52)

	Model 1			Model 2			Model 3		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Manual rating	1.08	0.49	.032	0.92	0.49	.067	1.00	0.48	.042
Abitur-grade				-0.60	0.29	.043	-0.50	0.30	.099
School type				-0.40	0.28	.159	-0.47	0.27	.087
Cognitive abilities							0.08	0.04	.068
Dispositional Interest							0.43	0.23	.066
Self-concept							0.38	0.26	.160
Self-control capacity							-0.28	0.21	.189
<i>R</i> ²		.092			.184			.342	
<i>F</i>		4.88*			3.46*			3.12**	

Note: Abitur-grade: lower values indicate better results according to the German grading system. * $p < .05$, ** $p < .01$, *** $p < .001$.

Reanalysis with Machine-Learning Approach

We applied our trained regression to test subjects at 1-s intervals and applied mean pooling to create a final estimation that summarized participants' engagement. Table 4 shows the performance of different modalities for estimating (dis)engagement in learning. The performance measures were mean squared errors in the regression and the Pearson correlation coefficient between the manual annotations' mean level and our models' prediction during the instructional session.

As shown in Table 4, the head pose modality exhibited a lower correlation with the manual ratings ($r = .29$) than the other features. By contrast, gaze information and facial expressions (AU intensities) were more strongly correlated with the manual annotations ($r = .44$). Combining head pose and gaze ($r = .61$) or all three modalities ($r = .61$) also led to substantial correlations with the manual annotations.

Table 4

Performance of different modalities in engagement in learning estimation depicted as mean squared error (MSE) for regression and Pearson correlations between manual ratings and our models' estimation (N = 30)

Modalities	MSE	r	p
Single students			
head pose	0.057	.29	.126
gaze	0.055	.44	.015
facial expressions	0.056	.44	.014
head pose + gaze	0.052	.61	.000
3-combined	0.051	.61	.000
Single students + cosine similarity			
head pose + gaze (sync)	0.029	.71	.000
3-combined (sync)	0.050	.70	.000

In addition, we tested the correlations between the posttest variables (i.e., the knowledge test and self-reported learning activities) and the different models for estimating the manual ratings (Table 5). According to these results, regression models, which perform better with respect to MSE and lead to higher correlations with the manual ratings, seem to contain more information that is relevant for the posttest variables, particularly with respect to involvement and cognitive engagement.

Table 5

Pearson correlations of different modalities in engagement in learning estimations with post-test variables (N = 30)

Modalities	Knowledge Test		Involvement		Cognitive Engagement		Situational Interest	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Single students								
manual ratings	.14	.468	.64	.000	.62	.001	.53	.003
head pose	-.17	.392	.05	.799	.02	.914	-.02	.913
gaze	.11	.582	.19	.335	.16	.414	.23	.236
facial expressions	-.09	.667	.37	.053	.23	.249	.30	.116
head pose + gaze	-.03	.867	.41	.029	.37	.053	.21	.286
3-combined	-.04	.827	.43	.023	.37	.055	.21	.277
Single students + similarity								
head pose + gaze (sync)	.08	.704	.43	.023	.39	.040	.26	.175
3-combined (sync)	-.01	.968	.45	.016	.38	.043	.26	.189

Addition of Synchrony to the Machine-Learning Approach

The cosine similarities of the manual annotations between neighboring students were strongly correlated with each neighbor's mean engagement level throughout the recording ($r = .78$). More specifically, taking the synchronization into consideration improved the correlation with the manual ratings by 9%, thus showing that synchronization information is helpful for understanding (dis)engagement in learning.

The correlations between the different models for estimating the manual ratings and students' self-reported learning activities and outcomes revealed that the best models were those in which head pose and gaze features were combined with neighbor synchrony ($r = .08, .43, .39$, and $.26$ for the knowledge test, involvement, cognitive engagement, and situational interest, respectively; Table 5). We calculated the mean correlation (based on Fisher's z-transformed correlations) of the three manual annotations (average $r = .74$) and the mean correlation of each

rater and the scores from a model combining head pose, gaze features, and neighbor synchrony (average $r = .64$) for the subsample.

Because the model in which head pose and gaze were combined with neighbor's synchrony had the highest correlation with the manual rating, we calculated a linear regression to predict the posttest variables (Table 6). In order to understand the contribution of neighbor's synchrony, we trained our regression models using the same features with and without synchronization information. Adding neighbor's synchrony improved the prediction of all posttest variables and explained at least 2% more variance. However, the manual rating remained superior.

Table 6

Prediction of post-test variables by fused head pose and gaze estimation, fused head pose and gaze estimation plus cosine similarity, and manual rating in subsample (N = 30)

	Estimated rating (head pose + gaze)					Estimated rating (head pose + gaze) + sync					Manual rating				
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>R</i> ²	<i>F</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>R</i> ²	<i>F</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>R</i> ²	<i>F</i>
Knowledge test	1.37	8.09	.867	.001	0.03	1.14	2.98	.704	.006	0.15	0.63	0.86	.468	.020	0.54
Cognitive Engagement	7.74	3.82	.053	.136	4.10	3.03	1.40	.040	.152	4.67*	1.38	0.35	.000	.380	15.91***
Involvement	13.94	6.05	.030	.170	5.31*	5.37	2.42	.023	.184	5.87*	2.33	0.54	.000	.414	18.34***
Situational Interest	5.64	5.17	.286	.044	1.19	2.63	1.88	.175	.070	1.95	1.54	0.48	.003	.286	10.42**

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

The present study reported key initial results from the development of a machine vision-based approach for assessing (dis)engagement in the classroom. We were able to find empirical support for the validity of our newly developed manual rating instrument. Furthermore, the machine-learning approach proved to be effective, as shown by its correlation with the manual annotations as well as its ability to predict self-reported learning activities. Finally, as expected, including an indicator of synchrony in the automated analyses further improved its predictive power. Next, we discuss our main results in more detail before turning to the limitations of the present study and the crucial next steps.

Empirical Support for the Newly Developed Approach

The manual rating of visible indicators for (dis)engagement in learning predicted achievement on a knowledge test following a university seminar session. This prediction was robust when we controlled for individual characteristics (Research Question 1). In terms of validity, self-reported cognitive engagement, involvement, and situational interest were strongly correlated with the manual rating. As these self-reported learning activities reflect students' cognitive processes during the seminar session, we concluded that our manual ratings capture visible indicators that are actually related to (dis)engagement in learning. Therefore, we inferred that it is reasonable to use these manual ratings as a ground truth for our machine vision-based approach.

In the automated analyses of engagement, we used several visible features (head pose, gaze, facial expressions). More specifically, we compared their contribution with visible indicators of (dis)engagement in learning separately and in combination. Our results showed that facial expressions were more strongly correlated with the manual rating than head pose or gaze alone; however, combining the engagement-related features and combining all three visible indicators improved the correlation with the manual annotations substantially, thus emphasizing the complexity of human rating processes. However, we were not able to replicate the prediction of the knowledge test scores by considering these visible features alone (Research Question 2).

We expected that additional information concerning interaction with peers and similar behavioral aspects would improve the estimated model. Indeed, adding synchrony by considering the engagement patterns of students' neighbors improved the correlations with the manual rating as well as the prediction of the posttest variables (Research Question 3). In line with Raca et al.'s (2013) correlative results, our findings indicated that considering neighbor

synchrony leads to a better understanding of engagement in predictive models. However, the manual ratings were still better at predicting the knowledge test results as well as self-reported cognitive engagement, involvement, and situational interest. Yet, the similarity between the three different manual raters ($r = .74$) differed from the similarity between the manual annotations and the machine-learning approach ($r = .64$). This difference obviously leaves some room for improvement; however, the approximation that was based on visual parameters and the synchrony with a neighbor's behavior appears to provide reliable results. This raises the question of whether human annotators should also include more than just a single person in their ratings and (unconsciously) consider additional information.

Possible Contributions of an Automated Approach for Assessing Engagement

Our machine-learning-based approach provides a promising starting point for reducing the effort involved in manual video inspection and annotation, which in turn would facilitate the analysis of larger numbers of individuals and longer videotaped lessons. In addition, such approaches enable the consideration of more complex information on synchronization across students in a way that goes beyond the ability of human observers. This approach is potentially fruitful for both research and practice.

Information from automated analyses of engagement can be used to provide feedback to teachers and improve their skills in monitoring and identifying relevant cues for students' attention in complex classroom interactions. When teachers can notice and identify a lack of engagement, they have the opportunity to adapt their teaching method accordingly and to encourage the students to deal with the learning content actively. Furthermore, by noticing and identifying distracting behavior, teachers get the chance to react to disruptions and ensure the effective use of instruction time. An automated analysis of videos can support novice teachers in developing professional vision skills, and it can provide feedback to teachers in general about the deep structure of their teaching. By making work with videos less effortful, this method could allow videos to be implemented in teacher training more systematically.

Moreover, the annotation of (dis)engagement in learning over time opens up new opportunities for further investigations of classroom instruction by adding a temporal component. This method allows for the detection of crucial events that accompany similar engagement-related behavior across students and provides deeper insights into different effect mechanisms during instruction. Furthermore, this approach can be combined with additional measures. For example, tracking human raters' eye movements can provide insights into where they retrieve their information and what kinds of visible indicators they actually consider. This

knowledge can further improve machine vision-based approaches by including the corresponding features. In addition, combining valid visible indicators of students' (dis)engagement in learning with eye-tracking data for the teacher, for example, makes it possible to analyze in more detail what kind of visible indicators attract novice teachers' attention (e.g., Sümer et al., 2018). This information can then be reported back in teacher training to support professional vision skills.

Challenges and Limitations

Our study has several notable limitations that need to be addressed in future research. First, face recognition was not possible for all students due to the occlusion of their faces some or most of the time. For this reason, we had to reduce the sample size for the automated analysis, which in turn reduced the statistical power. Limited data was also an issue in the study by Whitehill et al. (2014), who only found moderate correlations between engagement and learning for this reason. It can thus be assumed that increasing the number of participants recognized by face detection would further improve the linear regression models used to predict self-reported learning activities and learning outcomes. The use of mobile eye trackers for each student is an example of one solution that can provide data for individual students. However, the use of eye trackers is expensive, and when used with children who might touch the glasses too often, it deteriorates the gaze calibration and results in an erroneous analysis of attention. Besides, mobile eye trackers can affect the natural behavior of students, whereas field cameras are pervasive and do not create a significant intervention. To overcome the issue of students being occluded, different camera angles could be helpful in future studies.

Second, a challenging aspect of engagement estimation in our setting was the highly imbalanced nature of our data. Engagement levels on both outer ends of our rating scale were underrepresented. As a direct consequence of the learning setting (a teacher-focused session on statistics), few participants displayed active on-task behavior (e.g., explaining content to others); even less data were collected for visible indicators of disengagement in learning indicating active off-task behavior (e.g., walking around with the intention to interrupt). This imbalance has negative implications for the training of algorithms because greater variability in behavior typically leads to more accurate automated analyses. Whereas human raters are familiar with high levels of variance in an audience's on-task and off-task behavior and use this implicit knowledge in their annotation, the algorithms were trained using only the available data from our three sessions. However, this limitation can be overcome by recording real classroom situations, which will be part of our future work. Although it is not possible to control

the intensity of students' (dis)engagement in learning in natural classroom settings, completing more recording sessions and including more participants will eventually lead to a wider distribution of characteristics.

Third, additional research is necessary to validate our approach in schools due to the different target population. This is particularly important because high school students might exhibit a more diverse set of visible indicators of (dis)engagement in learning.

Conclusion

Remote approaches from the field of computer vision have the potential to support research and teacher training. For this to be achieved, valid visible indicators of students' (dis)engagement in learning are needed. The present study provides a promising contribution in this direction and offers a valid starting point for further research in this area.

References

- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system". *Conference on Human Factors in Computing Systems*, 383-390. doi:10.1145/985692.985741
- Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). *OpenFace 2.0: Facial behavior analysis toolkit*. Paper presented at the 13th IEEE International Conference on Automatic Face & Gesture Recognition, Lake Placid, NY.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35, 463–482. doi:https://doi.org/10.1016/S0883-0355(02)00004-6
- Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität. *Diagnostica*, 55(1), 2-10. doi:10.1026/0012-1924.55.1.2
- Bosch, N., D'Mello, S. K., Baker, R. S., Ocumpaugh, J., Shute, V., Ventura, M., . . . Zhao, W. (2016). *Detecting student emotions in computer-enabled classrooms*. Paper presented at the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY.
- Bosch, N., D'Mello, S. K., Ocumpaugh, J., Baker, R. S., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems*, 6(2), 1-26. doi:10.1145/2946837
- Brünken, R., & Seufert, T. (2006). Aufmerksamkeit, Lernen, Lernstrategien. In H. Mandl & H. F. Friedrich (Eds.), *Handbuch Lernstrategien* (pp. 27-37). Göttingen, Germany: Hogrefe.
- Büttner, G., & Schmidt-Atzert, L. (2004). *Diagnostik von Konzentration und Aufmerksamkeit*. Göttingen, Germany: Hogrefe Verlag.
- Carroll, J. (1963). A model for school learning. *Teachers College Record*, 64(723-733).
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. doi:10.1080/00461520.2014.965823
- Connell, J. P. (1990). Context, self, and action: A motivational analysis of self-system processes across the life span. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (Vol. 8, pp. 61-97). Chicago, IL: The University of Chicago Press.
- Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, E. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using

- mobile eye tracking with experts and novice teachers. *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-014-9610-5
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. doi: <https://doi.org/10.1177/0956797613504966>
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92(1), 53-78. doi:<https://doi.org/10.1348/000712601162103>
- Ehrhardt, K. J., Findeisen, P., Marinello, G., & Reinartz-Wenzel, H. (1981). Systematische Verhaltensbeobachtung von Aufmerksamkeit im Unterricht: Zur Prüfung von Objektivität und Zuverlässigkeit [Systematic behavior observation of attention in class: Proof of objectivity and reliability]. *Diagnostica*, 281-294.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, 11(1), 19-23. doi: <https://doi.org/10.1111/1467-8721.00160>
- Erickson, F. (2007). Ways of seeing video: Toward a phenomenology of viewing minimally edited footage. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (2 ed., pp. 145-155). New York, NY: Routledge.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. *Journal of machine learning research*, 9(Aug), 1871-1874.
- Frank, B. (2014). *Presence messen in laborbasierter Forschung mit Mikrowelten: Entwicklung und erste Validierung eines Fragebogens zur Messung von Presence*: Springer-Verlag.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1), 59-109. doi:<https://doi.org/10.3102/00346543074001059>
- Fujii, K., Marian, P., Clark, D., Okamoto, Y., & Rekimoto, J. (2018). *Sync class: Visualization system for in-class student synchronization*. Paper presented at the Proceedings of the 9th Augmented Human International Conference, New York, NY.
- Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary*

-
- Educational Psychology*, 48, 67-84.
doi:<https://doi.org/10.1016/j.cedpsych.2016.09.003>
- Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience*, 8(385). doi:10.3389/fnins.2014.00385
- Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software*, 2(1), e5. doi:10.5334/jors.ar
- Grafsgaard, J., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. (2013). *Automatically recognizing facial expression: Predicting engagement and frustration*. Paper presented at the Educational Data Mining 2013, Memphis, TN.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23-34.
- Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision: KFT 4-12+ R [Cognitive abilities test for 4th to 12th grade students]*. Weinheim, Germany: Beltz-Test.
- Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, 38, 130-141.
- Hommel, M. (2012). Aufmerksamkeitstief in Reflexionsphasen- eine Videoanalyse von Planspielunterricht. *Wirtschaft und Erziehung*, 1-2, 12-18.
- Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal*, 18(2), 157-171.
doi:<https://doi.org/10.3102/00028312018002157>
- Kleinknecht, M., & Gröschner, A. (2016). Fostering preservice teachers' noticing with structured video feedback: Results of an online- and video-based intervention study. *Teacher and Teacher Education*, 59, 45-56. doi:10.1016/j.tate.2016.05.020
- Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts "Pythagoras". In M. Prenzel & L. Allolio-Naecke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (pp. 127-146). Münster: Waxmann.

- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe 1: "Aufgabenkultur" und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Eds.), *TIMSS. Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Videodokumente* (pp. 33-49). Berlin, Germany: Max-Planck-Institut für Bildungsforschung.
- Knogler, M., Harackiewicz, J. M., Gegenfurtner, A., & Lewalter, D. (2015). How situational is situational interest? Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology*, 43, 39-50. doi:<https://doi.org/10.1016/j.cedpsych.2015.08.004>
- Krumpe, T., Scharinger, C., Rosenstiel, W., Gerjets, P., & Spueller, M. (2018). Unity and diversity in working memory load: Evidence for the separability of the executive functions updating and inhibition using machine learning. *Biological psychology*, 139, 163-172. doi: <https://doi.org/10.1016/j.biopsycho.2018.09.008>
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, 44(3), 197-203. doi: <https://doi.org/10.1007/s11251-016-9376-y>
- Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology*, 59(5), 320-324. doi:<http://dx.doi.org/10.1037/h0026223>
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality*, 46(3), 517-528. doi:10.2224/sbp.7054
- Lizdek, I., Sadler, P., Woody, E., Ethier, N., & Malet, G. (2012). Capturing the stream of behavior: A computer-joystick method for coding interpersonal behavior continuously over time. *Social Science Computer Review*, 30(4), 513-521. doi:10.1177/0894439312436487
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality* 74(2), 403-456. doi:10.1111/j.1467-6494.2005.00380.x
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59(1), 14-19. doi:<http://dx.doi.org/10.1037/0003-066X.59.1.14>

-
- McKinney, J. D., Mason, J., Perkinson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology*, 67(2), 198. doi:<http://dx.doi.org/10.1037/h0077012>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, 41(1), 49-100. doi:<https://doi.org/10.1006/cogp.1999.0734>
- Pauli, C., & Lipowsky, F. (2007). Mitmachen oder Zuhören? Mündliche Schülerinnen- und Schülerbeteiligung im Mathematikunterricht. *Unterrichtswissenschaft*, 35(2), 125-147.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom process: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119. doi:10.3102/0013189X09332374
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40. doi:10.1037/0022-0663.82.1.33
- Poh, M., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 57(5), 1243-1252. doi:10.1109/TBME.2009.2038487
- Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), *Master lectures in clinical neuropsychology* (pp. 173-202). Washington DC, WA: American Psychology Association.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*. doi:10.1007/s11858-018-0918-4
- Raca, M. (2015). *Camera-based estimation of student's attention in class*. EPFL, Lausanne, Switzerland.
- Raca, M., & Dillenbourg, P. (2013). *System for assessing classroom attention*. Paper presented at the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium.
- Raca, M., Kidzinski, L., & Dillenbourg, P. (2015). *Translating head motion into attention-towards processing of student's body-language*. Paper presented at the 8th International Conference on Educational Data Mining, Madrid, Spain.

-
- Raca, M., Tormey, R., & Dillenbourg, P. (2013). *Student motion and it's potential as a classroom performance metric*. Paper presented at the 3rd International Workshop on Teaching Analytics (IWTA), Paphos, Cyprus.
- Raca, M., Tormey, R., & Dillenbourg, P. (2014). *Sleepers' lag-study on motion and attention*. Paper presented at the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, IN.
- Rimm-Kaufman, S. E., Baroody, A. E., Larsen, R. A. A., & Curby, T. W. (2015). To what extent do teacher-student interaction quality and student gender contribute to fifth graders' engagement in mathematics learning? *Journal of Educational Psychology*, *107*(1), 170-185. doi:10.1037/a0037252
- 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. doi:10.1177/0022487108328155
- 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. doi:10.1007/s10857-007-9063-7
- Stipek, D. (2002). Good instruction is motivating. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 309-332). San Diego, CA: Academic Press.
- Stürmer, K., Seidel, T., Müller, K., Häusler, J., & Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft*, *20*(1), 75-92. doi:https://doi.org/10.1007/s11618-017-0731-9
- Sümer, Ö., Goldberg, P., Stürmer, K., Seidel, T., Gerjets, P., Trautwein, U., & Kasneci, E. (2018). *Teachers' perception in the classroom*. Paper presented at the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT.
- Trautwein, U., Gerjets, P., & Kasneci, E. (2017). *A cognitive interface for educational improvement: assessing students' attentional focus in the classroom*. University of Tuebingen.
- Whitehill, J., Serpell, Z., Lin, Y., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions.

-
- IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1), 86-98.
doi:10.1109/TAFPC.2014.2316163
- Wolff, C. E., 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, 243-265. doi:10.1007/s11251-016-9367-z
- Yamamoto, T., & Imai-Matsumura, K. (2012). Teachers' gazes and awareness of students' behavior: using an eye tracker. *Innovative Teaching*, 2(6). doi:10.2466/01.IT.2.6.
- Yoshida, R., Nakayama, T., Ogitsu, T., Takemura, H., Mizoguchi, H., Yamaguchi, E., . . . Kusunoki, F. (2014). *Feasibility study on estimating visual attention using electrodermal activity*. Paper presented at the 8th International Conference on Sensing Technology, Liverpool, UK.
- Zaletelj, J. (2017). *Estimation of students' attention in the classroom from kinect features*. Paper presented at the 10th International Symposium on Image and Signal Processing and Analysis (ISPA), Ljubljana, Slovenia.
- Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *EURASIP Journal on Image and Video Processing*, 2017(1), 80. doi:10.1186/s13640-017-0228-8
- Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., & Li, S. Z. (2017). *S3FD: Single shot scale-invariant face detector*. Paper presented at the The IEEE International Conference on Computer Vision (ICCV), Venice, Italy.

4

Study 2

Goldberg, P., Wagner, W., Seidel, T., & Stürmer, K. (2020). *Why do students exhibit different attention-related behavior during instruction? Investigating the effects of individual prerequisites, class membership, and classroom activities*. Manuscript to be submitted.

This version of the manuscript might not exactly replicate the final version published in the journal.
It is not the copy of record.

Abstract

It is well known that students do not always pay attention to the material that is being taught in class, and students' attention-related behaviors differ within and across classrooms. Research has yet to identify which factors can account for variability within individual students' attention-related behavior across a lesson and which factors determine differences between students in their attention-related behavior, as existing instruments cannot provide the necessary data. In the present study, we applied a continuous behavior-rating system to 10 classroom videos and based our analysis on 1,200 data points from a total of $N = 199$ students. We examined how individual characteristics, class membership, and classroom activities affected students' attention-related behavior during classroom instruction. Using dynamic structural equation modeling, students' attention-related behavior was primarily determined by factors that were specific to single classrooms, but within the same classroom, students' attention-related behavior appeared to be affected by their self-concept.

Keywords: attention-related behavior, classroom instruction, intensive longitudinal data, dynamic structural equation modeling

Students' attention is a central aspect of classroom instruction as it determines the success of knowledge construction (see Levine, 1990). Therefore, teachers aim to teach in a way that encourages students to pay attention, for example, by establishing a structure for the classroom that helps avoid disturbances or by engaging students in the desired learning activities by giving them challenging tasks (Hamre et al., 2013; Praetorius et al., 2018). To manage a classroom successfully and to provide an encouraging learning environment, teachers need to monitor their students continuously and identify relevant information (Wolff et al., 2016). Thus, teachers are required to detect observable cues in students' behaviors that indicate whether and how intensely students are paying attention (Goldberg et al., 2019). Previous research has suggested that students exhibit attention-related behavior that can approximate their covert cognitive processes (Chi & Wylie, 2014; Fredricks et al., 2004; Olney et al., 2015), thus allowing teachers to improve their quality of teaching on the basis of their observations of student behavior. However, research has indicated that notwithstanding the quality of teaching methods, students show different attention-related behaviors in the same instructional settings, making it more difficult for teachers to make inferences about students' needs (e.g., Kelly, 2007). These differences in students' attention-related behavior may be explained by students' individual learning prerequisites (e.g., cognitive and motivational-affective characteristics), components that are unique to individual classrooms and their processes (e.g., class culture and interaction dynamics), or certain requirements of classroom activities that encourage different levels of attention (Brühwiler & Blatchford, 2011; Chi & Wylie, 2014).

However, research has yet to examine the extent to which these factors actually affect students' attention-related behavior during instruction. Existing approaches commonly use aggregated measures of students' attention-related behavior that average out situation-specific effects and thus do not consider individual variability in students' momentary attention-related behavior across a lesson (Diener & Larsen, 2009). However, information about why some students demonstrate more variability in their behavior than others can help teachers interpret visible cues from individual students so teachers can adapt their teaching strategies to support all of their students' needs. Furthermore, previous research has commonly focused on active aspects of students' attention-related behavior, such as verbal contributions or hand-raising (Böheim et al., 2020; Kelly, 2008). Such research has usually failed to consider other behaviors that can also indicate whether students are paying attention but are more difficult for a teacher to notice (e.g., just focusing on the teacher or taking notes; Goldberg et al., 2021).

Given that situational aspects of students' attention-related behavior during classroom instruction have not yet been studied intensively, the present study seeks to explore how students' individual prerequisites, class membership, and classroom activities – as determinants to consider in addition to teachers' quality of instruction – can explain differences in variability within individual students' attention-related behavior across a lesson as well as differences between students. Therefore, we based our analysis on video recordings of real classroom lessons, in which students were introduced to a new topic in mathematics (Seidel et al., 2016). In contrast to already existing studies, we used an observation instrument with a continuous coding system that considers the entire spectrum of possible student behavior, ranging from on- to off-task behavior, and allows for continuous annotation throughout the instructional period. This approach allowed us to capture situation-specific variations in students' attention-related behavior to investigate within-student variability as well as between-student differences. We applied Dynamic Structural Equation Modeling (DSEM; Asparouhov et al., 2018) to the resulting multiple time series, a novel approach in the fields of psychology and education science that enabled us to model within-person dynamics for multiple individuals simultaneously while accounting for individual differences on the between-person level.

Observable Aspects of Students' Attentional Processes

Attention is a situation-specific mechanism that filters the ongoing stream of information and directs people's behavior (Cohen, 2014). In general, definitions of attention vary from rather cognitive-oriented (i.e., being attentive) to rather behavior-oriented (i.e., paying attention) interpretations, assigning attention the position of being located between covert cognitive processes and overt behaviors. This relation allows researchers and teachers to use the way students behave during instruction as an indication of the level of attention and engagement they are investing in the learning material.

Certain parts of otherwise covert cognitive processes are observable and can be seen in students' behavior. Posner (1988) showed that when we turn our visual orientation toward a certain stimulus, we improve our internal processing efficiency. Likewise, when a teacher is explaining something and a student is listening carefully, the student will most likely turn toward the teacher to process the respective material better. This connection between external, observable activities and internal, covert processes is also emphasized by the multidimensional construct of student engagement (Fredricks et al., 2004). Fredricks et al. (2004) proposed a three-dimensional structure of student engagement with cognitive, emotional, and behavioral components. The cognitive component covers investment in learning, and the emotional

component includes affective reactions to classroom situations. Whereas these two dimensions entail rather internal processes, the behavioral component is observable. Activities such as effort, asking questions, contributing to class discussions, and paying attention signal learning-related processes that can be recognized when focusing on students' overt behavior (Fredricks et al., 2004). Learning-related processes in turn have been demonstrated to be a decisive factor for academic achievement (Lahaderne, 1968; McKinney et al., 1975). Furthermore, already existing research has found that academic achievement is correlated with students' attention-related behavior (Helmke & Renkl, 1992; Hommel, 2012a; Karweit & Slavin, 1981; Stipek, 2002) as well as their behavioral engagement (Lei et al., 2018).

However, students' behavior can provide indications not only about the absence or presence of attention but also about the intensity with which students invest their attention. In their ICAP framework, Chi and Wylie (2014) proposed that different observable learning activities (passive, active, constructive, interactive) can signal different covert cognitive processes (receiving, manipulating, generating, dialoguing). On the basis of this assumption, Olney et al. (2015) argued that students' learning improves within this learning task structure as the quality of attention increases from passive to interactive learning activities. Conversely, when students cannot sustain the focus of their attention, they will exhibit different behaviors that result in a less elaborate and more superficial understanding (D'Mello, 2016). Consequently, also considering non-content-related behavior can improve estimations of students' internal processes on the basis of their overt behavior.

Nevertheless, it should be noted that students' behavior can serve only as an approximation of their covert cognitive processes, wherefore we will subsequently use the term *attention-related behavior* to refer to behavior that is associated with students' engagement and attentional processes.

Differences in Students' Attention-Related Behavior

Students have a broad spectrum of behavior to choose from, as their behavior can be content-related or not and can range from passive to interactive. Students exhibit different attention-related behavior within one lesson but also compared with other students. These differences can be attributed to variations in individual characteristics but also to the learning environment (Vygotsky, 1978). Within instructional settings, multiple factors that influence students' actions come into play (Baumert et al., 2002; Brühwiler & Blatchford, 2011; Helmke, 2009; Seidel, 2014). For example, Helmke and Renkl (1992) found that 56% of individual differences in attention-related behavior could be attributed to differences between classes.

They inferred that there must be factors that are unique to single classrooms and affect individual students' behavior. Social structures and teachers' socialization of participation have actually been shown to influence students' verbal participation in classroom discussions (Clarke et al., 2016).

Furthermore, teachers can create certain classroom settings that can support on-task behavior, but they might also create occasions for students to get distracted or become disruptive by their choice of classroom activities (e.g., classroom discussion or individual seatwork; Beyda et al., 2002). While some classroom activities, for example, require students to just passively listen to the teacher when giving a presentation, other activities such as classroom discussions request students to interact with each other and argue for a position (see Chi & Wylie, 2014). Beyda et al. (2002) discovered that students show more on-task behavior in settings in which they can interact with peers with minimal interruptions from the teacher compared with teacher-centered settings or phases of individual seatwork. Furthermore, early research found that students spent more time showing attention-related behavior during teacher-directed instruction compared with settings involving seatwork (Friedman et al., 1988).

Factors such as determinants that are unique to individual classrooms (e.g., interactions, social structures) or requirements of different classroom activities (e.g., classroom discussions, individual seatwork) have been assumed to affect all students and their attention-related behavior similarly in one classroom and thus might explain differences between students of distinct classes notwithstanding teachers' quality of teaching. However, previous research suggests not only that there seem to be differences between classes due to their unique character, but also that there are individual differences between students in behaviors shown. It appears that some students tend to make use only of a small subset of behaviors, whereas other students exhibit a wider range of behaviors during instruction. Kelly (2007) found a high level of variability across students in their classroom participation independent of teachers' dialogic practices (i.e., the quality of questions asked by the teacher, e.g., whether they required reiteration vs. deep thinking). Previous research has demonstrated that it is important to consider students' motivational-affective characteristics along with individual cognitive prerequisites regarding students' engagement in classrooms (Jurik et al., 2013; Turner & Patrick, 2004). Motivational-affective as well as cognitive prerequisites influence how students participate and engage in classroom instruction. Previous research has demonstrated that cognitive prerequisites, such as prior abilities, positively affect the number of verbal contributions students make (e.g., Kelly, 2008). Lau and Roeser (2002) found that motivational-

affective processes, such as students' self-concept of ability and task values (interest, perceived importance, and perceived utility) predicted students' self-reported attention and degree of participation in classroom experiments. Students who are more confident in their competence in a subject tend to raise their hand and participate more often in classroom discussions than their more doubtful peers (Böheim et al., 2020).

However, these studies focused on specific aspects of students' attention-related behavior (hand-raising; Böheim et al., 2020; verbal contribution; Kelly, 2008; degree of participation; Lau & Roeser, 2002) and did not consider subtle behaviors that also indicate a certain degree of attention. Students can exhibit passive but also rather interactive attention-related behaviors whereby passive behaviors are more difficult for teachers to notice (Goldberg et al., 2021). For example, students can signal that they are paying attention by raising their hands and verbally participating in classroom discussion. However, students can also follow the classroom discussion closely when acting passively and just focusing on the center of instruction. If the range of students' attention-related behavior is not considered entirely, no statements can be made about differences between students in their individual variability of attention-related behavior in the course of instruction. Therefore, it is important to decide carefully how to measure students' attention-related behavior and which aspects to consider.

Measuring Students' Attention-Related Behavior

Students' self-reports and external observer ratings are two popular assessment methods for measuring students' attention-related behavior during instruction. However, self-reports that ask about specific behaviors during instruction can be biased due to response tendencies or trouble with retrospective recall (e.g., when the self-report questionnaire is administered subsequent to a school lesson; see Fredricks & McColskey, 2012). In turn, systematic external behavioral observations can provide an objective assessment of students' behavior without instrumental effort (Helmke & Renkl, 1992).

The external observation of attention-related behavior is not a new idea. Beginning with direct observations in the classroom, Jackson and Hudgins (1965), Cobb (1972), Karweit and Slavin (1981), Helmke and Renkl (1992), and Lane and Harris (2015) developed approaches that can be applied to assess students' behavior during instruction. All of these approaches have in common that they are based on time-sampling and classify students' behavior into distinct categories. Technological advances made in recent decades offer new opportunities such as video analysis in contrast to classroom observations. Video analysis allows researchers to observe all students in one classroom for the same instructional period, instead of using time-

sampling of only a few students (Mayring et al., 2005). Additionally, as already mentioned above, students may exhibit a broad behavioral spectrum during instruction, which has to be taken into account. For example, when measuring only active participation (e.g., verbal contribution; Lipowsky et al., 2007), important information provided by the type, amount, and possible effect of off-task behavior gets lost. Thus, comprehensive behavioral definitions are crucial for drawing conclusions from specific results.

Existing approaches and studies commonly use mean values of students' attention-related behavior to make inferences about the relationship with individual outcomes (e.g., Cobb, 1972; Helmke & Renkl, 1992; Samuels & Turnure, 1974) or classroom activities (e.g., Beyda et al., 2002; Hommel, 2012b). However, Diener and Larsen (2009) argued that aggregating data for individuals can average out situational effects. In their research on the temporal stability and situational consistency of behavioral responses, they found short-term variations in individuals' responses in addition to long-term trends in mean-level responses (Diener & Larsen, 2009). This pattern can also be attributed to students' attention-related behavior during instruction. Students differ in terms of their average level of attention in a lesson (i.e., in the stable parts of their attention reflected by their mean levels) as well as their individual variability in momentary attention levels across a lesson (i.e., the variable part of each student's attention reflected by within-student dispersion). Imagine a student who is constantly focused on the teacher and is always taking notes but is not actively participating in classroom discussions. This student has a high attentional mean level with rather small deviations. By contrast, another student might be engaged in mind wandering, taking notes only occasionally, and disrupting others. This student would exhibit a rather low attentional mean level with striking deviations. When comparing different students with each other, students' mean level of attention-related behavior would be considered. However, by aggregating the measurements of students' behavior, situational effects on students' dispersion and mean-levels are averaged out.

To provide a data structure that offers researchers a way to study differences in students' attention-related behavior between students but also within students, it is necessary to apply a continuous measurement approach that can detect momentary changes in student behavior and allows suitable statistical models to be deployed. However, current approaches for investigating students' attention-related behavior tend to rely on categorical coding systems with time-samplings, do not account for the entire behavioral spectrum that students can exhibit during instruction, or aggregate the data to calculate percentage distribution scores.

Research Questions

Research has yet to examine which factors can account for individual within-student variability in attention-related behavior across a lesson and which factors determine differences between students' attention-related behavior, as existing instruments do not provide the necessary data structure. In the present study, we therefore deployed a continuous behavior-rating system, which continuously codes students' behavior on a scale that covers on- and off-task behavior with specifications that range from passive to interactive throughout the instructional session. We coded students' attention-related behavior in 10 school classes by re-analyzing video material from a study by Seidel et al. (2016). The resulting intensive longitudinal data structure for multiple people is still rather exceptional for research on education, and the respective methodology has not yet been extensively deployed.

As outlined above, multiple factors can determine how students behave during instruction (e.g., Brühwiler & Blatchford, 2011). When assuming that the degree to which teachers are able to implement high quality instruction is comparable between classes the questions that arise are how much students still differ within a lesson and between each other in their attention-related behavior and how much of the variation certain determinants can account for. Individual student characteristics can explain differences between students in their attention-related behavior during instruction. Factors such as the unique dynamics of a class and teacher-student relationships are further assumed to explain differences in students' attention-related behavior between different classes. Additionally, research has demonstrated that students show distinct attention-related behavior depending on classroom activities. By considering 1,200 data points for a total of $N = 199$ students, we aimed to examine how individual prerequisites, class membership (as representation for the uniqueness of single classrooms), and classroom activities can explain differences in individual within-student variability in attention-related behavior across a lesson as well as differences between students when teachers' quality of instruction is comparable between classrooms. We addressed the following research questions:

- 1) Do students' individual prerequisites (cognitive abilities, prior knowledge, interest, and self-concept), class membership, and classroom activities affect the within-student variability in attention-related behavior across one school lesson?

On the basis of the assumptions described above, we expected students' individual prerequisites to account for the extent to which individual students varied in their attention-related behavior during one school lesson. Additionally, we expected that components that are

unique to individual classrooms as well as the demands of certain classroom activities would affect the variability in individual students' behavior.

- 2) Do students' individual prerequisites (cognitive abilities, prior knowledge, interest, and self-concept), class membership, and classroom activities account for between-student differences in their attention-related behavior during instruction?

On the basis of previous findings, we expected that class membership with all determinants that are unique to individual classrooms would account for differences between students in their attention-related behavior. Furthermore, the relative proportions of certain classroom activities during one school lesson should explain why students behave differently. Finally, we examined the extent to which individual students' prerequisites could account for differences between students in their attention-related behavior.

Methods

Sample and Procedure. The present study is based on video material that was collected within the scope of a study by Seidel et al. (2016), which focused on teacher-student interactions and was registered with the Bavarian State Ministry for Education and Cultural Affairs. Seidel et al. (2016) collected videos of twenty eighth-grade school classes ($N = 503$ students) of eighteen German high schools during mathematics lessons. To ensure comparability, all videos show introductory lessons on either functional relations (e.g., proportionality) or the intercept theorem and similarity. Next to the video material, the dataset includes information about students' individual prerequisites (prior knowledge from mathgrade the year before; subject-specific interest; Baumert et al., 1997; cognitive abilities; Heller & Perleth, 2000; self-concept; OECD, 2014) as well as student-reported instructional quality (Waldis et al., 2002). Therefore, the video material provides an ideal base to carefully choose those classroom lessons that allow investigating our research questions.

We had to exclude two classes beforehand as data on students' prerequisites was missing, leaving $N = 18$ classes. We excluded one additional class as students' ratings of teachers' overall instructional quality have been conspicuous compared with the full sample. Additionally, we checked the quality of the video material and the visibility of the students in the video recordings, as this constitutes an important prerequisite to deploy a continuous behaviour observation rating. We had to exclude seven additional videos, as either the audio quality was insufficient or light and/or seating conditions did not allow to observe the majority of students in sufficient detail including those students in the back of the classroom.

In the remaining ten videos, we identified teacher-centered parts of the videotaped school lessons, as in those situations it is rather unambiguous where students' attentional focus is supposed to be. Either students should focus on the teacher or individual peers during classroom discussion, or the teacher gives clear statements about students' activities. In comparison, activities such as group work are characterized by rather unstructured and individual interactions. We excluded long periods where students only had to write book entries or did individual exercises and only considered the periods in which the teacher was supposed to be in the center of students' attention. To provide enough data, we chose video sections with teacher-centered classroom activities, so that the resulting video clips were approximately 20 minutes long.

In total, we analyzed 10 videos with $N = 209$ students (52.2% female, 46.4% male). Due to occlusions that made the observer ratings inconclusive, we had to exclude nine students and one additional student because he or she did not give consent to be analyzed in the videos, so that we based our analyses on $N = 199$ students. The descriptive statistics of students' individual prerequisites per class are depicted in Table 1. The teachers' mean age was 39.5 years ($SD = 8.11$) and they had 11 years ($SD = 7.79$) teaching experience on average. The teachers in our subsample were comparable to other participating mathematics teachers in gender distribution (subsample: 50% female; other: 60% female), age ($t = -0.52, p = .609$), and teaching experience ($t = -0.23, p = .820$). The descriptive statistics of the individual classes are presented in Table 2.

Table 3

Descriptive Statistics of Students Gender Distribution, Means (Standard Deviations) in Their Cognitive Abilities, Prior Knowledge, Interest, Self-concept, and Behavior Ratings per Class.

Class	N	Gender distribution		Cognitive abilities	Prior knowledge	Interest	Self-concept	Behavior rating
		n female	n male	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
1	25	19	6	18.58 (5.04)	2.56 (0.82)	1.93 (0.64)	2.26 (0.56)	0.78 (0.38)
2	16	8	8	15.27 (6.50)	2.94 (1.18)	1.86 (0.85)	2.44 (0.86)	0.70 (0.38)
3	17	10	7	19.00 (4.11)	2.53 (1.42)	2.54 (0.70)	2.75 (0.78)	0.68 (0.38)
4	19	11	8	19.89 (3.85)	2.58 (0.96)	2.47 (0.71)	2.72 (0.88)	0.73 (0.33)
5	23	9	14	18.14 (5.69)	2.57 (1.08)	2.38 (0.62)	2.60 (0.70)	0.63 (0.38)
6	19	11	8	16.95 (5.37)	2.79 (1.03)	2.47 (0.80)	2.73 (0.95)	0.63 (0.33)
7	17	6	10	17.20 (3.43)	2.50 (1.03)	2.20 (0.69)	2.44 (0.77)	0.58 (0.32)
8	29	6	23	15.42 (5.41)	3.34 (1.14)	2.32 (0.74)	2.72 (0.91)	0.60 (0.46)
9	22	22		17.48 (4.57)	1.91 (0.97)	1.49 (0.62)	1.69 (0.68)	0.68 (0.32)
10	22	7	13	17.83 (5.00)	2.90 (1.04)	2.29 (0.77)	2.74 (0.79)	0.62 (0.24)

Note: Cognitive abilities: 25 items. Prior knowledge: school grade from 1 to 5 (1 being the best grade). Interest: 4 items on Likert-scale with 1 = “fully applies” to 4 = “does not apply”. Self-concept: 5 items on Likert-scale with 1 = “fully agree” to 4 = “completely disagree”.

Table 4

Descriptive Statistics of Classroom Characteristics as in their Session Topic, Relative Frequency of Classroom Activities, and Students' Overall Teaching Quality Ratings.

Class	Session topic	Classroom activities (%)							Student rating of overall TQ <i>M (SD)</i>
		Teacher input	Classroom discussion	Collective exercise	Individual seatwork	Task evaluation	Book entry	Other	
1	similarity	5.76	8.77	13.73	10.64	17.55	43.54		2.62 (0.52)
2	rational functions		95.78				4.22		2.70 (0.33)
3	integer exponents		48.40	12.33	3.63		35.64		2.82 (0.36)
4	uniform scaling		26.14		0.96	43.89	29.02		2.70 (0.77)
5	uniform scaling		63.29		11.62	25.08			2.71 (0.47)
6	similarity		18.01	52.23			17.41	12.35	2.56 (0.41)
7	proportionality	3.85	30.71		12.78	26.45	7.53	18.67	2.48 (0.48)
8	rational functions		52.71		1.33		45.96		2.27 (0.62)
9	proportionality	2.90	61.89		1.41	0.99	32.81		2.32 (0.40)
10	intercept theorem		37.88	24.24	4.83	29.49	3.56		2.63 (0.39)

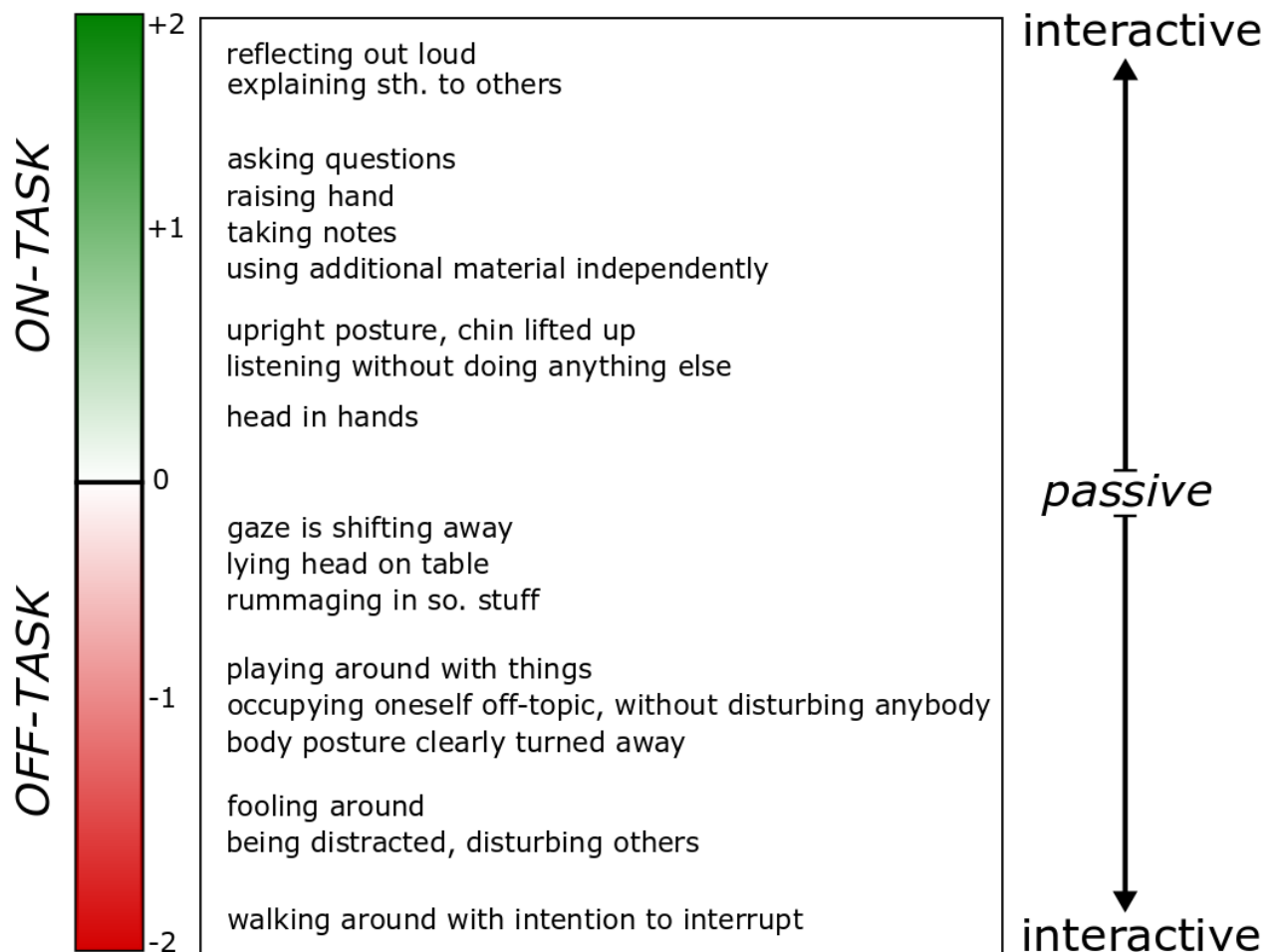
Note: TQ = Teaching Quality, sub dimensions: Socratic approach (5 items), learning techniques (5 items), error culture (4 items), cooperation and discussion (6 items), orientation towards understanding (2 items) on Likert-scale with 1 = “does not apply” to 4 = “fully applies”.

Behavior annotation. We manually annotated students' observable behavior on a one-dimensional scale along the video length in steps of seconds. This was done by using the free software CARMA (Girard, 2014), which enabled us to annotate intrapersonal behavior continuously with joysticks. We used a continuous coding system that has been demonstrated to capture students' behavior that is related to their attention in a valid way (Goldberg et al., 2019). This coding systems considers the whole spectrum of possible student behavior during instruction by combining already existing instruments and including the theoretical considerations of Chi and Wylie (2014). Behavior was annotated on a symmetric scale ranging from -2, indicating disturbing (i.e., interactive), off-task behavior, such as shouting across or walking around in the classroom, to +2, indicating highly engaged, interactive, on-task behavior in which, for example, learners ask questions and try to explain content to fellow learners (see Figure 1). Values closer to 0 indicated rather unobtrusive, passive behavior in which, for example, learners listened without participating (on-task) or rummaged through their belongings (off-task; Goldberg et al., 2019). In total three raters annotated students' behavior in random order so that each student was annotated by two raters. Inter-rater reliability was $ICC(1,3) = .780^3$ (consistency) on average (we report the average value here, as we had to calculate the average score agreement for each student separately; ICC values ranged between .484 and .946). The mean between those two raters was used for subsequent analysis and values between -0.1 and +0.1 were coded as missing values, as this interval was defined to be chosen in case a behavior rating was not possible (e.g., due to occlusions or student not being in the video).

³ We used Fisher's z-transformation before calculating the mean across all ICCs and re-transformed the values subsequently. ICC is considered poor for values < .40, fair for values between .40 and .59, good for values between .60 and .74, and excellent for values between .75 and 1.0. See Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>

Figure 6

Scale for Behavior Annotation with Exemplary Behavioral Indicators



Note: From “Attentive or not? Toward a machine learning approach to assessing students’ visible engagement in classroom instruction“ by P. Goldberg, Ö. Sümer, K. Stürmer, W. Wagner, R. Göllner, P. Gerjets, E. Kasneci, & U. Trautwein, 2019, *Educational Psychology Review*. CC BY.

Coding of classroom activities. Different classroom activities encourage different kinds of students’ attention-related behavior. For example, during phases of teacher input, students are supposed to listen whereas students should contribute verbally during classroom discussions. To consider differences in students’ behavior due to the requirements of specific classroom activities, we applied an event rating to all videos. Two raters coded the videos according to the category system displayed in Table 3 with an inter-rater reliability of $\kappa = .92$ (almost perfect agreement; Landis & Koch, 1977). The categories of the coding system were mutually exclusive, so that only one classroom activity could be coded at a time. We used the

relative proportions of classroom activities as they occurred within the 20 min video sequences for subsequent analyses.

Table 5

Categorical Coding System of Classroom Activities.

Category	Definition
Teacher input	Content is presented as oral presentation Commonly supplemented by visual material Usually at start of lesson
Classroom discussion	Teacher poses question or task Elaborating of underlying issues through discussion Teacher takes function of moderator Question-answer style
Collective exercise	Teacher gives task to all students Collective processing Can include book entry Direct feedback of teacher
Individual seatwork	Teacher give task instruction Individual processing Teacher walks around and gives individual support if necessary
Task evaluation	Evaluating the results of the individual seatwork Comparison/presentation of single answers
Book entry	Collective book entry Teacher is dictating or writing on black board Developing the book entry with alternating phases of discussion and writing
Other	Not relatable Conversion phase

Analysis. For subsequent analysis, we standardized the time-invariant covariates (cognitive abilities, prior knowledge, interest, and self-concept) to allow interpretation relative to the overall sample. As we annotated student behavior in steps of seconds, our dataset consisted of $N = 199$ individual time series with $1,166 \leq T \leq 1,251$ ($M = 1,207$) data points (i.e., seconds) on average ($SD = 26.62$). To consider the two-level structure of our data (i.e., behavior rating every second nested within students), multilevel location-scale autoregressive models were constructed within the Dynamic Structural Equation Modeling Framework (DSEM) in Mplus Version 8. By extending previous work (Asparouhov & Muthén, 2016), Asparouhov et

al. (2018) deployed the idea of time-series analysis to datasets featuring multiple persons and by this enable modeling the lagged relation of a large number of repeated measures (time-series analysis) of multiple individuals simultaneously while allowing for individual differences (multilevel modeling) with latent variables (structural equation modeling). We used an autoregressive Lag-1 for the behavior rating – denoted by AR(1) – which means that we predicted the rating value at time t by the rating value at the immediate preceding measurement occasion (i.e., one second earlier). Latent person-mean centering was used to allow interpretation of the predictor variable at the within-person level (i.e., Lag-1 rating) in a relative fashion for each student (*Model 1*). At the between level we considered cognitive abilities, prior knowledge, interest, and self-concept (*Model 2*), class membership (by using dummy variables; *Model 3*), and relative proportions of classroom activities (*Model 4*) as time-invariant covariates separately, or in combination (*Models 5-7*). We spread the rating scale by 100 enabling us to present the results with sufficient precision, as the results showed that the variance of the logarithmized within-student variances was numerically quite small.

In all models, we used Bayesian estimation based on Markov Chain Monte Carlo using Gibbs sampling. Point estimates were obtained by taking the median of the posterior distributions for each parameter. Statistical significance was determined by estimating a 95% credibility interval (*CI*) around each point estimate. Each model was specified using latent random intercepts, autoregressive slopes, and residual variance and was estimated using 10,000 iterations and two chains with a thinning factor of 10 iterations (meaning that only 1 in 10 iterations is saved). Model convergence was ensured by checking that the potential scale reduction was close to 1 and trace plots did not contain trends, spikes, or other irregularities. All data points were equidistant so there was no need to control for time (McNeish & Hamaker, 2019).

Results

The model estimates are provided in Tables 4 and 5. For the basic multilevel location-scale AR(1) model (*Model 1*), the average horizontal mean line across all students around which the behavior rating varies is 66.64 with $SD = 7.27$. This means that on average, students exhibited behavior that was rated at 0.67 (i.e., active on-task) with $SD = 0.07$ on the behavior scale⁴ and that some students had consistently higher or lower levels of attention-related behavior than the average. The expected person-specific residual variance was $\exp(4.36) = 78.26$ ($SD = 8.85$), meaning that “typical” students varied in their attention-related behavior for

⁴ Please note that the rating scale was spread by factor 100

0.09 points on average (95% Confidence Interval [0.04, 0.18]) around their horizontal mean line on the observation scale. The average autoregressive coefficient across all students was .96, meaning there was a strong carry-over effect from one second to the next.

With regard to students' within-person variance in their attention-related behavior, results do not show any effects of students' cognitive abilities, prior knowledge, interest, and self-concept (*Model 2*). We report the separate regressions results of the single individual prerequisites on students' mean attention-related behavior in Appendix Table A1. When considering class membership as dummy variable on the between-level (*Model 3*), we found that in some classes students had higher within-person variability than in others. Additionally, students exhibited different mean levels of attention-related behavior depending on their class membership. We found that students' behavior ratings were more difficult to predict the higher the proportions of classroom discussions, collective exercises, individual seatwork, book entries, and activities that fell in the category "Other" due to higher within-person variances (*Model 4*). Additionally, students demonstrated different mean level of attention-related behavior depending on the proportional distribution of classroom activities. When including individual prerequisites and class membership together on the between-level in *Model 5*, we found that a change in self-concept by one unit predicted a change in residual variance by the factor $\exp(0.158) = 1.17$. This means that students with higher values in their self-concept showed more within-person variance making behavior ratings for those students more difficult to predict. We found a similar pattern when including individual prerequisites and the relative proportions of classroom activities together on the between-level (*Model 6*): the higher students' self-concept the more within-person variance in students' attention-related behavior (a one unit increase in self-concept predicted a change in residual variance by factor 1.18). When including class membership and the relative proportions of classroom activities together on the between-level in *Model 7*, we did not find any effects on students' within-person variances.

With regard to differences between individual students, our results showed that between-level variances decreased in *Model 2* compared with the estimates of *Model 1*. This is because part of the random variation from person to person captured by the between level variance has been explained by the time-invariant covariates (i.e., individual prerequisites). The intercept variance dropped from 52.84 to 48.77, meaning that cognitive abilities, prior knowledge, interest, and self-concept accounted for 7.7% of the variability in students' means. Similarly, in *Model 3* between-level variances decreased compared with the estimates of *Model*

1. The intercept variance dropped from 52.84 to 6.59, meaning that class membership accounted for 87.52% of the variability between students attention-related behavior. When considering the relative proportions of classroom activities as time-invariant covariates on the between-level (*Model 4*), compared with *Model 1* the intercept variance dropped from 52.84 to 19.26, meaning that the relative proportions of classroom activities accounted for 61.16% of the variability between students attention-related behavior. Compared with individual prerequisites alone (*Model 2*), class membership explained additional 85.84% (*Model 5*) and the relative proportions of classroom activities additional 59.23% (*Model 6*) of between-person variances. When including classroom membership together with the relative proportions of classroom activities on the between-level (*Model 7*), we found less effects of individual classes and classroom activities on students' mean-levels of attention-related behavior. Intercept variance dropped from 19.883 to 7.107, meaning that class membership accounted for additional 64.27% of the variability in students' means compared with the relative proportion of classroom activities alone (*Model 4*).

Table 6

Estimates and 95% Credible Intervals for Multilevel Location-Scale AR(1) Model (Model 1) with Individual Prerequisites (Model 2), Class Membership (Model 3), and Relative Proportion of Classroom Activities (Model 4) as Time-Invariant Covariates Based on N= 199 Students in Ten Different Classes.

Effect		Model 1		Model 2		Model 3		Model 4	
		Estimate	95% Credible Interval	Estimate	95% Credible Interval	Estimate	95% Credible Interval	Estimate	95% Credible Interval
Between-level									
Intercepts	Rating	66.641*	[65.202, 68.083]	66.267*	[64.843, 67.692]	61.649*	[59.519, 63.797]	19.255*	[4.024, 34.129]
	phi	0.959*	[0.956, 0.962]	0.959*	[0.956, 0.962]	0.937*	[0.929, 0.944]	0.282*	[0.242, 0.323]
	logv	4.360*	[4.261, 4.457]	4.360*	[4.261, 4.455]	4.189*	[3.896, 4.476]	1.280	[-0.617, 3.121]
Rating on	Cog. Abilities			0.952	[-0.501, 2.413]				
	Prior Knowledge			1.150	[-0.673, 3.005]				
	Interest			0.127	[-1.764, 2.089]				
	Self-Concept			0.078	[-2.224, 2.287]				
	class_1					18.059*	[14.324, 21.577]		
	class_2					8.745*	[4.314, 12.932]		
	class_3					7.332*	[2.630, 12.211]		
	class_4					8.915*	[4.933, 12.939]		
	class_5					-1.332	[-5.066, 2.509]		
	class_6					1.264	[-3.181, 5.753]		
	class_7					-3.331	[-6.847, 0.128]		
	class_8					3.576	[-0.908, 7.828]		
	class_9					6.895*	[3.116, 10.563]		
	Event1							86.019*	[50.419, 122.236]
	Event2							41.346*	[26.273, 56.934]
	Event3							39.193*	[18.663, 59.090]
	Event4							65.361*	[33.457, 97.903]
Event5							45.620*	[28.800, 63.347]	
Event6							62.595*	[45.715, 78.034]	
Event7							-16.079	[-47.518, 14.502]	
logv on	Cog. Abilities			0.036	[-0.068, 0.139]				
	Prior Knowledge			-0.014	[-0.136, 0.109]				
	Interest			-0.102	[-0.231, 0.025]				
	Self-Concept			0.148	[-0.002, 0.298]				
	class_1					0.348	[-0.031, 0.720]		

	class_2			0.605*	[0.172, 1.035]		
	class_3			0.385	[-0.035, 0.802]		
	class_4			-0.144	[-0.555, 0.266]		
	class_5			0.044	[-0.352, 0.444]		
	class_6			-0.425*	[-0.826, -0.021]		
	class_7			0.383	[-0.033, 0.805]		
	class_8			0.522*	[0.146, 0.896]		
	class_9			-0.072	[-0.474, 0.329]		
	Event1					-2.218	[-5.291, 0.828]
	Event2					3.154*	[1.301, 5.041]
	Event3					3.520*	[1.274, 5.747]
	Event4					5.146*	[2.240, 8.185]
	Event5					1.916	[-0.067, 3.874]
	Event6					3.605*	[1.693, 5.518]
	Event7					5.167*	[2.187, 8.099]
phi on	Cog. Abilities	0.000	[-0.004, 0.003]				
	Prior Knowledge	- 0.001	[-0.004, 0.003]				
	Interest	0.001	[-0.003, 0.005]				
	Self-Concept	-0.001	[-0.005, 0.004]				
	class_1			0.023*	[0.014, 0.032]		
	class_2			0.016*	[0.005, 0.027]		
	class_3			0.026*	[0.016, 0.037]		
	class_4			0.030*	[0.020, 0.040]		
	class_5			0.027*	[0.017, 0.037]		
	class_6			0.039*	[0.029, 0.049]		
class_7			0.008	[-0.002, 0.019]			
class_8			0.030*	[0.020, 0.039]			
class_9			0.028*	[0.018, 0.038]			
	Event1					0.820*	[0.752, 0.889]
	Event2					0.674*	[0.632, 0.715]
	Event3					0.587*	[0.537, 0.636]
	Event4					0.686*	[0.619, 0.752]
	Event5					0.680*	[0.637, 0.723]
	Event6					0.696*	[0.653, 0.737]
	Event7					0.570*	[0.504, 0.638]

	Effect	Model 1		Model 2		Model 3		Model 4	
		Estimate	95% Credible Interval	Estimate	95% Credible Interval	Estimate	95% Credible Interval	Estimate	95% Credible Interval
Residual	Rating	52.840*	[35.643, 76.402]	48.770*	[44.984, 132.316]	6.593*	[0.875, 17.787]	20.521*	[8.607, 37.935]
Variances	phi	0.000*	[0.000, 0.000]	0.000*	[0.000, 0.000]	0.000*	[0.000, 0.000]	0.000*	[0.000, 0.000]
	logv	0.490*	[0.402, 0.603]	0.485*	[0.346, 0.627]	0.408*	[0.334, 0.505]	0.432*	[0.355, 0.536]
Inter-correlations	phi - logv	-0.006*	[-0.008, -0.003]	-0.006*	[-0.008, -0.004]	-0.004*	[-0.006, -0.003]	-0.004*	[-0.006, -0.003]
	phi - Rating	0.006	[-0.021, 0.033]	-0.007	[-0.019, 0.033]	0.001	[-0.016, 0.017]	-0.009	[-0.031, 0.012]
	logv - Rating	1.064	[-0.038, 2.231]	0.926	[-0.152, 2.061]	0.605	[-0.153, 1.394]	1.234*	[0.391, 2.168]
DIC			1,729,422.596		1,747,802.423		1,729,324.898		1,729,358.724
(pD)			(5,732.098)		(7,620.384)		(5,611.870)		(5,637.092)

Note: AR(1) = Autoregressive Lag-1; phi = Lag-1 slope; logv = log residual variance; DIC = Deviance Information Criterion; pD = Estimated Number of Parameters; * indicates significance.

Table 7

Estimates and 95% Credible Intervals for Multilevel Location-Scale AR(1) Models with Individual Prerequisites and Class Membership (Model 5), Individual Prerequisites and Relative Proportion of Classroom Activities (Model 6), and Class Membership and Relative Proportion of Classroom Activities (Model 7) as Time-Invariant Covariates Based on N= 199 Students in Ten Different Classes.

Effect		Model 5		Model 6		Model 7	
		<i>Estimate</i>	<i>95% Credible Interval</i>	<i>Estimate</i>	<i>95% Credible Interval</i>	<i>Estimate</i>	<i>95% Credible Interval</i>
Intercepts	Rating	61.096*	[58.875, 63.289]	19.010*	[3.817, 34.662]	13.933*	[2.552, 24.124]
	phi	0.937*	[0.930, 0.944]	0.282*	[0.244, 0.321]	0.206*	[0.161, 0.245]
	logv	4.179*	[3.891, 4.465]	1.301	[-0.622, 3.212]	0.928	[-0.971, 2.836]
Rating on	Cog. Abilities	0.301	[-0.752, 1.339]	0.394	[-0.808, 1.586]		
	Prior Knowledge	0.800	[-0.588, 2.215]	1.066	[-0.438, 2.605]		
	Interest	1.143	[-0.350, 2.604]	0.241	[-1.324, 1.851]		
	Self-Concept	0.659	[-0.982, 2.342]	0.406	[-1.465, 2.206]		
	class_1	18.351*	[14.589, 22.234]			11.866	[-4.750, 28.674]
	class_2	9.059*	[4.576, 13.540]			27.704*	[8.738, 44.260]
	class_3	5.979*	[1.156, 10.746]			13.334	[-0.634, 26.447]
	class_4	8.207*	[4.188, 12.250]			18.095	[-0.898, 36.663]
	class_5	-1.421	[-5.448, 2.723]			7.214	[-10.226, 25.302]
	class_6	1.589	[-2.971, 6.391]			9.117	[-16.677, 35.086]
	class_7	-2.576	[-6.184, 1.126]			1.333	[-24.729, 25.800]
	class_8	2.610	[-1.788, 6.905]			15.958	[-1.550, 31.032]
	class_9	9.644*	[5.574, 13.738]			19.557*	[4.165, 32.591]
	Event1			89.038*	[53.402, 125.367]	80.140	[-60.545, 220.493]
	Event2			41.653*	[25.919, 57.205]	26.696*	[5.790, 43.996]
	Event3			37.282*	[16.533, 57.405]	78.062*	[13.708, 132.432]
	Event4			66.986*	[34.820, 99.601]	97.551	[-34.015, 230.003]
Event5			45.079*	[27.675, 62.933]	40.117	[-4.363, 80.265]	
Event6			61.232*	[44.072, 77.037]	41.006*	[6.302, 70.726]	
Event7			-15.693	[-48.675, 14.922]	26.370	[-102.218, 157.897]	

	Effect	Model 5		Model 6		Model 7	
		Estimate	95% Credible Interval	Estimate	95% Credible Interval	Estimate	95% Credible Interval
logv on	Cog. Abilities	0.081	[-0.017, 0.178]	0.056	[-0.040, 0.153]		
	Prior Knowledge	-0.086	[-0.205, 0.032]	-0.041	[-0.155, 0.072]		
	Interest	-0.057	[-0.177, 0.069]	-0.084	[-0.206, 0.038]		
	Self-Concept	0.158*	[0.021, 0.296]	0.165*	[0.027, 0.310]		
	class_1	0.361	[-0.023, 0.750]			0.203	[-2.882, 3.308]
	class_2	0.661*	[0.237, 1.083]			1.905	[-1.279, 5.060]
	class_3	0.337	[-0.080, 0.763]			0.905	[-1.510, 3.255]
	class_4	-0.194	[-0.604, 0.207]			0.706	[-2.918, 4.272]
	class_5	0.037	[-0.361, 0.431]			0.631	[-2.642, 3.924]
	class_6	-0.423*	[-0.821, -0.015]			0.405	[-4.396, 5.169]
	class_7	0.399	[-0.012, 0.821]			0.598	[-4.117, 5.281]
	class_8	0.585*	[0.216, 0.971]			1.552	[-1.271, 4.356]
	class_9	-0.026	[-0.436, 0.390]			0.956	[-1.477, 3.431]
	Event1			-2.230	[-5.259, 0.818]	3.552	[-23.337, 29.500]
	Event2			3.138*	[1.221, 5.066]	1.901	[-1.419, 5.311]
	Event3			3.369*	[1.080, 5.647]	5.554	[-4.733, 15.790]
	Event4			5.344*	[2.353, 8.301]	6.969	[-18.430, 32.622]
	Event5			1.782	[-0.197, 3.801]	2.554	[-5.260, 10.390]
	Event6			3.613*	[1.633, 5.562]	2.422	[-3.155, 8.021]
	Event7			5.081*	[2.179, 8.025]	2.864	[-21.591, 27.827]
phi on	Cog. Abilities	-0.001	[-0.003, 0.002]	0.000	[-0.002, 0.002]		
	Prior Knowledge	0.001	[-0.002, 0.003]	0.000	[-0.003, 0.002]		
	Interest	-0.001	[-0.004, 0.002]	0.001	[-0.002, 0.004]		
	Self-Concept	0.000	[-0.003, 0.004]	0.001	[-0.003, 0.004]		
	class_1	0.024*	[0.014, 0.033]			0.003	[-0.065, 0.069]
	class_2	0.015*	[0.004, 0.025]			0.322*	[0.247, 0.387]
	class_3	0.027*	[0.017, 0.037]			0.168*	[0.110, 0.218]
	class_4	0.031*	[0.021, 0.041]			0.213*	[0.130, 0.287]
	class_5	0.028*	[0.018, 0.038]			0.175*	[0.097, 0.243]
	class_6	0.039*	[0.029, 0.049]			0.177*	[0.065, 0.278]
	class_7	0.008	[-0.002, 0.019]			0.058	[-0.042, 0.156]
	class_8	0.029*	[0.019, 0.038]			0.285*	[0.218, 0.344]
	class_9	0.028*	[0.018, 0.037]			0.273*	[0.215, 0.323]
	Event1			0.823*	[0.757, 0.891]	1.095*	[0.443, 1.652]
Event2			0.674*	[0.635, 0.713]	0.417*	[0.337, 0.487]	

	Event3			0.586*	[0.537, 0.635]	1.241*	[0.995, 1.448]
	Event4			0.694*	[0.627, 0.758]	1.377*	[0.782, 1.899]
	Event5			0.680*	[0.638, 0.721]	0.625*	[0.436, 0.787]
	Event6			0.695*	[0.656, 0.735]	0.512*	[0.372, 0.631]
	Event7			0.564*	[0.497, 0.631]	0.685*	[0.156, 1.190]
Residual	Rating	6.908*	[1.735, 18.048]	19.883*	[7.716, 37.379]	7.107*	[1.313, 21.703]
Variances	phi	0.000*	[0.000, 0.000]	0.000*	[0.000, 0.000]	0.000*	[0.000, 0.000]
	logv	0.389*	[0.325, 0.4990]	0.422*	[0.345, 0.523]	0.411*	[0.336, 0.530]
Inter-	phi - logv	-0.004*	[-0.006, -0.003]	-0.005*	[-0.006, -0.003]	-0.004*	[-0.006, -0.003]
correlations	phi - Rating	-0.001	[-0.018, 0.016]	-0.012	[-0.034, 0.008]	0.000	[-0.017, 0.020]
	logv - Rating	0.466	[-0.266, 1.240]	1.059*	[0.227, 1.979]	0.582	[-0.200, 1.445]
DIC			1,747,521.918		1,747,699.718		1,729,323.276
(pD)			(7,496.864)		(7,527.716)		(5,614.618)

Note: AR(1) = Autoregressive Lag-1; phi = Lag-1 slope; logv = log residual variance; DIC = Deviance Information Criterion; pD = Estimated Number of Parameters; * indicates significance.

Discussion

This study aimed to examine how students' individual prerequisites, class membership, and classroom activities could explain differences in individual within-student variability in attention-related behavior across a lesson as well as differences between students when teachers' quality of instruction is comparable between classrooms. We applied a continuous coding system to attention-related behavior in 10 videos showing teacher-centered settings of classroom instruction. Differences between students' attention-related behaviors should not have been affected by the teachers' ability to implement structured and cognitively activating instruction, as students' ratings of teacher' quality of instruction were comparable. We analyzed the resulting multiple time-series with a statistical approach that allowed us to model within-person dynamics while accounting for individual differences on the between-person level.

In accordance with previous research, we examined the effects of students' individual characteristics, class membership (as representation for determinants that are unique to single classrooms), and the relative proportions of classroom activities (e.g., classroom discussions, book entries, individual seatwork) on students' attention-related behavior within one lesson in a stepwise manner. We did not find effects of individual prerequisites on differences in individual students' attention-related behavior during one lesson with respect to variability in individual students' momentary attention-related behavior or their mean levels of attention-related behavior. This means that potential effects of students' cognitive abilities, prior knowledge, interest, and self-concept were not strong enough to become evident without controlling for other determinants of students' attention-related behavior. However, after controlling for class membership or classroom activities, respectively, our results indicated that students' attention-related behavior was more difficult to predict when students had higher levels of self-concept due to higher within-person variances. Within-student variability in attention-related behavior might be due to students shifting between passive on-task and active off-task behaviors (e.g., looking at the teacher and then rummaging through their belongings) but also by shifting between passive off-task and interactive on-task behaviors (e.g., looking out the window and then raising their hand). The latter pattern is in line with results from Böheim et al. (2020), who already found that self-concept predicted students' hand-raising in mathematics. However, we found an effect of students' self-concept only on variability in their individual behavior but not on their mean levels of attention-related behavior. Therefore, in line with previous research, our findings suggest that students with higher self-concept appear to engage in interactive on-task behaviors more often than their peers do during teacher-centered

instruction; however, students with higher self-concept might not per se pay more attention than their peers. This pattern further indicates that students with higher self-concepts might shift to non-content-related behavior occasionally but are able to regulate themselves and quickly focus on the learning material again when necessary.

With regard to effects of class membership, our results showed that students demonstrated higher within-person variability in their attention-related behavior in some classes compared with others. Therefore, determinants that are unique to single classrooms might cause students to display a larger range of attention-related behavior. As all classroom videos showed introductory lessons to new topics, the content of the school lesson should not have been decisive to cause substantial variations in students' attention-related behavior. Additionally, students considered teachers' ability to implement qualitative instruction comparable. Thus, that there must be additional factors within individual classrooms causing students to exhibit greater variability in their attention-related behavior during instruction in some classes compared with others. As the videos were recorded in the middle of the school year, teachers had enough time to establish certain structures that could encourage students to be (or prevent them from being) agitated and engaged in off-task behaviors. In turn, teachers could also have been able to implement a certain socialization of participation that influenced students' verbal participation in classroom discussions (see Clarke et al., 2016). In addition to teacher-student relationships, student-student relationships and dynamics can also explain differences between classes. Within one class, it can be assumed that students will eventually converge in their attention-related behavior. For example, if most students pay attention and follow the classroom rules, other students might be less likely to engage in behavior that deviates from the behaviors of their peers in order to avoid standing out. Additionally, if some students exhibit disruptive behaviors and the teacher does not manage to re-establish order, sooner or later more students will join in and engage in these rather undesirable behaviors. Findings by Goldberg et al. (2019) supported this idea by finding that the consideration of similarities in visual features between students sitting next to each other improved the computer-based estimations of students' attention-related behavior.

As the requirements of different classroom activities encourage different attention-related behaviors, we included information about the relative proportions of specific classroom activities in our analysis. Our results demonstrate that it was more difficult to predict students' attention-related behavior depending on, for example, the relative occurrence of classroom discussions or book entries. Whereas activities such as a teacher giving a lecture require the

students to just listen to the teacher without doing anything else, thus leaving less space for variation, activities such as classroom discussions enable students to actively participate and verbally contribute to the discussion or to remain rather passive and just listen as others discuss the material. Therefore, it would be easier to predict students' attention-related behavior while a teacher is lecturing compared with during classroom discussions. This means that differences between students become more evident during specific classroom activities, such as classroom discussions and collective exercises but also during book entries, as those activities allow a more diverse set of students' attention-related behavior. Thus, teachers should implement more of such classroom activities as it gets easier for them to infer about whether and how intensively students pay attention. However, in this context it is crucial to not only focus on interactive and striking behaviors but to also notice subtle cues that indicate students paying attention.

Even though class membership or classroom activities accounted for large differences between students, we still found effects of students' individual prerequisites. This indicates that students' individual characteristics appear to affect how much variation students exhibit in their attention-related behavior during instruction within single classrooms but not in general. This is in line with Diener and Larsen (2009), who claimed that specific behaviors in specific situations need to be explained by many variables in addition to individuals' predispositions. In this case, class membership covers multiple – yet partly unknown – factors that can help explain differences between students' behavior. In line with Helmke and Renkl (1992) and Böheim et al. (2020), our findings indicate that a substantial degree of the differences between students in their attention-related behavior can be explained by class membership. Even though students gave similar ratings to teachers' quality of instruction, there seem to be more classroom-specific components that affect students' behavior that go beyond special features of different classroom activities.

It should be noted that we only considered the comparability of teachers' ability to implement qualitative instruction and not the instructional quality of the respective lessons. To enrich the analyses, a teaching quality rating should be included throughout the video sequences. Furthermore, future research should locate even more comparable situations, as there are still unknown factors that cause variations in students' attention-related behavior.

The results of the present study point out the important role that teachers play as they create learning environments that are supposed to provide the best possible learning conditions for students. The degree to which they succeed in establishing meaningful interactions with their students can affect how students behave during instruction. For example, when the teacher

manages to create a safe and appreciative environment, insecure students might also participate more in classroom discussions. However, dynamics within classrooms are unique to the set of students and their teacher, and their prerequisites determine the interaction processes that occur during instruction.

Conclusion

The present study demonstrates the potential of intensive longitudinal data to investigate determinants of teacher-learner-interaction processes by considering students' attention-related behavior during instruction. Our results indicate that classroom-specific components have a great impact on students' attention-related behavior but that within the same classroom, variability in students' momentary attention-related behavior appears to be affected by their self-concept. However, future research is necessary to uncover more detailed knowledge about process-specific determinants within teacher-learner interactions.

References

- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359-388. <https://doi.org/10.1080/10705511.2017.1406803>
- Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163–192). Information Age. <http://www.statmodel.com/download/NCME12.pdf>
- Baumert, J., Artelt, C., Carstensen, C. H., Sibberns, H., & Stanat, P. (2002). Untersuchungsgegenstand, Fragestellungen und technische Grundlagen der Studie. In *PISA 2000—Die Länder der Bundesrepublik Deutschland im Vergleich* (pp. 11-38). Springer.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O., Schnabel, K.-U., Leven, I., Plaßmeier, N., Schleußer, R., Tauscher, N., & Thebis, F. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU): Dokumentation - Band 1*. Max-Planck-Institut für Bildungsforschung.
- Beyda, S. D., Zentall, S. S., & Ferko, D. J. K. (2002). The relationship between teacher practices and the task-appropriate and social behavior of students with behavioral disorders. *Behavioral Disorders*, 27(3), 236-255. www.jstor.org/stable/43153386
- Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction*, 65, 101250. <https://doi.org/10.1016/j.learninstruc.2019.101250>
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95-108. <https://doi.org/10.1016/j.learninstruc.2009.11.004>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. <https://doi.org/10.1080/00461520.2014.965823>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>

-
- Clarke, S. N., Howley, I., Resnick, L., & Penstein Rosé, C. (2016). Student agency to participate in dialogic science discussions. *Learning, Culture and Social Interaction, 10*, 27-39. <https://doi.org/10.1016/j.lcsi.2016.01.002>
- Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology, 63*(1), 74-80. <https://doi.org/10.1037/h0032247>
- Cohen, R. A. (2014). *The neuropsychology of attention* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-72639-7>
- D'Mello, S. K. (2016). Giving eyesight to the blind: Towards attention-aware AIED. *International Journal of Artificial Intelligence in Education, 26*(2), 645-659. <https://doi.org/10.1007/s40593-016-0104-1>
- Diener, E., & Larsen, R. J. (2009). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener* (pp. 7-24). Springer Netherlands. https://doi.org/10.1007/978-90-481-2354-4_2
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763-782). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_37
- Friedman, D. L., Cancelli, A. A., & Yoshida, R. K. (1988). Academic engagement of elementary school children with learning disabilities. *Journal of School Psychology, 26*(4), 327-340. [https://doi.org/10.1016/0022-4405\(88\)90041-6](https://doi.org/10.1016/0022-4405(88)90041-6)
- Goldberg, P., Schwerter, J., Seidel, T., Müller, K., & Stürmer, K. (2021). How does learners' behavior attract preservice teachers' attention during teaching? *Teaching and Teacher Education, 97*, 103213. <https://doi.org/10.1016/j.tate.2020.103213>
- Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., Kasneci, E., & Trautwein, U. (2019). Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-019-09514-z>

- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, *113*(4), 461-487. <https://doi.org/10.1086/669616>
- Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision: KFT 4–12+ R [Cognitive abilities test for 4th to 12th grade students]*. Beltz-Test.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (2 ed.). Klett Kallmeyer.
<http://www.worldcat.org/oclc/316016742>
- Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, *38*, 130-141.
- Hommel, M. (2012a). Aufmerksamkeitstief in Reflexionsphasen - eine Videoanalyse von Planspielunterricht. *Wirtschaft und Erziehung*, *1-2*, 12-18.
- Hommel, M. (2012b). Aufmerksamkeitsverhalten in verschiedenen Unterrichtskonzepten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *108*(4), 549-570.
- Jackson, P., & Hudgins, B. (1965). *Observation schedule for recording pupil attention*.
- Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, *23*, 33-42.
<https://doi.org/10.1016/j.learninstruc.2012.09.002>
- Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal*, *18*(2), 157-171.
<https://doi.org/10.3102/00028312018002157>
- Kelly, S. (2007). Classroom discourse and the distribution of student engagement. *Social Psychology of Education*, *10*(3), 331-352. <https://doi.org/10.1007/s11218-007-9024-0>
- Kelly, S. (2008). Race, social class, and student engagement in middle school English classrooms. *Social Science Research*, *37*(2), 434-448.
<https://doi.org/10.1016/j.ssresearch.2007.08.003>
- Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology*, *59*(5), 320-324.
<https://doi.org/10.1037/h0026223>

-
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. <https://doi.org/10.2307/2529310>
- Lane, E. S., & Harris, S. E. (2015). A new tool for measuring student behavioral engagement in large university classes. *Journal of College Science Teaching*, *44*(6), 83-91. www.jstor.org/stable/43632000
- Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment*, *8*(2), 139-162. https://doi.org/10.1207/S15326977EA0802_04
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality*, *46*(3), 517-528. <https://doi.org/10.2224/sbp.7054>
- Levine, M. D. (1990). *Keeping a head in school*. Educators Publishing Service.
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K., & Klieme, E. (2007). Gleicher Unterricht - gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, *35*(2), 125-147. https://www.pedocs.de/frontdoor.php?source_opus=5489
- Mayring, P., Gläser-Zikuda, M., & Ziegelbauer, S. (2005). Auswertung von Videoaufnahmen mit Hilfe der Qualitativen Inhaltsanalyse - ein Beispiel aus der Unterrichtsforschung. *MedienPädagogik*(9), 1-17. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-3414>
- McKinney, J. D., Mason, J., Perkinson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology*, *67*(2), 198. <https://doi.org/10.1037/h0077012>
- McNeish, D., & Hamaker, E. L. (2019). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*. <https://doi.org/10.1037/met0000250>
- OECD. (2014). *PISA 2012 Technical report*. OECD.
- Olney, A. M., Risko, E. F., D'Mello, S. K., & Graesser, A. C. (2015). Attention in educational contexts: The role of the learning task in guiding attention. In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 623-642). MIT Press.
- Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), *Master lectures in clinical neuropsychology* (pp. 173-202). American Psychology Association.

-
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM Mathematics Education*. <https://doi.org/10.1007/s11858-018-0918-4>
- Samuels, S. J., & Turnure, J. E. (1974). Attention and reading achievement in first-grade boys and girls. *Journal of Educational Psychology*, *66*(1), 29-32. <https://doi.org/10.1037/h0035812>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur-und Prozessparadigma. *Zeitschrift für Pädagogik*, *60*(6), 850-866. https://www.pedocs.de/frontdoor.php?source_opus=14686
- Seidel, T., Jurik, V., Häusler, J., & Stubben, S. (2016). Mikro-Umwelten im Klassenverband: Wie sich kognitive und motivational-affektive Schülervoraussetzungen auf die Wahrnehmung und das Verhalten im Fachunterricht auswirken. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer, & F. Schwabe (Eds.), *Bedingungen und Effekte guten Unterrichts* (pp. 65-87). Waxmann.
- Stipek, D. (2002). Good instruction is motivating. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 309-332). Academic Press.
- Turner, J. C., & Patrick, H. (2004). Motivational influences on student participation in classroom learning activities. *Teachers College Record*, *106*(9), 1759–1785. <https://doi.org/10.1111/j.1467-9620.2004.00404.x>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Waldis, M., Buff, A., Pauli, C., & Reusser, K. (2002). *Skalendokumentation zur Schülerinnen-und Schülerbefragung im schweizerischen Videoprojekt*. Universität Zürich: Pädagogisches Institut.
- Wolff, C. E., 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. <https://doi.org/10.1007/s11251-016-9367-z>

Appendix**Table A 1**

Regression Results of Individuals' Mean Attention-related Behavior Rating Score on Individual Prerequisites (N = 199).

	<i>b</i>	<i>SE</i>	<i>p</i>
Cognitive abilities	0.004	0.002	0.037
Prior knowledge	0.017	0.008	0.037
Interest	0.028	0.012	0.020
Self-concept	0.022	0.011	0.046

5

Study 3

Goldberg, P., Schwerter, J., Seidel, T., Müller, K., & Stürmer, K. (2021). How does learners' behavior attract preservice teachers' attention during teaching? *Teaching and Teacher Education*, 97, 103213. <https://doi.org/10.1016/j.tate.2020.103213>

Abstract

Teachers need to continuously monitor students' engagement in classrooms, but novice teachers have difficulties paying attention to individual behavioral cues in all learners. To investigate these interaction processes in more detail, we re-analyzed eye-tracking data from preservice teachers teaching simulated learners who engaged in different behaviors (Stürmer, Seidel, Müller, Häusler, & Cortina, 2017). With a new methodological approach, we synchronized the data with a continuous annotation of observable student behavior and conducted time series analysis on 3,646 seconds of video material. Results indicate that novice teachers' attention is attracted most often when learners show (inter)active learning-related behavior.

Keywords: teacher-student interaction; teachers' attention; student behavior; time series analysis; mobile eye-tracking research; multinomial regression

1. Introduction

A teacher's ability to provide students with sufficient learning time, engage all students in active learning processes, and elicit their cooperation comprise crucial prerequisites for enhancing students' achievement (Emmer & Stough, 2001). To manage the classroom successfully and provide appealing learning environments, teachers must direct their attention to relevant information and continually monitor students' learning (Wolff, Jarodzka, van den Bogert, & Boshuizen, 2016). Therefore, they must detect visual cues in students' behavior that indicate how learners pay attention and how engaged they are in learning content (Goldberg et al., 2019). When teachers are able to notice and identify a lack of engagement in students, they can adapt their teaching methods accordingly to encourage their students to actively engage with the learning content. However, novice teachers in particular have difficulties overseeing and distributing their attention evenly across learners (Cortina, Miller, McKenzie, & Epstein, 2015; Stürmer et al., 2017). It is assumed that they are often guided by conspicuous cues rather than an ability to monitor the classroom adequately (Wolff et al., 2016). Whereas experienced teachers observe more and notice more subtle cues compared to inexperienced teachers (e.g., Berliner et al., 1988; Carter, Cushing, Sabers, Stein, & Berliner, 1988; Sabers, Cushing, & Berliner, 1991), novice teachers are seen as lacking the required knowledge base that guides a professional view over the classroom (Berliner, 2001).

This is in line with research findings on teachers' professional vision, a concept that describes a teacher's ability to notice and interpret relevant features of classroom events for student learning (Goodwin, 1994; Sherin, 2007; van Es & Sherin, 2002). Professional vision is viewed as an indicator of knowledge representations that aid the preparation of effective teaching action (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Sherin, 2007). When it comes to noticing relevant features, previous findings indicate that novice teachers have difficulties identifying relevant cues for teaching and learning during classroom interactions while observing videotaped classroom situations (e.g., Santagata, Zannoni, & Stigler, 2007; van den Bogert, van Bruggen, Kostons, & Jochems, 2014; Wolff et al., 2016). However, extant research also shows that novice teachers can improve this ability as part of teacher training (e.g., Sherin & van Es, 2002; Star & Strickland, 2008). In this vein, it is assumed that the underlying professional knowledge structures develop over time (Stürmer, Seidel, & Holzberger, 2016). Grossman et al. (2009) point out that learning to recognize relevant elements of practice comprises a crucial part of professional development, a conclusion further supported by results from expertise research indicating that regardless of the domain in question, experts

have developed attentional skills that allow them to process visual information more effectively than novices (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010). However, extant research on the development of teachers' ability to notice relevant cues in the classroom while teaching (i.e., in-action) remains limited. Classroom instruction is based on teacher-student interaction processes characterized by their simultaneity, multidimensionality, and immediacy (Doyle, 1986). Investigating processes related to noticing while teaching poses additional challenges compared to assessments in which novice teachers observe videotaped classroom situations (i.e., research on-action). However, identifying the engagement-related cues in student behavior that preservice teachers are able to recognize while teaching and those they do not might further improve teacher training.

In the current study, we explore novice teachers' attentional focus during instruction and aim to uncover properties of visual cues in students' behavior on which teachers fixate. To systematically assess how behavioral cues influence novice teachers' attentional focus, comparable conditions across participants are necessary. Therefore, we based our analysis on a standardized experimental setting with videos conducted by Stürmer et al. (2017) and synchronized already-existing mobile eye-tracking data from preservice teachers with a continuous annotation of learners' behavior. Deploying continuous annotation gave us a unique opportunity to analyze teacher-student interactions during instruction and investigate what kind of behavior attracts novice teachers' attention.

2. Theoretical Background

Kounin (1970) identified teachers' ability to remain aware of what is going on in the classroom (*withitness*) as associated with student work involvement. Maintaining a functional overview is necessary to provide sufficient learning time, engage all students in active learning processes, and elicit their cooperation in creating a learning environment that enables all students to engage in relevant cognitive processes (Emmer & Stough, 2001). Teachers must engage in many cognitive activities to guide their students' learning (Duffy, Miller, Parsons, & Meloth, 2009). The development of so-called curriculum scripts facilitates the recognition of meaningful patterns in the classroom, which in turn enables teachers to improve their interactions with students (Putnam, 1987). Thus, as part of their expertise development, teachers need to integrate isolated knowledge structures and learn how to notice relevant cues and indicators, such as those that point out struggling students (Lachner, Jarodzka, & Nückles, 2016; Thiede et al., 2015).

2.1 Students' behavior as cues for teachers' attentional processes

When students engage in learning-relevant activities, some aspects of their cognitive processes are likely to be observable from the outside. For example, Posner (1988) demonstrated that visual orientation toward a certain stimulus improves processing efficiency. Thus, when a teacher is explaining classroom content and a student is listening, he or she might be more likely to turn and face the person speaking in order to better process the relevant information. This kind of student behavior, which can be described as external and observable activity, is viewed as an important element of the larger, multi-dimensional construct of student engagement (Fredricks, Blumenfeld, & Paris, 2004), as well as one of the key elements of learning and academic success. Three types of engagement have been defined: cognitive, emotional, and behavioral (Fredricks et al., 2004). While psychological investment in learning (cognitive component) and affective reactions to classroom situations (emotional component) are more internal processes, the behavioral component is observable. Concentration, attention, asking questions, and contributing to class discussions are activities that are already known to signal certain learning-related processes and become observable in students' behavior to some extent (Fredricks et al., 2004). As the three components are highly interrelated and do not occur in isolation, students' overt behavior can provide visible indicators of whether they are engaged in appropriate learning-related processes, which are in turn an important determinant of academic achievement (Lahaderne, 1968; McKinney, Mason, Perkerson, & Clifford, 1975). Previous research has found correlations between students' behavioral engagement and academic achievement (Lei, Cui, & Zhou, 2018), as well as between students' attention-related behavior and achievement (Helmke & Renkl, 1992; Hommel, 2012; Karweit & Slavin, 1981; Stipek, 2002). Opposing results finding no relation to achievement (e.g., Pauli & Lipowsky, 2007) might be due to the applied survey method (self-reports vs. observer ratings) and a restricted focus on certain facets of learning-related behavior. For example, measuring only active on-task behavior (Lipowsky, Rakoczy, Pauli, Reusser, & Klieme, 2007), without considering off-task behavior, does not account for the broad behavioral spectrum that students might demonstrate during classroom instruction, and thus does not allow for detection of possible effects of other kinds of behavior.

The challenge for teachers lies in noticing behavioral cues that are relevant for inferring individual students' needs. However, interpreting student behavior is not always straightforward and depends on both students' learning activities and their individual prerequisites. Learners can differ in their learning-related behavior, but still all be engaged in a

certain task. Simultaneously, a lack of certain behaviors can pinpoint a student who is distracted or whose mind is wandering. Therefore, students' learning-related behavior differs with respect to the learning activities in which they are engaged (Chi & Wylie, 2014). For example, previous research shows that high-achieving students typically engage more verbally than low-achieving students (e.g., Kelly, 2008; Sacher, 1995), and students with stronger beliefs in their own competence participate more often in classroom discussions than less-confident students (Böheim, Knogler, Kosel, & Seidel, 2020; Pauli & Lipowsky, 2007). Additionally, profiles based on students' general cognitive abilities, acquired knowledge in subject domains, interest, and subject-related self-concept (Seidel, 2006) can predict students' verbal participation (Jurik, Gröschner, & Seidel, 2013). Thus, the interplay between cognitive and motivational-affective prerequisites affects observable student behavior in teacher-student interactions. However, as previously mentioned, students' activities fall across a broad behavioral spectrum. Depending on their individual prerequisites, some students might display rather salient and active behavior, such as participation in classroom discussions or disruptions, whereas other students might remain unobtrusive and passive (Seidel, Schnitzler, Kosel, Stürmer, & Holzberger, 2020). Salient behavior is easier to observe, and teachers might have fewer difficulties inferring cognitive processes in more active students compared to quieter students with more subtle actions, even though the latter group might actually need the teacher's attention because they are struggling. Therefore, it is important that teachers not only react to salient student behavior, but also notice subtle cues that indicate problems and obstacles. Additionally, it is crucial that teachers are able not only to differentiate between attentive and non-attentive students, but also to determine the underlying cause of inattention (e.g., not interested vs. struggling; Seidel et al., 2020).

2.2 Measuring teachers' attention

To design effective teaching, teachers need to develop professional vision skills that allow them to identify important events and cues during teacher-student interactions (van Es & Sherin, 2002). However, previous research indicates that the required knowledge base is not yet present in novice teachers, but rather develops over time (Berliner, 2001). Novice teachers have been shown to have problems noticing relevant aspects of classroom instruction compared to more experienced teachers. For example, early research has demonstrated that expert teachers are better at noticing subtle differences in instructional strategies (Sabers et al., 1991) and that novices have difficulties focusing on students' actions (Carter et al., 1988). Following Blomberg, Stürmer, and Seidel (2011), *noticing* describes teachers' ability to pay attention to

important aspects in complex classroom environments. To measure teachers' noticing ability, video prompts (Seidel & Stürmer, 2014; Stürmer & Seidel, 2015), questionnaires (Steffensky, Gold, Holodynski, & Möller, 2015), and/or qualitative analysis of open questions (Kersting, 2008; van Es & Sherin, 2008) are deployed. However, using such non-physical measurements only can provide limited information on teachers' attentional focus, as these processes might happen rather unconsciously. Using attentional skills as an indicator of expertise, eye-tracking technology already has been used to study professional vision in various domains. The specialized way that members of a professional group view a scene of interest has been shown to be domain-independent and connected to expertise level. Due to their well-organized and structured schemata of concepts (Chi, Glaser, & Rees, 1982), experts possess attentional skills that allow them to focus on relevant rather than irrelevant visual information (Jarodzka et al., 2010). For example, experts were shown to fixate more often on relevant rather than irrelevant areas during chess games (Charness, Reingold, Pomplun, & Stampe, 2001). When viewing dynamic stimuli, experts exhibit longer, but fewer, fixations on relevant areas, indicating that experts might exhibit more selective search strategies because they know the visual cues that provide important information (Moreno, Reina, Luis, & Sabido, 2002). As these studies indicate, experts and novices differ in how they view certain situations and how they perceive visual information. Thus, teachers' visual perception can also provide important insights into their ability to notice relevant information within complex classroom interactions (Lachner et al., 2016). However, this complexity poses additional challenges in terms of attention allocation that distinguish research on teaching from the aforementioned studies (Cortina et al., 2015).

As teaching is defined as a process of teacher-student interaction, students also influence teachers' behavior. For example, they might interact through explicit behavior, such as asking questions or disturbing classroom instruction, or subliminal behaviors, such as showing a lack of understanding through their facial expressions. In this context, distinguishing relevant from irrelevant information becomes more complex, as teachers must interact with their students and react to contextual demands simultaneously. For example, during classroom discussions, teachers need to listen to student answers, consider the relevance and quality of these answers, and think about the next question, while simultaneously scanning the class for misbehavior and/or signs of miscomprehension (Doyle, 1986). Consequently, inexperienced teachers can easily become overwhelmed because they are not yet able to process all incoming information effectively and decide which visual cues are most relevant. Due to excessive demands, processes that direct novice teachers' eye movements might differ from those of experts.

Human eye movements in general are guided by two broad processes: bottom-up, through salient features in targets, and top-down, such as through plans and intentions derived from professional knowledge (Schütz, Braun, & Gegenfurtner, 2011; Seidel et al., 2020; Shulman, 1987). Therefore, it can be assumed that these processes also drive teachers' visual attention while teaching (Lachner et al., 2016). On one hand, salient features such as students raising their hands or disturbing the classroom can catch teachers' attention. On the other hand, teachers' attention also can be driven by specific tasks when observing certain students more closely, such as gathering information about their cognitive processes. This intentional distribution of attention requires more top-down mechanisms and has been shown to be associated with teaching expertise (Haataja et al., 2019; McIntyre, Mainhard, & Klassen, 2017). Psychological studies in the field of attention research further indicate that bottom-up processes initially guide visual attention, before intentional, top-down processes intervene and control the attentional focus (Theeuwes, Atchley, & Kramer, 2000). Therefore, it can be assumed that alongside expertise, a temporal component impacts how teachers' attention is guided during instruction.

2.3 How preservice teachers' attention is guided during instruction

By analyzing classroom videos, Lipowsky et al. (2007) found that teachers tend to interact with high-performing students and actively engage with them more often compared with weaker students. However, interaction with students alone does not capture the actual focus of teachers' attention. Past research has deployed eye-tracking technology to investigate teachers' ability to detect relevant events in classroom scenarios (van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2015). However, these studies' findings are limited with respect to external validity, as participants' eye movements are recorded while they look at a screen showing an instructional setting, as opposed to engaging in a real classroom with teacher-student interactions. As previous research demonstrates that people's gaze behavior in laboratory settings differs from that in the real world (Foulsham, Walker, & Kingstone, 2011), teachers might also perceive a classroom situation differently when watching it on a computer screen (on-action) compared to actually being in the situation (in-action).

Recent in-action research has deployed mobile eye-tracking technology to study teachers' cognitive load (Prieto, Sharma, Wen, & Dillenbourg, 2015) or compare teachers' gazes for information-seeking and information-giving across expertise and culture (McIntyre et al., 2017). Furthermore, Cortina et al. (2015) assessed expert and novice teachers' eye movements during teaching with mobile eye-tracking technology. Novice teachers tended to

give their undivided attention to particular students while providing feedback, while expert teachers were capable of monitoring the whole classroom simultaneously. These results are supported by Dessus, Cosnefroy, and Luengo (2016), who investigated teachers' strategies with respect to expertise. Experienced teachers were able to distribute their attention more frequently to a broader set of students than novice teachers. Stürmer et al. (2017) found similar results, as preservice teachers distributed their attention unevenly across four learners with different learning prerequisites while teaching in standardized settings. Notably, preservice teachers focused their attention on their instructional material 30.24% of the time (Stürmer et al., 2017). Furthermore, when looking at learners, all preservice teachers mainly focused on one learner, even though they did not focus consistently on learners who shared the same set of individual prerequisites (Stürmer et al., 2017). Similarly, Dessus et al. (2016) assumed that teachers' gaze might depend on certain salient student characteristics, and therefore considered students' current subject performance as well as self-reported and teacher-perceived behavioral self-regulation abilities in their analysis. Their results suggest that students' level of performance and self-regulation might affect experienced teachers' gaze, but not novice teachers' gaze. Taken together, other explanations besides student characteristics might guide novice teachers' attentional focus. According to findings by Wolff et al. (2016), inexperienced teachers' attentional processes might be driven rather bottom-up through salient features in student behavior rather than their intention to diagnose students' cognitive processes (top-down; Schütz et al., 2011). However, existing research has yet to examine what has happened in the classroom by the time students capture preservice teachers' attention.

2.4 Research questions

Current approaches do not consider teacher-student interactions in more detail, and research on how student behavior affects novice teachers' attention in particular during instruction is lacking. Therefore, in the present study, we investigate these interactions for the first time in an exploratory manner by analyzing preservice teachers' attentional processes contingent upon students' behavior in a small sample of video material. Despite the rather small sample size, the videos display standardized teaching situations with comparable behavior by learners. These standardized teaching situations involved preservice teachers instructing a small group of learners in a setting with reduced complexity on the same domain-independent topics. Learners acted in accordance with profile scripts so that the circumstances were the same for all preservice teachers (Seidel, Stürmer, Schäfer, & Jahn, 2015). Thus, the videotaped settings offer a unique opportunity to uncover properties of visual cues in learners' behavior that novice

teachers fixate on, and to examine the stability of these effects over the course of preservice teachers' teaching.

To control for possible confounding effects within the complexity of teaching, it is important to ensure standardized conditions. For research on-action (e.g., observing videotaped classroom situations), this implies, for example, using the same video material for all participants. However, providing similar situations in research in-action is more complicated, as much variation exists across the spectrum of students and their behavior. While Cortina et al. (2015) compared expert and novice teachers' attentional processes while instructing the same classrooms with the same students, Seidel et al. (2015) developed standardized "training" situations to provide comparable conditions for preservice teachers in their first teaching experiences. We based our analysis on Stürmer et al.'s (2017) video data, in which seven preservice teachers were asked to teach four simulated learners in a standardized teaching situation. The lesson topics were pre-defined (tactical game, public transportation system), and the instructional time lasted for a maximum of 20 minutes. Learners comprised university students who were carefully trained and systematically assessed to behave in accordance with either an uninterested (mixed cognitive abilities, low interest), underestimating (high cognitive abilities and prior knowledge, low self-concept, intermediate level of interest), struggling (low cognitive abilities, knowledge, and self-concept), or strong (high cognitive abilities, knowledge, self-concept, and interest; Seidel, 2006) profile. Acting scripts provided background information about each profile in terms of cognitive and motivational-affective characteristics, as well as observable behavioral indicators. The strong profile was instructed to interact with the preservice teacher in an active and motivational manner, whereas the underestimating profile would only participate actively when directly engaged and made comments indicating a lack of confidence. The uninterested profile was instructed to actively exhibit low interest and engage in disturbing behaviors and comments, while the struggling profile would exhibit avoidant, shy behavior and try not to become actively engaged in interaction with the teacher (see Seidel et al., 2015). The learners were taught to act using observable behavioral indicators and further instructed to interact naturally and adapt their behavior in line with the teaching-learning process taking place in the situation (Stürmer et al., 2017).

To identify specific interaction patterns between preservice teachers' attentional focus and what is occurring in the instructional setting, we applied a new methodological approach to the data sources in which we synchronized preservice teachers' mobile eye-tracking data with a continuous rating of visible cues in learners' behavior, ranging from salient to rather

unobtrusive indicators, and conducted time series analysis. The following research questions were addressed:

- 1) Are there behaviors in simulated learners that capture preservice teachers' attention? Does salient behavior capture pre-service teachers' attention relatively more often compared with less salient behavior?
- 2) Does the effect of learners' behavior on preservice teachers' attention change over time?
- 3) Are there profile-specific differences in how learners' behavior affects preservice teachers' attentional focus?

3. Method

3.1 Sample and Procedure

To answer our research questions, we based our analysis on the data from Stürmer et al.'s (2017) eye-tracking study, where seven preservice teachers taught one out of two pre-defined topics in a standardized teaching situation. The seven preservice teachers constituted a subsample of a full cohort of preservice teachers ($N = 89$, age: $M = 22.2$ years, $SD = 2.0$, 56 % female) from the teacher education program at the Technical University of Munich (TUM), Germany. The program focuses on training secondary school mathematics and science teachers. The full cohort participated in the standardized teaching situations in their third year of the teacher education program as part of a university course (see Seidel et al., 2015). At this point, the cohort had already gathered some teaching experience by successfully completing three short internships in schools and classrooms. However, as the preservice teachers were about to begin their professional teacher preparation program, they could be regarded as novices in teaching. The study by Seidel et al. (2015) investigated to what extent these novices display teaching skills in the standardized situations, identifying differences in preservice teachers' teaching quality (e.g., structuring, teaching support, and learning climate), and validated the shown teaching skills with real classroom performance. Within the sample, preservice teachers were asked to voluntarily participate in an eye-tracking study (Stürmer et al., 2017). A total of seven preservice teachers ($n = 5$ female) wore eye-tracking glasses while teaching in the standardized situations (age: $M = 22.19$ years, $SD = 2.3$). This subsample can be considered as representative for the full study cohort, as they did not deviate more than one standard deviation from the cohort means on measures of their motivational learning prerequisites (*ability self-concept with regard to teaching*: full cohort $M = 3.44$, $SD = 0.45$ / subsample $M = 3.67$, $SD = 0.32$, scale from 1 = does not apply to 4 = applies; *self-efficacy with regard to teaching*: full

cohort $M = 2.96$, $SD = 0.32$ / subsample $M = 3.23$, $SD = 0.39$, scale from 1 = does not apply to 4 = applies), the way they adapted to the teaching role in the situation (full cohort $M = 3.80$, $SD = 0.35$ / subsample $M = 3.94$, $SD = 0.10$, external rating from 1 = does not apply to 4 = applies) and with regard to their shown teaching skills (*structuring*: full cohort $M = 1.67$, $SD = 0.45$ / subsample $M = 1.46$, $SD = 0.35$; *teaching support*: full cohort $M = 1.92$, $SD = 0.66$ / subsample $M = 1.67$, $SD = 0.52$; *learning climate*: full cohort $M = 2.46$, $SD = 0.41$ / subsample $M = 1.42$, $SD = 0.13$, external ratings from 1 = does not apply to 4 = applies). For our data analysis, we had to reduce the sample size from the pool of seven videos, as two of the original eye-tracking datasets could not be synchronized. Furthermore, one preservice teacher's instructional time in the standardized setting was much shorter; thus, the range of behavior learners were supposed to provide was not comparable. The four videotaped sessions totaling $N = 3,646$ seconds on which our analysis is based are comparable in length (Table 1) and in the ways the simulated learners acted (see Figure 1). In the original data, each session was video-recorded with a complete view of the situation, and preservice teachers wore mobile eye-tracking glasses. Preservice teachers and simulated learners were placed around two tables, with the underestimating and uninterested learners sitting on the right-hand side of the preservice teacher and the strong and struggling learners on the left-hand side. The seating order was kept constant across participants. Each of the four simulated learners was defined as one area of interest (AOI).

Table 1

Video length for each preservice teacher

Preservice teacher	Total seconds	Total minutes
1	1,044	17.40
2	971	16.18
3	988	16.47
4	888	14.80

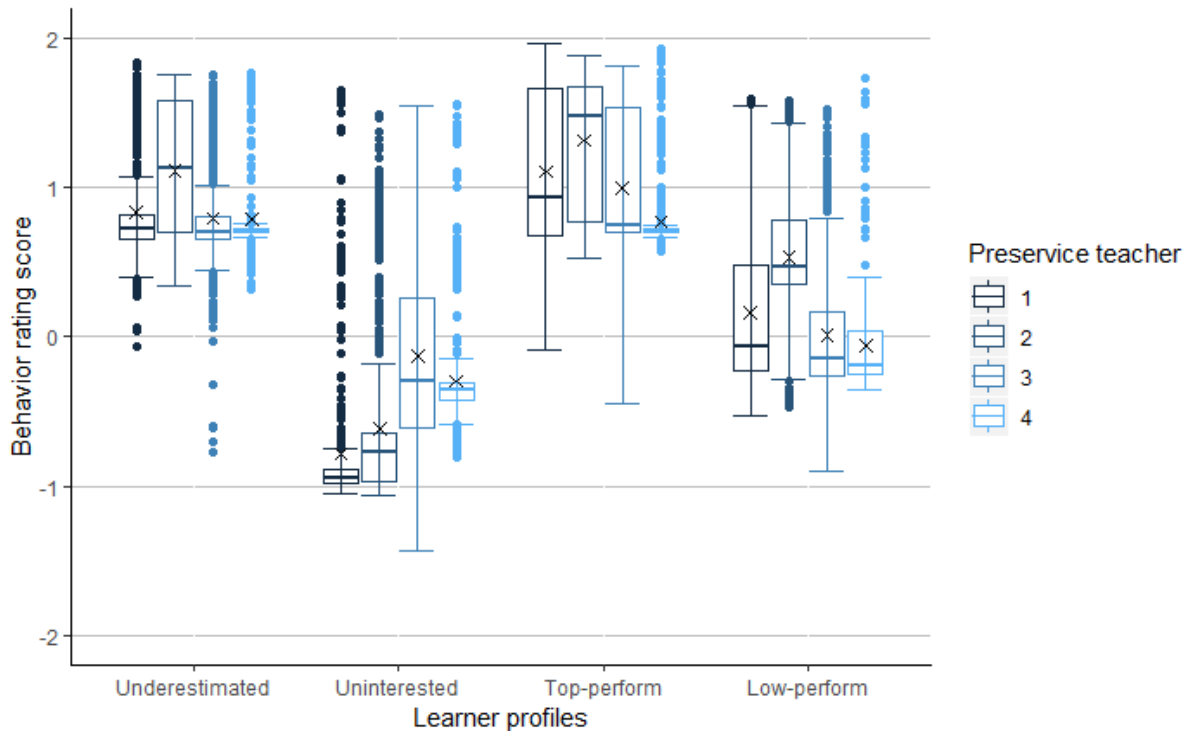


Figure 1 Behavior ratings per learner profile as boxplots separately for each preservice teacher (x indicating the mean)

3.2 Analysis

Behavior annotation. In the current study, we manually annotated learners' observable behavior on a one-dimensional scale over the entire instructional period in one-second steps. The free software CARMA (Girard, 2014) enables continuous interpersonal behavior annotation using joysticks (see Lizdek, Sadler, Woody, Ethier, & Malet, 2012). We combined the idea of on-task/off-task behavior (Helmke & Renkl, 1992; Hommel, 2012) with existing scales from the engagement literature and used the ICAP framework (Chi & Wylie, 2014) as inspiration to define more fine-grained differentiations within the spectrum of possible behaviors (passive, active, (de)constructive, and interactive). Thus, behavior was annotated on a symmetric scale ranging from -2, which indicated disruptive (i.e., interactive) off-task behavior, such as shouting across or walking around the classroom with the intention to interrupt, to +2, indicating highly engaged, interactive, on-task behavior in which, for example, learners ask questions and try to explain content to fellow learners (see Figure 2). Values closer to 0 indicated rather unobtrusive, passive behavior in which, for example, learners listened without participating (on-task) or rummaged through their belongings (off-task; Goldberg et al., 2019). Two raters annotated each learner in all videos in random order, with inter-rater reliability ICC(2,1) for each student profile ranging between .75 and .83 (absolute agreement).

For the subsequent analysis, the mean of the two raters was calculated for every learner at every second. In addition to the effect of behavior in general, we also investigated the impact of especially salient (i.e., active and interactive) behavior. To account for different effects of salient on-task and salient off-task behaviors, we defined behavioral annotation values above 1 and below -1 as salient behaviors and calculated two binary variables.

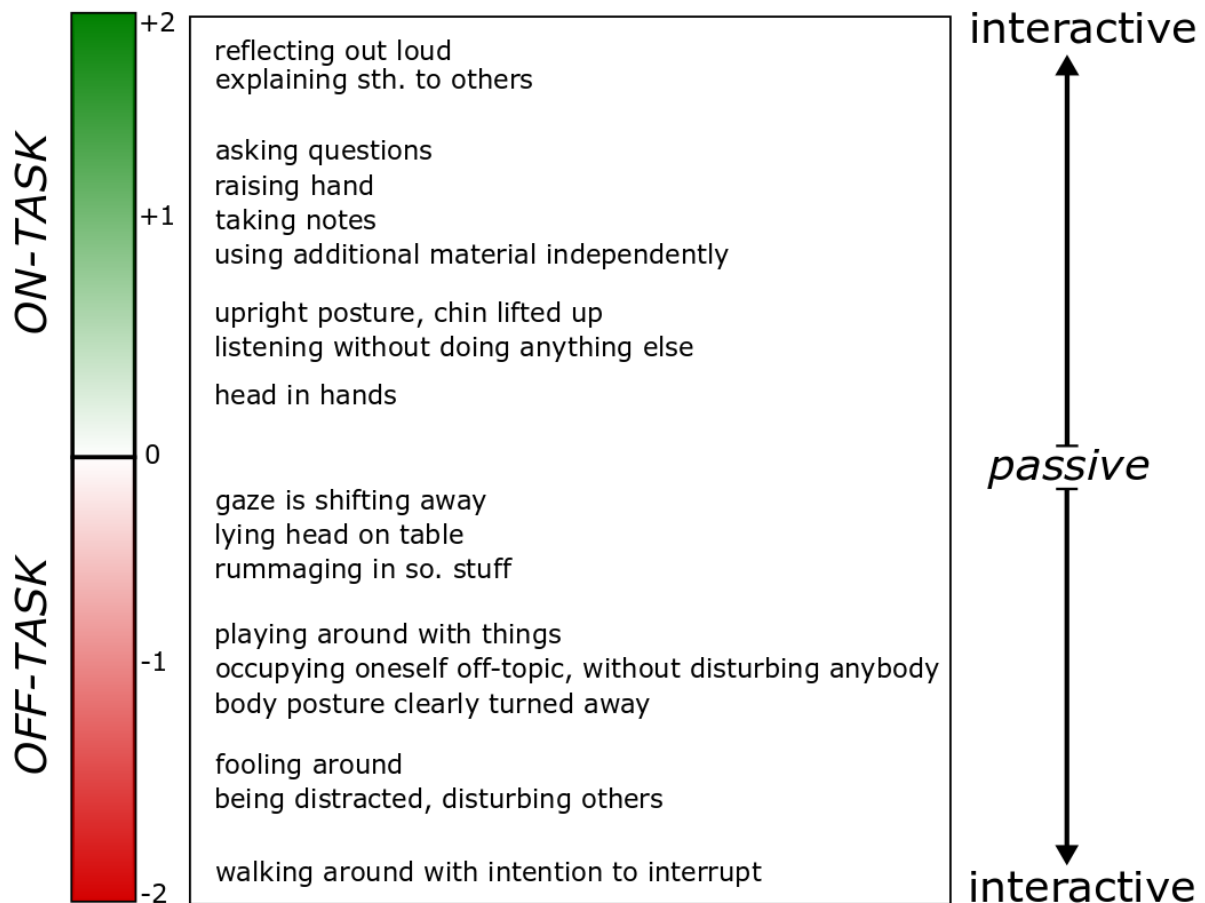


Figure 2 Scale for behaviour annotation with example behavioural indicators. From “Attentive or not? Toward a machine learning approach to assessing students’ visible engagement in classroom instruction“ by P. Goldberg, Ö. Sümer, K. Stürmer, W. Wagner, R. Göllner, P. Gerjets, E. Kasneci, & U. Trautwein, 2019, *Educational Psychology Review*. CC BY.

Teacher event rating. To enrich our analysis and control for the specific instructional setting, we conducted an event rating of preservice teachers' instructional practices (e.g., asking questions). Two raters coded the events according to the category system displayed in Figure 3, with an inter-rater reliability of $\kappa = .64$ (good agreement; Döring & Bortz, 2016). The raters applied a binary overall classification of preservice teachers' behavior (talks or does not talk) and also indicated whenever one of the students addressed the preservice teacher.

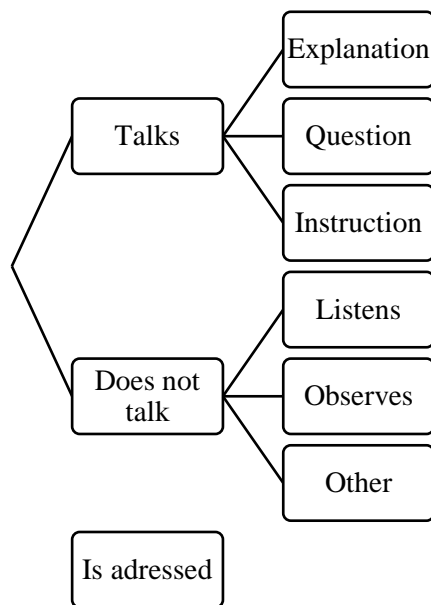


Figure 3 Event coding system of preservice teacher behavior

Preparing the time series. To reduce the information from the eye-tracking data and synchronize it with the manual annotations, we only used the fixation with the longest duration for each second. As a result, we conducted a time sampling of preservice teachers' AOIs and each simulated learner's behavioral information on a per-second basis. The resulting dataset is a time series that specifies preservice teachers' AOIs, the behavioral score for each learner, whether learners showed salient on- or off-task behavior, and what the preservice teacher did (i.e., teacher events) for each second.

Statistical analysis. We wanted to predict preservice teachers' AOIs, that is, whether they fixated on the underestimating, uninterested, strong, or struggling learner. Preservice teachers' AOIs are by nature a multinomial variable, as we cannot order the different profiles into a hierarchy of better or worse. Therefore, we applied multinomial regressions by using a mixed model with alternative-specific and alternative-unspecific variables. We predicted

preservice teachers' AOIs based on learners' behavioral ratings with a time lag. This decision was made to overcome the question: Which happens first, if both measures – AOI and behavioral rating – are used for the same second? By using time lags, we allow for a causal interpretation of the findings, as preservice teachers' AOIs should not influence learners' behavioral ratings one or more seconds earlier; that is, reverse causality is not an issue. We used the behavioral rating (first time lag; subsequently referred to as rating) together with the variables indicating salient on- and off-task behavior one second before preservice teachers' AOIs as our main variables of interest and also included the second and third time lag of the ratings to control for autocorrelation in our regressors.

Aside from the four students, the preservice teachers could choose not to look at any of the four learners, but rather somewhere in the room or at their instructional material. We used this option as the alternative in the multinomial regression, giving gaze towards the room/instructional material a rating score of zero. All variables of interest are alternative-specific, which eases the interpretation of the regression. This means that because we have individual ratings (and salient on- and off-task behavior) for all learners, we get one coefficient for the rating (and salient on- and off-task behavior respectively) for all alternatives (Cameron & Trivedi, 2005). By comparison, the teacher-event variables are not alternative-specific, but are the same for all learners. This leads to individual coefficients for each teacher event for each of the four alternatives.

To check for profile-specific effects, we conducted linear probability models for each learner profile separately. For this, we recoded the multinomial outcome variable as a series of binary variables, that is, a series of dummy variables equal to one if, for example, the preservice teachers' AOIs were directed towards the underestimating learner and zero otherwise. Regressions were calculated for the uninterested, strong, and struggling learners respectively, as well as for the alternative in which the preservice teachers looked anywhere but at one of the learners. We included all four preservice teachers in the analysis and controlled for general differences between the teachers by including dummy variables for each teacher. These analyses include a coefficient for salient off-task behavior only for the uninterested learner, as the other learners displayed no such behavior.

4. Results

4.1 Influence of learners' behavior on preservice teachers' attentional focus

In a first step, we included all preservice teachers in one multinomial regression. The coefficient of the manual annotation of learning-related behavior was significantly positive, b

= 2.04, $p < .001$, which means that the more learners' behavior moves toward the interactive on-task end of the behavioral continuum, the higher the likelihood that a learner will be looked at by the preservice teacher in the next second. Inversely, the more learners' behavior moves towards the interactive off-task end of the behavioral continuum, the lower the likelihood that a learner will be looked at by the preservice teacher in the next second. There was also a significant positive relationship with whether or not a learner showed salient on-task behavior, $b = 0.24$, $p < .05$. Thus, engaging in behavior such as asking questions or explaining something increased the probability of being in the preservice teachers' AOIs. Whether or not a learner displayed salient off-task behavior showed no significant relationship, $b = -0.85$, $p = .052$. Regarding gaze stability, the behavioral rating two seconds earlier was significantly negative, $b = -1.36$, $p < .001$, while the behavioral rating three seconds earlier was not significant, $b = 0.27$, $p = .263$. This means that if learners' behavior had a high rating score, the preservice teachers were less likely to keep looking at the respective learner two seconds later. The behavioral rating three seconds earlier did not exert any effect. To show that these effects are not sensitive to our choice of specification, Table 2 depicts the results when not controlling for the linear time trend (Model 2), preservice teachers (Model 3), teacher events (Model 4), or all of these (Model 5).

In multinomial regressions, only the direction and significance of the alternative-specific coefficient can be interpreted directly; the numerical value of the coefficient itself cannot because of the multinomial model's non-linearity. Therefore, we also calculated the marginal effects at the mean for the rating and for salient on-task behavior (Table 3). Values on the diagonal indicate the percentage increase in the likelihood of being looked at by the preservice teacher if the rating score rises by one unit or salient on-task behavior is shown. For example, if all variables are equal to their means, and the underestimating learner's rating score increases by one unit, the probability that the preservice teacher fixates on this learner increases by 18.09%. Values off the diagonal, in turn, indicate the percentage with which the likelihood of being in the teacher's AOI decreases when another learner's rating score rises, or if this other learner shows salient on-task behavior. The effect is symmetric, that is, an increase in the rating score of the underestimating learner, for example, leads to an equal decrease in the probability of the uninterested learner being in the preservice teacher's AOI (by 3.52%), as an increase in the rating score for the uninterested learner decreases the probability of the preservice teacher fixated on the underestimated learner. This is a general feature of alternative-specific regressions.

Table 2

Prediction of preservice teachers' AOI (N = 3618 sec)

	Model 1			Model 2			Model 3			Model 4			Model 5		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Rating t_{-1}	2.04	0.25	<.001	2.05	0.25	<.001	2.31	0.25	<.001	2.08	0.25	<.0001	2.34	0.25	<.001
Salient on-task behavior t_{-1}	0.24	0.10	.019	0.26	0.10	.009	-0.11	0.091	.023	0.48	0.10	<.001	0.24	0.08	.005
Salient off-task behavior t_{-1}	-0.85	0.44	.052	-0.915	0.43	.035	-0.76	0.43	.075	-0.68	0.43	.114	-0.63	0.42	.138
Rating t_{-2}	-1.36	0.41	.001	-1.39	0.40	.001	-1.45	0.41	<.001	-1.48	0.40	<.001	-1.63	0.40	<.001
Rating t_{-3}	0.27	0.24	.263	0.263	0.24	.135	0.34	0.24	.151	0.31	0.24	.194	0.48	0.24	.041
Controlled for teacher events		Yes			Yes			Yes			No			No	
Controlled for teacher		Yes			Yes			No			Yes			No	
Controlled for linear time trend		Yes			No			Yes			Yes			No	
χ^2		1790*			1719*			1475*			1422*			1019*	
Pseudo R^2		.157			.151			.129			.125			.089	

Note: χ^2 refers to the Likelihood Ratio Test. We calculated McFadden's Pseudo R^2 . * $p < .001$

Table 3

Marginal effects at the mean for the coefficients in percent

	Profile	underestimating	uninterested	strong	struggling
Rating	underestimating	23.62%	-4.73%	-6.56%	-3.86%
	uninterested		29.20%	-8.50%	-5.01%
	strong			37.19%	-6.94%
	struggling				24.76%
Salient on-task	underestimating	2.73%	-0.05%	-0.08%	-0.45%
	uninterested		3.38%	-0.98%	-0.58%
	strong			4.30%	-0.80%
	struggling				2.86%

Note: As the matrix is symmetrical, only the upper part is reported here.

To see whether our results are driven by just one preservice teacher and cannot be generalized, we ran the multinomial regressions separately for each preservice teacher. We found the same underlying patterns as in the aforementioned regression results (in which we included all teachers), with only minor deviations: When analyzing each preservice teacher separately, we again found a positive effect of the rating and a negative effect of the rating two seconds earlier. Therefore, we conclude that the effect of learners' behavior is not specific to one preservice teacher and thus is more generally valid. However, for salient on- and off-task behavior, we found mixed results in the teacher-specific regressions (for more information on the exact regression analysis, see Appendix A).

4.2 Influence of time on preservice teachers' attentional focus

Next, we investigated the impact of elapsed time during the course of instruction. Starting at the time point of 80 seconds⁵, we calculated regressions by adding data from the next second and continued the calculations over the time course. Figure 4 shows how the different coefficients help explain the teachers' AOIs over the course of instruction. When the full 95% confidence interval (as indicated by the blue area in the figure) is above or below zero, the coefficient's effect is significant at the 95% significance level.

The rating coefficient shows a stable positive effect, which increases only marginally after 500 seconds. As no changes exist over time and the rating's effect does not depend on the

⁵ We started with the 80th second to have a sufficient number of observations and enough variation in the data to calculate reasonable results.

time point within the instructional period, the effect of the rating can be viewed as stable over time. By comparison, the coefficient for salient on-task behavior shows some instability at the beginning of the instructional period, but this effect also stabilizes after about 500 seconds and appears to be similarly robust thereafter. This is not the case for the second time-lag coefficient of the rating. The estimation is less precise, and the effect is significant only after the teacher spent 700 seconds with the learners. Additionally, the estimation of salient off-task behavior is the most imprecise, as the confidence interval for this coefficient is the widest. The development over time here is also the least stable, given the different ups and downs of the coefficient. Thus, the effects of salient off-task behavior cannot be viewed as stable across instructional time.

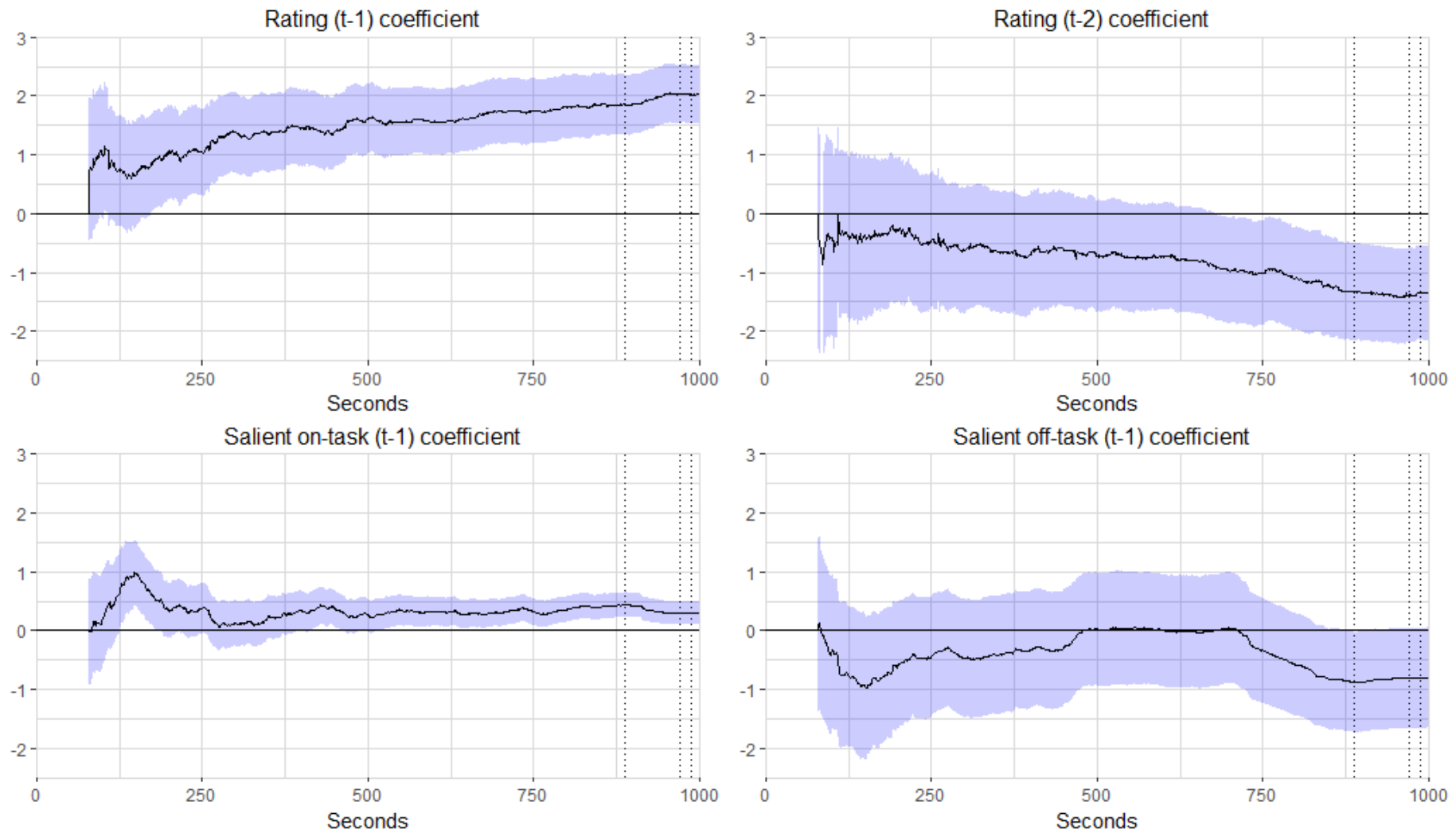


Figure 4 Plots displaying structural breaks for the different coefficients, with blue areas indicating the 95% CIs and vertical dotted lines indicating time points when no further data from one video was added

4.3 Profile-specific effects of learners' behavior

Finally, we calculated separate linear probability models for each student profile to investigate profile-specific effects of learners' behavior. The binary regression results are summarized in Figure 5, showing the variables of interest with their 95% confidence intervals (for exact values, see Appendix B).

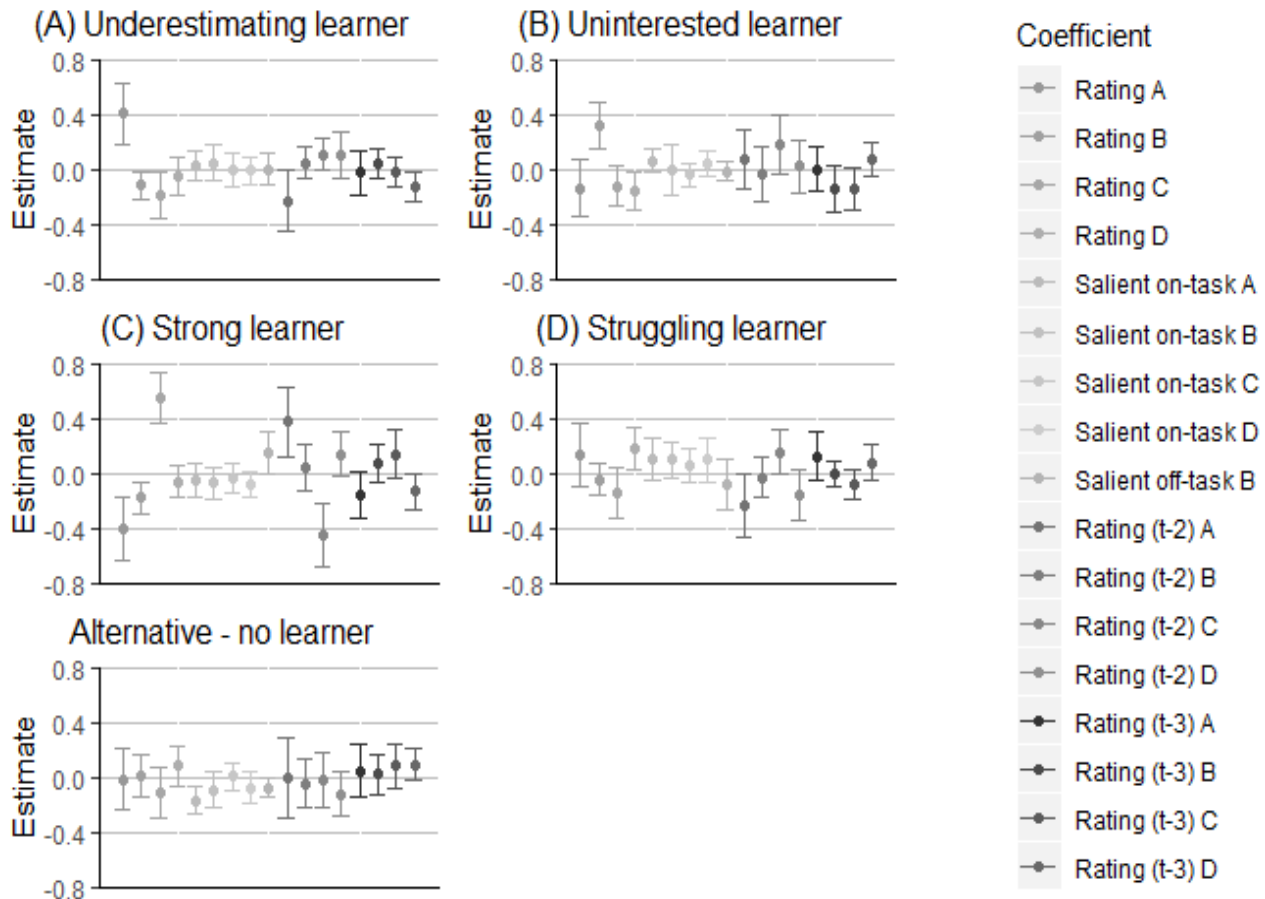


Figure 5 Linear probability models for each learner and the alternative with 95% CIs. Outcome is equal to one if the respective learner was looked at or (for the alternative) no learner was looked at and zero otherwise. In the legend, A indicates the underestimating learner, B the uninterested learner, C the strong learner, and D the struggling learner.

We found a significant effect of the rating for all profiles: If the rating increased by one unit, the probability of the preservice teacher focusing on that specific learner also increased. Thus, for example, asking questions or explaining something increased the probability of being in the preservice teachers' AOIs for all profiles. However, the rating exerted the greatest impact on the strong profile and the weakest impact on the struggling profile, suggesting the presence of profile-specific effects. Additionally, for some profiles, the other profiles' behavioral rating exerted a significant effect on the probability of being in the preservice teacher's AOI: When the uninterested and strong learners' ratings increased, the probability that the preservice

teacher would fixate on the underestimating learner decreased. Similarly, a rise in the underestimating and uninterested learners' behavioral ratings decreased the probability of the strong learner being in the preservice teachers' AOI, while an increase in struggling learner's rating decreased the probability of the uninterested learner being in the preservice teachers' AOIs. An increased behavioral rating for the other learners did not significantly affect the probability of the preservice teacher fixating on the struggling learner. The ratings two seconds earlier only exerted a significant effect on the strong learner. Similar to the results for the multinomial regression, a high rating score for the strong learner two seconds earlier decreased the probability that the preservice teacher would keep looking at him or her.

To cross-check our results, we also ran the linear probability model for the alternative case (i.e., the preservice teacher not looking at any learner). As expected, the learners' behavioral ratings did not explain when the preservice teacher looked elsewhere in the room or at the instructional material.

5. Discussion

In the present study, we aimed to more closely investigate student-teacher interactions by synchronizing preservice teachers' eye-tracking data with a continuous annotation of learners' behavior. We used time series analysis to examine whether certain behaviors in learners provoke preservice teachers' attentional focus and what role salient behaviors play in particular. Additionally, we evaluated the impact of the time point within instruction on novice teachers' attention. As students with different individual characteristics exhibit different kinds of visual cues, we further investigated profile-specific effects of learners' behavior on preservice teachers' attentional focus.

The patterns found in our results support previous research on preservice teachers' monitoring skills. Like Lipowsky et al. (2007), we found that preservice teachers focus their attention on students who are engaging in more (inter)active learning-related behavior, especially salient on-task behavior. Thus, active participation, such as asking questions or explaining something, increased the likelihood of the preservice teacher focusing on a learner who displayed this kind of behavior. Furthermore, the less actively learners participate and the more distracted their behavior becomes, the lower the probability of the preservice teacher focusing his or her attention on them. A possible explanation may concern teachers' need to control instructional progress (see Hofer, 1997). Novice teachers in particular might be more sensitive to this desire for control compared to experienced teachers. We found that preservice teachers focused more on learners who showed behavior that sustained the course of instruction

and tried to avoid misbehaving learners. These findings are particularly interesting, as research with stationary eye-trackers has demonstrated that novice teachers' attention is attracted by disruptive behavior and rather salient features (i.e., bottom-up influences) when watching a video rather than teaching themselves (Wolff et al., 2016). This mismatch is in line with research on people's gaze behavior finding different patterns in laboratory and real-world settings (Foulsham et al., 2011). According to Foulsham et al. (2011), these differences might be influenced by predictions of how the scene in the real world will change and the requirement to engage with the given task (i.e., top-down processes). The difference between our results and those of Wolff et al. (2016) could indicate, for example, that novice teachers' attention might be driven by the demands of the context (i.e., a more passive context when watching a video without the need to interact with learners vs. actual teaching in which they must interact with learners) and their underlying intention (i.e., observing a scenario vs. conveying learning content). Early research already found that novice teachers demonstrate certain inflexibilities when it comes to deviations from lesson plans (e.g., Livingston & Borko, 1989; Westerman, 1991). Therefore, top-down processes (i.e., plans and intentions) related to following their instructional agenda might guide novice teachers' attentional focus rather than top-down processes related to steady monitoring and the identification of problematic behavior. Nevertheless, it is important for inexperienced teachers to overcome the urge to focus mainly on actively engaged students and instead monitor the classroom evenly, as they have to identify inattentive students in order to encourage their active participation and support engagement and learning from all students (Seidel et al., 2020). By being more likely to react to salient behavior than rather unobtrusive cues, novice teachers might fail to identify students who need special attention because they are struggling (low-performing profile) and/or lack confidence in their skills (underestimating profile). Furthermore, it is important for teachers to be able to identify the underlying reasons for student behaviors, as a low-performing student needs different kinds of support than a student who underestimates him- or herself or a student who is simply not interested in the learning topic. Running the regressions separately for each preservice teacher revealed that the learners' salient behavior generally affected the preservice teachers' attentional focus. However, we found variations among individuals. For most preservice teachers in our sample, salient behavior exerted a positive effect on their attentional focus, meaning that they focused their attention on conspicuous rather than unobtrusive cues. However, one preservice teacher's attention was affected in the opposite way. Furthermore, while learners' behavior exerted a positive effect on preservice teachers' attentional focus, this

effect was insignificant for one preservice teacher. This disparity is in line with previous findings (Dessus et al., 2016; Stürmer et al., 2017) implying different processes of attention allocation and indicating varying stages of schema construction in preservice teachers. For example, trying to avoid focusing on salient behavior could indicate top-down, rather than bottom-up, processes of attention allocation, as attention is not guided by striking cues, but by the intention to avoid this kind of behavior and focus on more subtly acting learners.

We assumed that the time course would influence how preservice teachers distribute their attentional focus during instruction. We found no effect of instruction time on the relationship between learners' behavior and preservice teachers' attention in general. Learners' behavior guided preservice teachers' attentional focus throughout the time course. At all times, teachers were more likely to focus on actively engaged learners compared with rather passive or even disturbing behavior. Furthermore, preservice teachers focused on actively engaged learners who exhibited salient behavior, especially during the second half of the instructional time. This might be due to preservice teachers' intention to convey certain learning content during the instructional period. When experiencing pressure to finish in time, they might pay more attention to learners who can help them pursue their goals, and thus focus their attention on students who display salient on-task behavior. Finally, the behavioral rating two seconds earlier exhibited a negative effect during the last third of the instructional time, indicating that preservice teachers' gaze is not stable. This is in line with previous research showing that novice teachers' attention while teaching is dominated by the short term, involving quick changes between AOIs (Stürmer et al., 2017). On the other hand, it might also indicate that the novices attempted to monitor the classroom after they had some time to get accustomed to the situation. With respect to salient off-task behavior, our results indicate no significant effect and rather unstable estimations. One explanation from a technical point of view might be the comparatively fewer data points considered salient off-task behaviour, which made the estimations less precise. A more content-based explanation would be that preservice teachers did not react as deliberately and consistently to salient off-task behaviour as they did to salient on-task behaviour. Thus, the estimations were rather imprecise because the preservice teachers reacted in a non-systematic way when salient off-task behaviour occurred.

Previous research indicates that students exhibit different kinds of observable behavior depending on their individual characteristics (Jurik et al., 2013; Pauli & Lipowsky, 2007) and that teachers generally prefer to interact with actively engaged students (Lipowsky et al., 2007). Therefore, we investigated whether profile-specific effects exist that guide preservice teachers'

attentional focus. We found that learners' ratings exerted a generally positive effect. For example, asking questions increased the probability of being in the preservice teacher's attentional focus for all profiles. However, this effect was greatest for strong learners and weakest for struggling learners. This finding highlights a particular issue, as struggling students particularly need their teachers' attention. When teachers overlook students who are experiencing difficulties in understanding instruction, they fail to engage these students in the learning process, resulting in decreased and/or unsuccessful learning. Additionally, preservice teachers' attentional focus was affected differently by different profiles. For example, when the struggling learner participated more actively, only the uninterested learner's probability of being in the preservice teacher's AOI decreased, not those of the strong and underestimating learners. Moreover, only when the strong learner was participating actively did the probability of the preservice teacher continuing to look at him or her decrease. This might indicate that the preservice teachers knew that the strong learner was adequately engaged and were attempting to distribute their attention to other learners, as we did not find this effect with the other profiles. Taken together, our findings indicate profile-specific effects of learners' behavior on preservice teachers' attentional focus.

It should be noted that the number of data points in the category of salient off-task behavior was rather low, and the values less extreme compared with those in the salient on-task category (see Figure 1). Whereas learners displayed actions from the upper extreme of the behavioral spectrum, like explaining content to fellow students, they did not engage in activities on the lower extreme of the scale, such as walking around and actively disturbing others or instruction. Such behaviour also occurs rather rarely in real classroom situations involving university students (Goldberg et al., 2019). Even though one of the learners was instructed to behave in an uninterested manner, the instructions for this learner included behaviors such as playing on their smartphone or sometimes disturbing their neighbor but not the whole group (i.e., passive and active off-task behavior but not interactive off-task behavior). However, our rating instrument has to cover the entire possible spectrum of learners' behavior in order to be considered valid. Thus, the observed patterns might be driven by too little variation in the displayed behavior, meaning that interpretations regarding preservice teachers' attentional focus with respect to salient off-task behavior should be drawn carefully.

Nevertheless, by using standardized situations, we were able to ensure comparable conditions for all participants involving a similar set of observable behaviors. Differences in the profile-specific behaviors were due to the preservice teachers' individual methods of

interacting with the learners, as the learners were instructed to adapt to the situation naturally. Interestingly, even though all learners theoretically should have behaved in the same way, the variation in preservice teachers' interaction styles resulted in unequally pronounced behaviors.

Even though our sample size is rather small, the synchronization of the continuous data was based on almost 4,000 seconds of material (i.e., data points). This constitutes a vast amount of data and – to the best of our knowledge – analysis of triangulated data like ours has never before been performed. Therefore, our study provides a promising starting point for systematically investigating interactions between teachers and students by deploying mixed methods and time series analysis. Our next step will be to explore the effects of real classrooms containing more students – and thus more demanding interaction processes – on novice teachers' attention. Furthermore, in future studies, it would be of great interest to compare experts and novices in order to identify knowledge structures and competencies that inexperienced teachers do not yet possess. Insights like these could have critical implications to help teacher educators and mentors train novice teachers.

6. Conclusion

Conducting a time series analysis of teachers' eye-tracking data in combination with continuous ratings of student behavior is a promising approach to analyzing teacher-student interactions during instruction in more detail. We found that inexperienced teachers are more likely to focus their attention on students who exhibit actively engaged behavior compared with rather passive or even disruptive behavior, and that this effect is stable across the period of instruction. Our findings further support the distinction between on-action and in-action research, as novice teachers in particular might behave differently when faced with the demands of actual classroom instruction and interaction.

However, the rating procedure for such synchronized data is time-consuming. To further study such interaction processes with larger sample sizes and in real classroom settings in which teachers usually teach more than four learners, automated assessment seems to be a promising next step (Goldberg et al., 2019).

Appendix A

To see whether our results were driven by just one preservice teacher and cannot be generalized, we ran the multinomial regressions separately for each preservice teacher. However, not enough variation in salient off-task behavior existed to be estimated in the multinomial regression for Preservice Teachers 1 and 4. Thus, in our regression results, we did not differentiate between salient on- and off-task behaviors; instead, we used only one dummy variable indicating salient behavior in general (Figure A1). Regression results are displayed in Table A2. The rating is positively significant for Preservice Teachers 1, 2, and 3. Additionally, salient behavior is positively significant for Preservice Teachers 1, 3, and 4; however, it is negatively significant for Preservice Teacher 2. The second time lag of the rating is negatively significant only for Preservice Teachers 1 and 3.

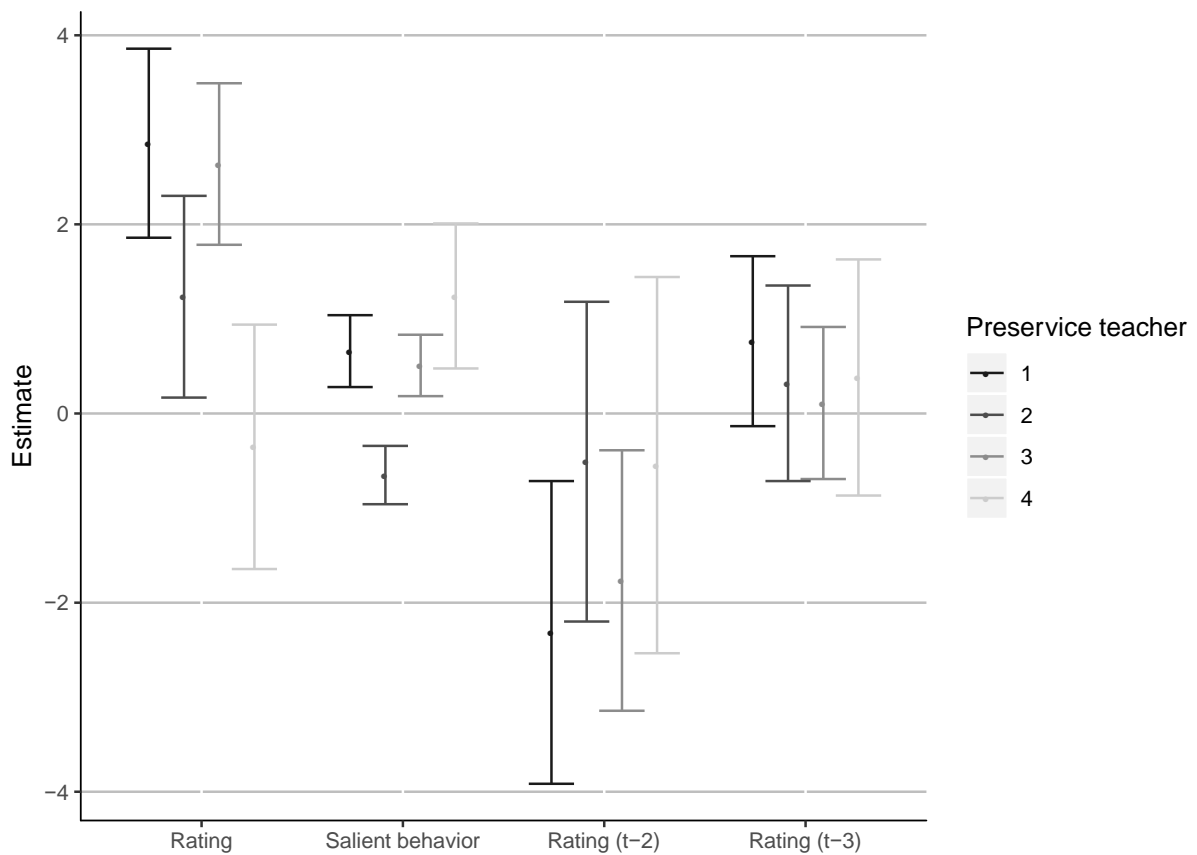


Figure A1. Teacher-specific regressions for all preservice teachers with one variable indicating salient behavior in general, as the model could not be estimated for Preservice Teachers 1 and 4 otherwise. Whiskers indicate 95% CIs.

Table A2

Teacher-specific regression for all preservice teachers with one variable indicating salient behavior in general, controlling for teacher events, teachers, and time trend.

	Preservice teacher 1			Preservice teacher 2			Preservice teacher 3			Preservice teacher 4		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Rating $t-1$	2.86	0.51	<.001	1.23	0.54	.023	2.64	0.44	<.001	-0.35	0.66	.593
Salient behavior $t-1$	0.66	0.19	.001	-0.65	0.16	<.001	0.51	0.17	.002	1.24	0.39	.001
Rating $t-2$	-2.32	0.82	.005	-0.51	0.86	.554	-1.77	0.70	.012	-0.55	1.02	.591
Rating $t-3$	0.77	0.46	.096	0.32	0.53	.544	0.11	0.41	.786	0.38	0.64	.550
χ^2		796*			305*			373*			64*	
Pseudo R^2		0.2565			0.1095			0.1277			0.031	
Observations		1011			918			929			760	

Note: χ^2 refers to the likelihood ratio test. We calculated McFadden's Pseudo R^2 . * $p < .001$

To investigate the specification including salient off-task behavior, we calculated multinomial regressions for Teachers 2 and 3. The estimation results are summarized in Figure A3, which presents the estimated coefficients of the variables of interest and the respective 95% *CI*s (for exact values, see Table A4).

The rating is significant for both teachers. Salient on-task behavior is negatively significant for Preservice Teacher 2, but positively significant for Preservice Teacher 3. Furthermore, salient off-task behavior is negatively significant for Preservice Teacher 2, but not significant for Preservice Teacher 3. The second lag of the rating is negatively significant only for Preservice Teacher 3. The third lag is statistically insignificant for both preservice teachers.

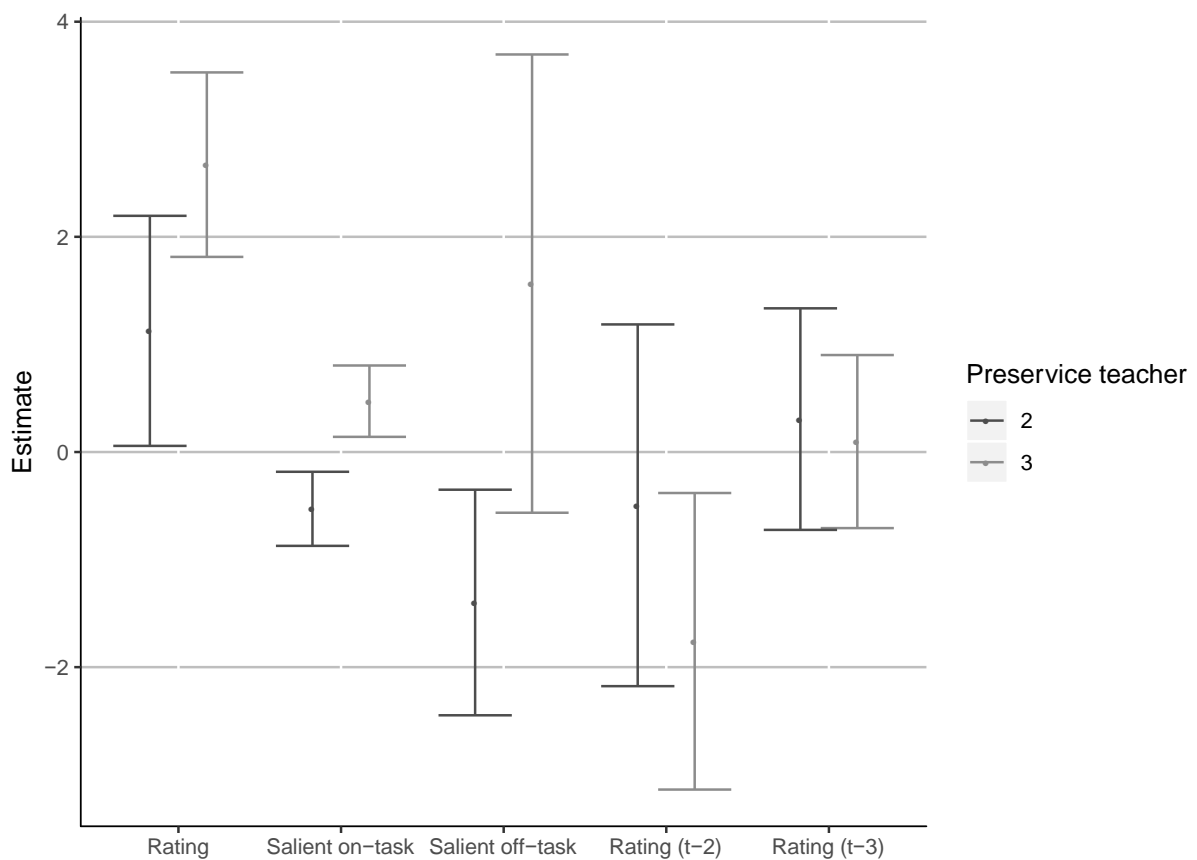


Figure A3. Teacher-specific regression for Preservice Teachers 2 and 3, as only these have enough variation in the salient on- and off-task behavior variables. Whiskers indicate 95% *CI*s.

Table A4

Teacher-specific regression for Preservice Teachers 2 and 3, controlling for teacher events, teachers, and time trend.

	Preservice teacher 2			Preservice teacher 3		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Rating $t-1$	1.13	0.55	.039	2.67	0.44	<.001
Salient on-task behavior $t-1$	-0.53	0.18	.003	0.47	0.17	.005
Salient off-task behavior $t-1$	-1.40	0.54	.009	1.57	1.09	.150
Rating $t-2$	-0.50	0.86	.564	-1.76	0.70	.012
Rating $t-3$	0.31	0.53	.561	0.10	0.41	.813
χ^2		307*			374*	
Pseudo R^2		0.1105			0.128	
Observations		918			929	

Note: χ^2 refers to the likelihood ratio test. We calculated McFadden's Pseudo R^2 . * $p < .001$

Appendix B

Linear probability models for each learner and the alternative with outcome equal to one if the respective learner was examined or (in the alternative model) no learner was examined and zero otherwise. Controlled for teacher events, teachers, and time trend.

	(A) Underestimating learner			(B) Uninterested learner			(C) Strong learner			(D) Struggling learner			Alternative – no learner		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
Rating A $t-1$	0.42	0.12	<.001	-0.14	0.11	.207	-0.40	0.12	.001	0.14	0.12	.254	-0.01	0.12	.926
Rating B $t-1$	-0.12	0.05	.024	0.32	0.09	<.001	-0.18	0.06	.002	-0.04	0.06	.465	0.01	0.08	.859
Rating C $t-1$	-0.19	0.09	.027	-0.12	0.07	.115	0.56	0.10	<.001	-0.14	0.10	.136	-0.11	0.09	.232
Rating D $t-1$	-0.05	0.07	.488	-0.16	0.07	.023	-0.06	0.06	.362	0.18	0.08	.023	0.09	0.07	.235
Salient on-task behavior A $t-1$	0.03	0.06	.566	0.07	0.05	.124	-0.05	0.07	.492	0.11	0.08	.164	-0.17	0.05	.001
Salient on-task behavior B $t-1$	0.05	0.07	.447	0.01	0.09	.981	-0.07	0.06	.268	0.10	0.07	.119	-0.09	0.07	.175
Salient on-task behavior C $t-1$	-0.01	0.06	.96	-0.04	0.04	.380	-0.03	0.06	.611	0.06	0.06	.320	0.01	0.05	.862
Salient on-task behavior D $t-1$	-0.01	0.05	.934	0.05	0.05	.319	-0.07	0.05	.125	0.11	0.09	.214	-0.08	0.06	.198
Salient off-task behavior B $t-1$	0.01	0.06	.907	-0.01	0.03	.741	0.15	0.08	.062	-0.08	0.10	.426	-0.07	0.03	.035
Rating A $t-2$	-0.23	0.12	.049	0.08	0.11	.486	0.38	0.13	.004	-0.23	0.12	.046	-0.01	0.15	.997
Rating B $t-2$	0.05	0.06	.398	-0.03	0.11	.760	0.05	0.09	.561	-0.03	0.08	.715	-0.04	0.09	.659

Rating C $t-2$	0.12	0.06	.047	0.19	0.11	.085	-0.45	0.12	<.001	0.16	0.08	.058	-0.01	0.10	.922
Rating D $t-2$	0.10	0.09	.244	0.03	0.10	.765	0.14	0.08	.08	-0.16	0.09	.091	-0.12	0.08	.143
Rating A $t-3$	- 0.02	0.08	.807	0.01	0.08	.989	-0.16	0.09	.066	0.13	0.09	.166	0.05	0.10	.603
Rating B $t-3$	0.04	0.06	.450	-0.14	0.09	.095	0.08	0.07	.290	<.01	0.05	.994	0.02	0.08	.752
Rating C $t-3$	- 0.02	0.05	.722	-0.14	0.08	.092	0.15	0.09	.101	-0.08	0.05	.155	0.09	0.09	.317
Rating D $t-3$	- 0.13	0.06	.025	0.08	0.06	.221	-0.13	0.07	.044	0.08	0.07	.226	0.10	0.06	.103
F		17.289*			20.6691*			38.7244*			23.5791*			21.759*	
R^2		0.115			0.134			0.225			0.150			0.140	
Observations		3633			3633			3633			3633			3633	

Note: * $p < .001$. A indicates the underestimating learner, B the uninterested learner, C the strong learner, and D the struggling learner.

References

- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35, 463–482. doi:[https://doi.org/10.1016/S0883-0355\(02\)00004-6](https://doi.org/10.1016/S0883-0355(02)00004-6)
- Berliner, D. C., Stein, P., Sabers, D. S., Clarridge, P. B., Cushing, K. S., & Pinnegar, S. (1988). Implications of research on pedagogical expertise and experience in mathematics teaching. In D. A. Grouws & T. J. Cooney (Eds.), *Perspectives on research on effective mathematics teaching* (pp. 67-95). Reston, VA: National Council of Teachers of Mathematics.
- Blomberg, G., Stürmer, K., & Seidel, T. (2011). How pre-service teachers observe teaching on video: Effects of viewers' teaching subjects and the subject of the video. *Teaching and Teacher Education*, 27(7), 1131-1140. doi:<https://doi.org/10.1016/j.tate.2011.04.008>
- Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction*, 65, 101250. doi:<https://doi.org/10.1016/j.learninstruc.2019.101250>
- Cameron, A., & Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge, England: Cambridge University 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. doi:<https://doi.org/10.1177/002248718803900306>
- Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory and Cognition*, 29, 1146-1152. doi:<https://doi.org/10.3758/BF03206384>
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. S. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (pp. 1-75). Hillsdale, NJ: Erlbaum.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. doi:<https://doi.org/10.1080/00461520.2014.965823>
- Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, E. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction

- using mobile eye tracking with experts and novice teachers. *International Journal of Science and Mathematics Education*, 13(2), 389–403.
doi:<https://doi.org/10.1007/s10763-014-9610-5>
- Dessus, P., Cosnefroy, O., & Luengo, V. (2016). “Keep your eyes on ’em all!”: A mobile eye-tracking analysis of teachers’ sensitivity to students. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Adaptive and Adaptable Learning. EC-TEL 2016. Lecture Notes in Computer Science* (Vol. 9891, pp. 72-84). Cham, Switzerland: Springer International Publishing.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Wiesbaden, Germany: Springer.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3 ed., pp. 392-431). New York, NY: Macmillan.
- Duffy, G. G., Miller, S., Parsons, S., & Meloth, M. (2009). Teachers as metacognitive professionals. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 240-256). New York, NY: Routledge.
- 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. doi:https://doi.org/10.1207/S15326985EP3602_5
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision research*, 51(17), 1920-1931. doi:<https://doi.org/10.1016/j.visres.2011.07.002>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. doi:<https://doi.org/10.3102/00346543074001059>
- Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software*, 2(1), e5. doi:10.5334/jors.ar
- Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., . . . Trautwein, U. (2019). Attentive or not? Toward a machine learning approach to assessing students’ visible engagement in classroom instruction. *Educational Psychology Review*. doi:<https://doi.org/10.1007/s10648-019-09514-z>
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96, 606-633. doi:10.1525/aa.1994.96.3.02a00100

-
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. Retrieved December, 2nd 2019 from <https://tedd.org/wp-content/uploads/2014/03/Grossman-et-al-Teaching-Practice-A-Cross-Professional-Perspective-copy.pdf>
- Haataja, E., Moreno-Esteva, E. G., Salonen, V., Laine, A., Toivanen, M., & Hannula, M. S. (2019). Teacher's visual attention when scaffolding collaborative mathematical problem solving. *Teaching and Teacher Education, 86*, 102877. doi:<https://doi.org/10.1016/j.tate.2019.102877>
- Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica, 38*, 130-141.
- Hofer, M. (1997). Lehrer-Schüler-Interaktion. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie. Bd.3. Psychologie des Unterrichts und der Schule* (pp. 213-252). Göttingen, Germany: Hogrefe.
- Hommel, M. (2012). Aufmerksamkeitstief in Reflexionsphasen - eine Videoanalyse von Planspielunterricht. *Wirtschaft und Erziehung, 1-2*, 12-18.
- Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146-154. doi:<https://doi.org/10.1016/j.learninstruc.2009.02.019>
- Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction, 23*, 33-42. doi:<https://doi.org/10.1016/j.learninstruc.2012.09.002>
- Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal, 18*(2), 157-171. doi:<https://doi.org/10.3102/00028312018002157>
- Kelly, S. (2008). Race, social class, and student engagement in middle school English classrooms. *Social Science Research, 37*(2), 434-448. doi:<https://doi.org/10.1016/j.ssresearch.2007.08.003>
- 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. doi:<https://doi.org/10.1177/0013164407313369>

-
- Kersting, N., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568-589. doi:<https://doi.org/10.3102/0002831212437853>
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. Oxford, England: Holt, Rinhart & Winston.
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science, 44*(3), 197-203. doi:<https://doi.org/10.1007/s11251-016-9376-y>
- Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology, 59*(5), 320-324. doi:<http://dx.doi.org/10.1037/h0026223>
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality, 46*(3), 517-528. doi:<https://doi.org/10.2224/sbp.7054>
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K., & Klieme, E. (2007). Gleicher Unterricht - gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft, 35*(2), 125-147. Retrieved from https://www.pedocs.de/frontdoor.php?source_opus=5489
- Livingston, C., & Borko, H. (1989). Expert-novice differences in teaching - A cognitive analysis and implications for teacher-education. *Journal of Teacher Education, 40*(4), 36-42. doi:<https://doi.org/10.1177/002248718904000407>
- Lizdek, I., Sadler, P., Woody, E., Ethier, N., & Malet, G. (2012). Capturing the stream of behavior: A computer-joystick method for coding interpersonal behavior continuously over time. *Social Science Computer Review, 30*(4), 513-521. doi:<https://doi.org/10.1177/0894439312436487>
- McIntyre, N. A., Mainhard, T., & Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze. *Learning and Instruction, 49*, 41-53. doi:<https://doi.org/10.1016/j.learninstruc.2016.12.005>
- McKinney, J. D., Mason, J., Perkinson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology, 67*(2), 198. doi:<http://dx.doi.org/10.1037/h0077012>

-
- Moreno, F., Reina, R., Luis, V., & Sabido, R. (2002). Visual search strategies in experienced and inexperienced gymnastic coaches. *Perceptual and Motor Skills, 95*(3), 901-902. doi:<https://doi.org/10.2466/pms.2002.95.3.901>
- Pauli, C., & Lipowsky, F. (2007). Mitmachen oder Zuhören? Mündliche Schülerinnen- und Schülerbeteiligung im Mathematikunterricht. *Unterrichtswissenschaft, 35*(2), 101-124. Retrieved from https://www.pedocs.de/frontdoor.php?source_opus=5488
- Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), *Master lectures in clinical neuropsychology* (pp. 173-202). Washington, DC: American Psychology Association.
- Prieto, L. P., Sharma, K., Wen, Y., & Dillenbourg, P. (2015). The burden of facilitating collaboration: towards estimation of teacher orchestration load using eye-tracking measures. In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the material conditions of learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015* (Vol. 1). Gothenburg, Sweden: International Society of the Learning Sciences.
- 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. doi:<https://doi.org/10.3102/00028312024001013>
- Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Difference among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal, 28*(1), 63-88. doi:<https://doi.org/10.3102/00028312028001063>
- Sacher, W. (1995). *Meldungen und Aufrufe im Unterrichtsgespräch. Theoretische Grundlagen, Forschungsergebnisse, Trainingselemente und Diagnoseverfahren*. Augsburg, Germany: Wissner.
- Santagata, R., Zannoni, C., & Stigler, J. W. (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. doi:<https://doi.org/10.1007/s10857-007-9029-9>
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision, 11*(5), 1-30. doi:<https://doi.org/10.1167/11.5.9>

-
- Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research*, 9(3), 253-271.
doi:<https://doi.org/10.1007/s10984-006-9012-x>
- Seidel, T., Schnitzler, K., Kosel, C., Stürmer, K., & Holzberger, D. (2020). Student characteristics in the eyes of teachers: Differences between novice and expert teachers in judgment accuracy, observed behavioral cues, and gaze. *Educational Psychology Review*. doi:<https://doi.org/10.1007/s10648-020-09532-2>
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in pre-service teachers. *American Educational Research Journal*, 51(4), 739-771.
doi:<https://doi.org/10.3102/0002831214531321>
- Seidel, T., Stürmer, K., Schäfer, S., & Jahn, G. (2015). How preservice teachers perform in teaching events regarding generic teaching and learning components. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 84-96.
doi:<https://doi.org/10.1026/0049-8637/a000125>
- Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 383-395). Mahwah, NJ: Lawrence Erlbaum.
- Sherin, M. G., & van Es, E. (2002). *Using video to support teachers' ability to interpret classroom interactions*. Paper presented at the Society for Information Technology & Teacher Education International Conference.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
doi:<https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- 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. doi:<https://doi.org/10.1007/s10857-007-9063-7>
- Steffensky, M., Gold, B., Holodyski, 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.
doi:<https://doi.org/10.1007/s10763-014-9607-0>

- Stipek, D. (2002). Good instruction is motivating. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 309-332). San Diego, CA: Academic Press.
- Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates- approaches to validate the observer extended research tool. *Zeitschrift für Psychologie*, 223(1), 54-63. doi:<https://doi.org/10.1027/2151-2604/a000200>
- Stürmer, K., Seidel, T., & Holzberger, D. (2016). Intra-individual differences in developing professional vision: preservice teachers' changes in the course of an innovative teacher education program. *Instructional Science*, 44(3), 293-309. doi:<https://doi.org/10.1007/s11251-016-9373-1>
- Stürmer, K., Seidel, T., Müller, K., Häusler, J., & Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft*, 20(1), 75-92. doi:<https://doi.org/10.1007/s11618-017-0731-9>
- Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 105-124). London, England: The MIT Press.
- Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., . . . Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education*, 49, 36-44. doi:<https://doi.org/10.1016/j.tate.2015.01.012>
- van den Bogert, N., van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education*, 37, 208-216. doi:<https://doi.org/10.1016/j.tate.2013.09.001>
- van Es, E., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571-596. Retrieved from <https://www.learntechlib.org/primary/p/9171/>
- van Es, E., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, 24(2), 244-276. doi:<https://doi.org/10.1016/j.tate.2006.11.005>
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42(4), 292-305. doi:<https://doi.org/10.1177/002248719104200407>

-
- Wolff, C. E., 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. doi:<https://doi.org/10.1007/s11251-016-9367-z>
- Yamamoto, T., & Imai-Matsumura, K. (2015). Teachers' gazes and awareness of students' behavior: Using an eye tracker. *Comprehensive Psychology, 2*(6). doi:<https://doi.org/10.2466/01.IT.2.6>

6

General Discussion and Conclusion

6 General Discussion and Conclusion

Students' (non)attention-related behavior can indicate whether students are focusing on the learning material and how intensely they are engaging in the desired cognitive processes. In turn, teachers need to consider individual students' behavior as a point of reference from which to identify potential problems in individual students' learning processes and the need for instructional improvement. The (non)attention-related behavior of all students in a classroom can then serve as an indicator of the current level of teaching quality. Therefore, students' (non)attention-related behavior can provide a great deal of information about students' individual level but also about the interaction level during instruction. However, research has not yet made use of the potential of students' (non)attention-related behavior for helping to provide a closer investigation of the processes involved in teacher-learner interactions. Already existing instruments are not yet suitable for capturing the highly dynamic and situation-specific nature of teacher-learner interactions or else they cannot be used to assess students' behavior in a sufficiently detailed or comprehensive manner.

To address the shortcomings of existing measurement approaches and to investigate the potential of students' (non)attention-related behavior for providing insights into teacher-learner interactions, the present dissertation presented and validated a new observation instrument (Study 1) that provides behavioral indicators that enable the continuous assessment of students' (non)attention-related behavior (CABI). Furthermore, the potential of students' (non)attention-related behavior was evaluated by using the CABI to investigate the mechanisms underlying students' behavior on the individual level (Study 2) as well as on the level of interactions between teachers and learners during instruction (Study 3).

In the following, I first discuss the results of the three empirical studies (Chapter 6.1). This chapter is divided into two parts, reflecting the two overarching research questions in this dissertation. In Chapter 6.1.1, I discuss the adequate measurement of students' (non)attention-related behavior during instruction using the new instrument that was developed in this dissertation. Subsequently, I elaborate on how the empirical studies used the potential of students' (non)attention-related behavior to provide new insights with respect to students' individual level as well as the level of interactions between teachers and learners (Chapter 6.1.2). In Chapter 6.2, I present the strengths and limitations of the present dissertation. The dissertation closes with implications for practice (Chapter 6.3.1) as well as implications for research and future directions (Chapter 6.3.2).

6.1 Discussion of the Results

6.1.1 Measurement of Students' (Non)Attention-related Behavior during Teacher-Learner Interactions

The present dissertation builds on a behavior-oriented definition of attention that conceptualizes attention as a situation-specific selection mechanism with variations in its intensity depending on the demands of the current activity. As outlined in Chapter 1.1.2, students' behavior during teacher-learner interactions can indicate whether students are paying attention. Furthermore, students' behavior can also provide hints about the intensity of their attention, as certain learning activities require more attentional resources than others (Chi & Wylie, 2014; Olney et al., 2015). This logic also holds true for activities that are not content-related because a conversation with the person sitting in the next seat might be more demanding (in terms of attentional load) than looking out the window. According to Lavie (2005), the demands of a given activity, content-related or not, determine how well competing information is processed. With regard to teacher-learner interactions, this means that if a learning activity is quite demanding, students need to invest more attentional resources in it and are less likely to react to distractions. In turn, if a non-content-related activity, such as chatting with a friend, requires a lot of attention, students will most likely fail to process the learning material adequately. Compared with previous work, Lavie's (2005) theory considers not only performance in the relevant activity but also the processing of irrelevant information (Lavie, 2000). Therefore, considering both students' on- and off-task behaviors can provide more valuable information about teacher-learner interactions with regard to (a) individual students' learning and (b) the teacher and the overall instructional process than specific aspects of behavior alone.

As the potential of already existing behavioral observation instruments has yet to be fully exploited, I developed a new behavioral observation scale (the CABI) by considering the opportunities for improvement as presented in Chapter 1.3.3. The continuous nature of the CABI provides intensive longitudinal data on students' (non)attention-related behavior that can be used to study the processes involved in teacher-learner interactions and can be combined with additional measures, such as information about classroom situations (Study 2) or teachers' eye-tracking data (Study 3).

The validity of the CABI was tested in Study 1 by relating students' (non)attention-related behavior to self-reported cognitive engagement, involvement, and situational interest as well as subsequent knowledge test results. The results supported the construct and predictive

validity of the CABI. In contrast to existing instruments, the CABI considers the entire spectrum of possible student behavior by using the ICAP framework (Chi & Wylie, 2014) and Lavie's (2005) load theory as its foundation to inspire a symmetrical scale that covers on- and off-task behaviors with passive to interactive manifestations. Studies 2 and 3 demonstrate how information not only about striking and easy to notice student behaviors, but also about subtle behaviors that carry additional information about students' attentional processes, can improve insights into the processes that determine teacher-learner interactions. The dimensional approach of the CABI reflects the theoretical considerations that underlie the interpretation of students' attention in teacher-learner interactions as presented in the current dissertation. Compared with other behavioral observation instruments (e.g., Cobb, 1972; Helmke & Renkl, 1992; Jackson & Hudgins, 1965; Karweit & Slavin, 1981; Lane & Harris, 2015), the CABI does not classify students' behavior into distinct categories but considers them to be arranged on a broad spectrum that ranges from *interactive off-task* to *interactive on-task* behavior. Considering students' (non)attention-related behavior on a spectrum is more in line with Lavie's (2005) load theory compared with categorical approaches, as Lavie assumed that excess attentional resources spill over to interfering activities until a substantial load of relevant information is reached (Lavie, 1995).

Taken together, the cross-disciplinary theoretical considerations that underlie the CABI and the empirical support of its validity (Study 1) make the CABI a useful tool for measuring students' (non)attention-related behavior to investigate teacher-learner interaction processes during instruction.

6.1.2 New Insights Offered by Students' (Non)Attention-related Behavior during Teacher-Learner Interactions

Existing observational instruments that are deployed in classroom settings are often restricted to time-sampling. Technological advances and the use of video cameras allow researchers to analyze the behavior of all students for the same instructional period. However, existing instruments commonly use interval-based approaches with a temporal granularity that does not fall below 5 s. Studies 2 and 3 evaluated the potential that the continuous assessment of students' (non)attention-related behavior has to provide new insights into the processes involved in teacher-learner interactions. In the following, I discuss the new insights that were provided on the (a) individual level and (b) interaction level.

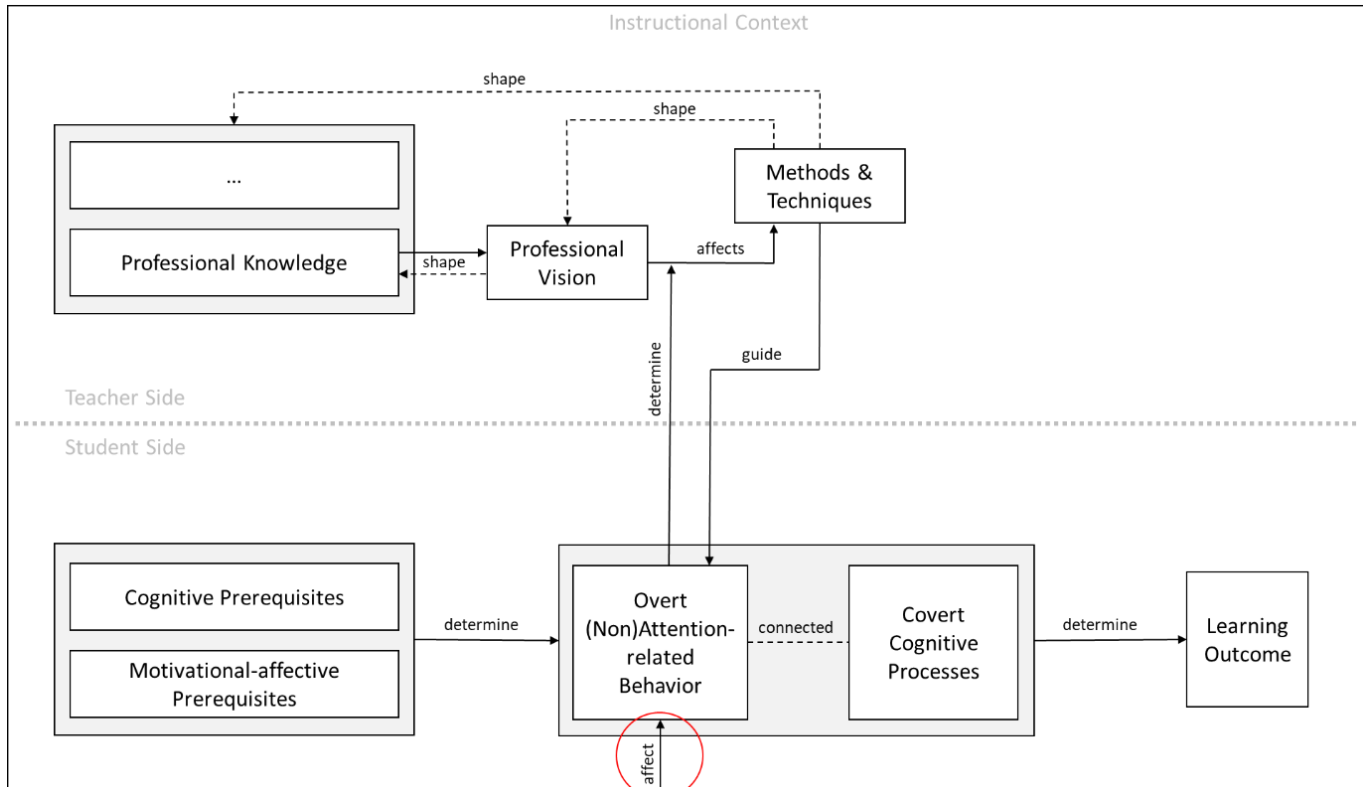
With regard to students' individual level, previous research has suggested that students demonstrate different patterns in their (non)attention-related behavior in the same instructional

settings (e.g., Kelly, 2007). For example, during classroom discussions, some students do not show much variation in their (non)attention-related behavior (e.g., focusing on the teacher all the time, taking notes occasionally). In comparison, other students might exhibit a great deal of variation in their (non)attention-related behavior because they may shift between on- and off-task as well as interactive and passive behaviors (e.g., chatting with friends about the weekend, looking out the window, raising one's hand and contributing to classroom discussions). These differences might be caused by individual characteristics as well as classroom-specific components (see Chapter 1.1.4). However, existing approaches commonly use aggregated measures of students' (non)attention-related behavior that average out situational effects, which may provide more accurate insights into the effects of specific determinants (Diener & Larsen, 2009). These insights were provided in Study 2. The analysis of students' (non)attention-related behavior during classroom instruction suggests that classroom-specific components have a great impact on students' (non)attention-related behavior but that within the same classroom, variability in students' momentary (non)attention-related behavior appears to be affected by their subject-specific self-concept. However, different demands of specific classroom situations (e.g., classroom discussion or book entries) and especially determinants that are unique to single classrooms appear to explain differences between students the most, emphasizing the strong impact of classroom-specific components. Study 2 pointed out that effects of additional factors – next to individual prerequisites and the teachers' choice of methods and techniques – that affect students' observable behavior but also their internal processes are not yet considered within the conceptual framework to systematize the mechanisms underlying students' (non)attention-related behavior during teacher-learner interactions (Chapter 1.2.3). To account for this incompleteness, I revised the framework depicted in Figure 3 and added an arrow from the instructional context to students' overt and covert learning processes (see Figure 5). The arrow serves as a placeholder for the external factors that are unique to individual contexts and have not yet been investigated in detail with respect to their effect on students in the process of teacher-learner interactions.

Focusing on the interaction level, teachers need professional vision skills to maintain a functional overview so that they can avoid disruptions and identify students who are having trouble understanding the material that is being taught in class (see Chapter 1.2.2). However, previous research has shown that especially novice teachers have trouble distributing their

Figure 7

Revised Conceptual Framework for Systematizing the Mechanisms Underlying Students' (Non)Attention-related Behavior during Teacher-Learner Interactions with the Red Circle Indicating the Additional Arrow.



attentional focus evenly across their students and are often guided by rather interactive behavior (Lipowsky et al., 2007; Wolff et al., 2016). Study 3 demonstrated that when inexperienced teachers are faced with the demand to interact with students while conveying learning material, they were more likely to focus on students who engaged in (inter)active content-related behavior. This means that they preferred to focus on students whose behavior sustained their instruction, such as active participation. This pattern is particularly problematic because teachers need to identify the students who are not paying attention or are at risk of not adequately processing the learning material. According to the theoretical considerations discussed above, students whose actions are rather subtle and who engage in passive learning behaviors might fail to direct their attentional resources toward the learning material. Passive behavior can serve as an indication that the relevant (i.e., learning-related) information is not demanding enough and that excess attentional capacity is being used to (unintentionally) process irrelevant information, thus making students more susceptible to becoming distracted (see Lavie, 1995). Therefore, it is important for teachers to identify students who are engaging in passive behavior and provide the required degree of cognitive activation. In this vein,

effective classroom management should ensure that students' primary activity is related to the learning material, and the degree of cognitive activation should attempt to ensure that students' attentional resources are "occupied" by this task. Students' (non)attention-related behavior serves as an important indicator for teachers to adapt their instructional methods and techniques accordingly. However, Study 3 suggested that inexperienced teachers might fail to provide adequate classroom management and cognitive activation. Contrary to previous research (e.g., Wolff et al., 2016), the results of Study 3 provide hints that novice teachers' attentional focus might be driven by top-down processes (i.e., plans and intentions) that cause them to pursue their instructional agenda instead of monitoring their students to identify problematic behavior. The distribution of novice teachers' attentional focus may be an indicator that classroom management plays a subordinate role for them and that they fail to succeed in providing adequate cognitive activation because they instead tend to focus on their very (inter)active students and tend to leave out the students whose behavior is more subtle. Therefore, inexperienced teachers may be more concerned with covering a certain topic during the given time and may be less concerned about aspects of teaching quality. Insights such as these provide critical implications for teacher training (see Chapter 6.3.1) as well as research on teacher-learner interactions (see Chapter 6.3.2) and are enabled by the combination of different process measures. The triangulation of teachers' eye-tracking data and information on their behavior with the continuous annotation of students' (non)attention-related behavior offers new insights into how students' (non)attention-related behavior affects where teachers focus their attention during instruction.

6.2 Strengths and Limitations

When evaluating the measurement and potential of students' (non)attention-related behavior to provide insights into teacher-learner interactions as was done in the three empirical studies, there are some general strengths and limitations that need to be considered. As all three studies used the CABI to provide information about students' (non)attention-related behavior during teacher-learner interactions, all of them benefitted from the continuous nature and intensive longitudinal data structure. This structure provided the opportunity to deploy interdisciplinary approaches that bore the potential to provide a new perspective on educational research. Study 1 demonstrated how continuous information about students' behavior can be used as a ground truth for training machine-learning algorithms. This fosters the collaboration with computer science, which will be necessary to make use of technological advances now and in the future (see Chapter 6.3.2). Studies 2 and 3 provide examples of how the statistical

methods that are common practice in economics (i.e., time series analysis) and are fairly new to research in psychology and education science (e.g., dynamic structural equation modeling) can be used to adequately analyze the respective data sets. The structure of the continuous data and the statistical methods involved allowed us to investigate processes and situation-specific mechanisms in a way that has not yet been possible. However, there are also some limitations that should be kept in mind.

First, students appear to exhibit (non)attention-related behavior in a somewhat unbalanced manner, meaning that some behaviors – especially from the lower end of the scale (i.e., interactive off-task) – only occur rarely during teacher-learner interactions. On a practical level, this is obviously a good sign, as extremely disruptive behavior is generally undesirable during instruction and teacher-learner interactions. However, this imbalance can cause problems for the statistical analyses. For example, in Study 1, students only sparsely displayed interactive on-task behaviors, such as explaining material to their peers, and even fewer interactive off-task behaviors, such as walking around with the intention to interrupt the class. Most of the (non)attention-related behavior that students exhibited was dispersed across the passive and active on-task areas on the behavioral scale, reducing the accuracy of the automated analysis. Similarly, too little variation in students' (non)attention-related behavior might have caused distortions in some statistical results in Study 3. However, this issue was not related to the CABI as such, as unbalanced data have already been reported in Helmke and Renkl (1992). This problem for statistical analysis is due to a certain degree of structure and regulation that appears to be present in most teacher-learner interactions and instructional settings. Nevertheless, approaches for measuring students' (non)attention-related behavior should still consider the entire spectrum of possible behaviors to make the instruments valid. Even though certain behaviors might only occur rarely, they are still within the realm of possibility.

Second, the approximation of students' attentional processes by means of their behavior might provide reasons for doubt. As discussed in Chapter 1.1.2, a theoretical and empirical foundation for the strong connection between overt behaviors and internal cognitions exists. However, as I already emphasized above, students' (non)attention-related behavior might be just an estimate of students' orientation and level of attention without absolute certainty. Even though the assessment of students' (non)attention-related behavior with the CABI works well for most students, but it most likely provides less exact estimations for some students. On the basis of characteristics such as personality, individual students will express their attention differently compared with others. For example, some students' behavior might be more

pronounced than other students', making it easier for external observers to notice variations. Additionally, some students (especially when they are younger and possess less impulse control) might display more noise in their behavior by being rather fidgety and restless. The challenge increases for the observer to estimate the current level of (non)attention-related behavior by identifying meaningful and important behavioral cues. The behavioral annotations of students that were conducted in the three empirical studies repeatedly showed that certain situations required more inferences than expected. Such situations involved conversations between students, as it was rather tricky to estimate the extent to which the conversations were related to the learning material. Even though there were some indicators that provided hints about the conversation topic (e.g., when students keep their upper body turned to the teacher or when students pointed to the blackboard while talking, we could infer that they were discussing the learning material), reliable estimations were more difficult in these situations.

Finally, Studies 2 and 3 were conducted in an exploratory manner. Therefore, the sample sizes were rather small, and the results need to be interpreted carefully. These studies should not be expected to provide stable and universally valid information about teacher-learner interactions. Rather, they provide initial insights into how students' continuous (non)attention-related behavior can be used to further investigate the processes involved in teacher-learner interactions.

6.3 General Implications and Future Directions

6.3.1 Implications for Practice

Students' (non)attention-related behavior and the insights it provides into teacher-learner interactions have practical implications, especially with regard to teacher training. Teacher training is supposed to prepare future teachers for professional practice in a way that goes beyond the theoretical education at university (Grossman et al., 2009). For example, teachers need to make adaptive decisions about their choice of methods and techniques by considering their students' behavior during instruction (Borko et al., 2008). In this context, learning to recognize relevant elements of practice (i.e., relevant cues in students' behavior) comprises a crucial part of professional development and requires teachers to acquire the theoretical knowledge of which instructional technique is suitable under which circumstances (Grossman et al., 2009). Whereas teachers' professional knowledge is covered by university-based curricula, a crucial question arises: How do novice teachers learn which behavioral cues are important for making decisions about their instructional process? Recent research found

that expert and novice teachers differ in how much attention they pay to students' behavior that might signal the need for adaptive teaching decisions (Seidel et al., 2020). It has been assumed that the underlying professional knowledge structures develop over time (Stürmer et al., 2016). Results from expertise research have indicated that regardless of the domain, experts have developed attentional skills that allow them to process visual information more effectively than novices (Jarodzka et al., 2010). Even though research has demonstrated that novice teachers can improve their ability to notice relevant classroom events with videos during teacher training, such events tended to be concerned with features of the environment, the learning material, or teacher-learner communication than with student behaviors that might signal internal processes (Sherin & van Es, 2002; Star & Strickland, 2008). Therefore, explicit information about the identification and interpretation of relevant cues in students' (non)attention-related behavior can further support novice teachers' ability to systematize the visual information that is provided during teacher-learner interactions. This includes knowledge about how behavior can provide indications for the direction and intensity of students' current attention as well as information about the determinants of students' (non)attention-related behavior, such as students' individual prerequisites but also the differences in behavior variations due to demands of the current situations (see Study 2). Furthermore, situations of practice should be provided during teacher training to enable novice teachers to learn how to notice even subtle cues in students' behavior. To avoid overwhelming novice teachers with the complexity and demands of classroom teaching, the approximations to practice should be implemented in a stepwise manner (see the Approximation of Practice Framework; Grossman et al., 2009). This approach allows novice teachers to systematically link their professional knowledge to corresponding elements of professional practice (Seidel et al., 2015). Grossman et al. (2009) suggested that the approximations to practice fall along a continuum with activities ranging from *less authentic* to *more authentic* regarding novice teachers' roles as the teachers in teacher-student interactions. As the results of Study 3 indicate, novice teachers distribute their attentional focus differently in authentic teaching situations with reduced complexity compared with less authentic settings (e.g., when watching videos; Wolff et al., 2016). Novice teachers were more likely to focus on salient student behaviors and tended to overlook subtle cues. Therefore, it is important to provide even more gradations and possible settings that approximate teachers' classroom practice to various degrees so that novice teachers can learn to notice and interpret the relevant information provided by students' (non)attention-related behavior.

One avenue for addressing this issue might be to utilize the technological advances recently made in fields such as immersive virtual reality (Blascovich et al., 2002). Immersive virtual realities are computer-generated simulated environments (Blascovich et al., 2002) that can be deployed to create teaching and learning settings with varying degrees of complexity and authenticity. This possibility for variation makes immersive virtual realities a useful tool for supplementing already existing approaches, such as standardized teaching situations, to make the transition into the real classroom easier for novice teachers. Furthermore, virtual students can be programmed to exhibit salient as well as subtle behaviors so that novice teachers can practice identifying the relevant information.

6.3.2 Implications for Research and Future Directions

The results of this dissertation demonstrate how continuous information on students' behavior can be used to implement process-oriented research on teacher-learner interactions. To achieve an overarching perspective on the mechanisms underlying students' (non)attention-related behavior, research needs to consider students' individual level as well as the level of interactions in the instructional context (see Chapter 1).

Determinants of students' (non)attention-related behavior identified at the individual level provide important implications for the interaction level. For example, Study 2 showed that students with a higher self-concept of ability appear to engage in interactive attention-related behavior more often. This finding supports the idea that teachers need to provide a safe and supportive environment so that all students feel equally encouraged to engage in the instruction process, for example, so that students with lower self-concepts feel safe participating in classroom discussions. When a teacher enables equal participation among students, regardless of their self-concept or other individual prerequisites (e.g., during classroom discussions), this can serve as an indicator of the overall level of teaching quality. The dimensions of teaching quality consider how well teachers succeed in keeping their students attentive and motivated and how well teachers promote their students' understanding (Praetorius et al., 2018). Therefore, students' (non)attention-related behavior can provide insights into how students perceive the current quality of instruction by considering how they respond to the techniques and methods the teacher uses to implement student support, classroom management, and cognitive activation. Additionally, deviations of individual students from the majority of the class in terms of (non)attention-related behavior can signal the need for teacher's additional attention, as these students might struggle or fail to understand the material. In turn, the greater the variation in students' behavior, the more conclusions can be drawn about the current level

of teaching quality. When all students are taking notes, for example, this might signal a high level of classroom management and a high degree of cognitive activation; but when various students are rummaging through their belongings or playing on their phones, classroom management and cognitive activation should be adapted. However, the situational context determines the adequacy of students' behavior and needs to be considered as well. Therefore, understanding the determinants of individual students' behavior can help derive process-specific mechanisms within teacher-learner interactions, thus providing the opportunity to improve classroom instruction. Future research needs to address the individual level as well as the interaction level to create a deeper understanding of teacher-learner processes during instruction.

In addition to research on the quality and determinants of instruction, moment-to-moment analysis can help provide further insights into teachers' internal processes and the determinants that guide their behavior with regard to research on teachers' professional vision within classroom instruction. Teachers' professional vision skills have found their application in the situated assessment of teachers' knowledge (Stürmer & Seidel, 2017). When teachers become good at noticing and interpreting the relevant cues involved in various classroom situations, this indicates integrated and distinct knowledge representations with the possibility for flexible application (Stürmer & Seidel, 2015). By contrast, when teachers fail to identify relevant features during instruction, they are assumed to possess rather disconnected knowledge structures without the ability to make versatile use of this knowledge (Stürmer & Seidel, 2017). However, the results from Study 3 indicate that depending on the circumstances in which teachers' professional vision is examined (i.e., on-action vs. in-action scenarios), they might underlie context-specific demands and intentions that distort the conclusions that can be drawn about teachers' knowledge structures. For example, the Observer tool (Seidel et al., 2010) provides a standardized and contextualized assessment of teachers' professional vision by using classroom videos (i.e., the on-action scenario). Whereas this approach works quite well for making inferences about teachers' integrated knowledge, teachers nevertheless might act differently during real classroom instruction (i.e., in-action scenarios). Depending on their level of expertise, teachers might have different priorities for how to act in teacher-learner interactions (e.g., following their instructional agenda vs. reacting to students' needs in adaptive ways). To reveal the processes that can account for these differences, the combination of information about students' (non)attention-related behavior and information about teachers' focus of attention provides a promising starting point (see Study 3). Therefore, more research

on the factors that determine teachers' professional vision during teaching is necessary so that such insights can be used to improve how teachers prepare to teach.

The present dissertation demonstrates that students' (non)attention-related behavior can serve as a useful approximation of students' attentional processes. When measured continuously and combined with additional process-oriented approaches, students' (non)attention-related behavior can provide new insights into the processes involved in teacher-learner interactions. This idea can similarly be transferred or extended by applying other measurement instruments that provide situation-specific data. Future research might thus add physiological measures, such as EDA or EEG on the individual level, to gather more profound data about students' cognitive processes. Recent research has further supported the advantage of multiperson process data for investigating the reciprocal aspects of teacher-learner interactions: Haataja et al. (2020) combined gaze-tracking data from teachers and students with a continuous coding of teachers' behavior to analyze situation-specific processes of nonverbal interactions. This is an excellent example of how the combination of different continuous measurement approaches can advance research on the level of interactions between teachers and learners.

With regard to future directions for research on the measurement of students' (non)attention-related behavior in teacher-learner interactions, the present dissertation also provides some implications. As computer-science research becomes more advanced and finds applications in various areas of our lives, the possibilities for automated observational measurements increase. In order to find alternatives and tackle feasibility problems, computer scientists recently developed machine-vision-based approaches that rely on visual features to measure students' attention and engagement, such as gaze, head movement, (upper) body posture, and facial expressions (e.g., D'Mello et al., 2017; Raca, 2015; Whitehill et al., 2014; Zaletelj & Košir, 2017). Study 1 also contributes to this growing field of research by providing a proof of concept for a machine-vision-based approach for assessing visible indicators of students' (non)attention-related behavior.

However, more research is needed to improve knowledge about how certain overt behavioral patterns are related to their covert intentions and underlying processes (Girard & Cohn, 2016). Human observers possess a natural understanding of how to interpret behaviors. The challenge lies in the translation of this natural understanding, which involves a certain degree of inference, into a set of highly objective features and patterns. To achieve this goal, research is faced with challenges on several fronts and requires interdisciplinary collaborations

that combine detailed knowledge about the underlying psychological constructs and the technical expertise to develop the respective computational approaches. First, the theoretical considerations underlying the construct of interest need to be revised and refined to gain a better understanding of exactly what to measure. As outlined in the theoretical background of this dissertation, the combination and integration of different theoretical approaches can help to specify, for example, the definition and scope of (non)attention-related behavior in teacher-learner interactions. Second, extensive research on the determinants of individual behavior is required. As Study 2 pointed out, there might be factors that affect the way learners behave that have not yet been investigated properly because the respective capacities were missing. When the factors that influence behaviors are known, they can be measured objectively with automated measurements too (e.g., computer-based estimation of personality and the consideration of this information in the interpretation of behavioral patterns). Therefore, third, additional indicators need to be incorporated into the automated analysis. This might include another set or combination of visual features as well as rather external information. For example, Study 1 showed that the estimations improved when a student's behavior was related to the behavior of the person sitting next to them. Information about how other people behave can thus provide additional indicators of the adequate interpretation of behavioral cues under specific circumstances. Fourth, more emphasis should be placed on investigating the observations. When more is known about how exactly humans observe behaviors, this knowledge can also be used to improve automated measurements. One avenue by which to address this undertaking would be to use think-aloud protocols or cued retrospective recall to verbalize an observer's thoughts and reference points. Another approach can involve eye tracking to investigate the visual areas that human observers take into account objectively.

Overall, recent developments in computer science seem promising but still have to be improved to achieve a reasonable amount of comparability to human observers. However, research should make use of the opportunities technology provides as behavioral observations conducted by humans will not be feasible in the long run.

6.4 Conclusion

The present dissertation provides a cross-disciplinary view on the importance of attention and its relation to observable behaviors during teacher-learner interactions. It further elaborates on the adequate assessment of students' (non)attention-related behavior and presents and validates a new observation instrument called the CABI. This instrument is based on current

theoretical models on the selectivity and capacity of attention in general as well as conceptual assumptions that are specific to the instructional context. In contrast to existing instruments, it allows researchers to capture the entire possible spectrum of students' (non)attention-related behavior in a continuous manner. The continuous data structure of student' (non)attention-related behavior can be used on the individual level of students but also on the level of interactions between teachers and learners to investigate processes and their underlying mechanisms. The integration of the two levels can provide insights that can in turn be used to improve teacher training and enable a new perspective for further research in this area. In addition to its rather content-related contribution to research on teacher-learner-interaction processes and the implications for teacher training, the present dissertation also provides important foundations for future work on fully automated assessments. The theoretical considerations can support a better understanding of aspects that are actually relevant for estimating students' attention during instruction.

Therefore, even though “every one knows what attention is” (p. 403; James, 1890), research on students' attention can still provide new insights and will continue to open up new insights into the processes involved in teacher-learner interactions. The present dissertation points out the potential that students' (non)attention-related behavior has for current and future work.

7

References

7 References

- Alimoglu, M. K., Sarac, D. B., Alparslan, D., Karakas, A. A., & Altintas, L. (2014). An observation tool for instructor and student behaviors to measure in-class learner engagement: a validation study. *Medical Education Online*, *19*(1), 24037. <https://doi.org/10.3402/meo.v19.24037>
- Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 12-153). Routledge & Kegan Paul.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.
- Anderman, L. H., & Anderman, E. M. (1999). Social predictors of changes in students' achievement goal orientations. *Contemporary Educational Psychology*, *24*(1), 21-37. <https://doi.org/10.1006/ceps.1998.0978>
- Anderson, J. R., & Funke, J. (2013). *Kognitive Psychologie*. Springer.
- Aronen, E. T., Vuontela, V., Steenari, M. R., Salmi, J., & Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance at school. *Neurobiology of Learning and Memory*, *83*(1), 33-42. <https://doi.org/https://doi.org/10.1016/j.nlm.2004.06.010>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of Learning and Motivation* (Vol. 2, pp. 89-195). Elsevier.
- Babiker, A., Faye, I., Mumtaz, W., Malik, A. S., & Sato, H. (2019). EEG in classroom: EMD features to detect situational interest of students during learning. *Multimedia Tools and Applications*, *78*(12), 16261-16281. <https://doi.org/10.1007/s11042-018-7016-z>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press <https://doi.org/10.1017/CBO9781139017343>
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system". *Conference on Human Factors in Computing Systems*, 383-390. <https://doi.org/10.1145/985692.985741>

- Baumert, J., Artelt, C., Carstensen, C. H., Sibberns, H., & Stanat, P. (2002). Untersuchungsgegenstand, Fragestellungen und technische Grundlagen der Studie. In *PISA 2000—Die Länder der Bundesrepublik Deutschland im Vergleich* (pp. 11-38). Springer.
- Beyda, S. D., Zentall, S. S., & Ferko, D. J. K. (2002). The relationship between teacher practices and the task-appropriate and social behavior of students with behavioral disorders. *Behavioral Disorders*, 27(3), 236-255. www.jstor.org/stable/43153386
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2), 103-124. https://doi.org/10.1207/S15327965PLI1302_01
- Blömeke, S., Gustafson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3-13. <https://doi.org/10.1027/2151-2604/a000194>
- Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction*, 65, 101250. <https://doi.org/10.1016/j.learninstruc.2019.101250>
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473-498. <https://doi.org/10.3102/00028312026004473>
- Borko, H., Roberts, S. A., & Shavelson, R. (2008). Teachers' decision making: From Alan J. Bishop to today. In P. Clarkson & N. Presmeg (Eds.), *Critical issues in mathematics education: Major contributions of Alan Bishop* (pp. 37-67). Springer US. https://doi.org/10.1007/978-0-387-09673-5_4
- Broadbent, D. E. (1958). *Perception and communication*. Elsevier.
- Bromme, R. (1992). *Der Lehrer als Experte*. Hans Huber.
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95-108. <https://doi.org/10.1016/j.learninstruc.2009.11.004>
- Brünken, R., & Seufert, T. (2006). Aufmerksamkeit, Lernen, Lernstrategien. In H. Mandl & H. F. Friedrich (Eds.), *Handbuch Lernstrategien* (pp. 27-37). Hogrefe.
- Büttner, G., & Schmidt-Atzert, L. (2004). *Diagnostik von Konzentration und Aufmerksamkeit*. Hogrefe.

- Carta, J. J., Greenwood, C. R., Schulte, D., Arreaga-Mayer, C., & Terry, B. (1988). *Code for Instructional Structure and Student Academic Response: Mainstream version (MS-CISSAR)*. Juniper Gardes Children's Project, Bureau of Child Research, University of Kansas.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.
<https://doi.org/10.1080/00461520.2014.965823>
- Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. Springer. <https://doi.org/10.1007/978-1-4614-2018-7>.
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology, 62*(1), 73-101.
<https://doi.org/10.1146/annurev.psych.093008.100427>
- Clarke, S. N., Howley, I., Resnick, L., & Penstein Rosé, C. (2016). Student agency to participate in dialogic science discussions. *Learning, Culture and Social Interaction, 10*, 27-39. <https://doi.org/10.1016/j.lcsi.2016.01.002>
- Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology, 63*(1), 74-80.
<https://doi.org/10.1037/h0032247>
- Cobb, J. A., & Hops, H. (1973). Effects of academic survival skill training on low achieving first graders¹. *The Journal of Educational Research, 67*(3), 108-113.
<https://doi.org/10.1080/00220671.1973.10884576>
- Cohen, R. A. (2014). *The neuropsychology of attention* (2nd ed.). Springer.
<https://doi.org/10.1007/978-0-387-72639-7>
- Connell, J. P. (1990). Context, self, and action: A motivational analysis of self-system processes across the life span. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (Vol. 8, pp. 61-97). The University of Chicago Press.
- Corbetta, M., & Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews, 3*, 201-215. <https://doi.org/10.1038/nrn755>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114.
<https://doi.org/10.1017/S0140525X01003922>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.
[https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)

-
- D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104-123. <https://doi.org/10.1080/00461520.2017.1281747>
- D'Mello, S. K. (2016). Giving eyesight to the blind: Towards attention-aware AIED. *International Journal of Artificial Intelligence in Education, 26*(2), 645-659. <https://doi.org/10.1007/s40593-016-0104-1>
- Darling-Hammond, L., & Bransford, J. D. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. Jossey-Bass.
- Dermody, S. S., Thomas, K. M., Hopwood, C. J., Durbin, C. E., & Wright, A. G. C. (2017). Modeling the complexity of dynamic, momentary interpersonal behavior: Applying the time-varying effect model to test predictions from interpersonal theory. *Journal of Research in Personality, 68*, 54-62. <https://doi.org/10.1016/j.jrp.2017.03.001>
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80-90. <https://doi.org/10.1037/h0039515>
- Dhamija, S., & Boulton, T. E. (2018). *Automated action units vs. expert raters: Face off* IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV. <https://doi.org/10.1109/Wacv.2018.00035>
- Diederich, J., & Tenorth, H.-E. (1997). *Theorie der Schule. Ein Studienbuch zu Geschichte, Funktionen und Gestaltung*. Cornelsen.
- Diener, E., & Larsen, R. J. (2009). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener* (pp. 7-24). Springer Netherlands. https://doi.org/10.1007/978-90-481-2354-4_2
- Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of Research in Education, 5*(1), 163-198. <https://doi.org/10.3102/0091732x005001163>
- Edelberg, R. (1972). Electrodermal recovery rate, goal-orientation, and aversion. *Psychophysiology, 9*(5), 512-520. <https://doi.org/10.1111/j.1469-8986.1972.tb01805.x>
- Ehrhardt, K. J., Findeisen, P., Marinello, G., & Reinartz-Wenzel, H. (1981). Systematische Verhaltensbeobachtung von Aufmerksamkeit im Unterricht: Zur Prüfung von Objektivität und Zuverlässigkeit [Systematic behavior observation of attention in class: Proof of objectivity and reliability]. *Diagnostica, 27*(4), 281-294.
- 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. https://doi.org/10.1207/S15326985EP3602_5

- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford University Press.
- Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, 95(5), 421-434. <https://doi.org/10.1086/461853>
- Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1981). Teaching behaviors, academic learning time, and student achievement: An overview. *The Journal of Classroom Interaction*, 17(1), 2-15. www.jstor.org/stable/43997772
- Foster, S. L., Bell-Dolan, D. J., & Burge, D. A. (1988). Behavioral observation. In A. S. Bellack & M. Hersen (Eds.), *Pergamon general psychology series* (3rd ed., Vol. 65, pp. 119-160). Pergamon Press.
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision research*, 51(17), 1920-1931. <https://doi.org/10.1016/j.visres.2011.07.002>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763-782). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_37
- Friedman, D. L., Cancelli, A. A., & Yoshida, R. K. (1988). Academic engagement of elementary school children with learning disabilities. *Journal of School Psychology*, 26(4), 327-340. [https://doi.org/10.1016/0022-4405\(88\)90041-6](https://doi.org/10.1016/0022-4405(88)90041-6)
- Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software*, 2(1), e5. <https://doi.org/10.5334/jors.ar>
- Girard, J. M., & Cohn, J. F. (2016). A primer on observational measurement. *Assessment*, 23(4), 404-413. <https://doi.org/10.1177/1073191116635807>
- Goldhammer, F., & Moosbrugger, H. (2006). Aufmerksamkeit. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp. 16-33). Springer. https://doi.org/10.1007/3-540-33020-8_2

- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96, 606-633.
<https://doi.org/10.1525/aa.1994.96.3.02a00100>
- Greenwood, C. R., Carta, J. J., Kamps, D., Terry, B., & Delquadri, J. (1994). Development and validation of standard classroom observation systems for school practitioners: Ecobehavioral assessment systems software (EBASS). *Exceptional Children*, 61(2), 197-210. <http://hdl.handle.net/1808/10954>
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2055-2100. <https://tedd.org/wp-content/uploads/2014/03/Grossman-et-al-Teaching-Practice-A-Cross-Professional-Perspective-copy.pdf>
- Grossman, P., & McDonald, M. (2008). Back to the future: directions for research in teaching and teacher education. *American Educational Research Journal*, 45(1), 184-205.
<https://doi.org/10.3102/0002831207312906>
- Haataja, E., Salonen, V., Laine, A., Toivanen, M., & Hannula, M. S. (2020). The relation between teacher-student eye contact and teachers' interpersonal behavior during group work: A multiple-person gaze-tracking case study in secondary mathematics education. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09538-w>
- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, 113(4), 461-487. <https://doi.org/10.1086/669616>
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in cognitive sciences*, 9(4), 188-194. <https://doi.org/10.1016/j.tics.2005.02.009>
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Sage.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (2 ed.). Klett Kallmeyer.
<http://www.worldcat.org/oclc/316016742>
- Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, 38, 130-141.

- Hommel, M. (2012). Kodierhandbuch des Beobachtungsinventars zur systematischen und videobasierten Erfassung der Aufmerksamkeit von Lernenden. In *Dresdner Beiträge zur Wirtschaftspädagogik* (Vol. 1). Technische Universität Dresden.
- Imhof, M. (2004). Aufmerksamkeit in der Schule. In G. Büttner & L. Schmidt-Atzert (Eds.), *Diagnostik von Konzentration und Aufmerksamkeit* (pp. 233-247). Hogrefe.
- Jackson, P., & Hudgins, B. (1965). *Observation schedule for recording pupil attention*.
- James, W. (1890). Attention. In *The principles of psychology* (Vol. 1, pp. 402-458). Dover.
- Janssen, M., Chinapaw, M. J. M., Rauh, S. P., Toussaint, H. M., van Mechelen, W., & Verhagen, E. A. L. M. (2014). A short physical activity break from cognitive tasks increases selective attention in primary school children aged 10–11. *Mental Health and Physical Activity*, 7(3), 129-134. <https://doi.org/10.1016/j.mhpa.2014.07.001>
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20, 146-154. <https://doi.org/10.1016/j.learninstruc.2009.02.019>
- Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, 23, 33-42. <https://doi.org/10.1016/j.learninstruc.2012.09.002>
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall Inc. https://scholar.princeton.edu/sites/default/files/kahneman/files/attention_hi_quality.pdf
- Kamps, P., & Oberdörster, M. (2002). Akustik in Klassenzimmern. In L. Huber, J. Kahlert, & M. Klatter (Eds.), *Die akustisch gestaltete Schule* (pp. 89-100). Vandenhoeck & Ruprecht.
- Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal*, 18(2), 157-171. <https://doi.org/10.3102/00028312018002157>
- Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, 74(6), 844-851. <https://doi.org/10.1037/0022-0663.74.6.844>
- Kelly, S. (2007). Classroom discourse and the distribution of student engagement. *Social Psychology of Education*, 10(3), 331-352. <https://doi.org/10.1007/s11218-007-9024-0>
- Kelly, S. (2008). Race, social class, and student engagement in middle school English classrooms. *Social Science Research*, 37(2), 434-448. <https://doi.org/10.1016/j.ssresearch.2007.08.003>

-
- Kersting, N., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568-589. <https://doi.org/10.3102/0002831212437853>
- Klieme, E. (2018). Unterrichtsqualität. In M. Gläser-Zikuda, M. Haring, & C. Rohlf's (Eds.), *Handbuch Schulpädagogik* (pp. 393–408). Waxmann.
- Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000 - Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 333-360). Leske + Budrich.
- Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. *Zeitschrift für Pädagogik*, 54(2), 222-237. https://www.pedocs.de/frontdoor.php?source_opus=4348
- Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P., & Lin, C.-T. (2017). Sustained attention in real classroom settings: An EEG study. *Frontiers in Human Neuroscience*, 11(388). <https://doi.org/10.3389/fnhum.2017.00388>
- Kollar, I., & Fischer, F. (2019). Lehren und Unterrichten. In D. Urhahne, M. Dresel, & F. Fischer (Eds.), *Psychologie für den Lehrberuf* (pp. 333-351). Springer. https://doi.org/10.1007/978-3-662-55754-9_17
- 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. <https://doi.org/10.1016/j.tate.2009.05.001>
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Rinhart & Winston.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neuband, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716-725. <https://doi.org/10.1037/0022-0663.100.3.716>
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, 44(3), 197-203. <https://doi.org/10.1007/s11251-016-9376-y>

- Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology*, 59(5), 320-324.
<https://doi.org/10.1037/h0026223>
- Lane, E. S., & Harris, S. E. (2015). A new tool for measuring student behavioral engagement in large university classes. *Journal of College Science Teaching*, 44(6), 83-91.
www.jstor.org/stable/43632000
- Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment*, 8(2), 139-162. https://doi.org/10.1207/S15326977EA0802_04
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology-Human Perception and Performance*, 21(3), 451-468.
<https://doi.org/10.1037/0096-1523.21.3.451>
- Lavie, N. (2000). Selective attention and cognitive control: Dissociating attentional functions through different types of load. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 175-194). MIT Press.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, 9(2), 75-82. <https://doi.org/10.1016/j.tics.2004.12.004>
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality*, 46(3), 517-528. <https://doi.org/10.2224/sbp.7054>
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19(6), 527-537.
<https://doi.org/10.1016/j.learninstruc.2008.11.001>
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K., & Klieme, E. (2007). Gleicher Unterricht - gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, 35(2), 125-147.
https://www.pedocs.de/frontdoor.php?source_opus=5489
- Lizdek, I., Sadler, P., Woody, E., Ethier, N., & Malet, G. (2012). Capturing the stream of behavior: A computer-joystick method for coding interpersonal behavior continuously over time. *Social Science Computer Review*, 30(4), 513-521.
<https://doi.org/10.1177/0894439312436487>
- Maini, R. D. S. (2011). *Teacher training in a proactive approach to classroom behaviour management: Teacher and student outcomes* <http://hdl.handle.net/1807/27591>

-
- Marliave, R., Fisher, C., Filby, N., & Dishaw, M. (1977). *The development of instrumentation for a field study of teaching* (Technical report 1-5). (Beginning Teacher Evaluation Study, Issue.
- Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, 41, 85-139. [https://doi.org/10.1016/S0079-7421\(02\)80005-6](https://doi.org/10.1016/S0079-7421(02)80005-6)
- McDowd, J. M. (2007). An overview of attention: Behavior and brain. *Journal of Neurologic Physical Therapy*, 31(3), 98-103. <https://doi.org/10.1097/NPT.0b013e31814d7874>
- Moran, A. (2012). Concentration: Attention and performance. In S. M. Murphy (Ed.), *The Oxford handbook of sport and performance psychology* (pp. 117-130). Oxford University Press.
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214-255. <https://doi.org/10.1037/0033-295X.86.3.214>
- O'Malley, K. J., Moran, B. J., Haidet, P., Seidel, C. L., Schneider, V., Morgan, R. O., Kelly, P. A., & Richards, B. (2003). Validation of an observation instrument for measuring student engagement in health professions settings. *Evaluation & the Health Professions*, 26(1), 86-103. <https://doi.org/10.1177/0163278702250093>
- Olney, A. M., Risko, E. F., D'Mello, S. K., & Graesser, A. C. (2015). Attention in educational contexts: The role of the learning task in guiding attention. In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 623-642). MIT Press.
- Pauli, C., Drollinger-Vetter, B., Hugener, I., & Lipowsky, F. (2008). Kognitive Aktivierung im Mathematikunterricht. *Zeitschrift für Pädagogische Psychologie*, 22(2), 127-133. <https://doi.org/10.1024/1010-0652.22.2.127>
- Pauli, C., & Lipowsky, F. (2007). Mitmachen oder Zuhören? Mündliche Schülerinnen- und Schülerbeteiligung im Mathematikunterricht. *Unterrichtswissenschaft*, 35(2), 101-124. https://www.pedocs.de/frontdoor.php?source_opus=5488
- Pennings, H. J. M., & Mainhard, T. (2016). Analyzing teacher–student interactions with state space grids. In M. Koopmans & D. Stamovlasis (Eds.), *Complex dynamical systems in education: Concepts, methods and applications* (pp. 223-271). Springer International Publishing. https://doi.org/10.1007/978-3-319-27577-2_12
- Pennings, H. J. M., van Tartwijk, J., Wubbels, T., Claessens, L., van der Want, A., & Brekelmans, M. (2014). Real-time teacher-student interactions: A dynamic systems approach. *Teaching and Teacher Education*, 37, 183-193. <https://doi.org/10.1016/j.tate.2013.07.016>

-
- Peterson, P. L., Swing, S. R., Stark, K. D., & Waas, G. A. (1984). Students' cognitions and time on task during mathematics instruction. *American Educational Research Journal*, 21(3), 487-515. <https://doi.org/10.3102/00028312021003487>
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom process: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119. <https://doi.org/10.3102/0013189X09332374>
- Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365-386). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_17
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40. <https://doi.org/10.1037/0022-0663.82.1.33>
- Piwowar, V., Thiel, F., & Ophardt, D. (2013). Training inservice teachers' competencies in classroom management. A quasi-experimental study with teachers of secondary schools. *Teaching and Teacher Education*, 30, 1-12. <https://doi.org/10.1016/j.tate.2012.09.007>
- Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), *Master lectures in clinical neuropsychology* (pp. 173-202). American Psychology Association.
- Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, 78(5), 391-408. <https://doi.org/10.1037/h0031333>
- Posner, M. I., & Rafal, R. D. (1987). Cognitive theories of attention and the rehabilitation of attentional deficits. In *Neuropsychological rehabilitation* (pp. 182-201). The Guilford Press.
- Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. Scientific American Library/Scientific American Books.
- Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58(1), 1-23. <https://doi.org/10.1146/annurev.psych.58.110405.085516>
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM Mathematics Education*. <https://doi.org/10.1007/s11858-018-0918-4>

-
- Praetorius, A.-K., Lauermann, F., Klassen, R. M., Dickhäuser, O., Janke, S., & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education, 65*, 241-254.
<https://doi.org/https://doi.org/10.1016/j.tate.2017.03.023>
- 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.
<https://doi.org/10.3102/00028312024001013>
- Raca, M. (2015). *Camera-based estimation of student's attention in class* École polytechnique fédérale de Lausanne]. Lausanne, Switzerland.
- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*(3), 700-712. <https://doi.org/10.1037/a0027268>
- Rollett, B. (2001). Die integrativen Leistungen des Gehirns und Konzentration: Theoretische Grundlagen und Interventionsprogramme. In K. J. Klauer (Ed.), *Handbuch Kognitives Training* (2nd ed., pp. 539-557). Hogrefe.
- Rushby, J. A., Barry, R. J., Clarke, A. R., & Vaez Mousavi, M. (2007). Arousal and activation effects on physiological and behavioral responding during a continuous performance task. *Acta Neurobiologiae Experimentalis, 67*(4), 461-470.
<https://ro.uow.edu.au/hbspapers/1963/>
- Sacher, W. (1995). *Meldungen und Aufrufe im Unterrichtsgespräch. Theoretische Grundlagen, Forschungsergebnisse, Trainingselemente und Diagnoseverfahren*. Wissner.
- Samuels, S. J., & Turnure, J. E. (1974). Attention and reading achievement in first-grade boys and girls. *Journal of Educational Psychology, 66*(1), 29-32.
<https://doi.org/10.1037/h0035812>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik, 60*(6), 850-866.
https://www.pedocs.de/frontdoor.php?source_opus=14686
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). "Observer" - Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (Vol. 56, pp. 296-306). Beltz.
https://www.pedocs.de/frontdoor.php?source_opus=3438

- Seidel, T., Jurik, V., Häusler, J., & Stubben, S. (2016). Mikro-Umwelten im Klassenverband: Wie sich kognitive und motivational-affektive Schülervoraussetzungen auf die Wahrnehmung und das Verhalten im Fachunterricht auswirken. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer, & F. Schwabe (Eds.), *Bedingungen und Effekte guten Unterrichts* (pp. 65-87). Waxmann.
- Seidel, T., & Reiss, K. (2014). Lerngelegenheiten im Unterricht. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie* (pp. 253-276). Beltz.
- Seidel, T., Schnitzler, K., Kosel, C., Stürmer, K., & Holzberger, D. (2020). Student characteristics in the eyes of teachers: Differences between novice and expert teachers in judgment accuracy, observed behavioral cues, and gaze. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09532-2>
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of educational research*, 77(4), 454–499. <https://doi.org/10.3102/0034654307310317>
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in pre-service teachers. *American Educational Research Journal*, 51(4), 739-771. <https://doi.org/10.3102/0002831214531321>
- Seidel, T., Stürmer, K., Schäfer, S., & Jahn, G. (2015). How preservice teachers perform in teaching events regarding generic teaching and learning components. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 84-96. <https://doi.org/10.1026/0049-8637/a000125>
- Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 383-395). Lawrence Erlbaum.
- Sherin, M. G., Jacobs, V. R., & Phillip, R. A. (2011). Situating the study of teacher noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 3-13). Routledge.
- Sherin, M. G., & van Es, E. (2002). Using video to support teachers' ability to interpret classroom interactions. Proceedings of SITE 2002--Society for Information Technology & Teacher Education International Conference, Nashville, TN.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14. <https://doi.org/10.2307/1175860>

-
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
<https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, 132(6), 946-958. <https://doi.org/10.1037/0033-2909.132.6.946>
- Stanley, S., & Greenwood, C. (1981). *CISSAR: Code for instructional structure and student academic response: Observer's manual*. Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.
- 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. <https://doi.org/10.1007/s10857-007-9063-7>
- Stipek, D. (2002). Good instruction is motivating. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 309-332). Academic Press.
- Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (2000). *The science of self-report: Implications for research and practice*. Lawrence Erlbaum Associates Publishers.
- Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates: Approaches to validating the Observer extended research tool. *Journal of Psychology*, 223(1), 54-63. <https://doi.org/10.1027/2151-2604/a000200>
- Stürmer, K., & Seidel, T. (2017). Connecting generic pedagogical knowledge with practice. In S. Guerriero (Ed.), *Pedagogical Knowledge and the Changing Nature of the Teaching Profession* (pp. 137-149). OECD Publishing.
- Stürmer, K., Seidel, T., & Holzberger, D. (2016). Intra-individual differences in developing professional vision: preservice teachers' changes in the course of an innovative teacher education program. *Instructional Science*, 44(3), 293-309.
<https://doi.org/10.1007/s11251-016-9373-1>
- Stürmer, K., Seidel, T., Müller, K., Häusler, J., & Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft*, 20(1), 75-92.
<https://doi.org/10.1007/s11618-017-0731-9>
- 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.
<https://doi.org/10.1023/A:1022193728205>

- Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *The American Journal of Psychology*, 77(2), 206-219. <https://doi.org/10.2307/1420127>
- Tryon, W. W. (1998). Behavioral observation. In *Behavioral assessment: A practical handbook* (4th ed., pp. 79-103). Allyn & Bacon.
- Turner, J. C., & Patrick, H. (2004). Motivational influences on student participation in classroom learning activities. *Teachers College Record*, 106(9), 1759-1785. <https://doi.org/10.1111/j.1467-9620.2004.00404.x>
- van de Grift, W. J. C. M., Chun, S., Maulana, R., Lee, O., & Helms-Lorenz, M. (2017). Measuring teaching quality and student engagement in South Korea and the Netherlands. *School Effectiveness and School Improvement*, 28(3), 337-349. <https://doi.org/10.1080/09243453.2016.1263215>
- van Es, E., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, 24(2), 244-276. <https://doi.org/10.1016/j.tate.2006.11.005>
- Van Zomeren, A. H., & Brouwer, W. H. (1994). *Clinical neuropsychology of attention*. Oxford University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Whitehill, J., Serpell, Z., Lin, Y., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1), 86-98. <https://doi.org/10.1109/TAFFC.2014.2316163>
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159-177. <https://doi.org/10.1080/14639220210123806>
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2 ed., Vol. 2, pp. 145-178). Taylor & Francis.
- Wolff, C. E., 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. <https://doi.org/10.1007/s11251-016-9367-z>
- Yoshida, R., Nakayama, T., Ogitsu, T., Takemura, H., Mizoguchi, H., Yamaguchi, E., Inagaki, S., Takeda, Y., Namatame, M., Sugimoto, M., & Kusunoki, F. (2014). Feasibility study on estimating visual attention using electrodermal activity. 8th International Conference on Sensing Technology, Liverpool, UK.

-
- Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *EURASIP Journal on Image and Video Processing*, 2017(1), 80. <https://doi.org/10.1186/s13640-017-0228-8>
- Zimbardo, P. G., Gerrig, R. J., & Graf, R. (2008). *Psychologie* (18 ed.). Pearson Studium.
- Zoelch, C., Berner, V.-D., & Thomas, J. (2019). Gedächtnis und Wissenserwerb. In D. Urhahne, M. Dresel, & F. Fischer (Eds.), *Psychologie für den Lehrberuf* (pp. 23-52). Springer. https://doi.org/10.1007/978-3-662-55754-9_2

8

Appendix

Coding Manual

CABI

(Continuous (non)Attention-related Behavior Indicators)

Materials for coding

The assessment of attention-related behavior should be fully recorded by multi-perspective cameras. The video technology should be of high quality, because the rater must pay special attention to facial expressions and gestures.

In addition, high-quality devices and storage media with a large capacity enable a smooth process without disturbing, attention-influencing changes of devices. The number of devices depends on the room, as well as the number of students and teaching staff. If no wireless microphones can be installed in the room, speakers and headphones facilitate the audio quality for subsequent encoding.

Using a seating plan or assigning codes to the learners, makes the coding easier. A simple aid, also for later coding, is to number the learners from the front to the back rows and to have place cards with their respective ID numbers in front of them. For the coding phase a playback medium (for CARMA the VCL media player) must be installed. The computer should work with an appropriate operating system (recommendation: Windows 10) and corresponding performance, because the program requires very high capacity. A large screen can also facilitate the Rating, as the video cannot be zoomed in. The evaluation is carried out via the media quotation system CARMA (Girard, 2014). This software has an integrated time display and is a very suitable software for collecting moment ratings.

CARMA software

CARMA is a media annotation software. It can be used to display both video and audio recordings. The aim of the program is to enable researchers to gather subjective experiences (here: teaching-learning milieus) and observable behavior over time (Girard, 2014).

The software is very easy to use and can be adapted to individual needs. Among other things, the labelling of the scale of upper and lower limits, as well as the numerical range and the visual representation can be adapted to the respective parameters (see Chapter 2.2, Figures 1 to 3). The file can be saved in a format that allows the data to be imported into statistical analysis software (e.g., SPSS or R).

Rating by joystick

The rating is carried out using a joystick. Therefore, how the rating can be done using a mouse will not be evaluated here. The advantages and disadvantages of using a joystick are listed in this chapter:

The Logitech Extreme™ 3D Pro joystick is used for the rating. Rating by joystick has the advantage that there is no necessity to press a button all the time, as when using a mouse. In addition, a better control over the range of the scale is given, as the hand does not have to move in "mini-movements" instead whole arm can be used. Both hands can be put at the joystick for better control over the device.

Unfortunately, using a joystick also has its disadvantages. Keeping the joystick steady is not easy despite using both hands. Therefore, it should be practiced. Moreover, in case of inattention, distractions etc., one can quickly slip on the device, which leads to measurements becoming inaccurate and one might have to repeat the whole rating. Unfortunately, even after consulting Logitech's customer service, no function could be found to lock the joystick in a certain position.

Overall, there are both advantages and disadvantages in using the joystick. Nevertheless, the advantages outweigh the disadvantages. Handling the joystick may require a certain amount of familiarization and practice, but if the rater is well trained it allows quick reactions and fluctuations on the scale.

Settings for CARMA

To work correctly with CARMA, the joystick must be connected before the program is started. When the software is started, a field appears in which you can select whether you want to use the joystick or the computer mouse for rating. In order to get a uniform axis of all ratings, the settings must be fixed before.

Opening *settings*, the axis labels were defined at *Set Axis Labels* as **On Task** for *Upper Axis Label* and as **Off Task** for *Lower Axis Label* (Figure 1). Further settings are defined for the axis numbers (*Set Axis Numbers*). *Axis Minimum Value* is fixed at -2 and *Axis Maximum Value* at 2. The number of axes' levels is set at 5 (Figure 2). Jet is selected as the *colormap* (Figure 3). The sample has a frequency of 20Hz, the values are determined using an interval of 1.0 seconds (Figure 4). After that the media file can be opened at "Open Media File". The rating can begin.

Figure 8

Screenshot of axis labelling

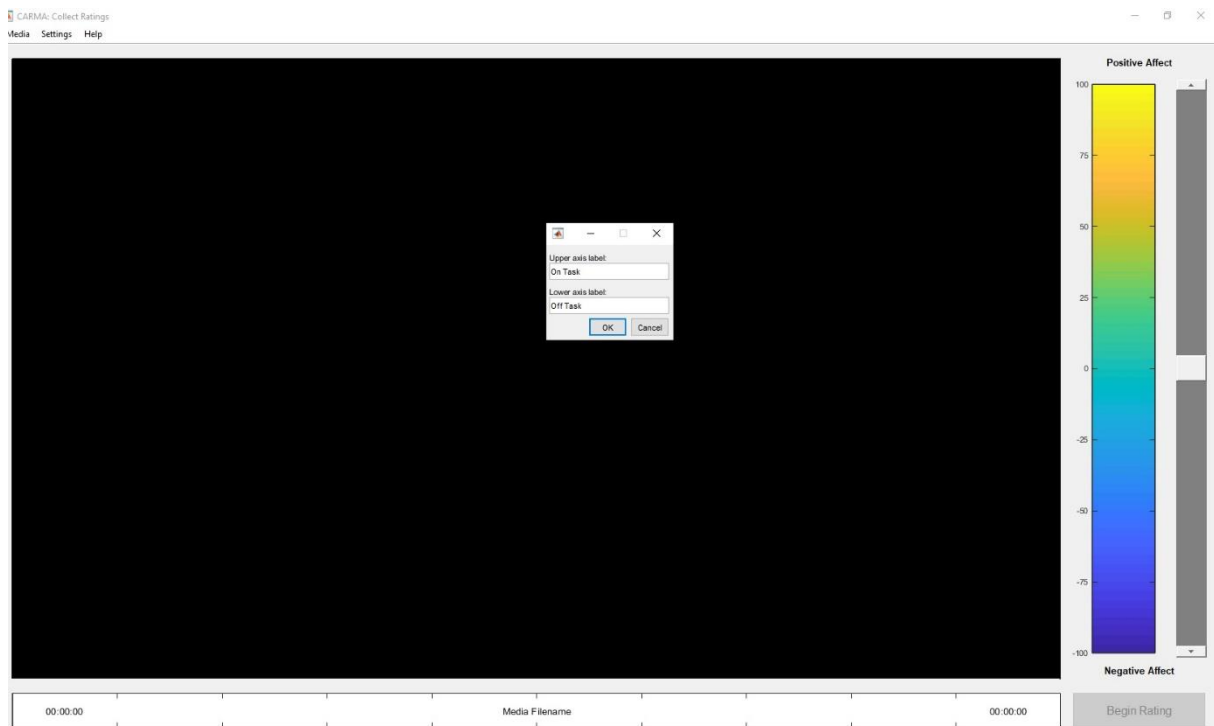


Figure 9

Screenshot Set Axis Numbers

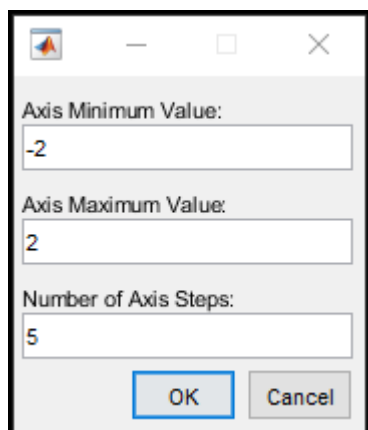
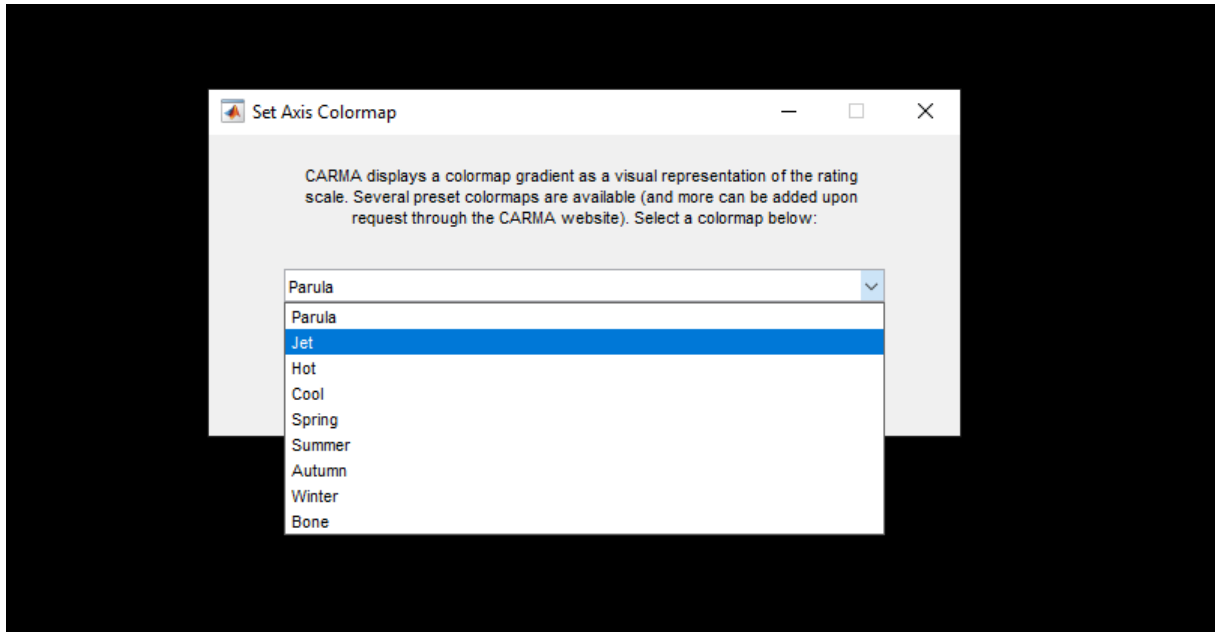
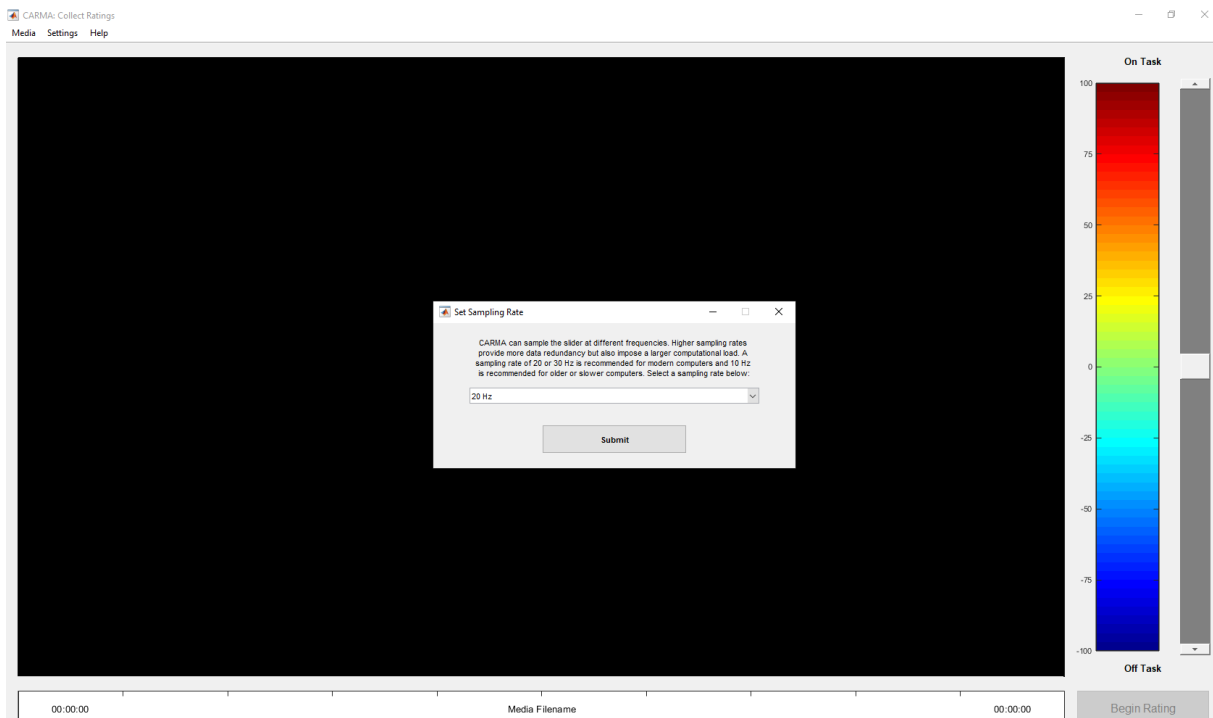


Figure 10*Screenshot Set Axis Colormap***Figure 11***Screenshot Set Sampling Rate*

Coding rules

In order to prevent sequential effects, learners are assessed in random order. For this purpose, a list of ID numbers is created in advance and the sequence is randomized.

During the rating process, the raters focus on only one person and include both verbal and non-verbal appearances of the person to be assessed in the rating. Each person is to be observed individually and, if possible, evaluated for himself/herself in each situation. No information from previous ratings should be included during the rating. In addition, interpretations and assumptions about the observable behavior of learners should be avoided as far as possible. In order to make the rating as precise as possible, the behavior of the person being observed should be responded to as quickly as possible. The corresponding default settings (see chapter 2) are used to code the behavior of the person being observed per second. The risk of assessment errors (see Chapter 2) must be considered. Assessment errors are to be avoided to the best of knowledge and belief.

The risk of assessment errors occurring occurs continuously for each rating. For this reason, the raters should be aware of this between and during the rating process and look it up if necessary. It is useful to keep this in mind and place it next to the workplace as a copy.

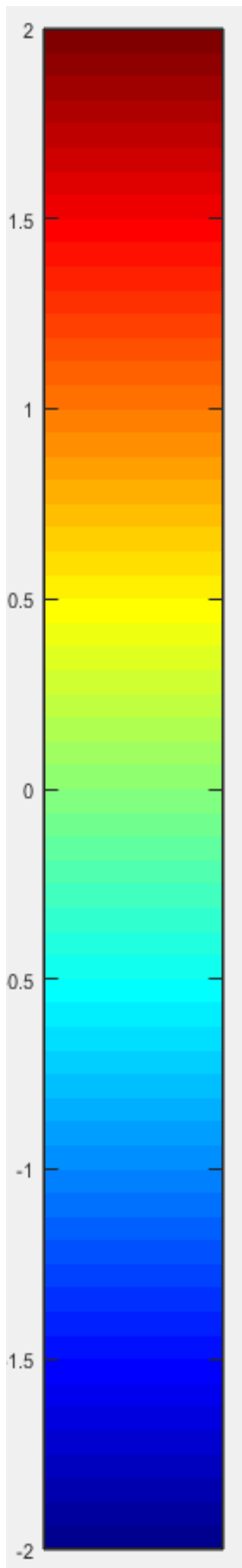
In the CARMA software, no ratings can be interrupted and reloaded at another time. However, there is the possibility to pause the video section briefly. This becomes necessary if an unexpected disturbance of the raters in occurs, or also, so that the raters in can look up something or take a short break, because the concentration is lost. However, the rating video must be worked through to the end so that it can be saved.

Rating scale

The continuous rating scale has been developed to allow assessment over time with as little interpretation by the raters as possible. A one-dimensional, continuous, bipolar scale was chosen, which allows a smooth transition between the two characteristics, on-task and off-task. The value range is divided from +2 to -2. The negative range indicates off-task behavior, the positive range indicates on-task behavior. The zero point indicates the period of time in which a person is absent or is covered in such a way that he or she cannot be recorded or the attention behavior cannot be clearly assessed (e.g., when indicators can be coded as both on-task and off-task). The advantage is that these zero values are coded as "missing values" in subsequent analyses.

The highest possible rating (2) is obtained when one reflects aloud, i.e. critically examines the lessons aloud. This is followed directly by the creation of concept maps and the attempt to explain technical issues to other learners (approx. 1.8). A question is rated with about 1.7, raising hands with 1.5.

Below the line of 1.5, non-verbal communication is recorded. Pointing and gesticulating at something on the blackboard/in the front is rated higher than non-verbal feedback on questions from the teacher, which means nodding, grinning at the teacher, or shaking the head. Repeating and practicing something quietly, as well as copying solution steps and taking notes, is rated just above 1. The rating of 1 is given when learners independently use further documents to look something up, no matter whether it is a book, a script or similar. If the person to be rated does not follow the lessons attentively at one moment (because questions are being discussed), but notes solution steps or similar, this is rated with +0.8, because the person still actively participates in the lessons. An upright posture is rated at around 0.6, depending on whether the chin is raised or not. If the person is turning his or her upper body towards the center of instruction and is not distracted by disturbing classmates, rating is at about 0.5. If the learner's head is supported on the hand while listening, this is rated at approx. 0.4. If the body is "slumping down" and



the head is no longer supported by the hand, but literally falls into the arm, this is rated at 0.25.

In the negative range below zero, behavior is rated that speaks against attention in class. At just under -0.2, it is rated when glances stray from the center of the lesson. If the learner dozes, yawns, or sleeps, or if the head is placed on the table, this is rated at -0.25 - -0.3. If the learner rummages through his or her belongings without disturbing other learners (e.g., to take learning material out of the backpack), this is rated at -0.5. Just under this value is everything that distracts the learner briefly but actively, such as looking at the clock, the mobile phone, or distracting oneself with a pen. If the learner is distracted, but no longer interacts, this can also be the case with classmates, this is rated at -0.6. If the learner is distracted, which can be observed by grinning or laughing, but nobody is disturbed, this is rated -0.75. The occupation should be close to the body and the posture often leans forward. A rating of -1 is given when the posture is strongly averted from the center of instruction. Shortly below this should be rated an (inconspicuous) walk to the wastebasket or the toilet. Fooling around or even "fidgeting around" is rated between -1.25 and -1.3. If the learner is distracted by another disruptive factor, interacts with other learners or even distracts them, the rating is -1.5. At -1.75, the attempt to attract the attention of other learners or the teacher in a non-professional way is rated, including shouting, waving one's hands or pranking obviously. The lowest level -2 is reached when the learner walks around the room with an intention to disturb.

Things not or especially considered in the rating

Sometimes things are not included in the rating because they influence the attention only slightly. Other things need special attention because they make the rating more difficult. It must be examined whether the behavior is decisive for attention behavior in a teaching-learning situation and how the learner deals with the respective topic.

If learners drink, this is not considered as inattentive behavior. However, it is important to consider whether the bottle is only used to quench thirst or whether it is played around with, for example plucking at labels, or constantly putting the bottle on and off, in which case drinking is considered a distraction.

A short yawn is not an indicator of the rating. Only if other movements are made, such as stretching strongly or placing the head on the table, it is rated as described in the manual. The same applies to a short scratching or rubbing through the face. In this case, it must be considered whether further body movements follow.

The use of laptops in the classroom is a challenge for the rating and should therefore be given special consideration. It is more difficult to observe attention related to behavioral

characteristics because the screen is not visible during the rating. Therefore, many indicators can be coded as on-task as well as off-task. In this case again, the rule applies that the gaze should wander between the screen and the center of the lesson. However, this often leads to misjudgments: An example would be if a learner looks something up in an e-book, but since it is not possible to understand exactly what he is doing, the activity is wrongly coded as off-task by a rater.

Other indicators that complicate the rating and require special attention are crossed arms and staring at a point. Crossed arms are normally considered a defensive attitude in society but it might be just comfortable for the learner to have both elbows on the table. Again, the rest of the body language must be considered. If the learner keeps the rest of his/her upper body upright and straight, the person seems to be attentive despite his/her crossed arms. However, if he/she leans against the back of his/her chair, it could indicate a rejection posture. If a person stares at a point, it is very difficult to judge his/her behavior ad-hoc. If the point the person stares at is at the center of the lesson, it could quickly be misinterpreted as listening. Often one notices the difference only when the center of the lesson shifts and the learner does not follow with his/her eyes. A good indicator for listening, compared to dreaming, is blinking one's eyes. "Dreamers" often blink less often than when they listen. However, this must be observed individually.

Examples of rating indicators

Table 8

Indicators for rating

	Specification	Description	Exemplary indicators
ON- TASK	2	Reflecting loudly upon s.th. (=2) Creating concept maps (=1.9-1.8)	Reproducing what has been said and mumbling quietly Discussing
	Interactive	trying to explain professional contents to other learners, repeating loudly (e.g. vocabulary), making jokes upon topics (=1.8) Asking questions (=1.7) Raising one's hand (=1.5)	Arguing, explaining, this can also happen during a conversation between learners. The gestures should be observed in particular: the fingers are showing on the documents from time to time; the learner is pointing at something in the front; the head is changing from his/her neighbor to the student in the center of the lesson Questions that are specifically addressed to the teacher
	Constructive	pointing at something (=1.4) gesticulating (=1.4) giving non-verbal feedback on questions from the teacher (=1.3) repeating/practicing something (=1,2) Copying solution steps, Making notes (=1,1 - 1)	Nodding or shaking one's head e.g. calculating tasks Looking alternately to the front and to the underlays

1	Active	<p>independent use of further documents by the learner to look up something (=1)</p> <p>writing something down while not following the lessons (=0.8)</p> <p>Upright posture / chin raised (=0.6)</p>	<p>Accessing to the book/script/glossary</p> <p>The hands are touching the face only slightly, the head is not supported on the hands, looking to the teaching center, chin is slightly raised</p>
	Passive	<p>Listening without doing anything else, looking at the center of the lesson, Body of the learner is inclined towards the respective teaching center, not reacting to disturbing learners, can only be distracted very briefly or not at all (=0.5)</p> <p>the head is supported on the hands (=0.4)</p> <p>"slumping down" (=0.25)</p>	<p>Posture is slightly bent forward; the gaze is directed rigidly to the front.</p>
0		<p>Concealed, out of place, not to be judged</p>	<p>Covered by other learners; has left the room or the video area</p>

OFF-TASK	Passive	<p>The gaze deviates from the center of the lesson and remains averted from what is happening in the classroom. looking out of the window (= -0,2)</p> <p>dozing, yawning or sleeping, Placing one's head on the table (= -0.25 - - 0.3)</p> <p>rummage through his/her things without disturbing other students (= -0.5)</p>	<p>Not turning when the teacher moves in the room</p> <p>Striking yawning, this also includes stretching and relaxing the body through stretching exercises</p> <p>to get out a book/drink</p>
	Active	<p>passively playing around with various things. Things are kept at a distance; posture is leaned on (= -0.55)</p> <p>Eating something, being distracted by other students, but no longer interaction (= -0.6)</p> <p>grinning / laughing softly to oneself while being is busy with something else; not disturbing anybody, occupation is close to the body / posture is bent forward (= -0,75)</p> <p>Posture is strongly averted (= -1)</p>	<p>Mobile phone, pen, watch It's happening more on the side</p> <p>Frequently observed after looking at a mobile phone</p>
-1			

	Deconstructive	<p>unobtrusive passage to the wastebasket or WC (= -1.1 - -1.2)</p> <p>Fooling around (= -1.25) "fidgeting" (= -1.3)</p> <p>being distracted by and interacting with another learner/ distracting other learners (= -1.5)</p>	<p>playing games</p> <p>Tilting, being in motion with the entire body</p> <p>Turning around to classmates sitting one row back; playing "four-wins" on a piece of paper</p>
-2	Interactive	<p>Attempting to attract the attention of other learners or the teacher in a non-professional way (= -1.75)</p> <p>Running around with malfunction intention (= -2)</p>	<p>pranking, calling for a classmate/colleague, including waving one's hands</p>

Rating problems

Facial expressions and gestures are not equally shown by every learner, making it difficult to recognize facial expressions and body language. In addition, assessment errors often occur, which make the rating more difficult. These so-called interpretation pitfalls are listed below.

Interpretation pitfalls

The focus of the rating should be kept as objective as possible and should allow little to no interpretation.

Assessment errors that occur in test psychology include:

A certain systematic error of judgment in the assessment of a person is called the *halo effect*. Here, one characteristic of a person is that effective that it pushes other characteristics into the background. The error could be that these suppressed features are no longer observed by the rater. This often happens when very quick judgments are made about a person. In addition, the dominant feature encourages to draw conclusions about the person and objective observation no longer takes place. An example of this can be physical attractiveness (Spektrum, 2000).

A further error is the *logical error*, which claims that one characteristic is inferred from another characteristic that is perceived as logically belonging to it. In this case, personal convictions are thus transferred to the traits (Wirtz & Strohmer, 2017).

The *leniency-severity error* is an error of judgement in which persons are systematically assessed too positively or too negatively (Spektrum, 2000)

Rater-ratee interactions are judgement distortions in which the rater compares himself or herself with the ratee. On the one hand, it includes the similarity error, in which certain characteristics overlap between the rater and the ratee. On the other hand, a contrast error can occur, when certain features are completely rejected and the rater does not identify with the person (Bortz & Döring, 1995).

Stereotyping is described as an error of judgement, which occurs when "stenciled and schematized imaginary content" is transferred to reality. In this case, characteristics or appearances are judged both negatively and positively. Depending on the judgment, personality traits are judged positively or negatively (Spektrum, 2000).

The *position effect* describes the summation of individual effects on the overall impression. Thus, the first person who is rated has no pre-evaluations, which is called the primary effect. In contrast to this, all following persons are rated differently, because there is always a pre-evaluation of the persons rated previously. However, the first person by Primary

Effect and the last person by Regency Effect are treated in a special way because he or she is particularly remembered (Spektrum, 2000).

Rater training

Prior to the actual rating, a rating training must be carried out. For this purpose, the trainer selects a video clip on which the procedure can be practiced. In order to develop a feel for the specific situation, choosing video clips of at least 15 minutes length are recommended. When selecting the persons to be evaluated, it is important to ensure that the classroom seating positions vary as much as possible. The behavior of the persons to be evaluated should also cover the rating scale as far as possible. In addition, the trainer needs to be prepared to discuss the rating scale in detail and to answer questions.

During the rating training, the rating scale and the software should be presented. The future raters will be familiarized with the rating procedure, get the opportunity to ask questions, and will be informed about possible difficulties and errors during assessment. In addition, one video is discussed together beforehand so that the raters can see exactly how the settings are made. It should be emphasized that the raters should interpret as little as possible and that they should evaluate the behavior of the person to be rated along the defined indicators. Afterwards each rater goes through a rehearsal and applies the coding with the joystick to the previously defined video sequence.

The test ratings are then compared with an expert rating and critical points, as well as errors and problems, are discussed. The training phase is considered complete when ICC (2.1; absolute agreement) is greater than 0.6.

References

- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Springer.
- Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software*, 2(1), e5. <https://doi.org/10.5334/jors.ar>
- Spektrum. (2000). *Lexikon der Psychologie*. Spektrum Akademischer Verlag. Retrieved 15.09.2020 from <https://www.spektrum.de/lexikon/psychologie/>
- Wirtz, M. A., & Strohmer, J. (2017). *Dorsch-Lexikon der Psychologie (18., überarbeitete Auflage)*. Hogrefe. <https://dorsch.hogrefe.com/ueber-den-dorsch>

Kodierhandbuch

CABI

(Continuous (non)Attention-related Behavior Indicators)

Materialien zur Kodierung

Die Beurteilung des Aufmerksamkeitsverhalten sollte mittels mehrperspektivischer Kameras vollständig aufgenommen werden. Die Videotechnik sollte hierbei von hoher Qualität sein, da die Rater*innen besonders auf Mimik und Gestik achten müssen.

Zudem ermöglichen hochqualitative Geräte und Speichermedien mit großer Kapazität, einen reibungslosen Ablauf ohne störendes, aufmerksamkeitsbeeinflussendes Wechseln der Geräte. Die Anzahl der Geräte ist durch den Raum, sowie die Anzahl der Schülerschaft und dem Lehrpersonal abhängig. Falls keine kabellosen Mikrofone im Raum angebracht werden können, erleichtern Lautsprecher und Kopfhörer die Audioqualität für die späterfolgenden Kodierungen.

Mit Hilfe eines Sitzplans beziehungsweise durch Zuordnung von Codes an die Lernenden, wird die Kodierung erleichtert. Eine einfache Hilfe ist es, auch für die spätere Kodierung, die Lernenden von den Reihen vorne bis hinten durch zu nummerieren und Platzkärtchen mit ihren jeweiligen ID-Nummern aufstellen zu lassen. Für die Kodierphase muss ein Wiedergabemedium (für CARMA der VCL Media-Player) installiert sein. Der Computer sollte mit entsprechendem Betriebssystem (Empfehlung: Windows 10) und entsprechender Leistung arbeiten, da das Programm sehr hohe Kapazität benötigt. Ein großer Bildschirm hilft zudem beim Raten, da das Video nicht rangezoomt werden kann. Die Auswertung findet über das Medien-Notierungs-System CARMA (Girard, 2014) statt. Diese Software hat eine integrierte Zeitanzeige und ist eine sehr geeignete Software zur Ermittlung von Moment-Ratings.

Software CARMA

CARMA ist eine Medien-Notierungs-Software. Sie ermöglicht es sowohl Video-, als auch Tonaufnahmen anzuzeigen. Ziel des Programms ist es, Forschern die Möglichkeit zu geben, subjektive Erfahrungen (hier: Lehr-Lern-Milieus) und beobachtbares Verhalten über die Zeit zu erfassen (Girard, 2014).

Die Benutzung der Software ist sehr einfach aufgebaut und individuell anpassbar. So kann unter anderem die Beschriftung der Skala von oberen beziehungsweise unteren Grenzen, sowie dem numerischen Bereich, sowie der visuellen Darstellung an die eigenen Parameter angepasst werden (siehe Kapitel 2.2, Abbildungen 1 bis 3). Die Datei kann so gespeichert werden, dass der Import der Daten in eine statistische Auswertungssoftware (z.B., SPSS oder R) möglich ist.

Rating mit Joystick

Das Rating wird mit dem Joystick durchgeführt. Eine Beurteilung über das Rating mit der Maus ist an dieser Stelle nicht vorgesehen. Die Vor- beziehungsweise Nachteile der Joysticknutzung führt dieses Kapitel auf:

Zum Raten wird der Logitech Extreme™ 3D Pro Joystick genutzt. Das Rating mit Joystick hat den Vorteil, dass man keine Taste, wie bei einer Maus gedrückt halten muss. Zudem hat man eine bessere Kontrolle über die Spannweite der Skala, da man nicht in "Minibewegungen" die Hand bewegt, sondern den ganzen Arm benutzt. Zudem kann man beide Hände an den Joystick legen, sodass man eine bessere Kontrolle über das Gerät hat.

Leider hat auch der Joystick seine Nachteile. Das Stillhalten ist trotz zweier Hände nicht sehr einfach und sollte geübt sein. Zudem kann man bei Unaufmerksamkeit, Ablenkungen etc. schnell am Gerät abrutschen, was dazu führt, dass Messungen ungenau werden und man gegebenenfalls das gesamte Rating wiederholen muss. Auch nach Rücksprache mit dem Customer Service von Logitech konnte leider keine Funktion gefunden werden, um den Joystick auf einer gewissen Position festzustellen.

Es zeigen sich also viele Vor- und Nachteile in der Bedienung des Joysticks. Trotzdem überwiegen die Vorteile in der Gewichtung der Nachteile, da das Handling mit dem Joystick zwar eine gewisse Eingewöhnungsphase und Übung benötigt, jedoch in einer geübten Hand schnelle Reaktionen und Schwankungen auf der Skala ermöglicht.

Einstellungen bei CARMA

Um mit CARMA korrekt zu arbeiten, muss der Joystick angeschlossen sein, bevor das Programm geöffnet wird. Es erscheint dann beim Öffnen der Software ein Feld, in welchem

man auswählen kann, ob man mit per Joystick, oder per Computermaus raten möchte. Um eine einheitliche Achse aller Ratings zu erhalten, sollten die Einstellungen vorher festgelegt sein.

*Bei den Einstellungen (Settings) werden die Kennzeichnungen der Achse unter Set Axis Labels auf **On Task** als Upper Axis Label und **Off Task** als Lower Axis Label bezeichnet (Abbildung 1). Weitere Einstellungen werden an den Achsennummern *Set Axis Numbers* festgelegt, *Axis Minimum Value* wird bei **-2** und *Axis Maximum Value* bei **2** eingestuft. Die Nummer der Achsen wird auf 5 Stufen festgelegt (Abbildung 2). Als *Colormap* wird **Jet** gewählt (Abbildung 3). Der Stichprobenanteil hat eine Frequenz von 20Hz, die Werte werden in einem Intervall von 1.0 Sekunden festgelegt (Abbildung 4). Anschließend kann die Mediendatei unter “Open Media File” geöffnet werden. Das Rating kann beginnen.*

Abbildung 1

Screenshot der Achsenbeschriftung

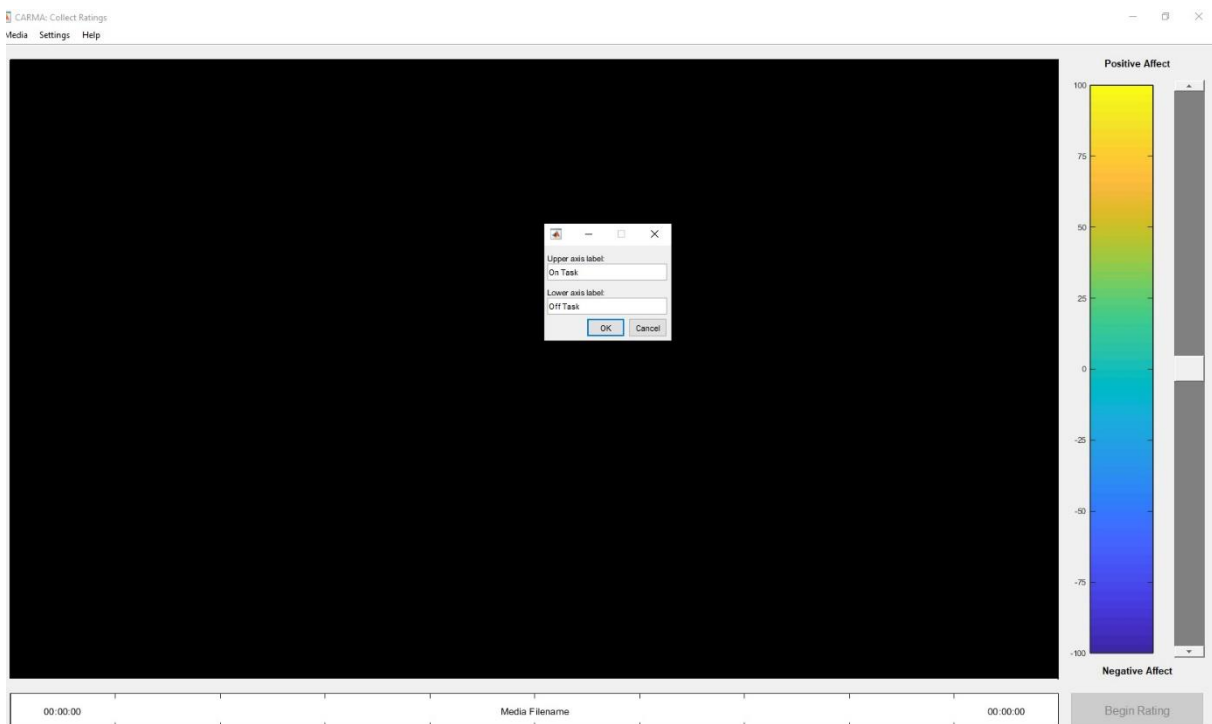
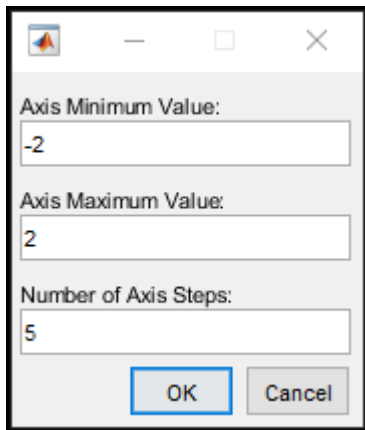


Abbildung 2

Screenshot Achsen-Nummerierung

**Abbildung 3**

Screenshot Einstellung der Colormap

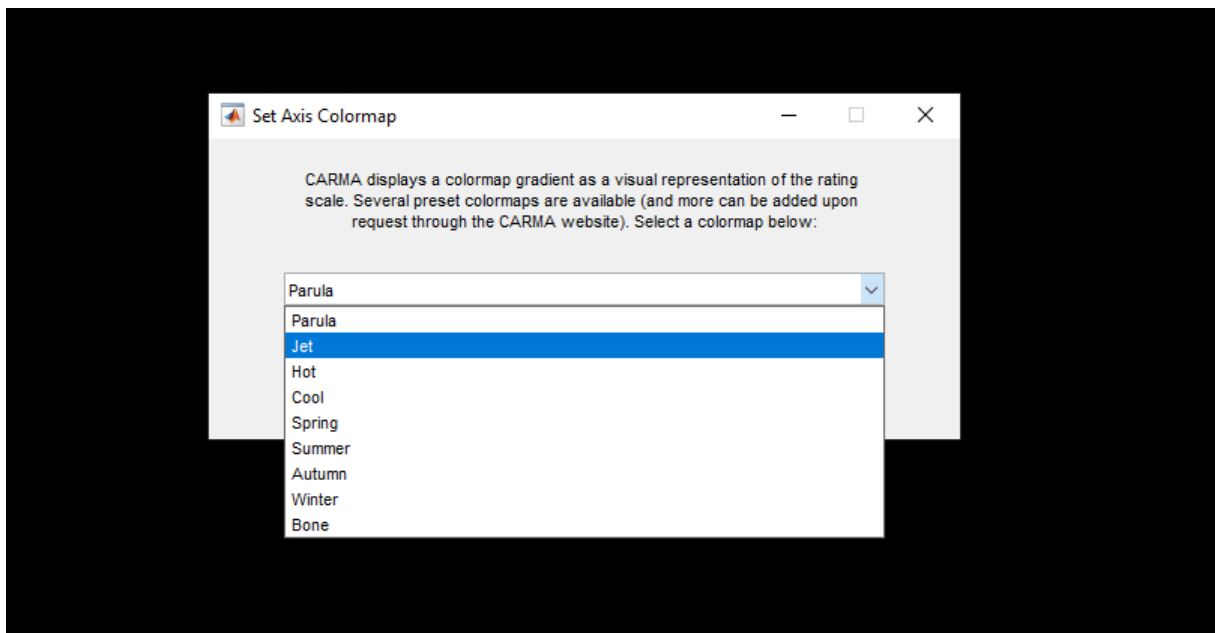
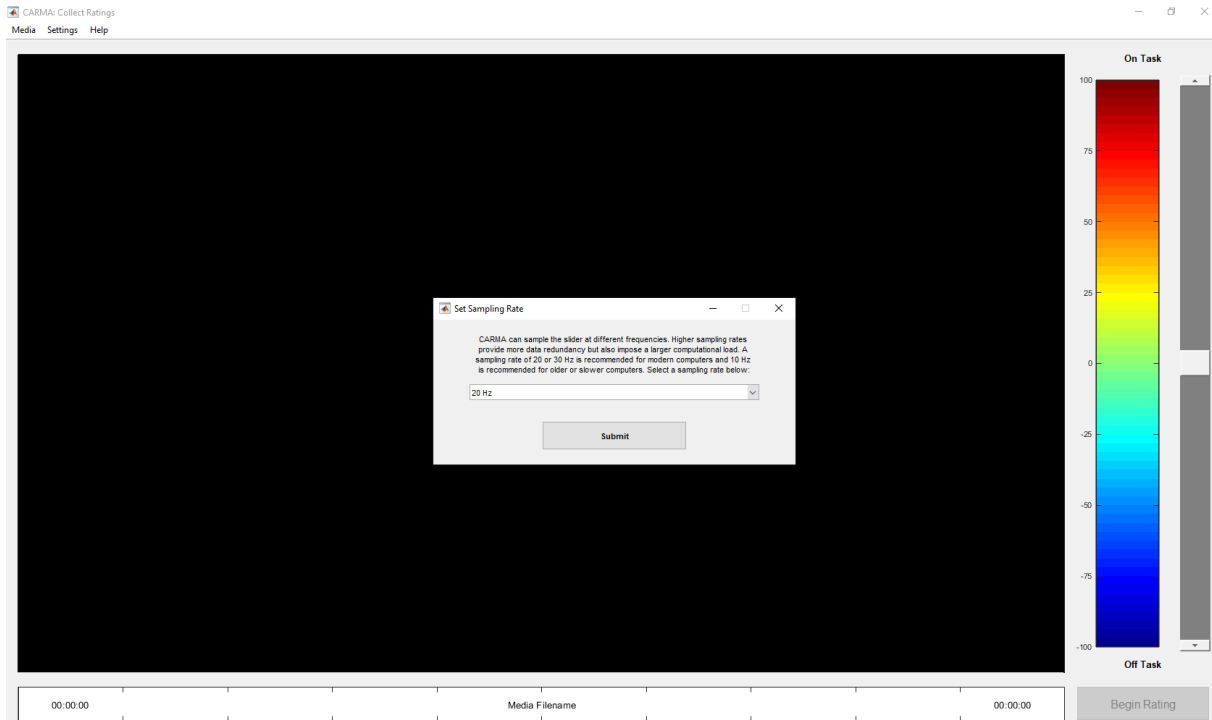


Abbildung 4

Screenshot Einstellung der Sampling Rate



Kodier-Regeln

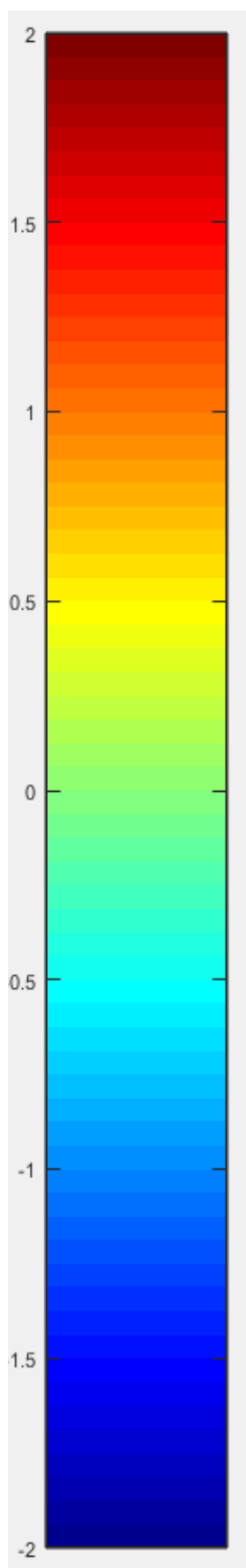
Um Reihenfolgeneffekten vorzubeugen, werden die Lernenden in zufälliger Reihenfolge bewertet. Dazu wird im Vorhinein eine Liste mit den ID-Nummern erstellt und die Abfolge randomisiert.

Während des Ratings fokussiert sich der/die Rater*in nur auf eine Person und bezieht dabei sowohl verbales als auch nonverbales Auftreten der zu beurteilenden Person in die Bewertung mit ein. Jede Person ist einzeln zu beobachten und in jeder Situation nach Möglichkeit für sich selbst zu bewerten. Es sollten während des Ratings keine Informationen von vorhergegangenen Ratings einbezogen werden. Zudem ist sind Interpretationen und Annahmen über das beobachtbare Verhalten der Lernenden so weit wie möglich zu vermeiden. Damit das Rating so präzise wie möglich gestaltet werden kann, ist so schnell wie möglich auf das Verhalten der zu beobachteten Person zu reagieren. Durch die entsprechenden Voreinstellungen (siehe Kapitel 2) wird die Kodierung des Verhaltens der zu beobachtenden Person pro Sekunde vorgenommen. Dabei ist die Gefahr von Beurteilungsfehlern (siehe Kapitel 2) zu beachten. Beurteilungsfehler sind nach bestem Wissen und Gewissen zu vermeiden.

Die Gefahr, dass Beurteilungsfehler entstehen können, tritt bei jedem Rating kontinuierlich auf. Deswegen sollte sich der/die Rater*in dies zwischen und während der Ratings immer wieder bewusst machen und gegebenenfalls nachschlagen. Es ist nützlich sich diese immer wieder vor Augen zu führen und als Ausdruck neben den Arbeitsplatz zu legen.

In der Software CARMA können keine Ratings unterbrochen und an einem anderen Zeitpunkt neu geladen werden. Allerdings gibt es die Möglichkeit den Videoabschnitt kurzzeitig zu pausieren. Dies wird notwendig, wenn eine unerwartete Störung des/der Rater*in auftritt, oder aber auch, damit der/die Rater*in etwas nachschlagen oder eine kurze Pause einlegen kann, da die Konzentration verloren geht. Das Ratingvideo muss allerdings bis zum Ende durchgearbeitet werden, damit es gespeichert werden kann.

Ratingskala



Die kontinuierliche Ratingskala wurde entwickelt, um die Beurteilung über den Zeitverlauf hinweg zu ermöglichen und dabei so wenig Interpretation von Seiten der Rater zu fordern wie möglich. Bei der Skala wurde sich für eine eindimensionale, kontinuierliche, bipolare Skala entschieden, die einen flüssigen Übergang zwischen den zwei Ausprägungen, On-Task beziehungsweise Off-Task ermöglicht. Der Wertebereich ist von +2 bis -2 unterteilt. Der negative Bereich zeigt dabei Off-Task-Verhalten an, der positive On-Task-Verhalten. Dabei kennzeichnet der Nullpunkt den Zeitraum, in dem eine Person abwesend ist, beziehungsweise so verdeckt ist, dass sie nicht aufgezeichnet werden kann oder das Aufmerksamkeitsverhaltens nicht eindeutig beurteilt werden kann (z.B. wenn Indikatoren sowohl als *On-Task* als auch als *Off-Task* kodiert werden können). Der Vorteil hierbei ist, dass dieser Nullwert in den späteren Analysen als „Missing Value“ kodiert werden.

Das höchstmögliche Rating (2) erhält man, wenn man laut reflektiert, sich also laut kritisch mit dem Unterricht auseinandersetzt. Direkt dahinter folgt das Erstellen von Concept-Maps und der Versuch fachliche Sachverhalte anderen Lernenden zu erklären (ca. 1,8). Eine Frage wird mit etwa 1,7 gewertet, Meldungen mit 1,5.

Unter der Linie von 1,5 wird die nonverbale Kommunikation aufgezeichnet. Auf etwas an der Tafel/vorne zu zeigen und zu gestikulieren wird hierbei höher geratet, als nonverbales Feedback auf Fragen der Lehrkraft, worunter das zu Nicken, aber auch zu Grinsen zur Lehrkraft, oder Kopfschütteln verstanden wird. Etwas leise zu wiederholen und üben, sowie Lösungsschritte abschreiben und sich Notizen zu machen, ratet man knapp oberhalb der 1. Das Rating bei 1 erfolgt, wenn die Lernenden selbstständig weitere Unterlagen nutzen, um etwas nachzuschlagen, dabei ist hier irrelevant, ob es sich dabei um ein Buch, das Skript oder ähnliches handelt. Folgt die zu ratende Person gerade nicht aufmerksam dem Unterrichtsgeschehen (da gerade Fragen besprochen werden), notiert aber Lösungsschritte o.ä. wird dies mit +0,8 geratet, da die Person trotzdem aktiv am Unterricht teilnimmt. Eine aufrechte Haltung, wird bei ca. 0,6 eingestuft, je nachdem ob das Kinn

angehoben ist, oder nicht. Wendet der/die Lernende den Oberkörper dem jeweiligen Unterrichtsmittelpunkt zu und lässt sich von störenden Klassenkameraden nicht bzw. nur sehr kurz ablenken, wird mit ca. 0,5 geratet. Das Abstützen auf der Hand während des Zuhörens führt zu einem Rating von ca. 0,4. Wird der Kopf nicht mehr nur mit der Hand gestützt, sondern fällt förmlich in die Hand und der Oberkörper sackt in sich zusammen, so wird das mit 0,25 geratet.

Im negativen Bereich unter der Null wird Verhalten, das gegen die Aufmerksamkeit im Unterricht spricht, geratet. Knapp bei -0,2 wird gewertet, wenn der Blick vom jeweiligen Unterrichtsmittelpunkt abschweift. Döst, gähnt, oder schläft der/die Lernende beziehungsweise wird der Kopf auf den Tisch abgelegt, so ratet man das bei -0,25 – -0,3. Wenn der/die Lernende in seinen/ihren Sachen kramt, ohne dabei andere Lernende zu stören (u.a. um Lernmaterial aus dem Rucksack zu holen), wird das mit -0,5 geratet. Knapp darunter fällt alles, womit sich der/die Lernende kurz, aber aktiv ablenkt, darunter fallen der Blick auf die Uhr, das Handy, oder eine Ablenkung mit dem Stift. Lässt sich der/die Lernende ablenken, interagiert aber nicht länger mit Klassenkameraden, wird mit -0,6 geratet. Beschäftigt sich der/die Lernende themenfremd, erkennbar durch Grinsen oder Lachen, dabei aber Niemand gestört wird, wird dies mit -0,75 geratet. Die Beschäftigung sollte hierbei körpernah erfolgen und die Haltung neigt sich häufig nach vorn. Bei -1 wird geratet, wenn die Körperhaltung stark vom Unterrichtsmittelpunkt abgewandt ist. Kurz darunter fällt ein (unauffälliger) Gang zum Papierkorb oder dem WC. Herumalbern oder sogar das „Herumzappeln“ wird zwischen -1,25 und -1,3 geratet. Lässt sich der*die Lernende von einem anderen Störfaktor ablenken und interagiert mit anderen Lernenden und lenkt diese gegebenenfalls sogar ab, wird bei -1,5 geratet. Hierbei ist besondere Aufmerksamkeit des/der Raters/in gefragt, da hier auch Spiele wie Tic-Tac-Toe gespielt werden könnten. Bei -1,75 wird der Versuch geratet, die Aufmerksamkeit anderer Lernenden beziehungsweise des Lehrenden auf nichtfachliche Weise auf sich zu ziehen. Darunter fallen Zurufe, mit den Händen winken oder offensichtliches „Faxen“ machen. Die niedrigste Stufe -2 ist erreicht, wenn der*die Lernende mit einer Störabsicht im Raum herumläuft.

Dinge, die nicht oder besonders im Rating berücksichtigt werden

Teilweise werden Dinge nicht im Rating berücksichtigt, da sie die Aufmerksamkeit kaum, nur wenig beeinflussen. Auf andere Dinge muss besonders geachtet werden, da sie das Rating erschweren. Es muss geprüft werden, ob das Verhalten für das Aufmerksamkeitsverhalten in einer Lehr-Lern-Situation entscheidend ist und wie sich die/der Lernenden mit der jeweiligen Thematik beschäftigt.

Wenn Lernende trinken, so wird dies nicht berücksichtigt. Dabei ist jedoch zu beachten, ob die Flasche nur zum Durststillen benutzt wird, oder doch an ihr herumgespielt wird, zum Beispiel das Herumzupfen an Etiketten, oder ständiges An- und Absetzen der Flasche., in diesem Falle wird das Trinken doch als Ablenkung geratet.

Kurzes Gähnen ist kein Indikator für das Rating. Nur, wenn andere Bewegungen, wie starkes Strecken oder den Kopf auf den Tisch auflegen erfolgen, wird es wie im Manual beschrieben geratet. Ähnlich wie ein kurzes Kratzen oder Reiben durch das Gesicht. Auch bei diesem Verhalten muss beachtet werden, ob weitere Körperbewegungen folgen.

Die Benutzung von Laptops im Unterricht stellt eine Herausforderung für das Rating dar und soll deswegen besonders im Rating berücksichtigt werden. Die Aufmerksamkeit bezüglich Verhaltensausrprägung zu beobachten ist schwieriger festzustellen, da man beim Raten den Bildschirm nicht sieht. Dadurch können viele Indikatoren sowohl als *On-Task* als auch als *Off-Task* kodiert werden. Hier gilt auch wieder die Regel, dass der Blick zwischen Bildschirm und Unterrichtsmittelpunkt schweifen sollte. Dies führt allerdings häufig auch zu Fehleinschätzungen: Ein Beispiel wäre, eine Lernende schlägt im e-Book etwas nach, da man jedoch nicht genau nachvollziehen kann, was sie tut, wird die Tätigkeit von einem Rater fälschlicherweise dabei als *Off-Task* kodiert.

Weitere Indikatoren, die das Rating erschweren und besondere Aufmerksamkeit erfordern sind verschränkte Arme, sowie das Starren auf einen Punkt. Während verschränkte Arme in der Gesellschaft eigentlich als Abwehrhaltung angesehen werden, könnte es für den/die Lernenden allerdings einfach angenehmer sein, beide Ellenbogen auf dem Tisch zu haben. Hierbei ist wieder auf die restliche Körpersprache zu achten. Hält die/der Lernende ihren Oberkörper ansonsten aufrecht und gerade, so scheint die Person trotz verschränkter Arme aufmerksam zu sein. Lehnt sie sich jedoch an der Lehne ihres Stuhles an, so kann es auf eine Ablehnungshaltung deuten. Starrt eine Person auf einen Punkt ist es sehr schwer dieses ad-hoc zu beurteilen. Liegt dieser Punkt beim Unterrichtsmittelpunkt könnte es schnell als Zuhören falsch geratet werden. Häufig bemerkt man den Unterschied erst, wenn sich der Unterrichtsmittelpunkt verschiebt, dass die/der Lernende nicht folgt. Ein guter Indikator für das Zuhören, im Vergleich zum Träumen ist das Blinzeln mit den Augen. „Träumende“ blinzeln häufig seltener, als wenn sie zuhören. Dies muss jedoch individuell beobachtet werden.

Beispiele für Indikatoren der Ratings

Tabelle 1

Indikatoren für das Rating

	Ausprägung	Kategorienbeschreibung	beispielhafte Indikatoren
ON- TASK	2	Laut reflektieren (=2)	Gesagtes wiedergeben, auch leise murmeln
		Konzept-Maps erstellen (=1,9-1,8) versuchen anderen Lernenden fachliche Sachverhalte zu erklären, Lautes Nachsprechen (z.B. von Vokabeln), Themenbezogene Witze machen (=1,8)	Diskutieren Argumentieren, Erklären, dies kann auch bei einer Unterhaltung zwischen den Lernenden passieren. Dabei sollte die Gestik besonders beobachtet werden: gehen die Finger auf die Unterlagen; wird vorne auf etwas gedeutet; Kopf wechselt von Sitznachbar zum Unterrichtsmittelpunkt
		Fragen stellen (=1,7) Sich melden (=1,5)	Nachfragen, die gezielt an die Lehrkraft gehen
	Konstruktiv	auf etwas zeigen (=1,4) gestikulieren (=1,4)	Nicken oder Kopfschütteln
		nonverbales Feedback auf Fragen der Lehrkraft geben (=1,3) etwas wiederholen/ üben (=1,2)	z.B. Aufgaben rechnen
		Lösungsschritte abschreiben, Notizen machen (=1,1 - 1)	Blick abwechselnd nach vorne und zu den Unterlagen

	1	<p>selbstständige Nutzung weiterer Unterlagen durch den Lernenden, um etwas nachzuschlagen (=1)</p> <p>etwas notieren und dabei nicht dem Unterrichtsgeschehen folgen (=0,8)</p> <p>Haltung aufgerichtet/ Kinn angehoben (=0,6)</p>	<p>Griff zum Buch/Skript/Glossar</p> <p>Hände berühren das Gesicht nur leicht, Kopf wird nicht abgestützt, Blick auf den Unterrichtsmittelpunkt, Kinn leicht angehoben</p>
	Aktiv	<p>Zuhören ohne etwas Anderes zu machen, Blick auf den jeweiligen Unterrichtsmittelpunkt gerichtet</p> <p>Körper des Lernenden ist dem jeweiligen, Unterrichtsmittelpunkt⁶ zugeneigt</p> <p>reagiert nicht auf störenden Lernende und lässt sich nur sehr kurz bzw. gar nicht ablenken (=0,5)</p> <p>Kopf in die Hand stützen (=0,4)</p> <p>„sich hängen lassen“ (=0,25)</p>	<p>Leicht nach vorne gebeugte Körperhaltung, Der Blick richtet sich starr nach vorn.</p>
	0	<p>Verdeckt, nicht am Platz, nicht zu beurteilen</p>	<p>Von anderen Lernenden verdeckt; hat den Raum bzw. den Videobereich verlassen</p>

⁶ Der Unterrichtsmittelpunkt ist Situationsabhängig und kann sowohl die Tafel/Frontal sein, als auch das Heft, oder MitschülerInnen, der Lehrende, der im Raum herum läuft.

OFF-TASK	Passiv	<p>Blick schweift vom jeweiligen Unterrichtsmittelpunkt ab und bleibt vom Unterrichtsgeschehen abgewandt. aus dem Fenster sehen (=0,2)</p> <p>döst, gähnt bzw. schläft, Kopf auf den Tisch legen (= -0,25 - -0,3)</p> <p>kramt in seinen Sachen, ohne dabei andere Lernende zu stören (= -0,5)</p>	<p>Nicht-mitdrehen, wenn der Lehrer sich im Raum bewegt</p> <p>Auffälliges Gähnen, dabei gehören auch sich zu strecken/recken und den Körper durch Streckübungen zu entspannen</p> <p>um Heft/Getränk rauszuholen</p>
	Aktiv	<p>spielt passiv mit diversen Dingen herum. Dinge werden in Distanz gehalten, Haltung angelehnt (= -0,55)</p> <p>isst etwas, lässt sich ablenken, interagiert aber nicht länger (= -0,6)</p> <p>grinst/lacht leise vor sich hin, während er sich themenfremd beschäftigt; stört aber keinen, Beschäftigung erfolgt Körpernah/ Haltung nach vorne gebeugt (= -0,75)</p> <p>Körperhaltung stark abgewandt (= -1)</p>	<p>Handy, Stift, Uhr Geschieht eher nebenbei</p> <p>Häufig nach dem Blick aufs Handy zu beobachten</p>

	Dekonstruktiv	<p>unauffälliger Gang zum Papierkorb oder WC (= -1,1 - -1,2)</p> <p>Herumalbern (= -1,25) “rumzappeln” (= -1,3)</p> <p>lässt sich von einem anderen störenden Lernenden ablenken und interagiert mit diesem/ lenkt andere Lernende ab (= -1,5)</p>	<p>Spiele spielen</p> <p>Kippeln, mit dem gesamten Körper in Bewegung sein</p> <p>Umdrehen zu MitschülerInnen, die eine Reihe weiter hinten sitzen; „Vier-gewinnt“ auf einem Zettel spielen</p>
-2	Interaktiv	<p>Versuch, die Aufmerksamkeit anderer Lernender bzw. des Lehrenden auf nichtfachliche Weise auf sich zu ziehen (= -1,75)</p> <p>Herumlaufen mit Störungsabsicht (= -2)</p>	<p>Faxen machen, Rufen nach einem Klassenkamerad/Kommilitonen, dazu zählt auch das Winken mit den Händen</p>

Probleme des Ratings

Die Mimik und Gestik ist nicht bei jedem/jeder Lernenden gleich ausgeprägt, sodass dies das Erkennen von Gesichtsausdrücken und Körpersprache erschwert. Zudem treten häufig Beurteilungsfehler auf, die die Aufgabe des Raterteams erschweren. Diese sogenannten Interpretationsfallen sind folgend aufgeführt.

Interpretationsfallen

Der Fokus des Ratings sollte so objektiv wie möglich gehandhabt werden und wenig, bis keine Interpretation ermöglichen.

Beurteilungsfehler, die in der Testpsychologie auftreten, sind unter anderem:

Ein bestimmter systematischer Urteilsfehler in der Beurteilung einer Person wird **Halo-Effekt** genannt. Hierbei wirkt ein Merkmal einer Person so sehr, dass es andere Merkmale in den Hintergrund drängt. Es könnte der Fehler entstehen, dass diese unterdrückten Merkmale nicht mehr gesehen werden. Dies geschieht häufig, wenn man sehr schnell ein Urteil über eine Person fällt. Zudem wird durch das dominante Merkmal auf die Person geschlossen und es findet keine objektive Beobachtung mehr statt. Beispielsweise kann man hier den Effekt der physischen Attraktivität gewählt werden (Spektrum, 2000).

Ein weiterer Fehler ist der **Logische Fehler**, in dem von einem Merkmal auf ein weiteres Merkmal geschlossen wird, das als logisch dazugehörig empfunden wird. Hier werden somit persönliche Überzeugungen auf die Merkmale übertragen (Wirtz & Strohmeyer, 2017).

Beim **Milde-Härte-Fehler** handelt es sich um einen Urteilsfehler, bei dem Personen systematisch zu positiv oder zu negativ bewertet werden (Spektrum, 2000).

Rater-Ratee-Interaktionen sind Urteilsverzerrungen, bei der sich der Rater mit dem Ratee vergleicht. Auf der einen Seite gibt es den Ähnlichkeitsfehler, indem sich gewisse Merkmale zwischen dem Rater und dem Ratee überschneiden. Auf der anderen Seite gibt es Kontrastfehler, indem gewisse Merkmale komplett abgelehnt werden und man sich nicht mit der Person identifiziert (Bortz & Döring, 1995).

Als **Stereotypisierung** wird der Urteilsfehler beschrieben, „schablonisierte und schematisierte Vorstellungsinhalte“ auf die Realität zu übertragen. Hierbei werden Eigenschaften, oder das Aussehen sowohl negativ, als auch positiv beurteilt. Je nach Urteil wird also positiv beziehungsweise negativ über Persönlichkeitsmerkmale geurteilt (Spektrum, 2000).

Der **Reihenfolge-Effekt** beschreibt die Aufsummierung von Einzeleffekten, auf den Gesamteindruck. So hat die erste Person, die geratet wird, keine Vorbeurteilungen. Dies wird als Primary-Effekt bezeichnet. Im Gegensatz dazu werden somit alle folgenden Personen anders

bewertet, da immer eine Vorbeurteilung der vorrausgehenden Personen einhergeht. Allerdings wird die erste Person im Primary-Effekt, und die letzte Person, im Recency-Effekt nochmal besonders gehandelt, da sie besonders in Erinnerung bleibt (Spektrum, 2000).

Ratertraining

Vor dem eigentlichen Rating ist eine Ratertraining durchzuführen. Der/die Trainer*in wählt dafür einen Videoausschnitt aus, an dem die Prozedur eingeübt werden kann. Um ein besseres Gefühl für die Situationen zu entwickeln, wird ein Videoausschnitt von mindestens 15 Minuten Länge empfohlen. Bei der Auswahl der zu bewertenden Personen ist auf möglichst unterschiedliche Sitzpositionen im Klassenzimmer zu achten. Das Verhalten der zu beurteilenden Personen sollten zudem die Ratingskala so weit wie möglich abdecken. Zudem ist der/die Trainer*in darauf vorbereitet, die Ratingskala im Detail zu besprechen und auf Rückfragen zu antworten.

In dem Ratertraining sind die Ratingskala und die Software vorzustellen. Die zukünftigen Rater*innen werden mit dem Ablauf des Ratings vertraut gemacht, haben die Möglichkeiten Rückfragen zu stellen und werden auf mögliche Schwierigkeiten und Beurteilungsfehler hingewiesen. Zudem wird vorher ein Video gemeinsam besprochen, sodass die Rater genau sehen, wie die Einstellungen gemacht werden. Es ist hervorzuheben, dass die Rater*innen möglichst wenig interpretieren und das Verhalten der zu beurteilenden Person entlang der festgelegten Indikatoren bewerten sollen. Anschließend durchläuft jeder/jede Rater*in einen Probedurchlauf und wendet die Kodierung mit dem Joystick auf die vorher festgelegten Videosequenz an.

Die Probe-Ratings werden anschließend mit einem Expertenrating verglichen und kritische Stellen, sowie Fehler und Probleme diskutiert und besprochen. Die Trainingsphase gilt als abgeschlossen, wenn ICC (2,1; absolute Übereinstimmung) größer als 0.6 liegt.

Quellen

- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Springer.
- Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software*, 2(1), e5. <https://doi.org/10.5334/jors.ar>
- Spektrum. (2000). *Lexikon der Psychologie*. Spektrum Akademischer Verlag. Retrieved 15.09.2020 from <https://www.spektrum.de/lexikon/psychologie/>
- Wirtz, M. A., & Strohmer, J. (2017). *Dorsch-Lexikon der Psychologie (18., überarbeitete Auflage)*. Hogrefe. <https://dorsch.hogrefe.com/ueber-den-dorsch>