Technische Universität München
TUM School of Education
Friedl Schöller Stiftungslehrstuhl für Pädagogische Psychologie

Teacher Vision

Exploring the Relationship Between Teachers' Visual Expertise and their Competence in Assessing Complex Student Profiles

Christian Kosel

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

Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Doris Holzberger

Prüfer*innen der Dissertation:

1. Prof. Dr. Christina Seidel
2. Prof. Dr. Kathleen Stürmer-Schröpfer, Universität Tübingen
3. Prof. Dr. Andreas Gegenfurtner, Universität Regensburg

Die Dissertation wurde am 01.03.2021 bei der Technischen Universität München eingereicht und durch die TUM School of Education am 10.05.2021 angenommen.

For my beloved children, Hendrik and Sophie.

Acknowledgments

First and foremost, I would like to offer my sincerest thanks and praise to my primary supervisor, Dr. Tina Seidel. Without her unwavering support over the last few years, I would not have stuck it out and completed my Ph.D. I truly appreciate Dr. Seidel's commitment to her research, students, and employees. I enjoyed every meeting and every discussion, and always appreciated her constructive comments on my research. Thank you Tina, for believing in me and for taking me on as a Ph.D. student in the first place.

I also like to thank my mentor and co-author, Dr. Doris Holzberger, who invested much time and effort to support me and my dissertation. I am grateful for all of her critical questions and comments that have helped me to continuously improve. She always kept the big picture in mind and often put me back on the right track. Doris, thank you for always being available for me.

A big thanks also goes out to my colleague and friend Katharina Schnitzler. Many thanks to Katharina for sharing her insights and the countless hours she spent discussing our research and my dissertation. In this context, I would also like to thank Dr. Ricardo Böheim, Elias Codreanu, and Ana Maria Fraile Castrillo, whom I have always been able to rely on over the last few years. I would also like to thank all of my other colleagues with whom I have worked in recent years at the Friedl-Schöller Endowed Chair for Educational Psychology at the Technical University of Munich. It was a pleasure working with you during this time.

I also want to thank Dr. Ilka Wolter, with whom I had the honor to work on a collaborative research article. I very much appreciate her competency, friendly manner, and wisdom.

I would also like to thank Dr. Kathleen Stürmer and Dr. Andreas Gegenfurtner, who kindly agreed to review this dissertation and serve on the examination committee, as well as Dr. Doris Holzberger for chairing the committee.

Finally, I would like to thank my family for their support and love over the past, sometimes stressful, few years. My deepest gratitude goes to my wife Stefanie, who always had my back during the Ph.D. years. She is my daily inspiration and one of the most intelligent people I know. Stefanie, thank you so much for your selfless support—I would not have come so far without you.

Abstract

Previous research on teachers' assessment accuracy found that teachers struggle to assess diversity among their students in terms of intra-individual combinations of cognitive and motivational-affective characteristics, especially when students show inconsistencies between these characteristics (e.g., high cognitive abilities but low interest). Based on these findings, a budding interest in studying cognitive processes associated with teachers' judgment formation emerged within the scientific community. This dissertation ties in with research investigating students' individual differences regarding their sets of cognitive and motivational-affective characteristics as reflected in so-called student profiles and research on cognitive processes underlying teachers' judgment when confronted with the challenge of assessing complex student profiles. To gain deeper insight into the process of teachers' judgment formation, this dissertation relies on process data collected by a modern eye-tracking device. Eye-tracking provides an appropriate method for gaining a deeper understanding of ongoing cognition during various tasks, such as judging student profiles.

The first study associated with this dissertation sets out to clarify which complex combinations of student characteristics are most common by using state-of-the-art statistical latent modeling techniques to identify student profiles. Based on self-report questionnaires and tests from the National Educational Panel Study (NEPS), intra-individual combinations of cognitive (cognitive ability and pre-knowledge) and motivational-affective (interest and self-concept) characteristics were investigated within the domains of mathematics and German language arts (GLA). The latent profile analysis (LPA) revealed five distinct student profiles in both domains. In mathematics, two consistent (strong and struggling) and three inconsistent (underestimating, overestimating, and average-uninterested) student profiles were identified. LPA results for GLA showed three consistent (strong, struggling [motivational], and struggling [cognitive]) and two inconsistent (overestimating and underestimating) student profiles. The underestimating student profile was identified most frequently in both domains (31.3% math and 35.3% GLA).

The second study aims to explore how teachers observe students when judging their underlying student profiles. To explore and understand cognitive processes during judgments, the study analyzes teachers' gaze patterns, so-called scanpaths. The main goals are to learn how teachers distribute their eye movements across students while assessing underlying student profiles, to learn more about this structure of gaze patterns, to explore how experts and novices differ in their gaze patterns, and to determine whether a particular gaze pattern leads to a more successful assessment of student profiles. The scanpath analysis revealed that teachers' gaze patterns were idiosyncratic and more similar within the same expertise group. The expert teachers observed all students to be diagnosed more regularly and performed recurrent scans to re-adjust their assessment. The results also showed that the experts' scanpaths were more complex - including more frequent revisits to all students and equally distributed attention across all students. The experts' visual behavior was statistically associated with higher judgment accuracy in terms of student profiles.

Overall, this dissertation highlights the individual differences among students in terms of their learning-related characteristics and the diversity and commonalities within these characteristics. Furthermore, it adds to previous knowledge about teachers' professional vision the fact that teachers' gaze patterns vary in relation to their professional experience and that experienced teachers' gaze patterns were associated with higher judgment accuracy.

Zusammenfassung

Bisherige Forschungsergebnisse zur Beurteilungsgenauigkeit von Lehrkräften zeigen die Schwierigkeit bei der Beurteilung von kognitiven und motivational-affektiven Schülervoraussetzungen (z.B. Interesse). Für Lehrkräfte scheint herausfordernd zu sein, intra-individuelle Kombinationen von kognitiven und motivational-affektiven Schülervoraussetzungen richtig zu beurteilen, insbesondere wenn Schüler*innen Inkonsistenzen zwischen ihren Schülervoraussetzungen aufweisen (z.B. hohe kognitive Fähigkeiten, aber geringes Interesse). Basierend auf diesen Erkenntnissen entstand in der Wissenschaft ein wachsendes Interesse an der Untersuchung kognitiver Prozesse, die mit der Urteilsbildung von Lehrern einhergehen. Diese Dissertation knüpft zum einen an Forschung zu individuellen Unterschieden von Schülern*innen hinsichtlich ihrer kognitiven und motivational-affektiven Schülervoraussetzungen an, die sich in sogenannten Schülerprofilen widerspiegeln. Zum anderen setzt sie die Forschung zu zugrundeliegenden kognitiven Prozessen von Lehrkräften während der Beurteilung komplexer Schülerprofile fort. Um tiefere Einblicke in den Prozess der Urteilsbildung bei Lehrkräften zu erhalten, stützt sich diese Dissertation auf experimentelle Prozessdaten, die mit einem modernen Eye-Tracker erhoben wurden. Die erhobenen Prozessdaten ermöglichen eine tiefergehende Untersuchung von Aufmerksamkeits- und Diagnoseprozessen von Lehrkräften.

Die erste Studie, die mit dieser Dissertation verknüpft ist, betrachtet, welche komplexen Kombinationen von Schülervoraussetzungen am häufigsten vorkommen, indem Schülerprofile mithilfe latenter Profilanalysen (LPA) identifiziert werden. Auf der Basis von Fragebogendaten und Tests aus dem Nationalen Bildungspanel (NEPS) wurden intra-individuelle Kombinationen von kognitiven (kognitive Fähigkeiten und Vorwissen) und motivational-affektiven (Interesse und Selbstkonzept) Schülervoraussetzungen innerhalb der Domänen Mathematik und Deutsch untersucht. Die latenten Profilanalysen ergaben fünf verschiedene Schülerprofile in beiden Domänen. In Mathematik wurden zwei konsistente (stark und schwach) und drei inkonsistente (unterschätzend, überschätzend und durchschnittlich-uninteressiert) Schülerprofile identifiziert. Die LPA-Ergebnisse für Deutsch zeigten drei konsistente (stark, schwach

[motivational] und schwach [kognitiv]) und zwei inkonsistente (überschätzend und unterschätzend) Schülerprofile. Das unterschätzende Schülerprofil wurde in beiden Domänen am häufigsten identifiziert (31,3 % Mathematik und 35,3 % Deutsch).

Basierend auf diesen Erkenntnissen war das Ziel der zweiten Studie, Aufmerksamkeitsprozesse von Lehrkräften während des Diagnostizierens zu erfassen – insbesondere um den Prozess zu verstehen, wie Lehrkräfte Schüler*innen beobachten um beurteilungsrelevante Informationen zu sammeln. Um Aufmerksamkeitsprozesse zu erforschen, wurden sogenannte Blickbewegungspfade (Scanpaths) der Lehrkräfte untersucht. Die Hauptziele lagen darin, zu beleuchten, wie Lehrkräfte ihre Aufmerksamkeit auf die verschiedenen Schüler*innen verteilen, während sie ihre zugrundeliegenden Schülerprofile beurteilen; mehr über die Struktur der Blickmuster zu erfahren; zu erforschen, wie sich Experten und Novizen in ihren Blickmustern unterscheiden; und festzustellen, ob ein bestimmtes Blickmuster zu einer erfolgreicheren Beurteilung der Schülerprofile führt. Die Ergebnisse der Scanpath-Analyse zeigten idiosynkratrische Blickmuster der Lehrkräfte und, dass sich Blickmuster innerhalb der gleichen Expertise-Gruppe stärker ähnelten. Experten beobachteten alle zu diagnostizierenden Schüler regelmäßiger und führten wiederkehrende Scans durch, um ihre Einschätzungen neu zu justieren. Darüber hinaus führten Experten komplexere visuelle Scanpaths durch – Experten zeigten häufig wiederkehrende Fixationen auf allen Schülern und verteilten ihre Aufmerksamkeit gleichmäßiger auf alle Schüler. Das visuelle Verhalten der Experten war statistisch mit einer höheren Beurteilungsgenauigkeit in Bezug auf Schülerprofile verbunden.

Insgesamt stellt diese Dissertation die individuellen Unterschiede zwischen den Schüler*innen in Bezug auf ihre lernrelevanten Charakteristika sowie die Vielfalt und die Gemeinsamkeiten innerhalb dieser Charakteristika heraus. Darüber hinaus ergänzt sie das bisherige Wissen über die visuelle Expertise von Lehrkräften und zeigt auf, dass sich Blickmuster von Lehrern in Abhängigkeit von ihrer Berufserfahrung verändern, wobei Blickmuster erfahrener Lehrkräfte mit einer höheren Urteilsgenauigkeit verbunden sind.

Included Publications

The present dissertation was written cumulatively and consists of two articles published in international peer-reviewed journals. The author of this dissertation is the first author of both articles and played the leading role in developing, conceptualizing, and writing these journal articles, as well as for performing the statistical data analysis and publication-based presentation associated with them.

The first journal article was submitted to the peer-reviewed journal *Learning and Instruction* (ISSN: 0959-4752) in August 2019 and was accepted for publication in December 2020.

> Kosel, C., Wolter, I., & Seidel, T. (2020). Profiling secondary school students in mathematics and German language arts using learning-relevant cognitive and motivational-affective characteristics. *Learning and Instruction*, *73*. https://doi.org/10.1016/j.learninstruc.2020.101434

The first author played the leading role in the conceptualization, preparation, data analyses, and publication-based presentation of this paper (70%), while the co-authors, Dr. Ilka Wolter (15%) and Prof. Dr. Tina Seidel (15%) guided the development of the manuscript with critical and helpful reviews.

The second journal article was submitted to the peer-reviewed journal *Frontiers in Education, Section Assessment, Testing and Applied Measurement* (ISSN: 2504-284X) in September 2020 and was accepted for publication in December 2020.

> Kosel, C., Holzberger, D., & Seidel, T. (2021). Identifying Expert and Novice Visual Scanpath Patterns and their Relationship to Assessing Learning-Relevant Student Characteristics. *Frontiers in Education*, *5*:612175. https://doi.org/10.3389/feduc.2020.612175

The first author played the leading role in the conceptualization, preparation, data analyses, and publication-based presentation of this paper (80%), while the co-authors, Prof. Dr. Doris Holzberger (10%) and Prof. Dr. Tina Seidel (10%) guided the development of the manuscript with critical and helpful review

Contents

1 Introduction

As early as the late 1970s, Richard J. Shavelson, in his paper "Teachers' Estimates of Student 'States of Mind' and Behavior" (Shavelson, 1978), outlined that one of the essential tasks in teachers' everyday professional work is assessing learning-relevant cognitive and motivational-affective student characteristics. He wrote, "teachers' estimates of students 'states of mind'—cognitive, emotional, motivational—provide primary information in deciding how to teach" and "during teaching itself, new information can be obtained bearing on the student's current state of mind" (Shavelson, 1978, p.37). Almost forty years and much research later, the relevance of Shavelson's statement remains unchanged and is the basis of many current research approaches. In their day-to-day teaching, teachers continuously collect moment-to-moment data about their students to provide tailored instructions, adapt the difficulty of current tasks at hand, plan further classroom activities and materials, provide feedback, or grade students (Corno, 2008). When teachers plan their lessons or teach and interact with their students, they are confronted with a classroom full of individuals who have acquired unique combinations of learning-relevant student characteristics throughout their educational careers. Some students might be interested in mathematical questions but do not have the necessary knowledge to answer them, while others display the opposite combination of traits. Still other students are aware of their abilities and hold high cognitive as well as motivational-affective characteristics. Researchers have gradually begun to investigate this complex intra-individual interplay of cognitive and motivational-affective characteristics to identify which combinations, so-called student profiles, are most predominant among students (e.g., Linnenbrink-Garcia et al., 2012; Seidel, 2006). However, compared to traditional variable-centered research (i.e., the statistical correlation between two student characteristics or a learning outcome), only a few scholars in education have used this so-called person-centered research approach. Although previous person-centered research has provided some exciting findings (e.g., that the number of students who have good cognitive prerequisites but low interest and self-concept is very high in physics, see Seidel, 2006), it is still challenging to underpin general statements on student profiles with empirical findings. Comprehensive knowledge of student profiles is still limited, for example, in terms of the number of learning domains included in the research or in terms of the focus of specific school track levels. Hence, there is still considerable uncertainty about the generalization of the various student profiles identified in the literature.

To come back to the opening statement by Richard J. Shavelson—he continued his remarks by raising an important question that is still highly relevant for many researchers today, which is, "do [teachers] evaluate the information accurately?" (Shavelson, 1978, p.37). When Shavelson's paper was originally written and published in the late 1970s, the answer would probably have been: *we do not know for sure.* Not only does he attest that research on teachers' judgment accuracy is a "subject ripe for research" (Shavelson, 1978, p.38), but he also specifies that "research [is] needed" (Shavelson, 1978, p.38). Forty years later, this call for research was addressed by various scholars in education. Several studies investigated teachers' judgment accuracy on different student characteristics relevant to education, such as students' academic self-concept (Praetorius et al., 2013) or students' cognitive abilities (Machts et al., 2016). However, the studies outlined have exclusively focused on teachers' judgment accuracy concerning isolated student characteristics—an approach that might be limited because teachers perceive students holistically and include more than one student characteristic in the process of forming judgments (Kaiser et al., 2013). An emerging research strand aims to overcome this limitation and explore how accurately teachers can assess complex student profiles (e.g., Huber & Seidel, 2018; Seidel et al., 2020). In terms of judgment accuracy, there is preliminary evidence that teachers tend to overestimate the consistency of student profiles and have difficulty identifying student profiles with conflicting information on cognitive and motivational-affective characteristics (e.g., a student with high cognitive abilities but low self-concept). It is not yet sufficiently understood why some teachers are better at assessing student profiles than others. Therefore, moderators like teacher characteristics—teachers' intelligence (Kaiser et al., 2012), teachers' judgment confidence (Praetorius et al., 2013), or teachers' beliefs (Gralewski & Karwowski, 2016)—may help to explain some of the variances in teachers' judgments of student characteristics. Within research on teachers' judgment accuracy, however, is an increasing theoretical interest in capturing the cognitive processes associated with teachers' judgments (Loibl et al., 2020) of student profiles (Schnitzler et al., 2020; Seidel et al., 2020) rather than relying exclusively on the judgment outcome (Artelt & Gräsel, 2009; Spinath, 2005). To gain deeper insight into the process of teachers' judgment formation, some researchers in recent years have increasingly relied on process data collected by modern eye-tracking devices (Schnitzler et al., 2020; Seidel et al., 2020). Eye-tracking provides an appropriate method for gaining a deeper understanding of cognitive and ongoing mental processes during various tasks, such as judging student profiles. However, the connection between teachers' perceptual processes and a professional outcome, such as assessing student profiles, has rarely been made in existing literature (Schnitzler et al., 2020; Seidel et al., 2020), and numerous questions are still unanswered.

Aim of the dissertation. This dissertation responds to Shavelson's call for research and aims to contribute new knowledge in both of the research strands briefly outlined in the introduction. The following theoretical considerations and empirical results are summarized in this thesis overview—the more interested reader will find additional and more detailed passages in the two publications associated with this dissertation (Kosel et al., 2021; Kosel et al., 2020).

Students' learning-relevant cognitive and motivational-affective characteristics and teachers' professional competencies, such as their abilities to assess student characteristics, can be represented in the *supply-use model* (Helmke, 2012), which is one of the most causal models in instructional effectiveness research and has been adopted by many researchers, including Seidel & Reiss (2014). The present dissertation is guided by the adapted supply-use model depicted in Figure 1. In general, supply-use models divide teaching mechanisms into a supply level, a use level, and an outcome level. At the *supply level*, teachers create and provide learning opportunities for students to use. Teaching and the provided learning opportunities depend mainly on teacher characteristics, like their professional competencies. Influencing factors for teaching quality are, for example, the teacher's professional knowledge, such as his/her subject-specific knowledge and didactic competencies. The *use level* of teaching and learning includes student learning prerequisites, labeled as individual student characteristics. Individual student characteristics influence learning *outcomes* as well as individual learning activities. The intra-individual interplay of these cognitive and motivational-affective characteristics is the focus of Journal Article I associated with this dissertation. Besides, factors of the *learning environment* (e.g., migration background) are used as validation variables in Journal Article I.

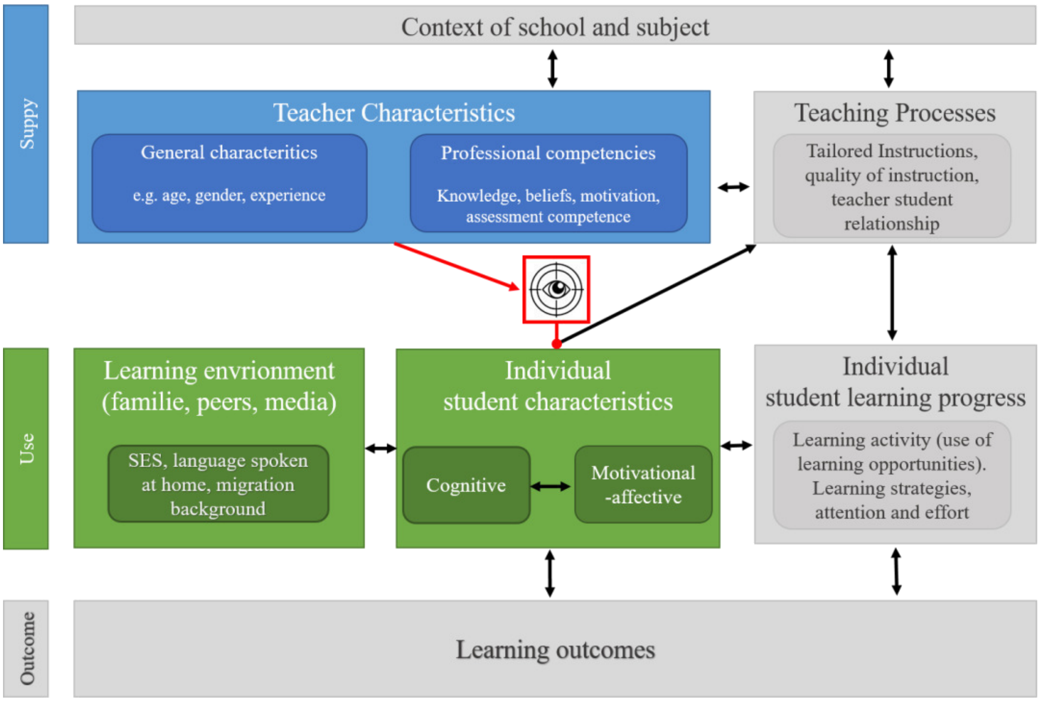

Figure 1. Framework model of this dissertation - Supply-Use Model adopted from Seidel & Reiss (2014) and Huber (2017)

In traditional supply-use models, the link between teachers' professional competencies and student characteristics exists only through the instructional process they initiate to stimulate students' individual learning activities according to their characteristics (i.e., tailored instructions). However, before teachers can provide tailored instructions, they have to assess their students' characteristics. As this dissertation focuses on the process of teachers' judgment formation, this dissertation's used framework model follows Huber's (2017) adaptation and highlights a particular link between teachers' professional competencies and students' characteristics by adding an eye-arrow. The added eye arrow represents the teacher's perception of their students' learning relevant characteristics. In this sense, teachers' are required to notice diagnostic-relevant student engagement behaviors and linking these perceptions to their interpretation of underlying student characteristics. The ability to successfully accomplish this task is recognized as a central component of teachers' professional competence (Blömke et al., 2015) and is often labeled as professional vision (Goodwin, 1994, Seidel & Stürmer, 2014). Moreover, teachers' visual perception is seen as an essential aspect of the noticing component of professional vision (Lachner et al., 2016). In this course, Journal Article II aimed to use process data (eye-tracking data) to explore teachers' noticing abilities in the context of teachers observing student engagement and judging underlying student profiles.

Overall, this dissertation seeks to shed more light on the intra-individual relationship of learning-relevant cognitive and motivational-affective student characteristics—manifested in student profiles—and to examine the cognitive processes involved in the teacher's professional task of assessing these complex student profiles. First, the dissertation aims to overcome the problem of limited generalizability of previous findings on student profiles by using a large, nationally representative dataset, including student data from all secondary school tracks in Germany, and examining student profiles within the two domains of mathematics and German language arts (GLA). Furthermore, another focus was on a comprehensive validation of the student profiles to increase their empirical robustness. Second, the dissertation endeavors to account for the budding interest in studying cognitive processes associated with teachers' judgments (Loibl et al., 2020) of complex student profiles (Seidel et al., 2020). With this aim in mind, the thesis is explicitly interested in analyzing *how* teachers judge student profiles instead of investigative moderators, like teachers' characteristics (e.g., teachers' motivation), that might explain differences in judgment outcomes. Because teaching is a visually intensive profession, there is a growing interest in eye-tracking data from teachers. However, current knowledge about teachers' perceptual processes when assessing student profiles is limited (Schnitzler et al., 2020; Seidel et al., 2020), and there is currently no research available that has directly linked teachers' visual behavior to their judgment accuracy. This thesis aims to bridge this gap. Inspired by other research areas (medicine: O'Neill et al., 2011; or chess: Charness et al., 2001) and in particular by research on teachers' gaze patterns by McIntyre and Foulsham (2018), this thesis analyzes teachers' gaze patterns during the assessment of complex student profiles. The main aim is to learn how teachers distribute their eye movements over the students (and their underlying student profiles) being assessed, learn more about the structure of this gaze pattern, explore how experts and novices differ in their gaze patterns and determine if any gaze pattern led to a more successful judgment of student profiles.

This dissertation's overview begins by reviewing the literature of single cognitive and motivational-affective student characteristics and their relevance in explaining differences in the development of academic achievement. This is followed by a paragraph on previous studies that have used the person-centered approach and examined student profiles. This work then demonstrates the current state of research on teachers' ability to assess individual student characteristics as well as complex student profiles, followed by theories that can help explore the cognitive process underlying teachers' judgments. The dissertation then deals with the professional vision of teachers and current research findings. It then iterates how the analysis of eye movement patterns can further contribute to understanding teachers' cognitive processes involved in teaching and judging and explores whether specific eye movement patterns can be associated with better judgment accuracy. The subsequent chapter provides an overview of the methodological approach, followed by a summary of the two journal articles associated with this dissertation. The final chapter presents a general discussion of principal findings from both journal articles.

2 Theoretical Background

2.1 Identifying Learning-relevant Student Characteristics

In educational research, the relevance of learning-relevant cognitive and motivational-affective student characteristics for learning and achievement has been extensively studied (Alexander et al., 1994; Carroll, 1993; Snow, 1989). Traditionally, cognitive student characteristics are most prominent. Cognitive student characteristics comprise students' general cognitive abilities and their subject-specific knowledge (e.g., pre-knowledge). Motivational-affective student characteristics comprise students' perceptions of their subject-specific self-concept or subject-specific interest.

2.1.1 Cognitive Learning-relevant Student Characteristics

In the mid-nineties, a task force convened by the American Psychological Association addressed the emerging debate about the issues of what is known and unknown about human cognitive abilities (Neisser et al., 1996). In their final report, cognitive abilities were described as the ability "to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, [and] to overcome obstacles by taking thought" (Neisser et al., 1996, p. 77). In a similar vein, Carroll (1993) defined human cognitive abilities as any skill involving cognitive tasks in which the correct processing of mental information is essential for successful performance. Several meta-analyses and studies have demonstrated that cognitive abilities predict school achievements (Deary et al., 2007; Fleming & Malone, 1983), such as SAT scores (Roth et al., 2015; Wilson et al., 2014) and test performance (Lauermann et al., 2020). This extensive body of robust empirical studies has consistently shown that cognitive abilities account for substantial variance in school performance (Deary et al., 2007; Roth et al., 2015).

Domain-specific pre-knowledge is another crucial cognitive student characteristic. In line with Alexander, Jetton, and Kulikowich's (1995) perspective, students' pre-knowledge embodies all the information they have acquired on a specific topic or in a particular domain (e.g., understanding and knowledge about basic concepts of a topic or domain, specific vocabulary). Domain-specific pre-knowledge is the foundation and a strong predictor for acquiring new knowledge

(Alexander, 2003). Students who lack relevant pre-knowledge have difficulty learning new information (Ausubel, 2012).

Despite this uncontroversial impact of cognitive characteristics on academic achievement, research increasingly highlights that motivational-affective characteristics are also critical to student learning and school achievement (Dai & Sternberg, 2004).

2.1.2 Motivational-affective Learning-relevant Student Characteristics

Self-concept of ability refers to students' perceptions of their own skills and abilities in academic tasks (Shavelson et al., 1976). The effect of students' self-concept on achievement has been studied several times (Huang, 2011; Lauermann et al., 2020). In his meta-analysis, Huang (2011) found a moderate correlation between students' self-concept and academic achievement. In a recent study, Lauermann and colleagues (2020) found that students' self-concept of ability is the best motivational-affective predictor of mathematics and German language arts achievement.

Furthermore, subject-related interest, defined as a tendency to like and value a specific content or domain, predicts academic achievement (Alexander et al., 1994; Schiefele et al., 1992). In interest research, a distinction is usually made between situational and individual interest (Schiefele, 2009). According to the first perspective, situational interest can be described as a momentary experience in which an object, context, or feature attracts the individual's attention for a limited time (external causes—e.g., a particular topic). In the second perspective, individual interest is perceived to be a more enduring and relatively stable disposition toward a particular domain or content, where the origin of interest lies within the individual—the likelihood of engaging and re-engaging with this specific domain or topic is high (Schiefele, 2009). Schiefele and colleagues' (1992) meta-analysis shows a moderate overall correlation between interest and academic achievement. Moreover, interest is not only associated with achievement but also with academic effort and learning quality (Schiefele et al., 1992).

2.2 Complex Combinations of Learning-relevant Student Characteristics

A substantial body of variable-centered research has found that cognitive and motivational-affective student characteristics are related to each other (Denissen et al., 2007; Marsh et al., 2005; Schrader & Helmke, 2008; Tobias, 1994). For example, Tobias (1994) found that students' interests and knowledge are positively connected because students interested in specific contents or domains are more engaged and therefore accumulate more knowledge. Other research found a substantial relation between students' interest and self-concept (Denissen et al., 2007). Self-concept and interest are very positively related and develop mutually so that their levels converge over time. Domains in which students perceive themselves as competent are generally those in which they have also developed an interest (Denissen et al., 2007). Both student characteristics have a similar structure with a more general higher-order component and multidimensional subject-specific components (Gogol et al., 2016; Shavelson et al., 1976).

Another research strand has increasingly investigated the complex interactions of cognitive and motivational-affective characteristics by following a person-centered approach (Lau & Roeser, 2008; Linnenbrink-Garcia et al., 2012; Südkamp et al., 2018). The goal of person-centered research is to identify potential subpopulations that exhibit differential configurations, so-called profiles, of a set of indicator variables that would be difficult to detect or interpret using more traditional variable-centered approaches (Lazarsfeld, 1957; Lazarsfeld & Henry, 1968). Previous work has shown that some combinations of cognitive and motivational-affective student characteristics are most predominant among students (Linnenbrink-Garcia et al., 2012; Seidel, 2006; Südkamp et al., 2018). For example, Südkamp et al. (2018) analyzed cognitive ability, academic self-concept, motivation to learn, and achievement-related anxiety in elementary school students. Latent profile analysis yielded a three-profile solution that included two consistent profiles: a) students with high scores on cognitive and motivational-affective variables and low performance-related anxiety, and b) students with average to low scores on cognitive and motivational-affective variables and high performance-related anxiety. The remaining profile, c), showed a more inconsistent pattern that included students with average cognitive ability, low academic self-concept, very low learning motivation, and moderate anxiety. In physics, Seidel (2006) identified five complex student profiles using latent cluster analysis. Students' cognitive abilities, pre-knowledge, domain-specific interest, and self-concept were used as a set of indicator variables. Two of these student profiles can be described as being consistent, in which students showed either overall high (strong student profile) or low (struggling student profile) values for cognitive and motivational-affective characteristics. Other students displayed inconsistent combinations: "Overestimating" students showed relatively low values for cognitive characteristics but reported a high domain-specific interest and self-concept. "Underestimating" students showed the opposite pattern—these students had good cognitive abilities combined with low values in domain-specific interest and self-concept. The last identified student profile was characterized by students considered to be "uninterested" due to their good to average cognitive skills but low domain-specific interest. Altogether, Seidel (2006) found that 57% of the students were assigned to inconsistent profiles. Since the number of inconsistent profiles was also high in other studies (Linnenbrink-Garcia et al., 2012; Seidel et al., 2016; Südkamp et al., 2018), this suggests that these inconsistent profiles are meaningful not only for researchers, but also for teachers. Teachers are likely to have multiple students in their class with inconsistent combinations of cognitive and motivational-affective characteristics. Therefore, it is alarming that previous research has found that teachers have trouble identifying inconsistent combinations of cognitive and motivational-affective characteristics.

2.3 Assessment as an Everyday Task for Teachers

Being able to adequately assess the current state of students' cognitive (e.g., pre-knowledge) and motivational-affective (e.g., self-concept) characteristics and appropriately estimate learning requirements (Artelt & Gräsel, 2009) are essential components of professional teacher competencies (Herppich et al., 2018). They are also a prerequisite for teachers to provide tailored instruction (Corno, 2008). Therefore, assessing individual learners can be seen as the main challenge in teaching (Corno & Snow, 1986). This raises the question of how well teachers are able to identify individual relevant student characteristics—or their complex combinations—in the form of student profiles.

2.3.1 Teacher's Judgment Accuracy of Student Characteristics and Student Profiles

Much of the research on teacher judgment accuracy has focused exclusively on teachers' ability to judge single cognitive or motivational-affective characteristics without accounting for the intrapersonal connectedness of student characteristics. Previous meta-analyses have emphasized that teachers assess students' cognitive abilities (Machts et al., 2016) and achievement (Demaray & Elliot, 1998; Südkamp et al., 2012) relatively accurately. However, some authors have investigated the accuracy of teacher judgments of motivational-affective characteristics and recognized that teachers tend to have more problems judging such characteristics, such as self-concept (Praetorius et al., 2013; Spinath, 2005) or interest (Karing, 2009), in students. Kaiser and colleagues (2013) provided evidence that teachers perceived students holistically and intermingled distinct student characteristics when asked to separately judge cognitive abilities and motivational characteristics. In the light of reported findings of Kaiser et al. (2013), the question emerges as to the extent to which teachers are able to assess student profiles that combine cognitive and motivational-affective student characteristics. A few studies have explored this question (Huber & Seidel, 2018; Seidel et al., 2020; Südkamp et al., 2018). For example, Südkamp et al. (2018) found that teachers consistently judge students as average, over-average, or under-average, indicating that cognitive and motivational-affective characteristics typically go hand-in-hand for many teachers. Südkamp et al. (2018) interpreted these results as showing how teachers might apply heuristics by clustering students according to pre-defined prototypes. However, student profiles from previous studies (Linnenbrink-Garcia et al., 2012; Seidel et al., 2016) yielded a much larger variety, including inconsistent student profiles, such as underestimating, overestimating, or uninterested. In a recent study by Seidel et al. (2020), teachers were asked to judge five student profiles established in a previous study (Seidel et al., 2016). Following an expert-novice comparison approach, it was shown that experienced teachers were more accurate in judging inconsistent student profiles when compared to novice teachers (Seidel et al., 2020). However, no significant differences in judgment accuracy of the consistent profiles (strong, struggling) were found.

In summary, previous studies on the assessment accuracy of student profiles indicate that teachers have particular difficulty correctly identifying inconsistent student profiles. However, it remains relatively unclear why some teachers can reach a high judgment accuracy while others fail to judge student profiles correctly. Seidel et al. (2020) provided initial findings that teachers' professional experience is a critical factor explaining differences in judgment accuracy of student profiles. Nevertheless, to understand why expertise might influence this type of assessment task, it is necessary to examine the process of judgment formation more closely. Along this research line, Schnitzler et al. (2020) analyzed teachers' differential use of diagnostic student cues, such as engagement (e.g., hand-raisings), to examine the judgment formation process. She found that student teachers with high judgment accuracy of student profiles utilized combinations of diagnostic student cues that pointed to a specific student profile. In contrast, student teachers with a low judgment accuracy had difficulty using distinct combinations of relevant diagnostic cues. These findings may highlight the importance of the cognitive process in judgment processes for explaining why some teachers are better at judging than others. However, as past research has mainly focused on judgment accuracy, far less is known about which specific cognitive and information processes underlie teachers' assessments (Leuders et al., 2018; Loibl et al., 2020; Schnitzler et al., 2020).

2.3.2 Cognitive Processes That Underlie Teachers' Assessments

Empirically-proven cognitive models of information processing from social and general psychology (Brunswik, 1955; Chaiken & Trope, 1999; Croskerry, 2009; Fiske & Neuberg, 1990) can provide a theoretical basis for cognitive modeling of teacher judgments. For example, Brunswik's lens model of perception is based on the paradigm that humans continuously evaluate various latent and not directly observable distal traits in their daily life (e.g., the emotional state of strangers). In order to evaluate the latent traits as accurately as possible, humans are required to observe and interpret so-called information cues (e.g., facial expression, body posture). Brunswik's lens model of perception can also be used to explain teachers' judgment and decision-making processes. In trying to apprehend a not directly observable student characteristic (e.g., a student's current level of interest), the teacher only has recourse to imperfect indicators or cues (i.e., number of hand-raisings as an indicator for interest). Since there are typically multiple information cues available, the teacher's task is to combine information collected from these ambiguous cues to reach the best possible judgment. Teachers' judgment accuracy, therefore, depends first on whether critical student cues are noticed, then on whether noticed cues are related to the actual distal trait, or—if cues are misleading (cue validity)—the degree to which teachers build their judgment on noticed cues (cue utilization) (Brunswik, 1955; Leuders et al., 2018).

Other theories can be described as dual-process theories of judgment formation under uncertainty (Croskerry, 2009; Fiske & Neuberg, 1990). The dual-process theories have in common the fact that mental processes are divided into two primary forms of reasoning, depending on whether they are automatic or controlled. For example, Croskerry's (2009) universal diagnostic reasoning model from his research on medical reasoning is a promising model that can be applied to better understand teachers' underlying cognitive processes during diagnostic tasks. This model posits two decision-making systems, which run partly parallel: *System 1* (heuristic, intuitive) processes very quickly and is based on teachers' professional knowledge and heuristics. Teachers can use this fast and automated system when their internal pattern processor has detected critical cues that are already stored in memory and available for recall in the diagnostic situation. In contrast, *System 2* (systematic, analytical) is engaged when teachers have not readily recognized a student's behavior (cues) as belonging to a specific student characteristic or if cues do not follow regular scripts. Teachers then perform a slow and resource-intensive systematic search of decision-relevant cues and must therefore tease out various possibilities from one another. Returning to the example of the distal characteristic, "student's interest," from above: if a teacher wants to judge a student's level of interest, and if they have stored knowledge about critical cues related to a student's level of interest (e.g., number of hand-raisings or body language), then the heuristic System 1 is more engaged—the teacher can then take mental shortcuts because they have immediately and automatically recognized the pattern of cues (e.g., high frequency of hand-raising + positive body language = high interest). In contrast, if a teacher has not stored critical cues in long-term memory (which indicates the student's level of interest), then the decision-making process is more cognitively intense. System 2 is more engaged in this situation, and the teacher must notice all available cues and select and weigh critical ones by making inferences as to whether they are related to the distal characteristic. It is important to note that System 2 is built by learning and conscious activation through repeated exposure (professional experience) that promotes the acquisition and storage of knowledge in the long-term memory (i.e., repeated systematic processing with the same exposure leads to automated processing).

The outlined theories have also been applied in recent models of teacher judgment competence by Herppich et al. (2018) or Loibl et al. (2020). The authors suggest that teachers form their judgments about students on the continuum of the two processing modes described in the theories. However, the increasing theoretical interest in capturing the cognitive processes associated with teachers' judgments (Loibl et al., 2020) has increased the need to implement new methods for capturing the judgment process better. One of these methods is eye tracking.

2.4 Teacher Gaze in the Context of Professional Vision

When teachers gain information and collect behavioral cues (like information and behavioral cues described earlier in the Lens Model; Brunswik, 1955) to make inferences (or reasoning) about underlying student characteristics, the judgment and decision-making process relies heavily on attentional processes (Borko et al., 2008). The primary source of information for teachers in this scenario is student observation during instruction. Therefore, analyzing teachers' attentional processes has the potential to help us to better understand the cognitive processes and behavioral activities underlying teachers' assessments. Teachers' eye gaze is a direct indicator of the allocation of visual attention. The ways in which teachers selectively direct their attention to specific events to make inferences are critical features of the concept of *professional vision* (Goodwin, 1994). The concept of professional vision implies two interconnected processes: (1) *noticing*, describing teachers' ability to direct their attention to relevant classroom events and information cues; and (2) *knowledge-based reasoning*, referring to teachers' ability to interpret these events based on their professional knowledge and anticipate consequences for further learning (Goodwin, 1994; Seidel & Stürmer, 2014). Despite an increased focus on teachers' noticing ability and the use of qualitative analyses of think-aloud protocols or interview transcripts (Clarridge & Berliner, 1991; van Es & Sherin, 2010), fewer studies have investigated teachers' perceptual processes, which are linked to the noticing component of teachers' professional vision (Seidel et al., 2020). Existing research on teacher perception primarily uses eye-tracking metrics, such as the number of fixations or average fixation duration, to determine where teachers focus their attention and process visual information (Seidel et al., 2020; van den Bogert et al., 2014; Wolff et al., 2016). These studies adopted an expert-novice approach and showed that experienced teachers exhibit notable differences in eye movement behavior compared to novices. Over the years, several eye-tracking parameters have emerged that are associated with the teacher's level of expertise: Studies have shown that experts a) have more fixations on task-relevant areas and fewer on task-irrelevant ones (Wolff et al., 2016); b) have shorter mean viewing times (shorter fixations) (Wolff et al., 2016), indicating that they are faster at encoding information; and c) distribute their fixations more evenly (van den Bogert et al., 2014).

The question remains as to why teachers change their eye movement behavior throughout professional development. When teachers observe students, their eye movements are controlled by bottom-up and top-down mechanisms. Bottom-up attention is mainly driven by salient features of the visual stimuli, such as brightness, color, or motions, while top-down attention is driven by the teacher's task-related plans and goals (Gegenfurtner et al., 2011; Goldberg et al., 2020). Professional experience can affect these task-related plans and goals as teachers are likely to restructure their knowledge base throughout their professional development. Teachers' knowledge base includes practice-based cognitive schemas that guide their actions during instruction (Boshuizen et al., 1995) and is a critical top-down driver of teachers' professional vision (Gegenfurtner et al., 2011; Seidel & Stürmer, 2014). Other prominent theories can also explain why professional experience changes the way teachers see and notice events in the classroom. With increasing experience, teachers encounter the same situations repeatedly and embed this knowledge in their long-term memory. In this context, Ericsson and Kintsch (1995) postulated in their long-term working memory theory that experts increase their working memory capacity by building retrieval structures in long-term memory. The embedded knowledge inside this retrieval structure is readily available for use in the capacity-limited working memory and allows experts to process more visual information and larger perceptual chunks. Furthermore, the information reduction hypothesis (Haider & Frensch, 1996) posits that experienced teachers learn to separate task-relevant from task-redundant information. The outlined findings and theories can also be embedded in recent theoretical frameworks, such as the cognitive theory of visual expertise (CTVE; Gegenfurtner, 2020). This overarching framework allows for an understanding of why information processing changes as expertise develops and incorporates many empirically-tested theories about cognitive processes underlying visual information processing (Ericsson & Kintsch, 1995; Haider & Frensch, 1996). Among other aspects, CTVE assumes that experts have developed the capacity for parafoveal and holistic information processing in the visual register, which allows experts to extend their visual span (i.e., experts show longer saccadic amplitudes). Furthermore, experts are better at selecting relevant information for a specific task and can actively ignore task-irrelevant information (Gegenfurtner, 2020; Haider & Frensch, 1996). Ignoring redundant information results in experts having more capacity in their working memory to process relevant information. Furthermore, CTVE assumes that prior knowledge—in terms of declarative, procedural, and metacognitive knowledge—is associated with experts' ability to notice relevant information or cues, which are then selected or ignored (Gegenfurtner, 2020).

Taken together, the outlined theories and empirical studies show that teachers develop and restructure their cognitive schemas (top-down processes) as they gain experience, leading to different ways of perceiving and processing visual information compared to less experienced teachers. However, the body of research about teachers' visual processes mainly focuses on average eye-tracking metrics like the number of fixations. Such metrics are important for revealing how teachers perceive their students during instruction but fail to capture the processual nature of eye-tracking behavior. Therefore, looking more closely at the eye-tracking sequence, so-called scanpaths, can provide rich information and overcome this problem (McIntyre & Foulsham, 2018).

2.4.1 Analyzing (Teachers') Scanpaths

Scanpaths represent the pattern of fixations and saccades constructed from the path of eye movements over a specific time (Holmqvist et al., 2015) and reflect the unfolding of visual attention, indicating exactly which contents of visual information a human has attended to. The scanpath theory was defined by Noton and Stark in two major publications (Noton & Stark, 1971a; Noton & Stark, 1971b) and has become a highly-relevant theory for understanding human eye movements and gaze patterns. Scanpath theory states that eye movements are generated in a top-down fashion to facilitate correct recognition of an image by comparing it to prior experience. Individuals who view an image or a particular scene store both the scene's features and the gaze sequence with which they inspect the scene. Noton and Stark hypothesized that individuals who recognize a previously viewed scene follow a scanpath similar to the one that resulted from the initial viewing. In related research and based on their study of human face recognition, Kanan et al. (2015) presented a less strict version of scanpath theory called *scanpath routines*. This adapted version of the original scanpath theory considers that individuals rarely come across the same visual stimuli twice in real-world situations. Therefore, it is more likely that individuals' eye movements are similar between viewings of scenes or images from the same stimulus class. In terms of the teaching profession, repeated observation of students in the classroom could be considered as such a stimulus class. Scanpath routines in a given stimulus class evolve to enable enhanced visual processing (Kanan et al., 2015), for example, by filtering important and unimportant information (Haider & Frensch, 1996).

Scanpath analyses have already been carried out in various professional domains (e.g., medicine: O'Neill et al., 2011; art: Antes & Kristjanson, 1991; chess: Charness et al., 2001) to analyze and compare the visual behavior of experts and novices. However, in the literature on teachers' perception, the analysis of scanpaths has received very little attention and is limited to research on teachers' classroom management skills (McIntyre & Foulsham, 2018). Using mobile eye-tracking, McIntyre and Foulsham (2018) found that experienced teachers prioritized and ordered the way in which they scanned the classroom during instruction. The authors found that experienced teachers followed a sequential pattern of observation. For example, they found that experienced teachers observed a particular student first and returned to the initial fixated student more regularly after being distracted by another student. In contrast, novices did not routinely return to the initially fixated student after a distraction and continued to observe other students. This first demonstration that differences in teachers' scanpaths are related to expertise is essential to developing knowledge about perceptual sequences in the teaching profession.

However, there is not yet a great deal of knowledge about how scanpaths and gaze behavior are related to professional outcomes—such as teachers' ability to assess learning-relevant student characteristics. Yet, to get a sense of the extent to which experts' enhanced visual processing affects a task's outcome, it is worth reviewing results from other research areas. For example, O'Neill et al. (2011) explored physicians' gaze behavior during optic disc examination and found that expert physicians (glaucoma subspecialists) showed a more systematic and circumferential scanpath pattern, which was related to higher diagnostic accuracy. In contrast, novices' (ophthalmology trainees) gaze patterns were more local, less systematic, and lacked diagnostically-relevant features of the optic disc, which may have been the cause of their less accurate diagnoses. In another study, Kasarskis et al. (2001) demonstrated that experienced aircraft pilots had less complex scanpaths and exhibited a well-defined visual scanning pattern during a landing task. The experts omitted task-redundant instruments and focused only on the runway and airspeed indicator. Experts were able to reduce complexity primarily by hiding task-redundant instruments or regions, resulting in less complex scanpaths. Together, in both of the studies briefly outlined here, experts' enhanced visual processing directly affected how well they performed the task at hand. Analyses of teachers' scanpaths may also be valuable to identifying levels of expertise since they would consider the sequential nature of eye movements. Scanpath analyses could provide important information on how teachers perceive multiple students and the systematic ways in which they compare diverse students or the extent to which they consistently seek new information from individual students for the purposes of assessment.

3 The Present Research

This dissertation seeks to shed more light on the intra-individual relationship of learning-relevant cognitive and motivational-affective student characteristics—manifested in so-called student profiles—and to examine the cognitive processes involved in the teacher's professional task of assessing these complex student profiles.

With this overarching aim, the dissertation encounters researchers' increasing interest in latent modeling using a person-centered approach (Seidel et al., 2016; Südkamp et al., 2018) and in exploring the cognitive processes associated with teachers' judgment formation (Loibl et al., 2020). First, the dissertation aims to move beyond the limitations of previous person-centered research and offer a more comprehensive look at student profiles across all three secondary school track levels in Germany and the two domains of mathematics and German language arts. Furthermore, the most up-to-date statistical method (BCH three-step approach; Asparouhov & Muthén, 2014) was used to identify student profiles and validate them statistically. This advanced understanding of student profiles was fundamental to accomplishing the second objective of this dissertation—to analyze the process of how teachers judge student profiles. To uncover differences in the process of judgment formation, not only was modern eye-tracking used, but new techniques derived from other research areas were also applied. Thus, this is the first work in teacher research that has conducted more advanced scanpath analyses (i.e., examining scanpath complexity) during a teacher's professional task of assessing students. The two journal articles related to this dissertation and the research questions examined herein are presented in more detail below.

Journal Article A. The focus of the first journal article was to uncover student profiles following a person-centered approach (Lazarsfeld, 1957). Therefore, the main goal was to replicate results from a study carried out in the context of physics (Seidel, 2006) in the domains of mathematics and German language arts. Moreover, to strengthen the generalization of student profiles, a large, nationally representative sample of 10,025 ninth graders from all three main secondary school tracks was used (Blossfeld & Roßbach, 2019). The present research also provided an extensive validation of the resulting student profiles using the information on students' socio-demographic variables (gender, migration background, parental education level) as well as end-of-year grades for both domains as educational outcome variables.

Finally, the journal article compared how students switched between student profiles depending on the domain in question.

Altogether, the journal article addressed the following three research questions:

(1) Among secondary school students from different tracks, which student profiles in (1a) mathematics and (1b) GLA can be systematically differentiated in terms of cognitive and motivational-affective characteristics?

Identified student profiles were expected to range from overall positive characteristics (strong) to a more challenging composition of student characteristics (struggling). Inconsistent student profiles were also expected to vary between students with low cognitive ability but high self-concept and interest (overestimating) and vice versa (underestimating). There could also have been a student profile with average cognitive and affective-motivational characteristics (Seidel et al., 2016). Furthermore, student profiles derived from the mathematics context were anticipated to more closely resemble those that emerged from the physics classes (Seidel, 2006) because mathematics and physics are perceived as being more closely-related domains and less similar to the verbal one (Goetz et al., 2014).

(2) Is it possible to validate latent student profiles with regard to socio-demographic factors and learning outcomes by using sub-profile analysis?

- (2a) Gender. Do males and females differ in their student profiles?
- (2b) Migration background. Do native students and students with a migration background differ in their student profiles?
- (2c) School track. Are there differences in the student profiles of students from different school tracks?
- (2d) Parental education level. Are there differences in student profiles between students depending on their parents' educational level?
- (2e) End-of-year Grade. Are there differences between student profiles in their domain-specific end-of-year grades?

The respective hypothesized relationships between the validation variables and student profiles were presented in detail in the journal article (Kosel et al., 2020) and will only be highlighted here briefly. Based on a large body of previous research, it was first hypothesized that female students would be more likely to underestimate their math ability but show a strong profile in GLA (OECD, 2016). Male students were anticipated to show an overestimating profile in mathematics and an average profile in GLA (Skaalvik & Skaalvik, 2004). Second, it was assumed that students with a migration background would show a struggling or overestimating profile in mathematics as well as in German language arts (OECD, 2016; Shajek et al., 2006). Third, it was hypothesized that students with stronger (cognitive) characteristics should be more likely to be high school and middle school students given Germany's ability-tracked educational system (Baumert, 2006). Fourth, it was hypothesized that students of parents with lower educational backgrounds would be more likely to be characterized by lower self-concept and/or interest and moderate to low prior knowledge (Davis-Kean, 2005; Senler & Sungur, 2009). Finally, it was assumed that students with a consistently positive profile were more likely to have received the best end-of-year grades than students with a consistently weak profile.

(3) Do individual students have distinct latent profiles in mathematics and GLA?

The last research question was more explorative in character. Deduced from theoretical considerations (e.g., Internal/External Frame of Reference Model; Marsh, 1986), it was expected that more of the students would belong to different profiles in mathematics and GLA, primarily due to domain-dependent differences in motivational-affective characteristics. It was unclear whether there would be systematic changes between student profiles if different domains were considered.

Journal Article B. The second journal article addressed cognitive processes during teachers' professional tasks of assessing complex student profiles. Eye-tracking provides an appropriate method for understanding cognitive and ongoing mental processes during different kinds of tasks. Because teaching is a visually intense profession, there has been a growing interest in teachers' eye-tracking data. However, the current knowledge about teachers' perceptual processes while assessing complex student profiles is limited (Schnitzler et al., 2020; Seidel et al., 2020), and there are currently no studies available that have directly linked teachers' visual behavior to their judgment accuracy. Therefore, the journal article aimed to bridge this research gap and explored how eye-movement patterns (scanpaths) differed across experienced and novice teachers, and whether a specific visual behavior was related to higher judgment accuracy. In this context, it should be noted that the main objective was to explore the process of judgment formation and not to analyze which student profile was correctly assessed and which was not. This was analyzed in another published paper, the main results of which are presented therein detail (Schnitzler et al., 2020; Seidel et al., 2020).

Altogether, the journal article addressed the following three research questions:

(1) Are teacher scanpaths (a) of an idiosyncratic nature and (b) more similar within expertise groups?

First, it was investigated whether teachers' scanpaths are idiosyncratic (as described in the scanpath theory; Noton & Stark, 1971a; 1971b) in an assessment situation. The results were expected to support the idea that teachers' visual perception is primarily a top-down process. It was therefore expected that (a) teachers' scanpaths would be significantly more similar within individual teachers than between teachers. Second, if cognitive schemata indeed guide scanpaths in a top-down process, then (b) scanpaths of individual experts were anticipated to be more similar to scanpaths of other experts than to those of a group of novices, and scanpaths of individual novices were expected to be more similar to those of other novices.

(2) Do experts' scanpaths include recurring sub-patterns—a consistent visual strategy—that differ from recurring sub-patterns in novice scanpaths?

If experts' scanning paths are more comparable within the group of experts than the group of novices, then there may be specific patterns that indicate a visual strategy of experts that differs from that of novices. Based on earlier evidence demonstrating that experts spread their gaze more evenly across all students in the classroom (van den Bogert et al., 2014), it was expected that this might be a first indication that experts in the process of assessing students use a visual strategy that is more consistent across all students. In this context, Fiske et al. (1983) noted that experts' added capacities (see also the ideas behind the theory of the long-term working memory; Ericsson & Kintsch, 1995) potentially frees them to process additional relevant information, whereas novices cannot yet handle the amount of information available, and might be cognitively overwhelmed.

(3) Is there (a) a relationship between teachers' visual strategy and their judgment accuracy, and (b) are there systematic differences between experts and novices?

Drawing on theoretical perspectives on teachers' professional competencies (Blömeke et al., 2015), professional vision, and visual expertise (Gegenfurtner et al., 2011; Gegenfurtner, 2020), and considering the limited previous research (e.g., Schnitzler et al., 2020; Seidel et al., 2020), it was assumed that there is a relationship between teachers' gaze behavior and their judgment accuracy of student profiles.

4 Methodology

The present dissertation was embedded into the research project “Interaction II – Students through teacher eyes” founded by the German Research Foundation (Grant No. SE1397/7-3), in which teachers’ assessments of student characteristics and underlying perceptual processes were explored. The funders had no role in the study’s design, data collection and analysis, the decision to publish, or preparation of the dissertation or journal articles. In the following sections, samples and methodological details of both journal articles are briefly outlined. Two methodological features, latent profile analysis (Journal Article I) and scanpath analysis (Journal Article II), will be described in more detail, as both approaches are so far relatively unknown within the field of teacher research.

4.1 Journal Article I

4.1.1 Participants and data collection

The first journal article used data provided by the German National Educational Panel Study (NEPS), a longitudinal study on educational trajectories in Germany, following a multi-cohort sequence design (Blossfeld & Roßbach, 2019). The NEPS aims to collect longitudinal data on the development of competencies, educational processes, and educational decisions across an individual’s life span. The journal article used data from the NEPS starting cohort four (SC4), which began to follow students in the ninth grade. The subsample consisted of N = 10,025 ninth-graders with a mean age of M = 14.74 years (SD = 0.73). The study used information from the first waves of this longitudinal panel study conducted in the 2010/2011 school year in Grade 9 (i.e., Wave 1, the first half of the 2010/2011 school year, and Wave 2, the second half of the school year in 2011).

4.1.2 Measures

For the creation of latent student profiles, so-called indicator variables are required. In the journal article, the chosen cognitive characteristics for latent profile analysis were students’ cognitive abilities and pre-knowledge. Furthermore, motivational-affective indicators were students’ interests and self-concept of abilities. General cognitive abilities were measured with two tests (Haberkorn &

Pohl, 2013) in the NEPS. To measure students' perceptual speed, students were required to match numbers with graphical symbols as quickly as possible. Furthermore, in a matrices test (Raven Matrices; Raven, 1989), students' reasoning skills were assessed. Each matrix task consisted of several geometrical elements arranged horizontally and vertically, whereby one field of the matrix remained free. The students' task was to derive the logical rules that the arrangement of the geometric elements followed to fill the free field.

Students' domain-specific pre-knowledge was assessed with competence tests in mathematics (Duchhardt & Gerdes, 2013) and German language arts (Haberkorn et al., 2012). For both tests, Weighted Maximum Likelihood Estimates (WLE; Warm, 1989) were used. Moreover, both tests showed high reliability (mathematics .79; GLA .75). Using WLE scores has the advantage that they represent the best point estimator for the particular facet of knowledge (Rost, 2004).

Students' domain-specific self-concept of ability in mathematics and GLA was measured with one set of three items each (e.g., "In German, I am a hopeless case."). All items were scored on a four-point Likert scale, where higher scores indicated higher self-concept in the particular domain (range: 1= "doesn't apply at all" to 4 = "completely applies"; internal consistency: $\alpha_{math} = 0.87$; $\alpha_{GLA} = 0.86$).

Furthermore, students' domain-specific interest in mathematics and GLA was measured with one set of four items each (e.g., "When I am working on a mathematical problem, it may happen that I do not notice how time flies."). Again, all items were rated on a four-point Likert scale and averaged, with higher scores indicating a higher interest (range: 1 = "doesn't apply at all" to 4 = "completely applies"; internal consistency: $\alpha_{math} = 0.91$; $\alpha_{GLA} = 0.86$).

4.1.3 Data analysis

Latent profile analysis (LPA; Lazarsfeld, 1957; Lazarsfeld & Henry, 1968) is a categorical latent variable approach that aims to identify latent subpopulations within a sample based on a specific set of measured indicator variables. Thus, LPA assumes that people can be classified into latent categories, called profiles, with varying probability degrees. LPA offers many advantages in comparison to traditional methods like cluster analysis or median split techniques. By contrast, with these traditional methods, LPA results do not vary according to the retained clustering algorithm (Muthén, 2004). Furthermore, LPA permits the simultaneous inclusion of several measurement scales (e.g., continuous and categorical) and the direct inclusion of predictor variables in the models (Muthén & Muthén, 2007). Finally, LPA permits the direct specification of alternative models that can be compared with various model fit statistics.

In the journal article, LPA was performed using the statistical modeling software Latent Gold 5.1 (Vermunt & Magidson, 2016) and R (R Core Team, 2020). As recommended by Asparouhov and Muthén (2014) and Vermunt (2010), a three-step approach was followed.

First, models with one to ten profiles were estimated separately for mathematics and GLA. The resulting profile solutions were compared using various model fit statistics (AIC; Akaike, 1973, BIC; Schwarz, 1978, BRLT; McCutcheon, 1987, entropy; Celeux & Soromenho, 1996) to find the adequate number of profiles. For example, the entropy measure is attributed to classification uncertainty (Celeux & Soromenho, 1996). The uncertainty of classification is rather considerable when the posterior probabilities are alike across profiles. The normalized version of entropy, which has an interval scale [0, 1], is generally used as a model selection criterion. This entropy measure suggests the level of separation between profiles. When a value of normalized entropy is higher, it portrays a better fit; values > 0.80 reveal that latent profiles are rather discerning (Muthén & Muthén, 2015).

Second, once the appropriate profile solution was identified, each student was assigned entirely to the latent profile to which they had the highest estimated probability of belonging based on their values in the observed indicator variables (Vermunt & Magidson, 2016).

Third, the association between latent profiles and covariates (socio-demographic variables) and the distal learning outcome (end-of-year grade) were analyzed. The three-step procedure's significant advantage lies in this step, as the resulting logistic regression coefficients are adjusted for the classification error made in step one. Finally, the differences in the end-of-year grades across latent profiles were tested for significance using the Wald test (Vermunt & Magidson, 2016). Figure 1 illustrates the full LPA model.

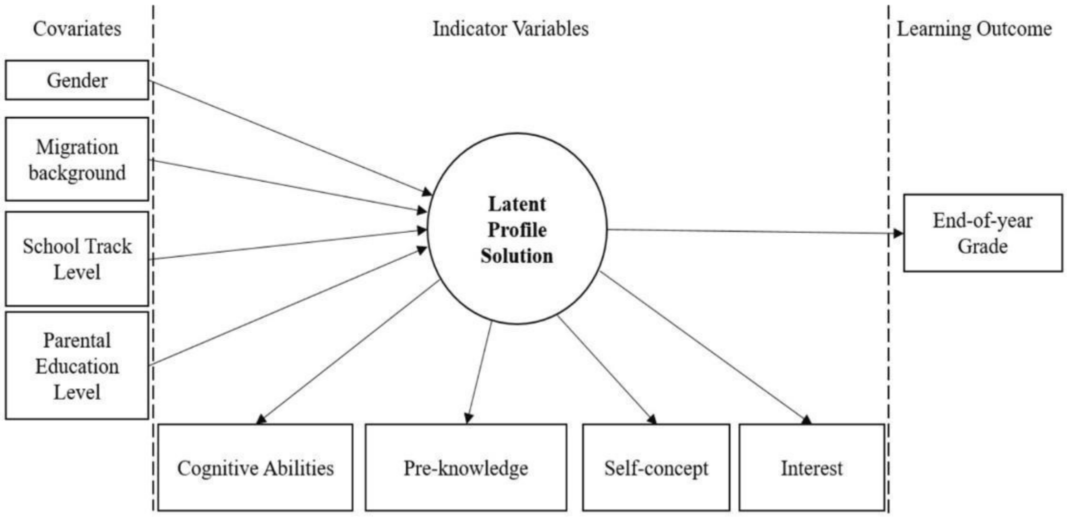

Figure 2. Complete LPA model for mathematics and German language arts (published in Kosel et al., 2020)

4.2 Journal Article II

4.2.1 Participants and data collection

In the following described eye-tracking experiment, teachers watched an authentic video of a classroom lesson and were subsequently asked to assess five different students. High-quality eye-tracking data ($M_{\text{trackingratio}}$ = .94; average deviation x-axis = .58°, y-axis = 57°) were available for 44 participants. Among them were 35 novice teachers (female =60.5%) enrolled at the university level in a bachelor's teacher training program to become teachers in German high-track secondary schools for science or mathematics. Furthermore, the sample included nine in-service teachers (female = 70.5%) with an average teaching experience of 12.40 years (SD = 8.25) ranging from 1.5 to 25.0 years.

Eye movements were recorded using the static and laboratory-based binocular eye tracker SMI RED 500 with Experiment Center 3.7 software (SensoMotoric Instruments GmbH, 2017) on a 22-inch monitor and a sampling frequency of 500 Hz. Eye-tracking conditions were standardized for all participants (constant ceiling light, 65 cm distance to the eye tracker, use of a chin rest). In addition, an automatic 9-point calibration followed by validation was performed before the start of eye-tracking to ensure data quality. Calibration was performed again if the 9-point automatic procedure failed.

Experimental procedure. In a first step, participants were introduced to learning-relevant cognitive and motivational-affective student characteristics (i.e., self-concept, pre-knowledge), as well as their complex interplay in student profiles (Seidel, 2006). Next, participants viewed a 10-minute video of an eighth-grade mathematics lesson showing 23 students. Participants watched the video twice and were instructed to carefully observe the students in order to assign them to student profiles (Seidel, 2006). In the second viewing, five critical students, each holding a different profile (determined by a prior latent profile analysis; Seidel et al., 2016), were marked throughout the video to ensure that participants were always aware of which students to observe and assess. After participants had watched the video and eye tracking had been recorded, the evaluation began, with participants assessing the underlying student profile of each of the five marked students.

4.2.2 Measures

The following is a brief overview of the three central measures used in the journal article.

Judgment accuracy. Participants' judgment accuracy of student profiles ranged between zero (no correct judgment) and five points (only correct judgments). Participants received one point if the assigned profile matched the underlying student profile and zero points if the incorrect student profile was assigned. If a participant first assigned an incorrect profile but stated the correct profile in their alternative choice, they received half a point. The points for each correct assignment were added up so that a higher mean indicated higher judgment accuracy.

String edit distance. To study the *order* of fixations (scanpaths) of different Areas of Interest (AOI), a unique identifier was assigned to each AOI—in this case, five different letters were chosen, one letter for each of the five marked students. The order of fixation was thus translated into a string of characters, representing the individuals' scanpaths. In order to compare the scanpaths between individual participants and between both expertise groups, Levenshtein distances (LDs; Levenshtein, 1966) were calculated. Levenshtein distance is a string edit algorithm applied to measure the dissimilarity of strings. In this technique, a sequence of basic mathematical operations (delete, insert, or replace) is used to convert one sequence of strings into another. The more similar two scanpaths are, the fewer mathematical operations need to be performed, and the lower the cost of converting one string to another. Since it is a common practice to represent scanpaths as a sequence of strings (Davies et al., 2016; Mathôt et al., 2012; McIntyre & Foulsham, 2018), the LD is suitable for comparing scanpaths.

Scanpath entropy. To capture and analyze the complexity of scanpaths, Shannon's entropy of information (Shannon, 1948) was used. The entropy of information is grounded in information theory (Shannon, 1948) and measures the information in a variable or system in terms of ordering and complexity. For example, the entropy of a given *control system* (i.e., the human saccadic eye movement system) is comparative to the *amount of information* necessary to describe the behavior of that system. Therefore, the more information needed to specify the system, the more complex the system is. The following example illustrates the idea behind entropy: A simple coin flip with a fair coin is a situation with maximum uncertainty (i.e., high complexity) (probability p(h) = probability b(h)); it is difficult to predict the outcome of the coming coin flip. However, if the coin is not fair and the probability p(h) is higher than b(h), where $p \neq b$, then there is less uncertainty and consequently less complexity. In research about teachers' attentional processes, the entropy of information can be used to describe teachers' gaze distribution across multiple students or students and teaching-related objects (e.g., a board) (McIntyre et al., 2017). For example, a high entropy occurs when the teacher distributes their attention equally among many students, and, after fixing upon one student, all other students have a *similar probability* of being looked at.

4.2.3 Data analysis

In order to obtain the scanpaths of the participants, some preparations had to be made. First, five AOIs, each representing a target student, were drawn manually using SMI BeGaze 3.4 (SensoMotoric Instruments GmbH, 2017). Scanpaths were created for eight teaching events extracted from the 10-minute video sequence with an average duration of 43 seconds. The scanpaths were generated for each of the eight teaching events using the built-in saccade and fixation detector of SMI Begaze 3.4 (SensoMotoric Instruments GmbH, 2017). The raw eye-tracking data were then converted into strings using the conversion application smi2ogama (Dolezalova & Popelka, 2016). In this step, the fixation sequence was recoded into a sequence of strings representing fixation locations. Finally, multiple scanpaths were obtained as strings (e.g., ABCDE) for each participant and the eight instruction events.

After the outlined preparations, the primary analyses were carried out. The LDs were used to measure the similarity of scanpaths (a) of each participant compared to other participants and (b) across the group of experienced and novice teachers. LDs were calculated for all pairwise scanpaths using the R package stringdist 0.9.5.5 (van der Loo, 2014). For statistical analyses of the various sets of LDs, a repeated-measures ANOVA was performed. To identify visual strategies and uncover how experienced and novice teachers differ in detail, the R package GrpString 0.3.2 (Tang et al., 2018) was used. The function *common sub-patterns* was used to uncover repetitive scanpath patterns within the data. Common sub-patterns are defined as sequences within scanpaths that are found more than once, with a minimum length of three characters (Tang et al., 2018). To investigate gaze transitions, substrings with only two characters (i.e., gaze change from student A to student B), transition matrices were computed. In the next step, participants' scanpath complexity was analyzed using Shannons' entropy score to measure the complexity of scanpaths. Finally, multiple regression analysis was performed to investigate the relationship between scanpath entropy, teachers' judgment accuracy, and expertise level.

5 Summary of Publications

5.1 Journal Article I

The following is a summary of the journal article “Profiling secondary school students in mathematics and German language arts using learning-relevant cognitive and motivational-affective characteristics” (see Kosel, Wolter & Seidel, 2020).

Journal Article I aimed to profile secondary school students based on their learning-relevant cognitive and motivational-affective student characteristics. Students’ cognitive abilities, pre-knowledge, self-concept, and interest were identified as student characteristics that are fundamental to explaining academic achievement and were used as indicator variables for LPA. The LPA results for mathematics and GLA revealed five distinct profiles in both domains. In mathematics, two consistent (strong and struggling) and three inconsistent (underestimating, overestimating, and average-uninterested) student profiles were identified. The results of the LPA for GLA indicated three consistent (strong, struggling [motivational], and struggling [cognitive]) and two inconsistent (overestimating, underestimating) student profiles. The underestimating student profile was most frequently identified in both domains (31.3% mathematics and 35.3% GLA). Central results from the validation analyses were as follows: It could be shown that girls, compared to boys, had a higher chance of being profiled as underestimating students in mathematics, whereas the chance of girls being profiled as strong in GLA was higher than for boys. Furthermore, results indicated that students with a migration background, compared to peers without a migration background, had a higher probability of being profiled as struggling in both domains. In both domains, students with a migration background suffered mainly from reduced subject-specific self-concept and interest. When the school track level was taken into account, the results showed that almost half of the students from the highest school track had low subject-specific self-concept and interest but had good cognitive prerequisites. Students from the basic school track had higher chances of falling into one of the more problematic student profiles (e.g., struggling). The parental educational level also had a significant impact on student profiles. It was found that students’ whose family educational backgrounds were high also had a higher chance of being profiled as strong students. This relationship was found for mathematics as well as for GLA. Finally, when looking at the learning outcome (end-of-year grade), the results showed that students who were profiled as strong received the highest end-of-year grades in both domains. The third research question examined the extent to which students showed different profiles in mathematics and GLA. The results showed that students tended to show different profiles in the domains, except for those profiled as underestimating students—almost one-fifth of all students showed this pattern in both domains. Taken together, the results support findings from previous person-centered research that also identified consistent as well as inconsistent student profiles (Seidel, 2006; Seidel et al., 2016; Südkamp et al., 2018).

5.2 Journal Article II

The following is a summary of the journal article "Identifying Expert and Novice Visual Scanpath Patterns and Their Relationship to Assessing Learning-Relevant Student Characteristics" (see Kosel et al., 2021).

Journal Article II addressed teachers' judgment formation process during a teachers' task of assessing student profiles. The way in which eye-movement patterns (scanpaths) differed across novice and expert teachers was explored. In an eye-tracking experiment, teachers watched an authentic video of a mathematic lesson showing 23 students and were subsequently asked to assess five of these students and their underlying student profiles. In order to analyze teachers' gaze patterns, scanpaths were extracted and then compared qualitatively (common sub-patterns) and quantitatively (scanpath entropy) between novices and experts. Finally, whether or not specific scanpaths patterns are related to judgment accuracy was explored. First, the repeated measures ANOVA results revealed that teachers' scanpaths are idiosyncratic. This finding indicates that the different scanpath patterns were more similar *within* a teacher than *between* teachers, thus supporting original assumptions of the scanpath theory (Noton & Stark, 1971a;b). Second, the results show that scanpath patterns were more similar within expertise groupings, indicating that teachers' professional knowledge affected how they monitored the five students in a top-down process. Qualitative sub-pattern analyses revealed systematic differences between the scanpath patterns of experts and novices.

Experts' visual behavior maintained up-to-date information on the students by monitoring all of them more regularly. In contrast, novices made frequent transitions between just two students. The experts' visual behavior also resulted in more complex scanpaths (higher entropy) than that of novices. The results of regression analysis demonstrated that experts' visual behavior was statistically related to higher judgment accuracy. The more teachers followed the "expert-like" visual strategy (more complex scanpaths), the better their judgment of the five student profiles. However, the overall judgment accuracy of experts was not significantly higher compared to that of novices.

6 General Discussion

In this chapter, four principal findings of the dissertation are reflected, discussed, and related to other research works. Moreover, the most critical limitations of this dissertation will be identified, and further implications for practice and research proposed.

6.1 Discussion of Central Results

6.1.1 Principal Finding 1: Comparable Student Profiles Found in Mathematics and German Language Arts

First, the thesis aims to identify cognitive and motivational-affective student profiles composed of students' information about their cognitive abilities, pre-knowledge, self-concept, and interests. In mathematics and GLA, the model solutions with five student profiles achieved the best fitting result. In mathematics, two consistent (strong and struggling) and three inconsistent (underestimating, overestimating, and average-uninterested) student profiles were identified. The five-profile solution for mathematics was remarkably similar to an earlier study in physics instruction (Seidel, 2006), which is reasonable because mathematics and physics are domains associated with the STEM (science, technology, engineering, mathematics) field and were, for example, perceived by students as two rather challenging domains (Goetz et al., 2014). Therefore, student profiles that represent students who suffer from low self-concept in these two areas are plausible. This makes the finding that student profiles for GLA are comparable with student profiles found in mathematics all the more compelling. The results of the LPA for GLA indicated three consistent (strong, struggling [motivational], and struggling [cognitive]) and two inconsistent (overestimating, underestimating) student profiles. Students being further classified by virtue of deficient cognitive performance or low motivation notwithstanding, the student profiles are comparable in terms of their internal structure. Thus, the first principal finding of this thesis is that the configuration of the cognitive and motivational-affective characteristics in most students can be described by patterns that show either high, low, and mixed values in the respective set of cognitive and motivational-affective characteristics, but there are very few students (average-uninterested) who have contradictory and fragmented information *within* cognitive or motivational-affective characteristics. For example, no students were identified in mathematics and GLA as having low cognitive abilities but high pre-knowledge.

Overall, two facts strengthen the generalizability and robustness of the student profiles identified in this study: the fact that both cognitive and motivational-affective characteristics did not further subdivide the profiles, and the fact that the results back up results of previous studies in physics, for example (Seidel, 2006), that used similar indicator variables.

6.1.2 Principal Finding 2: High number of underestimating students

The second principal finding of this thesis was that the underestimating student profile was most frequently identified in both domains (31.3% mathematics and 35.3% GLA). High cognitive ability and good pre-knowledge do not correlate with high self-concept of ability and interest in underestimating students. This substantiates previous findings in the literature (Seidel, 2006; Südkamp et al., 2018). However, this thesis goes beyond previous research by showing that the underestimating student profile was (a) the most stable student profile across mathematics and GLA when individual switching patterns were examined (16.8% of all students underestimating themselves in both domains), and (b) found in nearly half of the students from the highest school track level. In terms of finding (a), it seems plausible that many students equally underestimate their abilities within the STEM cluster (i.e., mathematics and physics); however, it was somewhat unexpected that the number of underestimating students was even high in GLA and that many students showed an underestimating profile in both domains. It is plausible to assume that these students may have distanced themselves from school across domains or generally carry low school-related self-esteem, which carries over to these underestimating cross-domain profiles (Huber et al., 2015; Huber & Seidel, 2018). Regarding finding (b), the “big-fish-little-pond effect” (Marsh et al., 2008) could be a meaningful interpretation for this finding. In light of this effect, students in higher school tracks demonstrated lower self-concept due to a higher-performing environment than students in basic school tracks with a lower performing environment, mainly due to social comparison with their peers. This effect is even more likely because this study showed that only a small percentage of students in higher school tracks had a struggling profile in mathematics or GLA.

6.1.3 Principal Finding 3: Teachers' Scanpath are Idiosyncratic and Influenced by Expertise

The second overarching aim of the present dissertation is to contribute to knowledge about cognitive processes underlying teachers' assessments of student profiles. Therefore, gaze patterns were analyzed in order to understand the process of judgment formation. A third principal finding of this thesis was that the participating teachers observed the video sequence in their own manner. The extracted and analyzed scanpaths were consistently most similar *within* a teacher than *between* teachers. This result ties in well with previous research across diverse disciplines like face recognition research (Kanan et al., 2015) or imagery (Foulsham et al., 2012). Therefore, the dissertation confirmed the original descriptions of the scanpath theory (Noton & Stark, 1971a; b) and extended the literature by showing that top-down processes (i.e., knowledge and schemata-driven gaze) seemed to primarily guide teachers' attention during the observation of the video sequence.

Furthermore, the finding outlined above was a necessary condition for the second finding—it was demonstrated that teachers' scanpaths were more similar to other teachers with the same expertise. Because cognitive schemas primarily guide gaze in a top-down process, the finding suggests that experts' cognitive schemas change over the course of professional experience and tend to converge in a professionally-shared cognitive schema within the group of experts, in contrast to the cognitive schemas of the group of novice teachers. This principal finding goes beyond previous reports on teachers' visual expertise (Gegenfurtner et al., 2011; van den Bogert et al., 2014) by showing that not only do averaged eye-tracking metrics change as expertise develops, but also that the underlying structure of the gaze—the way teachers order their eye movements—changes. Therefore, the principal finding of this thesis regarding teachers' visual expertise ties in with prior research of teacher scanpaths (McIntyre & Foulsham, 2018).

This dissertation seeks to go a step further and determine how eye movement patterns differed exactly. The qualitative sub-patterns analysis revealed that experienced teachers' most identified and recurring sub-patterns covered more individual students (i.e., four students) than those of novices (i.e., two students). Consequently, the experts' visual strategy kept the information about the target students up-to-date by monitoring all students more regularly. In contrast, the novices' sub-pattern analysis revealed a distinct visual strategy, as they made repetitive transitions between only two students. This result now provides new evidence supporting assumptions of the CTVE (Gegenfurtner, 2020); in the light of this model, the outlined findings might indicate that the experienced teachers from this study were able to process more and longer visual chunks and, therefore, were better at monitoring multiple students. Furthermore, experts might have quickly noticed what was relevant in a specific situation and observed more students because of their faster information encoding skills. Due to their advanced memory structures and extra capacities in the working memory, experts' also included more judgment-relevant information from each of the students; novices, however, had to reduce the amount of incoming information to only two students. However, the results can be interpreted in light of other eye-tracking studies showing that experts were more engaged in revisits, which is a strong indication of monitoring activities (Wolff et al., 2015). The authors interpreted their finding to mean that experienced teachers also looked for activity between students and tracked behavioral movements. The results also tie in with another study by van den Bogert et al. (2014), which found a positive relationship between scanpath length (interpreted based on broader monitoring behavior of experts) and level of expertise. Tan (1996) similarly reported that experienced teachers were aware of a more significant number of student behavior (cues) than novice teachers during physical education lessons.

Furthermore, the finding can also be linked to and interpreted in terms of relevant teacher judgment models and dual-process theories (Croskerry, 2009; Fiske et al., 1999). In light of dual-process theories, the novices' visual behavior may indicate that they could not collect and combine all critical cues and that they did not yet have stored chunked knowledge in their memory. This may indicate that System 2 was more engaged in novices. On the contrary, experts were able to make mental shortcuts by identifying critical cues of a specific student, combining these critical cues, and making an initial judgment. In this case, the more automated System 1 was engaged. Experts might know more about the validity of the collected cues (i.e., whether the identified cue is related to the target characteristic or multiple characteristics) and weighted their cues according to their knowledge about the validity (Croskerry, 2009). As a result, experts may have then been able to process holistically and had more time to monitor all students, make comparisons between them, and re-adjust their judgment until they terminated the search for cues—

ultimately finding a satisfying solution. In contrast, novices could not take mental shortcuts, and a more systematic and non-intuitive search strategy was required in which fewer students may have played a role.

Based on this principal finding, the question arose as to which strategy was more successful in assessing students and their underlying characteristic profiles.

6.1.4 Principal Finding 4: More Complex Scanpaths were Associated with Higher Judgment Accuracy

The visual strategy first had to be quantified with a suitable measure to analyze the relationship between different visual strategies (scanpaths) and the accuracy with which teachers judged student profiles. Teachers' visual strategies were quantified using Shannon's entropy (Shannon, 1948), wherein higher entropy values display more complexity. In this context, more complexity indicates that a teacher distributes their attention *equally* across all relevant students (AOIs) and when, after fixing upon one student, all the other students have the same *probability* of being looked at. On the contrary, less complexity could signal less monitoring of the group of students due to over- or under-focusing on some specific students. In the context of judgment accuracy, it was found that experts were more accurate in judging students and their underlying student characteristic profiles, but this difference did not reach statistical significance. However, with the focus on visual processes, experts' gaze patterns were more complex in comparison to those of novices. The experts' significantly higher entropy values suggest that they observed each student with a more constant frequency and transferred their gaze between all possible combinations of students with approximately equal frequency. However, this dissertation's fourth principal finding is that more complex scanpaths were associated with higher judgment accuracy of student profiles. The more a teacher followed an "expert-like" strategy (in the form of complex visual behavior), the better their judgment of student profiles.

The open question that remains to be discussed is, why was the "expert-like" strategy more successful? Evidence from research about human perception indicates that higher entropy indicates a perceivers' interest in a more detailed exploration of a scene (Krishna et al., 2018; Shic et al., 2008). In their comprehensive review, Shiferaw et al. (2019) found that gaze entropy can be used to measure visual scanning efficiency and that entropy increases relative to scene complexity and with more top-down processing. Because experts tend to mainly control *top-down* while novices are more subject to *bottom-up* cognitive processes (Hershler & Hochstein, 2009), experts' higher gaze entropy may reflect their advanced top-down processing mainly driven by task-related plans, scripts, and schemata derived from their professional knowledge (Gegenfurtner et al., 2011; Gegenfurtner, 2020; Goldberg et al., 2020). Therefore, their strategy to include more of the target students in their recurring scanpaths was more successful. The question of how judgments depend on how attention is distributed among diverse students is only partially understood. Dessus et al. (2016), for example, found that novice teachers faced a higher cognitive load compared to experts, and that the number of students that they were able to observe was related to the level of experience. Their findings showed that experienced teachers were able to scan a larger group of students and had a more comprehensive monitoring scheme that allowed them to gather more fine-grained information about their students. In another study, Karst and Bonefeld (2020) used a simulated classroom setting and teachers' click frequency (more clicks indicated that the teacher was gathering more information) as an indicator of attention allocation, and found that teachers judged an isolated student within a group of students better when they paid more attention (more clicks) to that particular student. However, teachers' *overall judgment* accuracy increased when teachers distributed their attention more evenly across all students.

However, while the thesis showed that higher complexity (high entropy) of teachers' scanpath patterns tended to contribute to the more accurate assessment of underlying student profiles, previous literature has also reported rather opposite results. For example, Kasarskis et al. (2001) demonstrated that experienced airplane pilots had lower entropy during scans of aircraft instruments, and exhibited a clearly-defined visual scanning pattern during a landing task, resulting in a better task outcome. Similarly, Di Stasi et al. (2016) found that gaze complexity decreased with increasing task complexity among helicopter pilots. However, pilots have a significant advantage over teachers in this context, as pilots can actively reduce complexity by hiding task-redundant instruments or regions. Aircraft instruments are static, standardized, and prioritized in importance (Brams et al., 2018), while teachers are typically instructed to monitor all of their students equally.

6.2 Implications for Teacher Education

This dissertation's findings are of practical relevance and have implications for teacher education and professional development. First, this work contributes additional empirical evidence of the richness of *intra-individual* differences in cognitive and motivational-affective characteristics, highlighting, in particular, the prevalence of inconsistencies, which can be mentioned along with the findings of Seidel (2006) or Linnenbring-Garica et al. (2012). Identifying replicable types of student profiles allows teachers to understand the strengths and weaknesses of different subgroups of students and subsequently design interventions that address their specific needs. For example, students with an underestimating profile are unlikely to improve positively when focusing on their cognitive skills, as their profile suggests that the problems lie in motivational deficits. Therefore, interventions should focus primarily on motivational training. In this context, by making pedagogical adjustments to their instruction (e.g., avoiding criticism or emphasizing the positive) or by reflecting on their feedback in terms of classroom reference norms (Köller, 2004) (e.g., focusing on temporal frames of comparison and reflecting on their feedback to avoid social comparison processes), teachers can help increase students' interest and improve their self-concept.

Furthermore, it is alarming that research on teachers' judgment accuracy of educationally-relevant student characteristics has continuously revealed that teachers perceive only consistent profiles and ignore students with inconsistent ones (Huber & Seidel, 2018; Südkamp et al., 2018). For teachers, these inconsistent profiles are meaningful and quite likely to be present in every classroom. Therefore, the results of this thesis can be used to raise teachers' awareness of the existence of students with inconsistent profiles. For example, teacher education programs can promote student teachers' declarative knowledge with respect to educationally-relevant student characteristics, as well as their intra-individual combination manifested in consistent and inconsistent student profiles; fostering this knowledge might improve teachers' judgment accuracy in this area.

Moreover, as outlined in this thesis theory section, teachers need to observe and notice student behavioral cues that provide diagnostically-relevant information about students' cognitive and motivational-affective characteristics to make accurate judgments. In this thesis, teachers who could *see like an expert* had a higher judgment accuracy than those who could not follow the experts' visual strategy. It was argued that experts' visual behavior might be indicative of their knowledge-driven observation and rapid information processing skills. Knowledge about experts' visual expertise might be meaningful for novice teachers. In this sense, novices could use recordings of expert teachers' gazes as a model for learning effective attention allocation (van Gog et al., 2009), just as the gaze of experienced surgeons is used to train and inform trainee surgeons (Khan et al., 2012). Hence, novices may learn effective strategies for allocating attention by adopting the gaze of experienced teachers.

6.3 Limitations and Further Research

Numerous limitations should be considered when interpreting the dissertations' findings. In terms of the first study associated with the thesis, the identified student profiles included only information from secondary school students in grade 9, and the article did not provide a longitudinal perspective on the temporal stability of student profiles. Future research would benefit from including multiple consecutive measurement time points to more deeply capture students' cognitive and motivational-affective changes—and thus membership in various student profiles—over time. In addition, further research should explore the underestimating profile—the student profile most frequently identified in mathematics and GLA—in more detail by including additional domains or covariates that might be predictive of this specific underestimation configuration (e.g., general school self-concept). It is plausible to assume that these students may have distanced themselves from school across domains or incorporated low school-related self-esteem in general, which carries over to these underestimating domain-general profiles.

In terms of the second study associated with the thesis, teachers' gazes were examined while watching a video, which is not the same as monitoring teachers' visual attention while they are actually teaching. For example, teachers are required to be more acutely aware of their surroundings in real-life teaching situations than when observing a video sequence. Future research might try to replicate the thesis findings using mobile eye-tracking to study teachers' scanpath patterns during real-life teaching. In this context, it is essential to note that the eye-tracking experiment had no variation in the visual stimulus (authentic classroom video). Although the selected video has been judged to be very authentic in a piloting study, the findings may be explicitly related to the chosen video footage. Further research should address this issue and support the findings with different video sequences. Furthermore, an event-based scanpath comparison method was used to identify differences and similarities between experts and novices. Therefore, it did not account for the amount of time a teacher spent at each AOI, which should be considered in future research. The differences in scanpath similarity could be even more significant if fixation time were taken into account, as previous research has repeatedly shown that experts process information more quickly than novices (Gegenfurtner et al., 2011). Also, scanpath length might play a vital role in better understanding experts' scanpaths (van den Bogert et al., 2014). Furthermore, the analyses showed that expert and novice teachers differed in their visual behavior, but little knowledge is available about how they differed in their interpretation of what they saw. Future research should focus on a more extensive pairing of eye-tracking and think-aloud protocols or subjective reports. Combining multiple data streams can increase knowledge about which student cues teachers have noticed.

6.4 Conclusion

This dissertation provides further empirical evidence of individual differences among students in terms of their learning-related characteristics along with diversity and commonness within these characteristics. By considering four essential educationally-relevant cognitive and motivational-affective characteristics for latent profiling, this work uncovered four domain-generalizable student profiles that teachers face in their everyday teaching: Students who have a) high overall, b) low overall, c) high cognitive but low motivational-affective, and d) low cognitive but high motivational-affective characteristics. Furthermore, this thesis revealed that teachers differ in their perceptional processes when observing students in order to make inferences about underlying student profiles. It adds to prior knowledge about teachers' professional vision the fact that teachers' gaze patterns during observation vary with respect to their professional experience. Experienced teachers were more able to continuously monitor a larger group of students in comparison to novice teachers. The distribution of attention among several students and the concomitant collection of judgment-relevant informative cues was related to a higher judgment accuracy of student profiles. Altogether, this thesis has presented important fundamentals that could be relevant to subsequent research as well as for teacher education.

7 References

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. Budapest, Hungary: Akadémiai Kiadó.

Alexander, P. A. (2003). The Development of Expertise: The Journey From Acclimation to Proficiency. *Educational Researcher*, *32*(8), 10–14. https://doi.org/10.3102/0013189X032008010

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology*, *87*(4), 559–575. https://doi.org/10.1037/0022-0663.87.4.559

Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The Role of Subject-Matter Knowledge and Interest in the Processing of Linear and Nonlinear Texts. *Review of Educational Research*, *64*(2), 201. https://doi.org/10.2307/1170694

Antes, J. R., & Kristjanson, A. F. (1991). Discriminating artists from nonartists by their eye-fixation patterns. *Perceptual and Motor Skills*, *73*(3 Pt 1), 893–894. https://doi.org/10.2466/pms.1991.73.3.893

Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften [Diagnostic competence of teachers]. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 157–160. https://doi.org/10.1024/1010-0652.23.34.157

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus web notes*, *21*(2), 1–22.

Ausubel, D. P. (2012). *The acquisition and retention of knowledge: A cognitive view*: Springer Science & Business Media.

Baumert, J. (Ed.) (2006). *Herkunftsbedingte Disparitäten im Bildungswesen: differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit: Vertiefende Analysen im Rahmen von PISA 2000 [Origin-related disparities in education: differential educational processes and problems of distributive justice: in-depth analyses in the context of PISA 2000]*. (1. Aufl.). Wiesbaden: VS Verl. für Sozialwiss. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=747619

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Blossfeld, H.-P., & Roßbach, H.-G. (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS)* (2nd ed.). *Edition ZfE: Vol. 3*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

Borko, H., Roberts, S., & Shavelson, R. (2008). Teachers' Decision Making: From Alan J. Bishop to Today. In (pp. 37–67). https://doi.org/10.1007/978-0-387-09673-5_4

Boshuizen, H.P.A., Schmidt, H. G., Custers, E.J.F.M., & van de Wiel, M. W. (1995). Knowledge development and restructuring in the domain of medicine: The role of theory and practice. *Learning and Instruction*, *5*(4), 269–289. https://doi.org/10.1016/0959-4752(95)00019-4

Brams, S., Hooge, I. T. C., Ziv, G., Dauwe, S., Evens, K., Wolf, T. de, . . . Helsen, W. F. (2018). Does effective gaze behavior lead to enhanced performance in a complex error-detection cockpit task? *PloS One*, *13*(11), e0207439. https://doi.org/10.1371/journal.pone.0207439

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (Vol. 1): Cambridge University Press.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212. https://doi.org/10.1007/BF01246098

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. *Dual-process theories in social psychology*. New York, NY, US: The Guilford Press.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, *29*(8), 1146–1152. https://doi.org/10.3758/BF03206384

Clarridge, P. B., & Berliner, D. C. (1991). Perceptions of student behavior as a function of expertise. *The Journal of Classroom Interaction*, 1–8.

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

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In *Handbook of research on teaching* (pp. 605–629). New York, NY u.a.: Macmillan u.a.

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges*, *84*(8), 1022–1028. https://doi.org/10.1097/ACM.0b013e3181ace703

Dai, D. Y., & Sternberg, R. J. (Eds.) (2004). *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*. *The educational psychology series*. Mahwah, N.J: Lawrence Erlbaum Associates. Retrieved from http://lib.myilibrary.com/detail.asp?id=237498

Davies, A., Markel, V., Simon, H., & Caroline, J. (2016). *The visualisation of eye-tracking scanpaths: What can they tell us about how clinicians view electrocardiograms?*

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology : JFP : Journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, *19*(2), 294–304. https://doi.org/10.1037/0893-3200.19.2.294

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8–24. https://doi.org/10.1037/h0088969

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, *78*(2), 430–447. https://doi.org/10.1111/j.1467-8624.2007.01007.x

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, P. Dessus, O. Cosnefroy, & V. Luengo (Eds.), *"Keep Your Eyes on 'em all!": A Mobile Eye-Tracking Analysis of Teachers' Sensitivity to Students: Adaptive and Adaptable Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-319-45153-4_6

Di Stasi, L., Diaz-Piedra, C., Rieiro, H., Ruiz-Rabelo, J., Cardenas, D., Cherino, A., . . . Catena, A. (2016). Gaze-entropy as a task load index for safety-critical operators: Military pilots and surgeons. *Journal of Vision*, *16*(12), 1341. https://doi.org/10.1167/16.12.1341

Dolezalova, J., & Popelka, S. (2016). ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques. *Journal of Eye Movement Research*, *9*(5), 1–13. https://doi.org/10.16910/jemr.9.4.5

Duchhardt, C., & Gerdes, A. (2013). NEPS Technical Report for Mathematics–Scaling results of Starting Cohort 4 in ninth grade. *Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study*, 1–28.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Fiske, S., & Neuberg, S. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. In Mark P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Academic Press. https://doi.org/10.1016/S0065-2601(08)60317-2

Fiske, S. T., Kinder, D. R., & Larter, W.M. (1983). The novice and the expert: Knowledge-based strategies in political cognition. *Journal of Experimental Social Psychology*, *19*(4), 381–400. https://doi.org/10.1016/0022-1031(83)90029-X

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. T. Fiske, M. Lin, & S. L. Neuberg (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York, NY, US: The Guilford Press.

Fleming, M. L., & Malone, M. R. (1983). The relationship of student characteristics and student performance in science as viewed by meta-analysis research.

Journal of Research in Science Teaching, *20*(5), 481–495. https://doi.org/10.1002/tea.3660200510

Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach. *Journal of Eye Movement Research*, *5*(4), 1–14.

Gegenfurtner, A. (2020). *Professional vision and visual expertise* (Habilitation Thesis). University of Regensburg, Regensburg.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552.

Goetz, T., Haag, L., Lipnevich, A. A., Keller, M. M., Frenzel, A. C., & Collier, A. P. M. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology*, *5*, 1153. https://doi.org/10.3389/fpsyg.2014.01153

Gogol, K., Brunner, M., Preckel, F., Goetz, T., & Martin, R. (2016). Developmental Dynamics of General and School-Subject-Specific Components of Academic Self-Concept, Academic Interest, and Academic Anxiety. *Frontiers in Psychology*, *7*, 356. https://doi.org/10.3389/fpsyg.2016.00356

Goldberg, P., Schwerter, J., Seidel, T., Müller, K., & Stürmer, K. (2020). How does learners' behavior attract preservice teachers' attention during teaching? *Teaching and Teacher Education*, *97*(7), 103213. https://doi.org/10.1016/j.tate.2020.103213

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Gralewski, J., & Karwowski, M. (2016). Are Teachers' Implicit Theories of Creativity Related to the Recognition of Their Students' Creativity? *The Journal of Creative Behavior*, *52*(2), 156–167. https://doi.org/10.1002/jocb.140

Haberkorn, K., & Pohl, S. (2013). Cognitive basic skills (non-verbal): Data in the scientific use file. *Bamberg: University of Bamberg, National Educational Panel Study (NEPS). Available online:* https://www.*neps-data.*

de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/com_cogbasic2013_en. pdf (accessed on 1 December 2020).

Haberkorn, K., Pohl, S., Hardt, K., & Wiegand, E. (2012). NEPS technical report for reading–Scaling results of starting cohort 4 in ninth grade (NEPS Working Paper No. 16). *Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel*.

Haider, & Frensch (1996). The Role of Information Reduction in Skill Acquisition. *Cognitive Psychology*, *30*(3), 304–337. https://doi.org/10.1006/cogp.1996.0009

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., . . . Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*(4), 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Hershler, O., & Hochstein, S. (2009). The importance of being expert: Top-down attentional control in visual search with photographs. *Attention, Perception & Psychophysics*, *71*(7), 1478–1486. https://doi.org/10.3758/APP.71.7.1478

Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts [Teaching quality and teacher's profession: Diagnosis, evaluation and improvement of teaching]* (4th ed.) Seelze: Klett-Kallmeyer.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2015). *Eye tracking: A comprehensive guide to methods and measures* (First published in paperback). Oxford: Oxford University Press.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Huber, S. (2017). Individual Student Learning Dispositions – Teacher and Student Perspectives on Incoherences and Their Development. Dissertation. http://mediatum.ub.tum.de/?id=1356687

Huber, S. A., Häusler, J., Jurik, V., & Seidel, T. (2015). Self-underestimating students in physics instruction: Development over a school year and its connection to internal learning processes. *Learning and Individual Differences*, *43*(2), 83–91. https://doi.org/10.1016/j.lindif.2015.08.021

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

Kaiser, J., Helm, F., Retelsdorf, J., Südkamp, A., & Möller, J. (2012). Zum Zusammenhang von Intelligenz und Urteilsgenauigkeit bei der Beurteilung von Schülerleistungen im Simulierten Klassenraum [On the relationship between intelligence and judgment accuracy in assessing student performance in the simulated classroom]. *Zeitschrift für Pädagogische Psychologie*, *26*(4), 251–261. https://doi.org/10.1024/1010-0652/a000076

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, *28*(4), 73–84. https://doi.org/10.1016/j.learninstruc.2013.06.001

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, *108*, 67–76. https://doi.org/10.1016/j.visres.2015.01.013

Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen [Diagnostic competence of elementary and high school teachers in the areas of achievement and interests]. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 197–209. https://doi.org/10.1024/1010-0652.23.34.197

Karst, K., & Bonefeld, M. (2020). Judgment accuracy of preservice teachers regarding student performance: The influence of attention allocation. *Teaching and Teacher Education*, *94*(1), 103099. https://doi.org/10.1016/j.tate.2020.103099

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C. (2001). Comparison of Expert and Novice Scan Behaviors During VFR Flight. *Proceedings of the 11th International Symposium on Aviation Psychology*.

Khan, R. S. A., Tien, G., Atkins, M. S., Zheng, B., Panton, O. N. M., & Meneghetti, A. T. (2012). Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical Endoscopy*, *26*(12), 3536–3540. https://doi.org/10.1007/s00464-012-2400-7

Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen [Consequences of performance groupings]. Pädagogische Psychologie und Entwicklungspsychologie: Vol. 37*. Münster: Waxmann.

Kosel, C., Holzberger, D., & Seidel, T. (2021). Identifying Expert and Novice Visual Scanpath Patterns and Their Relationship to Assessing Learning-Relevant Student Characteristics. *Frontiers in Education*, *5*, 3. https://doi.org/10.3389/feduc.2020.612175

Kosel, C., Wolter, I., & Seidel, T. (2020). Profiling secondary school students in mathematics and German language arts using learning-relevant cognitive and motivational-affective characteristics. *Learning and Instruction*, *73*(4), 101434. https://doi.org/10.1016/j.learninstruc.2020.101434

Krishna, O., Helo, A., Rämä, P., & Aizawa, K. (2018). Gaze distribution analysis and saliency prediction across age groups. *PloS One*, *13*(2), e0193149. https://doi.org/10.1371/journal.pone.0193149

Lachner, A., Jarodzka, H. & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science. 44.* https://doi.org/10.1007/s11251-016-9376-y.

Lau, S., & Roeser, R. W. (2008). Cognitive abilities and motivational processes in science achievement and engagement: A person-centered analysis. *Learning and Individual Differences*, *18*(4), 497–504. https://doi.org/10.1016/j.lindif.2007.11.002

Lauermann, F., Meißner, A., & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology*, *112*(2), 364–383. https://doi.org/10.1037/edu0000377

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lazarsfeld, P. F. (1957). *Latent structure analysis*: Columbia University, Bureau of Applied Social Research New York.

Leuders, T., Philipp, K., & Leuders, J. (2018). *Diagnostic Competence of Mathematics Teachers*. Cham: Springer International Publishing.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, *10*, 707.

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

Loibl, K., Leuders, T., & Dörfler, T. (2020). A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education*, *91*(3), 103059. https://doi.org/10.1016/j.tate.2020.103059

Machts, N., Kaiser, J., Schmidt, F. T.C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, *19*(2), 85–103. https://doi.org/10.1016/j.edurev.2016.06.003

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

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

Mathôt, S., Cristino, F., Gilchrist, I., & Theeuwes, J. (2012). A simple way to estimate similarity between pairs of eye movement sequences. *Eye*, *5*, 1–15.

McCutcheon, A. L. (1987). *Latent class analysis* (Vol. 64): Sage.

McIntyre, N. A., & Foulsham, T. (2018). Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instructional Science*, *46*(3), 435–455. https://doi.org/10.1007/s11251-017-9445-x

Muthén, B. (2004). Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 346–369). 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc. https://doi.org/10.4135/9781412986311.n19

Muthén, L. K., & Muthén, B. O. (2015). *Mplus User's Guide. Seventh Edition.* Los Angeles.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Noton, D., & Stark, L. (1971a). Eye movements and visual perception. *Scientific American*, *224*(6), 35–43.

Noton, D., & Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision research*, *11*(9), 929-IN8. https://doi.org/10.1016/0042-6989(71)90213-6

OECD, P. (2016). *Results (Volume I): Excellence and equity in education*: Paris: OECD Publishing.

O'Neill, E. C., Kong, Y. X. G., Connell, P. P., Ong, D. N., Haymes, S. A., Coote, M. A., & Crowston, J. G. (2011). Gaze behavior among experts and trainees during optic disc examination: Does how we look affect what we see? *Investigative Ophthalmology & Visual Science*, *52*(7), 3976–3983. https://doi.org/10.1167/iovs.10-6912

Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment Confidence and Judgment Accuracy of Teachers in Judging Self-Concepts of Students. *The Journal of Educational Research*, *106*(1), 64–76. https://doi.org/10.1080/00220671.2012.667010

R Core Team (2020). R [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raven, J. (1989). The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States. *Journal of Educational Measurement*, *26*(1), 1–16. https://doi.org/10.1111/j.1745-3984.1989.tb00314.x

Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion [Textbook Test Theory - Test Construction]* (2. ed.). *Psychologie Lehrbuch*. Bern: Huber.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*(2), 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Schiefele, U. (2009). Situational and individual interest. In *Educational psychology handbook series. Handbook of motivation at school* (pp. 197–222). New York, NY, US: Routledge/Taylor & Francis Group.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, A. Krapp, & A. Renninger (Eds.), *The Role of interest in Learning and Development* (pp. 183–212). Hoboken: Taylor and Francis.

Schnitzler, K., Holzberger, D., & Seidel, T. (2020). Connecting Judgment Process and Accuracy of Student Teachers: Differences in Observation and Student Engagement Cues to Assess Student Characteristics. *Frontiers in Education*, *5*, 1016. https://doi.org/10.3389/feduc.2020.602470

Schrader, F.-W., & Helmke, A. (2008). Determinanten der Schulleistung [Determinants of school performance]. In *Lehrer-Schüler-Interaktion* (pp. 285–302). Springer.

Schwarz, G. (1978). Estimating the Dimension of a Model, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

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

Seidel, T. & Reiss, K. M. (2014). Lerngelegenheiten im Unterricht [Learning opportunities in class] In Seidel. T & Krapp, A. (Eds), Pädagogische Psychologie (pp.253-276). Weinheim: Julius Beltz.

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 [Micro-environments in the classroom: how cognitive and motivational-affective student dispositions affect cognition and behavior in the subject classroom]. *Bedingungen und Effekte guten Unterrichts*, 65–87.

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, *15*(7), 98. https://doi.org/10.1007/s10648-020-09532-2

Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, *51*(4), 739–771. https://doi.org/10.3102/0002831214531321

Senler, B., & Sungur, S. (2009). Parental influences on students' self-concept, task value beliefs, and achievement in science. *The Spanish Journal of Psychology*, *12*(1), 106–117. https://doi.org/10.1017/s1138741600001529

SensoMotoric Instruments GmbH (2017). BeGaze (Version 3.7) [Computer software]. Berlin.

Shajek, A., Lüdtke, O., & Stanat, P. (2006). Akademische Selbstkonzepte bei Jugendlichen mit Migrationshintergrund [Academic self-concepts among adolescents with an immigrant background]. Unterrichtswissenschaft 34:125-145.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shavelson, R. J. (1978). Teachers' Estimates of Student 'States of Mind' and Behavior. *Journal of Teacher Education*, *29*(5), 37–40. https://doi.org/10.1177/002248717802900511

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Shic, F., Chawarska, K., Bradshaw, J., & Scassellati, B. (2008). Autism, eye-tracking, entropy. In F. Shic, K. Chawarska, J. Bradshaw, & B. Scassellati (Chairs), *Autism, eye-tracking, entropy*.

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience and Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007

Skaalvik, S., & Skaalvik, E. M. (2004). Gender Differences in Math and Verbal Self-Concept, Performance Expectations, and Motivation. *Sex Roles*, *50*(3/4), 241–252. https://doi.org/10.1023/B:SERS.0000015555.40976.e6

Snow, R. (1989). Cognitive-Conative Aptitude Interactions in Learning. In R. Kanfer (Ed.), *Abilities, Motivation and Methodology*. Routledge.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teachers' assessment of student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, *19*(1/2), 85–95. https://doi.org/10.1024/1010-0652.19.12.85

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

Südkamp, A., Praetorius, A.-K., & Spinath, B. (2018). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*, *76*, 204–213. https://doi.org/10.1016/j.tate.2017.09.016

Tan, S. K. S. (1996). Differences between experienced and inexperienced physical education teachers' augmented feedback and interactive teaching decisions. *Journal of Teaching in Physical Education*, *15*(2), 151–170.

Tang, H., Day, E., Atkinson, M., & Pienta, N. (2018). GrpString: An R Package for Analysis of Groups of Strings. *The R Journal*, *10*(1), 359. https://doi.org/10.32614/RJ-2018-002

Tobias, S. (1994). Interest, Prior Knowledge, and Learning. *Review of Educational Research*, *64*(1), 37–54. https://doi.org/10.3102/00346543064001037

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*(1), 208–216. https://doi.org/10.1016/j.tate.2013.09.001

Van der Loo, M.J. (2014). The stringdist Package for Approximate String Matching. *The R Journal*, *6*(1), 111. https://doi.org/10.32614/RJ-2014-011

Van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, *13*(2), 155–176. https://doi.org/10.1007/s10857-009-9130-3

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, *25*(3), 785–791.

Vermunt, J. K. (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Vermunt, J. K., & Magidson, J. (2016). *Upgrade Manual for Latent GOLD 5.1*. Belmont Massachusetts.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450. https://doi.org/10.1007/BF02294627

Wilson, H. E., Siegle, D., McCoach, D. B., Little, C. A., & Reis, S. M. (2014). A Model of Academic Self-Concept. *Gifted Child Quarterly*, *58*(2), 111–126. https://doi.org/10.1177/0016986214522858

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

8. Appendix

Appendix A

Kosel, C., Wolter, I., & Seidel, T. (2020). Profiling secondary school students in mathematics and German language arts using learning-relevant cognitive and motivational-affective characteristics. *Learning and Instruction*, *73*. https://doi.org/10.1016/j.learninstruc.2020.101434

Appendix B

Kosel, C., Holzberger, D., & Seidel, T. (2021). Identifying Expert and Novice Visual Scanpath Patterns and their Relationship to Assessing Learning-Relevant Student Characteristics. *Frontiers in Education*, *5*:612175. https://doi.org/10.3389/feduc.2020.612175

Note:

For copyright reasons, supplements are not included in this online publication of the dissertation.