Understanding Effective Teaching From the Student Perspective: Exploring Dynamics in Teaching Quality

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

M.Sc. Johanna Sophie Marder

Tübingen

2022

1. Betreuer:	Prof. Dr. Richard Göllner
2. Betreuer:	Prof. Dr. Benjamin Fauth
3. Betreuer:	Prof. Dr. Ulrich Trautwein

Tag der mündlichen Prüfung:	11. Februar 2022

Dekan:	Prof. Dr. Josef Schmid
1. Gutachter:	Prof. Dr. Richard Göllner
2. Gutachter:	Prof. Dr. Lisa Bardach

DANKSAGUNG

Viele Menschen haben in unterschiedlichster, aber essentieller Weise zum Gelingen dieses Vorhabens beigetragen. Euch möchte ich an dieser Stelle herzlich danken.

Mein ganz besonderer Dank gilt meinen Betreuern Prof. Dr. Richard Göllner, Prof. Dr, Benjamin Fauth und Prof. Dr. Ulrich Trautwein. Vielen Dank, dass ihr mir die Möglichkeit zur Promotion gegeben habt und den Weg so mit mir gegangen seid, wie er gekommen ist. Ihr habt mich in allen Phasen der Promotion begleitet, unterstützt und inspiriert. Danke, dass ich so viel von Euch lernen durfte.

Darüber hinaus gilt mein großer Dank Dr. Wolfgang Wagner, der immer ein offenes Ohr für meine methodischen Fragen hatte. Ebenso haben mich Dr. Karen Aldrup und Prof. Dr. Felicitas Thiel durch ihre wertvollen Anmerkungen zu den jeweiligen Artikeln sehr unterstützt.

Danke auch an meine Kolleg*innen und Freund*innen am Hector-Institut. Ihr habt die Zeit in Tübingen für meine Familie und mich zu etwas ganz besonderem gemacht. Danke für den Austausch, die Gespräche und die schönen Momente in den letzten Jahren.

Ein großer Dank gilt schließlich auch meiner Familie. Danke für alles was ihr mir mit auf den Weg gegeben habt. Steve, danke für deine Unterstützung und Zuversicht. Gabriel und Sebastian, ihr zeigt mir was wichtig ist.

ABSTRACT

How to promote adolescent students' learning and socio-emotional functioning in secondary school is a central question in empirical educational research. Hence, a large body of research has examined which specific factors and processes within the complex school context are most important for predicting students' learning and socio-emotional functioning (Eccles & Roeser, 2011). Overall, research highlights the important role of the classroom environment and processes that happen therein. Specifically, the quality of teaching provided by the teacher is one of the most central characteristics that affect students' experience of their classroom (e.g., Wang et al., 1993; Hattie, 2009). However, empirical research tends to treat teaching quality as a rather "static" factor within the classroom (Way et al., 2007) and little is known about dynamics that operate over time and between the teacher and the specific students in the classroom. For example, lower secondary school coincides with early adolescence, during which students experience critical developmental processes (Eccles & Midgley, 1989; Eccles et al, 1993). Furthermore, high-quality teaching is defined and established within the classroom context, which is in turn shaped by the teacher and the specific students being taught (Doyle, 2007).

The overarching aim of the present dissertation was to explore the dynamic and contextual nature of teaching quality. With respect to the dynamic nature of teaching quality, the present dissertation examined changes in teaching quality over time and its associations with students' learning and socio-emotional functioning from a longitudinal perspective. To illuminate the contextual nature of teaching quality, the present dissertation explored the extent to which the specific students in the classroom contribute to teaching quality via their background characteristics and their own agentic behavior (i.e., disruptive behavior) in the classroom. A subordinate aim of the present dissertation concerned the assessment of teaching quality via student reports. Student reports provide unique insight into what happens in the classroom, but are suspected of being affected by factors unrelated to teaching quality (Lüdtke et al., 2009). Therefore, the present dissertation began with an examination of whether students at different grade levels are able to report on teaching quality in a reliable and valid way.

Study 1 (*Ask me, I (Dis)agree! Acquiescence in Student Ratings of Teaching Quality in German Vocational Schools*) examined the extent to which student reports on teaching quality provided by students at different grade levels are affected by an acquiescent response style

(acquiescence: “yay-saying”, dis-acquiescence: “nay-saying”) while additionally controlling for an extreme response style. Using a large cross-sectional sample of n = 2,234 fifth-grade and n = 1,832 eighth-grade students from vocational-track schools in Germany, Study 1 a) investigated the degree to which acquiescence differs at the student and at the classroom level and b) evaluated the impact of acquiescence on the psychometric properties of teaching quality reports by systematically comparing factor means and factor inter-correlations before and after statistically controlling for acquiescence. The results of Study 1 suggest acquiescence was not counterbalanced at the classroom level. In addition to that, acquiescence and extreme responding primarily affected reports provided by younger students (i.e., fifth grade) and when negatively worded items were used. Nonetheless, acquiescence affected the psychometric properties of student data on teaching quality to only a minor degree, which can be considered largely unproblematic for practical use. Overall, the results support students’ ability to provide valid and reliable information on teaching quality.

Study 2 *(Student Development in Early Adolescence: Does Teaching Quality Shape Students’ Academic Achievement, Academic Engagement, and Their Social and Emotional School Adjustment?)* addressed the dynamic nature of teaching quality by longitudinally linking the trajectories of multiple aspects of teaching quality to the trajectories of students’ development (i.e., academic achievement, academic engagement and their social and emotional school adjustment). In this study, teaching quality was assessed from both the student and teacher perspectives. To this end, groups of students forming specific school classes and their homeroom teachers were followed over three years of lower secondary school (i.e., fifth to eighth grade). The pooled sample consisted of N = 3,880 students and N = 126 homeroom teachers from German vocational-track schools. The results showed that students and their homeroom teachers reported negative changes in teaching quality across lower secondary school. Moreover, teaching quality buffered the downward trend in most aspects of students’ academic engagement and promoted students’ academic achievement at the classroom level. In contrast, students’ individual perceptions of teaching quality (i.e., student level) were most important for their social and emotional school adjustment, indicating that students’ social and emotional school adjustment during adolescence is heavily driven by individual processes. Overall, the revealed associations were more pronounced when examining student reports of teaching quality compared to teacher self-reports. The results highlight that teaching is variable over time and underscore the important role of teachers in fostering young adolescent students’ development.

Study 3 (*Classroom Management: Can it be too Much of a Good Thing?*) focused on the contextual nature of teaching quality. Using classroom management as an example, the study examined the extent to which the specific students in the classroom contribute to the classroom management progress through a) their background characteristics and b) their own agentic behavior (i.e., disruptive behavior). To this end, the study examined key aspects of classroom management referring either to students' actions (i.e., disruptive behavior), or teachers' actions (i.e., monitoring, structure, clarity of instruction). Study 3 used data from two independent large-scale data sets. The first data set consistent of N_1 = 4,645 German tenth-grade students enrolled in the academic or intermediate school tracks. The second data set consisted of N_2 = 6,298 German students from Grades 6 to 10 enrolled in either academic-track or in different types of vocational-track schools. Overall, the results suggested that the specific students in the classroom significantly contribute to the classroom management process. Across both data sets, classroom management measures referring to students' behavior in the classroom (i.e., disruptive behavior) were more closely related to students' background characteristics than measures referring to teacher actions were. Moreover, after accounting for the average level of disturbances in the classroom, both students' disruptive behavior and teachers' monitoring activity were negatively associated with students' pre-adjusted math achievement. This finding suggests that students' disruptive behavior in the classroom contributes to the association between teachers' management actions and students' academic achievement.

In the end, the findings of the three empirical studies are summarized and discussed in light of their contributions to empirical educational research. Moreover, theoretical and practical implications for educational research and practice are derived. From a theoretical perspective, the findings of the present dissertation are highly relevant for conceptualizing and measuring teaching quality in educational research. From a practical perspective, the findings of the present dissertation contribute to the discussion on assessing teaching quality via student reports and underscore the critical role of teachers for students' long-term academic development, and their social and emotional school adjustment.

ZUSAMMENFASSUNG

Schülerinnen und Schüler in ihrem Lernen und in ihren sozio- emotionalen Fähigkeiten zu fördern ist ein zentrales Anliegen empirischer Bildungsforschung. Eine Vielzahl an Forschungsarbeiten hat sich dabei mit der Frage beschäftigt, welche Faktoren und Prozesse im komplexen schulischen Umfeld für das Lernen und die sozio-emotionalen Fähigkeiten von Schülerinnen und Schülern relevant sind (Eccles & Roeser, 2011). In dieser Hinsicht haben sich vor allem jene Prozesse, welche innerhalb des Klassenraumes stattfinden, als besonders bedeutsam erwiesen. Dazu zählt unmittelbar die Qualität des Unterrichts, den die Schülerinnen und Schüler täglich erleben (e.g., Wang et al., 1993; Hattie, 2009). Jedoch wird die Qualität des Unterrichts in empirischen Forschungsarbeiten hauptsächlich als „statischer" Faktor (Way et al., 2007) im Klassenraum behandelt und folglich ist wenig über die Dynamiken bekannt, welche sich im Unterricht über längere Zeiträume hinweg und zwischen der Lehrkraft und den Schülerinnen und Schülern im Klassenraum abspielen. Beispielsweise überschneidet sich die Sekundarstufe 1 mit der Phase der frühen Adoleszenz, in welcher Schülerinnen und Schüler grundlegende Entwicklungs- und Veränderungsprozesse durchlaufen (Eccles & Midgley, 1989; Eccles et al, 1993). Weiterhin entwickelt sich die Qualität des Unterrichts erst im Kontext der zu unterrichtenden Klasse, wozu sowohl die Lehrkraft als auch die jeweiligen Schülerinnen und Schüler beitragen (Doyle, 2006).

Das übergeordnete Ziel dieser Dissertation bestand daher in der Erforschung der dynamischen und der kontextuellen Natur der Unterrichtsqualität. Um die dynamische Natur der Unterrichtsqualität zu erforschen, wurde die längsschnittliche Entwicklung der Unterrichtsqualität, sowie deren Zusammenhang mit der Lern- und der sozio-emotionalen Entwicklung der Schülerinnen und Schüler untersucht. Um die kontextuelle Natur der Unterrichtsqualität zu erforschen, wurde untersucht, in wie weit die zu unterrichtenden Schülerinnen und Schüler zur Qualität des Unterrichts beitragen. Dabei wurden die Hintergrundmerkmale und das eigene Verhalten (d.h., das Störverhalten) der Schülerinnen und Schüler im Klassenraum in den Blick genommen.

Ein weiteres Anliegen dieser Dissertation betraf die Erfassung der Unterrichtsqualität mit Schülerurteilen. Schülerurteile ermöglichen einzigartige Einblicke in die Prozesse, welche innerhalb des Klassenraums ablaufen. Jedoch wird Schülerurteilen auch unterstellt, von Faktoren

jenseits der Qualität des Unterrichts beeinflusst zu sein (e.g., Lüdtke et al., 2009). Aus diesem Grund untersuchte die vorliegende Dissertation zunächst, in wie weit Schülerinnen und Schüler unterschiedlicher Altersstufen in der Lage sind, valide und zuverlässige Informationen über Unterrichtsqualität zu liefern.

Studie 1 (*Ask me, I (Dis)agree! Acquiescence in Student Ratings of Teaching Quality in German Vocational Schools*) untersuchte, in wie weit die Urteile über die Unterrichtsqualität von Schülerinnen und Schülern unterschiedlicher Altersstufen von einer akquieszenten Antworttendenz beeinflusst sind (Akquieszenz: „Ja-Sage“ Tendenz; Dis-akquieszenz: „Nein-Sage“ Tendenz). Zusätzlich wurde für extremes Antwortverhalten kontrolliert. Unter der Verwendung einer großen Stichprobe basierend auf Daten von n = 2.234 Fünftklässlern und n = 1.832 Achtklässlern der nicht-gymnasialen Schulformen in Deutschland untersuchte Studie 1 zum einen, in wie weit Akquieszenz auf der Schülerebene und auf der Klassenebene variiert. Zum anderen untersuchte Studie 1 den Einfluss der Akquieszenz auf die psychometrischen Eigenschaften der Daten, indem systematisch Faktormittelwerte und Faktorinterkorrelationen vor und nach der Kontrolle der Akquieszenz verglichen wurden. Es zeigte sich, dass der Einfluss der Akquieszenz durch die Aggregation der Daten auf der Klassenebene nicht ausgemittelt werden konnte. Weiterhin legen die Ergebnisse nahe, dass Akquieszenz und extremes Antwortverhalten hauptsächlich die Urteile jüngerer Schülerinnen und Schüler betreffen, insbesondere wenn negativ formulierte Items verwendet werden. Trotz allem übte Akquieszenz einen eher geringen Einfluss auf die psychometrischen Eigenschaften der Daten aus, der als trivial für den praktischen Gebrauch der Schülerdaten erachtet werden kann. Insgesamt sprechen die Ergebnisse dafür, dass Schülerinnen und Schüler valide und zuverlässige Informationen über Unterrichtsqualität geben können.

Studie 2 (*Student Development in Early Adolescence: Does Teaching Quality Shape Students' Academic Achievement, School Engagement, and Their Social and Emotional School Adjustment?)* befasste sich mit der dynamischen Natur der Unterrichtsqualität. Dazu wurde untersucht, in wie weit der Entwicklungsverlauf der Unterrichtsqualität mit dem Verlauf der Lernentwicklung sowie dem Entwicklungsverlauf der sozio-emotionalen Fähigkeiten der Schülerinnen und Schüler zusammen hängt. In Studie 2 wurde Unterrichtsqualität aus der Perspektive der Schülerinnen und Schüler, sowie der Klassenlehrkraft erfasst. Dazu wurden die gleichen Schülerinnen und Schüler einer Schulklasse und deren Klassenlehrkräfte über drei

aufeinanderfolgende Schuljahre der Sekundarstufe 1 hinweg (d.h., fünfte bis achte Klasse) befragt. Die Gesamtstichprobe setzte sich aus Daten basierend auf N = 3.880 Schülerinnen und Schülern sowie N = 126 Klassenlehrkräften zusammen. Den Ergebnissen zufolge berichteten die Schülerschaft und deren Klassenlehrkräfte negative Veränderungen in der wahrgenommenen Unterrichtsqualität über den Verlauf der Sekundarstufe 1 hinweg. Darüber hinaus hing die längschnittliche Entwicklung der Unterrichtsqualität mit der Lernentwicklung sowie der Entwicklung der sozio-emotionalen Fähigkeiten der Schülerinnen und Schüler zusammen: Veränderungen der Unterrichtsqualität standen am konsistentesten mit Veränderungen der Lernentwicklung der Schülerinnen und Schüler auf der Schüler- und Klassenebene in Verbindung. Die individuelle Wahrnehmung der Unterrichtsqualität der Schülerinnen und Schüler (d.h., Schülerebene) war vor allem für deren sozio-emotionalen Entwicklung relevant. Diese Zusammenhänge ergaben sich hauptsächlich aus der Perspektive der Schülerinnen und Schüler, nicht jedoch aus der Perspektive der Klassenlehrkräfte. Die Ergebnisse verdeutlichen, dass die Qualität des Unterrichts zeitlichen Veränderungen unterliegen kann. Weiterhin betonen die Ergebnisse die einflussreiche Rolle der Lehrkräfte, jugendliche Schülerinnen und Schüler in ihrer Entwicklung zu unterstützen

Studie 3 (*Classroom Management: Can it be too Much of a Good Thing?*) befasste sich mit der kontextuellen Natur der Unterrichtsqualität. Am Beispiel der Klassenführung untersuchte die Studie, in wie weit die zu unterrichtenden zu unterrichtenden Schülerinnen und Schüler zur Qualität des Unterrichts durch a) ihre Hintergrundmerkmale und b) ihr eigenes Verhalten (d.h., Störverhalten) beitragen. Zu diesem Zweck verwendete die Studie Kernaspekte der Klassenführung, welche entweder das Verhalten der Schülerschaft (d.h., deren Störverhalten) oder das Managementverhalten der Lehrkraft (d.h., Monitoring, Regelklarheit, Struktur) im Klassenraum adressierten. Dazu wurden in Studie 3 zwei große, unabhängige Stichproben aus Deutschland verwendet. Die erste Stichprobe basierte auf Daten von auf N_1 = 4.645 Zehntklässlern, die entweder ein Gymnasium oder die Realschule besuchten. Die zweite Stichprobe basierte auf Daten von N_2 = 6.298 Schülerinnen und Schülern der Klassenstufen sechs bis zehn, die entweder das Gymnasium oder eine nicht-gymnasiale Schulform besuchten. Insgesamt weisen die Ergebnisse darauf hin, dass die zu unterrichtenden Schülerinnen und Schüler im Klassenraum zur Qualität des Unterrichts beitragen. In beiden Stichproben waren jene Aspekte der Klassenführung, welche das Verhaltend der Schülerschaft adressierten (z.B. das Störverhalten), stärker mit den Hintergrundmerkmalen der Schülerschaft assoziiert, als Aspekte der Klassenführung, welche das

Managementverhalten der Lehrkraft adressierten. Darüber hinaus zeigte sich, dass unter der Berücksichtigung des durchschnittlichen Anteils von Störungen im Unterricht, sowohl das schülerseitige Störverhalten als auch das Monitoring der Lehrkraft in einem negativen Zusammenhang mit der Matheleistung der Schülerinnen und Schüler stand. Die Ergebnisse legen nahe, dass das schülerseitige Störverhalten im Klassenraum zur Effektivität des Monitorings hinsichtlich der Matheleistung der der Schülerinnen und Schüler beiträgt.

Die Ergebnisse der drei empirischen Studien werden zusammengefasst und hinsichtlich ihres Beitrags zum aktuellen Stand der empirischen Bildungsforschung diskutiert. Weiterhin werden Implikationen für die empirische Bildungsforschung und Bildungspraxis abgeleitet. Aus theoretischer Sicht sind die generierten Ergebnisse der Dissertation für a) die Konzeptualisierung und b), die Erfassung der Unterrichtsqualität relevant. Aus praktischer Sicht leisten die generieten Ergebnisse einen wichtigen Beitrag zur Diskussion über die Erfassung der Unterrichtsqualität mit Schülerurteilen und zeigen die wichtige Rolle der Lehrkraft, Schülerinnen und Schüler in ihrer langfristigen akademischen und sozio-emotionalen Entwicklung zu unterstützen.

CONTENTS

1 INTRODUCTION AND THEORETICAL BACKGROUND

In many ways, school plays an important role in the life of adolescent students. School is the place where adolescent students spend the majority of their time outside the home (Hamre & Pianta, 2006), and can provide numerous opportunities for students to practice and learn academic and socio-emotional skills. As such, students learn in school how to make friends and develop healthy social relationships outside the family, with peers and teachers (Emmer & Gerwels, 2006; Furman & Buhrmester, 1992). Moreover, students gain knowledge, skills and competences and start to differentiate their interests (Baumert & Köller, 1998; Hidi & Ainley, 2002), all of which is associated with the developmental process of establishing oneself as a competent and autonomous being and developing a unique identity (Roeser et al., 2000). Lastly, students' academic and socio-emotional functioning in school often sets the stage for their later life prospects and success (Chernyshenko et al., 2018; OECD, 2021). That is, students who experience academic and socio-emotional struggles at school are at higher risk for early school drop-out (Battin-Pearson et al., 2000; Archambault et al., 2009), developing mental and physical health issues (Becker et al., 2014; Orth et al., 2008), receiving criminal convictions (Bäckman, 2017) and having difficulties in the labor market (National Research Council, 2012), which in turn has economic and financial consequences for both individuals and society in general (Rumberger, 2001).

Taken together, schools are crucial socio-cultural contexts in which students' learning and socio-emotional functioning develops over time (e.g., Eccles & Roeser, 2011). Given the social and organizational complexity of the school context, a multitude of factors and processes could potentially be relevant for students' learning and socio-emotional functioning. In general, there is large consensus in theoretical and empirical works that the classroom is the most immediate learning environment within the school context (Bronfenbrenner, 1979, 1989; Eccles & Roeser, 2011). Naturally, teachers shape the classroom environment through the quality of their teaching. That is, the ways in which teachers support students both academically and emotionally, structure and manage their classrooms can be expected to exert a strong influence on students' academic development and socio-emotional functioning (Trautwein et al., 2015). From a theoretical perspective, high-quality teaching provides students with a sense of security and confidence in the classroom, which in turn facilitates students' engagement in learning activities and supports their socio-emotional functioning in the classroom.

However, empirical research tends to treat teaching quality as a rather "static" factor in the classroom (e.g., Way et al., 2007), and little is known about dynamics that operate over time and between the teacher and the specific students in the classroom. To gain a more comprehensive understanding of how teaching quality exerts its influence on students' learning and socio-emotional functioning in the classroom, the present dissertation explores the contextual and dynamic nature of teaching quality. To capture the dynamic nature of teaching quality, changes over time in students' perceptions of teaching quality across three consecutive school years (i.e., Grade 5 to Grade 8) will be examined. Examining such longitudinal changes in teaching quality permits a deeper understanding of how teaching quality relates to students' learning and socio-emotional functioning from a longitudinal perspective. To illuminate the contextual nature of teaching quality, the present dissertation explores the extent to which the specific students in the classroom contribute to the quality of teaching. When examining teaching quality, it needs to be considered that high-quality teaching is established and defined within the classroom context, to which the teacher and the specific students in the classroom contribute. This notion implies that students are more than just passive recipients of teaching in the classroom, but also interact with their peers and with the teacher and thus actively contribute to the processes that happen therein (Doyle, 2007).

Before examining the dynamic and contextual nature of teaching quality, it is first necessary to address the assessment of teaching quality. Given that teaching is directed at students, who are thus directly involved in the teaching process, student reports are a promising approach to gain unique information on teaching quality. However, using student reports requires that students be able to report on teaching quality in a reliable and valid way. A frequently posed objection is this regard concerns the influence of response tendencies. Response tendencies refer to systematic patterns in students' response behavior that are unrelated to the actual quality of teaching provided (e.g., Baumgartner & Steenkamp 2001; Spooren et al., 2012). Thus, as a preliminary step, the present dissertation examined the extent to which teaching quality reports provided by students at different grade levels (i.e., fifth grade and eighth grade) are affected by a response tendency.

The present dissertation is structured as follows: The introductory chapter provides a comprehensive overview of the theoretical foundation underlying teaching quality by first presenting the origins and conceptualization of teaching quality in empirical educational research (Section 1.1). The second part of the introductory chapter addresses the dynamic and contextual

nature of teaching quality by considering changes in teaching quality over time and the degree to which students' "co-construct" teaching quality (Section 1.2). The third section of the chapter concerns the assessment of teaching quality via student reports, placing special focus on the unique information on teaching quality the student perspective can provide (Section 1.3). The introductory chapter is followed by a brief summary and outline of this dissertation's three guiding research questions (Chapter 2), which are addressed in three empirical studies (Chapters 3 to 5). Finally (Chapter 6), the findings of the empirical studies are discussed with regard to their relevance for teaching quality research. After outlining particular strengths and limitations of the present dissertation, the dissertation closes with an overall conclusion and practical implications.

1.1 Teaching Quality: Origins and Conceptualizations

Students' experience of their learning environment has concerned researchers for decades and has been addressed in many different fields of inquiry. In order to better understand the role of teaching quality for students' learning and socio-emotional functioning in the classroom, the following chapter provides a comprehensive overview of the theoretical foundation of teaching quality. In the first section, teaching quality will be embedded into a broader conceptual framework that describes the complex school context via a multilevel structure. Second, the origins and conceptualization of what constitutes high-quality teaching in empirical educational research will be reviewed, before closing the chapter with a discussion of the theoretical and empirical relevance of teaching quality for students' learning and socio-emotional functioning.

1.1.1 The School and Classroom Context

One theoretical lens for understanding students' learning and socio-emotional functioning in school is ecological theory, which posits that socio-cultural contexts, such as families and schools, shape human development through their characteristics and processes that happen therein (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 1998). Guided by this perspective, it is necessary to first clarify what constitutes the school context. A theoretical framework by Eccles and Roeser (2009; 2011) conceptualizes the school context as a complex, nested social system that "*can influence children's cognitive, socio-emotional, and behavioral development through organizational, social, and instructional processes that operate at several different levels of the overall school system"* (Eccles & Roeser, 2011, p. 567). As shown in Figure 1.1, this framework conceptualizes the school context as comprising seven different organizational levels. These levels are ordered hierarchically, ranging from the most basic level of the classroom to the most complex level of school as an organizational system, which is embedded in a larger cultural system. Moreover, the framework identifies core characteristics of each level that affect students' experiences of their learning environment on a daily basis.

Within this broader school context, the most immediate educational environment for students is the classroom. Students' classroom experiences are affected by several core characteristics, which are described in the framework's first three organizational levels. The nature of the academic work (i.e., design and content of the curriculum) and the structure of learning activities and grouping (i.e., whole-group vs. individualized instruction, individual vs. group work)

are general conditions for teaching and learning in the classroom. Another important feature of the classroom is the teacher, specifically the way the teacher interacts with students in the classroom. Thus, the third organizational level refers to the teacher's characteristics (e.g., level of proficiency, beliefs), the quality of teaching provided and the classroom climate (e.g., social, instructional and organizational). The next two levels capture characteristics of the school as an institution that shape the school environment as a whole. The track in which students enroll reflects curricular differentiation, which in turn determines the types of courses students take and the classmates and teachers they are exposed to. In addition, the school's general organizational structure (i.e., student body size, public vs. private, available resources) as well as the school culture (i.e., general school climate) should affect the experience of the school context for all involved individuals. The highest and most complex organizational levels reflect the fact that schools are embedded in larger social systems. That is, schools vary with respect to the involvement of students' families and the local community and as a result of their location within a specific district, state and/or national structure.

A large body of research has addressed the question of what characteristics at the different organizational levels of the school context are most important for predicting students' learning and socio-emotional functioning. In particular, research points to the important role of the classroom environment and the processes that happen therein. Specifically, the quality of teaching provided by the teacher is one of the most central characteristics affecting students' experience of their classroom, and has been found to exert a stronger influence on students' learning and development than more distal influences, such as structural characteristics (i.e., school track) or the general school climate (e.g., Wang et al., 1993; Hattie, 2009). However, to better understand the critical role of teaching quality within the school context, the pathways via which the school context impacts students' experiences of their learning environment need to be taken into account. These pathways reflect dynamics operating across time and between the teacher and the students in the classroom and thus can be used to explore the dynamic and contextual nature of teaching quality.

The ecological framework includes a temporal dimension that addresses changes over time in both students and in the school context (see Figure 1.1). Specifically, the framework emphasizes that as students transition to and progress through different types of schools (i.e., from elementary to secondary school and on to higher education), they experience changes in the characteristics of the school context in terms of structural aspects and organizational, social and instructional processes. That is, students' development over time (e.g., from late childhood to late adolescence)

is closely linked to changes in the nature of students' learning environments. With regard to the dynamic nature of teaching quality, the extent to which teaching quality is variable over time (i.e., as students progress through secondary school) needs to be explored. Understanding how teaching quality differs at different age or grade levels can provide insight into how students' experiences of teaching quality relate to their learning and socio-emotional functioning from a longitudinal perspective.

Another important pathway considers the role of the individual student within the school context. According to the ecological framework, students are not only influenced by the school context, they also actively "co-construct" their learning and general development within the school context. That is, the influence of the school context is mediated – at least to some degree – by the students themselves. As such, students' own appraisals and interpretations of their learning environment are assumed to be stronger determinants of their learning and general development than objective or third-perspective accounts of the same. This is particularly true with regard to processes within the school context directed toward or involving the students themselves, such as the teacher-student relationship or the appropriateness of various tasks. In addition, students interact with their learning environment is specific ways shaped by their personal characteristics (i.e., background characteristics, personality), and own agentic behavior (i.e., disruptive behavior) which in turn reciprocally affects the nature of their learning environment and its effects on their learning and general development. Hence, the interplay between students and their classroom environment needs to be considered, which can provide insight into the contextual nature of teaching quality. To this end, it needs to be explored whether and to what degree students "co-construct" teaching quality in their classroom through their background characteristics and own agentic behavior.

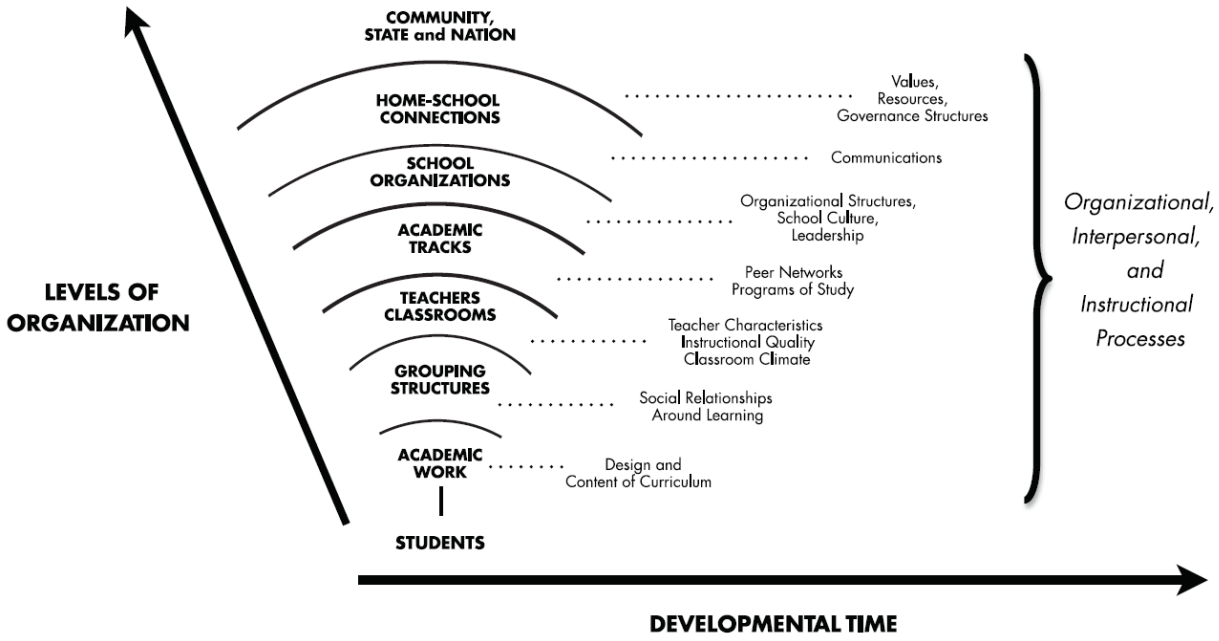

Figure 1.1 Schematic Model of the context of schooling (from Eccles & Roeser, 2011)

1.1.2 Teaching Quality in Educational Research

The question of what constitutes high-quality teaching has concerned researchers, scholars and practitioners for decades and has been addressed quite controversially across many different research areas and disciplines. For example, "good teaching" (e.g., Berliner, 2005) can be guided by a diversity of normative ideals and hypothetical concepts. As such, the degree to which the learning materials and topics are "child-friendly" (Klafki, 1963), the students and teacher have formed a learning and working alliance (Meyer, 2004), or the degree to which teaching is rooted in democratic principles (Meyer, 2004) have been identified as characteristics of good teaching. Within the field of teaching methodology, contemporary models of teaching describe it as an interplay between the students, the teacher and the subject matter (e.g., Reusser, 2008). However, while this model has a rich theoretical foundation, it seeks to describe the complexity of teaching and learning in the classroom from a phenomenological perspective, rather than find empirical evidence for the assumed mechanisms (for a discussion, see Reusser et al., 2010).

In contrast, the field of empirical educational research conceptualizes high-quality teaching primarily in terms of its effectiveness: Guided by the idea that high-quality teaching must be quantifiable with regard to student outcomes, empirical educational research seeks to find adequate

ways of gauging the effects of schooling and teaching (Berliner, 2005; Reusser et al., 2010; Trautwein et al., 2015; Weinert et al., 1989). The discipline of empirical educational research is deeply rooted in the process-product line of research that began to flourish in the 1960s (for an overview, see Brophy, 2006). In order to validate theoretical assumption about what constitutes high-quality teaching with empirical data, process-product studies aimed at finding empirical support for which processes (i.e., which aspects of teaching) can be associated with desirable student outcomes, such as students' achievement. Eventually and in response to the growing body of research highlighting the importance of students' perception and interpretation of their learning environment (e.g., Walberg et al., 1977; Eccles & Roeser, 2011), students' perception and interpretations were included in the concept as a mediating factor between processes and outcomes (see Brophy, 2006).

Process-product studies have had a profound influence on educational research, policymaking and teacher education. The methodological approach underlying process-product research is nowadays considered "state of the art" (Helmke, 2017; Gruehn, 2000), and the large number of national and international large-scale empirical studies (for an overview, see Praetorius et al., 2018) systematically assess teaching quality as well as a broad range of student outcomes. Moreover, process-product studies made it possible to draw concrete conclusions and provide implications for how to foster students' learning and positive development in the school environment.

The Product: Defining the Objectives of Teaching

According to process-product studies, the extent to which specific aspects of teaching can systematically be associated with student outcomes is an indicator for high-quality teaching. Consequently, determining the effectiveness of teaching is necessarily tied to defining appropriate outcome criteria. On the one hand, schools are primarily concerned with students' education, so students' learning (e.g., in terms of standardized achievement tests or grade point averages), or learning-related outcome variables (e.g., motivation, subject-related interest) are most commonly used to determine the quality of teaching (e.g., Klieme et al., 2003). On the other hand, schools seek to equip students with the competences and skills they need to shape their own lives and to successfully participate in society (Chernyshenko et al., 2018; OECD, 2021), which indicates that students' socio-emotional functioning (e.g., well-being, self-esteem, school belonging) needs to be considered in the discussion of high-quality teaching as well. Thus, in order to arrive at a

comprehensive understanding of teaching quality, it is necessary to consider the multidimensional goals of schooling and evaluate teaching quality on the basis of multiple outcomes (e.g., Aldrup et al., 2018; Eccles et al., 2011).

The Process: Identifying the Characteristics of High-quality Teaching

Since the 1960s, a growing body of empirical research has examined a wide range of factors and processes within the classroom context with regard to their associations with student outcomes (e.g., achievement). Hattie's (2009) prominent review of the most important determinants of students' academic achievement is indicative of the large number of studies in the field and highlights the educational and political significance of examining the impact of teaching on students' learning. Even more importantly, several overview studies and reviews (Hattie, 2009; Brophy & Good, 1986; Wang et al., 1993) have helped to organize the large base of accumulated information and research findings into a more coherent body of knowledge. To date, research has converged on identifying proximal processes in terms of teachers' organizational, social and instructional actions in the classroom as having a stronger impact on students' learning and development than structural aspects of the school or the classroom (e.g., student body size, class size, or school resources; Seidel & Shavelson, 2007; Pianta & Allen, 2008, Trautwein et al., 2015). Within the field of educational research, today's understanding of high-quality teaching generally refers to the actions the teacher takes that provide students with opportunities to engage in insightful learning processes (e.g., Kunter & Baumert, 2006; Kunter & Voss, 2013). Across several lines of inquiry in both the US and Europe, there is growing consensus that high-quality teaching comprises an overarching organizational as well as a supportive dimension (e.g., Hamre & Pianta, 2010; Klieme, et al., 2009). The organizational component concerns the organization and management of the classroom, which sets the stage for teaching and learning. The supportive dimension encompasses two broad areas of teacher support referring to teachers' care and responsiveness to affective-emotional matters (i.e., emotional support) and instructional matters (i.e., academic support; see Aldrup et al., 2018).

From a theoretical perspective, the teaching quality dimensions are considered "generic in nature" (Praetorius et al., 2018, p. 408). That is, the underlying organizational, social and instructional actions are considered to exert their influence on students' learning and socio-emotional functioning regardless of surface aspects of teaching, such as instructional methods or learning formats (e.g., group vs. individual work; teacher-centered instruction vs. cooperative

learning methods). For example, both group and individual work can stimulate students' learning; their effectiveness largely depends on how the teacher implements and guides students through them. In a much broader sense, the dimensions of high-quality teaching are even considered to be "applicable across school subjects, grade levels, and potentially even countries and cultures" (Praetorius et al., 2018, p. 408).

From a developmental perspective, teaching quality in terms of classroom management and teacher support are considered to meet students' developmental needs to feel autonomous rather than being controlled, to feel competent and to feel socially connected with others, which are in turn central for fostering students' growth and positive development (e.g., Ainsworth et al., 1978; Bowlby, 1969; Eccles & Midgley, 1989; Eccles et al, 1993; Ryan & Deci, 2000).

1.1.3 Teaching Quality, Students' Learning and Socio-emotional Functioning

Classroom Management

In general, classroom management refers to the actions the teacher takes to establish optimal conditions for teaching and learning in the classroom (Kounin, 1970; Doyle, 2006). Broadly speaking, this includes the organization of the classroom and the management of time, students' behavior and attention in the classroom (Emmer & Stough, 2001). On a more profound level, classroom management comprises a set of strategies and actions that aim at promoting positive student behavior and preventing disturbances during lessons, such as communicating clear rules and behavioral expectations, establishing stable routines, monitoring what is happening in the classroom, and redirecting misbehavior before it escalates (e.g., Hamre & Pianta, 2010; Klieme et al., 2001). Thus, the current understanding of classroom management emphasizes the importance of proactive or preventive actions rather than reactive responses to student misbehavior or disturbances during the lesson.

In a well-managed and orderly classroom, teachers spend a limited amount of time on non-instructional activities and transitions, which allows for an optimal use of the allocated learning time. Since the amount of "time-on-task" is a critical condition for students' learning, classroom management has most consistently been associated with students' academic achievement (see Brophy 2000; Seidel & Shavelson, 2007; Wang et al. 1993). Less research on its associations with students' socio-emotional functioning exists. From a developmental perspective (e.g., Eccles et al., 1993; Ryan & Deci, 2000), a classroom environment with adequate structure and orderliness that

provides sufficient opportunities for learning is believed to facilitate students' experiences of autonomy and competence in the classroom (Wigfield et al., 2006). For example, organizing the classroom so that it is meaningful to students (e.g., by providing clear and explanatory rationales) should allow students to experience autonomy in the classroom, rather than feeling forced to comply with classroom rules (Assor et al., 2002; Leon et al., 2017). More specifically, the provision of clear and consistent structures has been shown to be associated with students' development of cognitive and behavioral self-control (Emmer & Strough, 2001; Rimm-Kaufman et al., 2009; McCaslin et al., 2006). Self-regulatory skills, in turn, are seen as key competences for success in all areas of schooling (Hamre & Pianta, 2010; Raver, 2004).

Emotional Support and Academic Support

The notion that learning in the classroom is a socially mediated process (e.g., Vygotsky, 1978) indicates that the nature and quality of teachers' and students' interactions with one another reflect an important dimension of teaching quality. With regard to *teachers' emotional support*, key aspects include the overall emotional tone in the classroom, the quality of the teacher-student relationship, and the degree to which the teacher shows a personal interest in students and respect for their feelings and points of view (Patrick et al., 2007; Patrick et al., 2003; Pianta & Allen, 2008). With regard to instructional matters (i.e., academic support), teachers' sensitivity to and care for students' academic needs, such as providing constructive feedback and helping students overcome challenges with regard to the learning content, are key aspects of teachers' academic support (Hamre & Pianta, 2010; Klieme et al., 2009).

From the perspective of attachment theory and self-determination theory (Ainsworth et al., 1978; Bowlby, 1969; Ryan & Deci, 2000), these aspects are pivotal for both students' learning and their socio-emotional functioning. In essence, it is assumed that caring and supportive teacher behavior contributes to creating a classroom environment in which students experience emotional security and self-confidence, which in turn helps students devote energy and attention to academic activities and develop a positive attitude towards school (e.g., Birch & Ladd, 1997; Burchinal et al., 2002; Crosnoe et al., 2004; Spilt et al., 2012). Moreover, it has been argued that supportive teachers provide students with more opportunities for autonomy, for establishing social connections with others and for experiencing competence, which is critical for engaging students in learning (Brophy, 2006; Eccles & Roeser, 1998; Niemiec & Ryan, 2009; Skinner et al., 2008; for an overview, see Ruzek et al., 2016). Overall, teacher support has most consistently been

associated with students' socio-emotional functioning. In particular, teacher support has been found to enhance students' feelings of self-esteem, well-being, valuing of school and school belonging (e.g., Baumeister & Leary, 1995; Hall-Lande et al., 2007; Roeser et al., 1996) and has also been identified as a protective factor for problematic student behaviors (Buyse et al., 2008; Gazelle, 2006; Mashburn et al., 2008; Wang & Dishion, 2012). However, the literature reports rather mixed results regarding the link to students' academic achievement. Whereas some studies report positive associations (e.g., Cornelius-White, 2007; Downer et al., 2015; Roorda et al., 2011; Spilt et al., 2012), other studies found teacher support be less important for students' academic achievement (e.g., Kunter et al., 2013; Scherer et al., 2016; Wagner et al., 2016).

1.2 The Dynamic and Contextual Nature of Teaching Quality

Taken together, there is a strong theoretical and empirical rationale that how teachers define the classroom setting through the quality of their teaching exerts a strong influence on students' learning and socio-emotional functioning. However, empirical research tends to treat teaching quality and student development as rather "static" factors within the classroom (Way et al., 2007), and less attention has been given to a) examining teaching quality across longer time periods and b) the specific classroom context in which high-quality teaching is defined and established. To gain a more comprehensive understanding of how teaching quality exerts its influence on students' learning and socio-emotional functioning in the classroom, dynamics that operate across time and between the teacher and the specific students in the classroom must be taken into account.

1.2.1 Longitudinal Perspective on Teaching Quality

Teaching quality in terms of teachers' organizational, social and instructional actions with students are reoccurring, proximal processes, which should have a greater impact on students' development than processes that are more episodic or inconsistent (e.g., Bronfenbrenner & Morris, 1998). However, the large body of research investigating the impact of teaching quality on students' learning and socio-emotional functioning has largely focused on analyses at single time points or short-term longitudinal analyses. Unfortunately, such "snapshot" views of the classroom may not adequately explain how students' learning and socio-emotional functioning develop over longer time periods (e.g., across lower secondary school), nor do they capture changes in teaching quality as a reoccurring process over time. Thus, applying a longitudinal perspective to teaching quality is a promising approach to explore the dynamic nature of teaching quality. Moreover, in order to better understand the classroom as a context in which learning and development take place over longer time periods (e.g., across several school years), it is necessary to consider both students' development and changes in teaching quality over time.

Arguably, the secondary school years are critical for investigating how student development is related to teaching quality over time. On the one hand, the lower secondary school years largely coincide with the significant developmental period of early adolescence. The early adolescent years are considered a vulnerable phase in which students are at risk of becoming detached from school and experiencing negative changes in their academic engagement and their social and emotional school adjustment (Engels et al., 2017; Morinaj & Hascher, 2019; Roeser et al., 2000; Virtanen et

al., 2021). On the other hand, adolescent students experience drastic changes in their developmental needs. As such, the developmental task of establishing oneself as an independent and autonomous being is accompanied by a desire for increased autonomy and less control by parents or teachers. As adolescents seek to become more independent from their parents (Gutman & Eccles 2007; Steinberg & Silk, 2002), high-quality relationships with peers and non-parental adults become increasingly important sources of social and emotional support (e.g., Way & Greene, 2006). This shift in students' social relationships is also accompanied by a heightened sensitivity to peers' reactions. Relatedly, students experience a heightened self-consciousness during the process of forming their own identity (Soto et al., 2008, 2011).

From a developmental perspective, students' development is influenced by the extent to which their needs for autonomy, social relatedness and competence are aligned with the opportunities present within their learning environment as they mature (Eccles et al., 1991; Roeser, et al., 2000; Ryan and Deci, 2000; Connell & Wellborn, 1991). In this regard, teachers must adjust their teaching to meet their students' developmental needs, which should result in variability in teaching quality over time. For example, walking around to check students' homework at the beginning of each lesson might be an appropriate strategy to establish stable routines in lower grades (e.g., Grade 5), but could have a different effect when applied in higher grades (e.g., Grade 10) given students' need for more autonomy and less control during this phase (e.g., Patall et al., 2010).

Support for the link between changes in students' classroom experiences and changes in their academic development and their social and emotional school adjustment largely stems from short-term longitudinal research on school transitions (e.g., from elementary to secondary school; Evans et al., 2018; Longobardi et al., 2019; Roeser & Eccles 1998). For example, compared to elementary school, secondary school is characterized by less personal teacher-student relationships and lower levels of support (e.g., Eccles et al., 1991). This seems to be at odds with students' developmental needs and has thus been suspected to account for the negative changes in academic development and their social and emotional school adjustment many students reportedly experience after their transition to secondary school (e.g., Engels et al., 2017). However, less research has focused on changes in students' perceptions of teaching quality during secondary school itself, in large part because longitudinal data on the same students across several time points over several school years would be required. To foster students' learning and development, it is

necessary to investigate how changes in students' learning and development are associated with changes in teaching quality as they proceed through secondary school.

Examining the Co-development of Teaching Quality and Student Development: Longitudinal Growth Curve Models

Longitudinal growth curve models (LGCM; Bollen & Curran, 2006¸ McArdle & Epstein, 1987; Meredith & Tisak, 1990) are a well-established and flexible method for examining changes over time as well as individual differences in such changes. In a LGCM, the trajectories of teaching quality and student development can be described by specifying a latent intercept factor, which indicates the initial level at the first measurement point (e.g., beginning of secondary school) and as well as a latent slope factor, which reflects the average amount of change across subsequent measurement points. One particular advantage of LGCM is that it separates measurement error from true change over time (Sayer & Cumsille, 2001). In addition, changes over time can be represented by different time trends: One possibility is to specify a linear rate of change across equidistant time points. Another possibility is to estimate a model with an unspecified rate of change, in which the time function is not fixed a priori but estimated based on the data (Meredith & Tisak, 1990). A further advantage is that multilevel LGCMs take into account the hierarchical structure of student data (i.e., students nested within classes) and consequently provide information about within- and between-classroom differences in the intercepts and slopes.

Specifying correlations between the intercepts and slopes of teaching quality and student development (e.g., students' learning and socio-emotional functioning) ultimately provides information on both concurrent and longitudinal associations between the two variables. As such, a correlation between intercepts represents an association between the initial level of teaching quality and the initial level of student development at the beginning of secondary school and can be interpreted similarly to a concurrent correlation. A correlation between slopes indicates a longitudinal association between changes in teaching quality over time and changes in students' learning and socio-emotional functioning over time. It is important to note that concurrent and longitudinal associations between the same variables can differ in magnitude, directionality and/or statistical significance. For example, a significant intercept-intercept correlation would indicate a concurrent association between teaching quality and student development, whereas a non-significant slope–slope correlation would indicate that teaching quality and student development are unrelated from a longitudinal perspective (i.e., change over time in teaching quality is unrelated

to change over time in student development). Moreover, correlations between the initial level of teaching quality and the slope of student development suggest that, for example, a higher (or lower) level of teaching quality at the beginning of secondary school is related to a greater increase (or decrease) in student development across subsequent school years. Consequently, from a theoretical perspective, different scenarios for the co-development of teaching quality and student development need to be considered. For example, some students' secondary school years may be marked by ongoing patterns of academic success and positive development, in which initially high levels of teaching quality may set the stage for ongoing growth and positive school adjustment. Other adolescent students might experience decreases in teaching quality over time, challenging their adaptive capacities. Still other students may initially experience struggles and insufficient levels of teaching quality, followed by continued negative perceptions of teaching quality or gradually increasing adjustment to school (Way et al., 2007, Widlund et al., 2021). In summary, applying a longitudinal perspective to teaching quality is a fruitful avenue for examining the degree to which students' learning and socio-emotional functioning are shaped by students' experiences of teaching quality across lower secondary school.

1.2.2 Teaching Quality in the Classroom Context

Theoretical perspectives (e.g., Eccles & Roeser, 2011; Doyle, 2006) and conceptualizations of teaching (e.g., Fend, 1998; 2008; Helmke, 2003) emphasize that high-quality teaching is defined and established within the classroom context, to which the teacher and the specific students in the classroom contribute. In this regard, the complex, social process of teaching can be understood as an interplay between the teacher and the students in the classroom. This notion is consequential, as it implies that high-quality teaching is not determined solely by the teacher, but emerges through the interactions between the teacher and students in the classroom (e.g., Fauth et al., 2020b). To better understand the contextual nature of teaching quality, the role of the students being taught in the classroom needs to be explored.

In particular, supply-use (or offer-usage) models have classified a range of factors and processes within the classroom context that contribute to effective teaching (see Figure 1.2.2). In essence, supply-use models (e.g., Brühwiler and Blatchford, 2011; Helmke, 2007; Fend, 2008; Seidel, 2015) acknowledge that neither the teacher nor teaching itself has a direct impact on students' learning. Instead, teaching is understood as an "offer" that provides students with opportunities for learning. Students, in turn, need to actively use these opportunities and turn them

into actual learning activities (i.e., "use"). Moreover, the model assumes that the extent to which students make use of learning opportunities increases along with the quality of the provided opportunities (i.e., quality of teaching, quality of learning materials).

In addition to identifying teaching activities and students' learning activities as central processes of effective teaching, the model classifies a range of factors that influence these processes in the classroom (Figure 1.2.2). According to the supply-use model, students' learning activities are affected by individual factors (e.g., ability level, prior knowledge, self-related beliefs), which in turn are shaped by family-related factors (e.g., social background). The quality of teaching largely depends on teachers' professional competences (e.g., self-related beliefs, teaching methodologies, pedagogical and content knowledge), which are in turn shaped by teachers' more general characteristics (e.g., age, gender, personality, level of teaching experience). Lastly, the school and classroom context (e.g., school type, school and class size, student composition, student behavior) is expected to influence both the teacher-generated opportunities for learning as well as students' learning activities in the classroom.

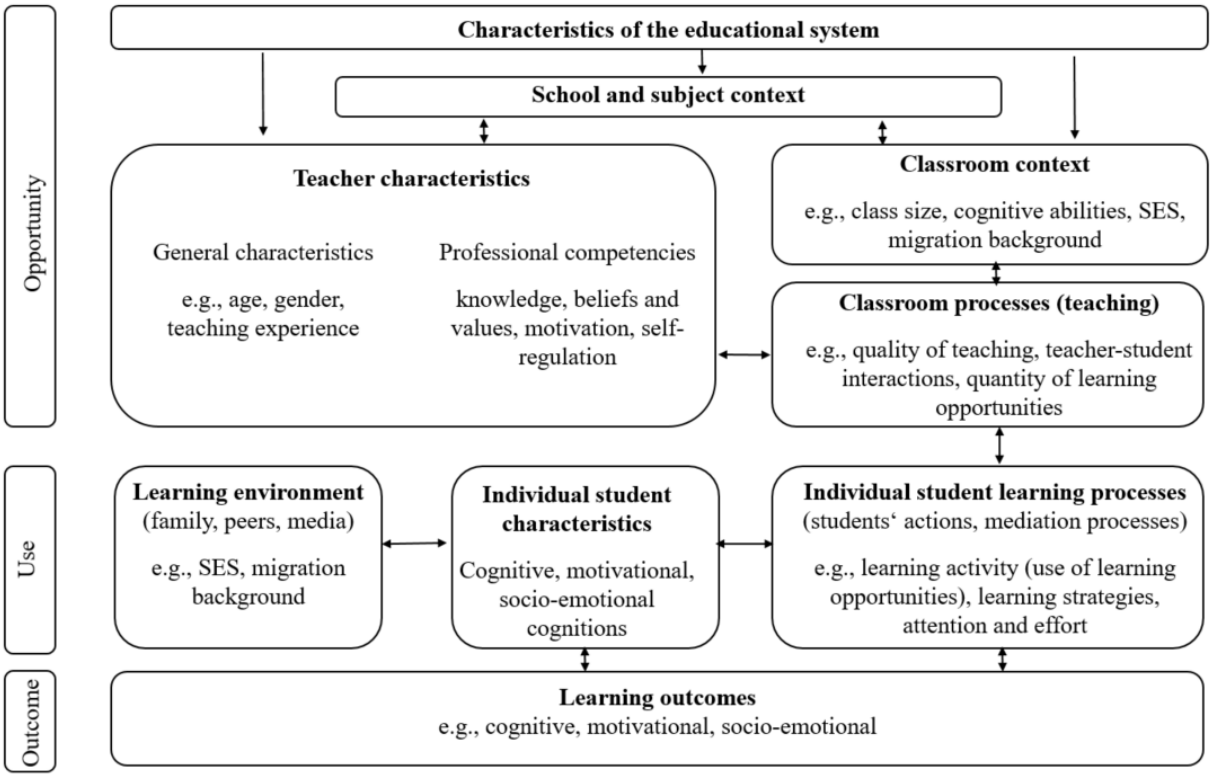

Figure 1.2.2 Supply-use model of teaching and learning (adapted from Brühwiler and Blatchford, 2011; Seidel, 2015)

Teaching Quality in the Classroom Context in Educational Research

From a theoretical point of view, students themselves mediate the effects of teaching quality on their learning and development through their perception and interpretation of events, their background characteristics and their own agentic behavior in the classroom (e.g., Eccles & Roeser, 2011). While students' perception and interpretation of teaching quality is reflected in the constructivist understanding of learning and the conceptualization of teaching as an offer-use process (e.g., Fend, 1998; 2008; Helmke, 2003), less is known about the degree to which students' background characteristics and agentic behavior contribute to the quality of teaching in the classroom. However, there is initial evidence that teaching quality seems to be related to the specific students in the class. Fauth et al. (2020b) demonstrated that the same teacher's quality of teaching varied considerably across classes. Given that the same teacher taught the same, scripted unit to several classes at the same grade level, this study's findings suggest that the specific students in the classroom account for at least some of the variability in teaching quality. More specifically, studies by Campbell & Ronfeldt (2018), Fauth et al. (2021), and Göllner et al. (2020) have shown that teaching quality is systematically related to students' background characteristics in terms of their level of achievement, gender, socio-economic and migration background.

Importantly, the theoretical view of teaching quality as an interactive process between the teacher and the students in the classroom seems to be reflected in various measures of teaching quality in empirical research (Fauth et al., 2020a; Göllner et al., 2020). That is, some measures clearly refer to the teacher's actions in the classroom, which might provide information about teacher-generated conditions for engaging students in the learning process (e.g., monitoring: "Our teacher realizes immediately when students are engaged in something else"). Other measures refer to students' actions, or to the interplay between the teacher and the students in the classroom, which might tap more into the consequences of the teacher's actions in the classroom (e.g., disturbances: "In class there is constant loud talking"). Thus, teaching quality can be conceptualized both as a set of teacher-generated conditions for teaching and learning in the classroom as well as an indication of successful teaching and learning in the classroom. For example, according to Praetorius et al. (2018, p. 409), classroom management is "*a condition for students getting attentive (e.g., through teacher monitoring) and an indication of students being attentive (e.g., lack of interruptions)*".

In this regard and in line with the conceptualization of teaching as an offer-use process (e.g., Fend, 1998; 2008; Helmke, 2003), there are at least two mechanisms by which the specific students in the classroom may contribute to teaching quality: First, students actively contribute to the degree to which teachers' actions are converted into learning opportunities during the lesson. This refers to processes on the usage side of the offer-use model. For example, providing high-quality teaching might be easier in a class of highly motivated and well-behaved students compared to uninterested and poorly behaved ones. Second, in order to provide optimal conditions for teaching and learning, teachers must adapt their social, instructional and organizational actions to the needs, abilities and interests of the specific students in the classroom (e.g., Doyle, 1985; Stroet et al., 2013). For example, with respect to students' achievement, it can be assumed that students will be more motivated to learn if the lesson's tasks and affordances correspond to their current level of competence (Midgely, 2002). Thus, the specific students being taught in a class may affect teaching processes located on both the offer and use sides. In summary, it can be assumed that students are not just passively impacted by teaching quality, but actually "co-construct" the processes happening on both the usage and offer sides of teaching and learning in the classroom to at least some extent.

1. 3 Measuring Teaching Quality: Students as Informants about Teaching Quality

Teaching quality assessment is challenged by the fact that teacher-student relationships and interactions, teacher instruction and student behavior in the classroom can hardly be measured in an objective and neutral way, unlike students' academic achievement (e.g., through standardized achievement tests). Information on what happens in the classroom must be gained in order to evaluate the degree to which students receive high-quality instruction. In this regard, student reports are a promising and well-established approach to assessing teaching quality (Anderson, 1982; Clausen 2002; Fraser, 1991; Turner and Meyer, 2000, Weinstein, 1983), alongside teacher self-reports and reports by external observers. Student reporters can be considered "experts" due to their large amount of experience with different teachers in different subjects across several school years (e.g., Clausen, 2002; De Jong & Westerhof, 2001). In contrast to external observers, who typically rely on "snapshot" views of the classroom, students observe and participate in classroom processes on a daily basis and thus have a comprehensive understanding of what happens in the classroom (Downer et al., 2015). Compared to teacher self-reports, student reports ought to be less affected by self-serving bias (Clausen, 2002; Wubbels et al., 1992). Moreover, student reports are also appealing economically, as a time- and cost-efficient method of gaining information from large group of respondents (Fraser & Walberg, 1991).

A large body of research findings has shown that student reports on teaching quality are associated with a broad range of outcome measures (i.e., student achievement, learning-related and socio-emotional aspects), confirming the strong predictive validity of student reports (e.g., Marsh, 2007). However, student reports have also been criticized for their low convergence with other data sources on the same lesson, specifically with teacher self-reports. Research on this issue (e.g., Aldrup et al., 2018; Kunter & Baumert, 2006) has converged to assume "perspective-specific validities" (Kunter & Baumert, p. 234) for each approach. That is, teachers and students take on different roles and positions in the classroom and interact differently with the classroom environment. Thus, teachers and students presumably focus on different aspects when evaluating teaching quality. In this regard, students have the potential to provide unique information on what happens in the classroom that may not be gleaned from other data sources (Feldlaufer et al., 1988; Fraser & O'Brien, 1985; Schenke et al., 2018).

1.3.1 Uniqueness of Students' Teaching Quality Reports

Interest in student reports on teaching quality in educational research is deeply rooted in the school and classroom climate research of the 1960s (e.g., Anderson, 1982; Fraser and Walberg, 1981; Moos, 1979), which emphasized students' view on classroom processes (e.g., Weinstein, 1982; for an overview, see Fraser, 1991). In line with the turn from a behaviorist to a more constructivist understanding of learning and new research on social cognition ("the cognitive turn"; e.g., Bandura, 1986), students were more strongly conceived as actively constructing their knowledge, rather than passively receiving instruction and taking on information (Reusser, 2006). This implies that how students perceive and interpret the classroom environment and the quality of instruction more strongly influences their learning and development than a more objective account of teacher behavior (Lüdtke et al., 2009; Waxman, 1991). For example, consider a situation in which a teacher praises a student in order to appreciate the student's effort. However, this does not necessarily mean that the student feels praised; for example, the student might feel that he or she is not being taken seriously (Lenske, 2016). On the one hand, this example demonstrates that students' views on teacher behavior might diverge from observed or intended instruction. On the other hand, it shows that students' individual experiences are most relevant for their learning and development (Lüdtke et al., 2009; Maehr & Midgley, 1991; Ryan & Grolnick, 1986). That is, teaching can only enhance students' learning and socio-emotional functioning to the degree that it is perceived as supportive and helpful by the students in the classroom.

This notion has several consequences for educational research: First, with regard to the conceptualization of teaching as an offer-use situation (e.g., Fend, 1998; 2008; Helmke, 2003), focusing more strongly on how students perceive aspects on the offer side (e.g., the quality of instruction, teacher-generated opportunities for learning, and availability of learning time) may provide insight into the conditions under which students effectively take up the generated learning opportunities. That is, students' perception of teaching can be considered as an important intermediary between instruction (i.e., offer/input) and students' learning and socio-emotional functioning (i.e., use/output; Den Brok et al., 2006; Göllner et al., 2018). Second, since teaching is directed to students, some aspects of teaching quality may actually be hard to evaluate from an outside perspective. For example, the degree to which the task level or the pacing of the lesson felt appropriate, whether students felt supported or were engaged in the task may be best evaluated by the involved students themselves (Kunter & Baumert, 2006; Kunter & Voss, 2013). In summary,

student reports offer uniquely valuable insights into their view on teacher-student relationships, teacher-student interactions, the meaningfulness of instruction, and student behavior in the classroom. Examining how students perceive and interpret teaching quality may be a gateway to better understanding how teaching quality relates to students' learning and socio-emotional functioning in the classroom, which in turn is crucial for optimizing instruction and fostering learning processes among all students (Hattie, 2009; Gruehn, 2000).

1.3.2 Students' Shared and Non-shared Perceptions of Teaching Quality

From a phenomenological point of view, students' own perceptions and interpretations of teaching quality can be assumed stronger drivers of their learning and development than any objective account of the same (e.g., Anderson et al., 1989; Eccles & Roeser, 2011; Lüdtke et al., 2009). At the same time, students' interpretations of events are shaped by a multitude of individual factors, such as their personality, prior experiences, ability level or even response tendency (Maehr & Midgley, 1991; Ryan & Grolnick, 1986). Consequently, what happens within a specific classroom can be perceived, interpreted and evaluated quite differently by each individual student in that classroom. Thus, from a theoretical perspective, student reports on teaching quality contain two different components that must be considered separately (Lüdke et al., 2009): shared perceptions among all students in a classroom (e.g., on average, students in a specific class agree that the teacher helps students with learning-related difficulties), and an idiosyncratic component reflecting students' individual or non-shared perceptions (e.g., inter-individual differences in the extent to which students in the classroom agree that their teacher is supportive; Göllner et al., 2018). This distinction is consequential for the analysis of student reports on teaching quality: modeling approaches that consider the hierarchical structure of student data (i.e., the clustering of students within classes) are needed to appropriately deal with the different components inherent to student reports, (e.g., Marsh et al., 2011, Morin et al., 2014).

Multilevel Modeling of Student Reports on Teaching Quality

Because single-level analyses ignore the hierarchical structure of student reports (i.e., the clustering of students in classes) and thus are likely to lead to biased parameter estimates (for further discussion, see Goldstein, 2010; Hox, 2010; Raudenbush & Bryk, 2002), multilevel modeling is an appropriate methodological approach for the analysis of student reports on teaching quality. In a multilevel analysis, the clustering of students results in two distinct levels of analyses,

which affects the interpretation of results. At the class level, students' individual reports are aggregated to represent their shared perception of teaching quality and provide information about the classroom as a whole. Analyses at the classroom level indicate the extent to which the average level of teaching quality differs across classrooms, which helps to explain differences in student outcomes between classrooms. In contrast, analyses at the student level provide information about students' individual perceptions of teaching quality, indicating the extent to which individual student perceptions diverge from the class average. Thus, analyses at the student level explain variation in student outcomes within the same classroom (Lüdtke et al., 2009; Raudenbush & Bryk, 2002).

In educational research, a large number of studies have focused on teaching quality at the classroom level (e.g., Kunter & Baumert, 2006). Teaching quality can be conceptualized theoretically as a feature of the classroom that concerns all students within that classroom to a comparable extent. Consequently, analyses at the classroom level have been conducted to gain systematic insights into the effects of the learning environment on student outcomes. Moreover, aggregating individual student data at the classroom level has the methodological advantage of counterbalancing students' idiosyncrasies to at least some extent (Lüdtke et al., 2009; Marsh et al., 2012).

In contrast, analyses of teaching quality at the student level have been conducted to a much smaller extent, based on the assumption that these analyses mainly address students' "phenomenology" (Lüdke et al., 2009, p. 122). Yet, a growing body of recent studies has highlighted that student-level analyses of teaching quality hold valuable information (Aldrup et al., 2018; Bardach et al., 2021; Göllner et al., 2018; Schenke et al., 2017; Schenke et al., 2018; Schweig 2016). Fueling this growing body of research was the recent finding that the largest amount of variability in student reports on teaching quality can be found within classrooms, rather than between classrooms (see Bardach et al., 2021; Göllner et al., 2018). Thus, conceptualizing within-class differences in student reports on teaching quality as pure measurement error may underestimate the complexity of classroom processes. In support of this, a more differentiated view of within-class variability in teaching quality suggests that teaching quality aspects addressing the extent to which the teacher provides students with individual support or promotes individual students' understanding have the widest within-class variability (e.g., Aldrup et al., 2018; Wagner et al. 2016). Moreover, students' idiosyncratic perceptions of teacher support have been found to

be associated with individual students' achievement and socio-emotional functioning (e.g., Aldrup et al., 2018; Rucinski et al., 2018), whereas teacher support was largely unrelated to student outcome measures at the classroom level. Because a teacher's supportive actions often address individual students rather than the whole class, within-class differences in teacher emotional or instructional support seem to reflect differences in the extent to which the teacher addresses each student's individual needs (Eccles et al., 1989; Göllner et al., 2018; Nurmi & Kiuru, 2015) and can be of practical value in understanding the effects of supportive teacher behavior.

In summary, previous research has shown that both students' individual and shared perceptions of teaching quality are related to student outcomes (Aldrup et al., 2018; Downer et al., 2015; Göllner et al., 2018). Therefore, multilevel modeling of student reports on teaching quality offers a fruitful approach for understanding how within- and between-classroom differences in teaching quality are related to students' learning and socio-emotional functioning.

1.3.3 Assessing Teaching Quality with Student Questionnaires: Student Response Tendencies

A key finding in educational research is that teaching quality varies within and across classrooms, which in turn accounts for variation in student outcomes (e.g., Aldrup et al., 2018; Downer et al., 2015; Kunter et al., 2013; Wagner et al., 2013). However, variation in student reports on teaching quality might arise to a certain degree from inter-individual differences in response tendencies. Thus, one important aspect that needs to be explored is the degree to which the common and well-established approach of administering questionnaires to students is related to inter-individual differences in use of the response scale.

When teaching quality is assessed via student questionnaires, students are asked to indicate how strongly they agree or disagree with a given statement on what happens in their classroom. However, such response formats (i.e., a rating scale) are suspected to be susceptible to response tendencies (Podsakoff et al., 2012; van Vaerenbergh & Thomas, 2013; Moors, 2008). In general, response styles are conceptualized as a respondent's systematic tendency to use certain categories of a response scale, irrespective to the construct in question (Baumgaertner & Steenkamp, 2001; Paulhus, 1991). One common response tendency reflecting such an inter-individual difference in scale use is acquiescence, and its counterpart disacquiescence, which refer to an individual student's tendency to answer in the affirmative (acquiescence; "yea-saying") or the negative (dis-

acquiescence; “naysaying”) regardless of item content (Baumgartner & Steenkamp 2001; Paulhus, 1991; Winkler et al., 1982). When evaluating teaching quality, an acquiescent student would exhibit a preference for categories reflecting agreement, whereas a dis-acquiescent student would exhibit a preference for categories reflecting disagreement. For example, the students in a specific class may agree that the teacher helps students overcome learning-related difficulties, but individual students may differ in their tendencies to agree with items on a rating scale. Consequently, some students will systematically assign higher ratings than others, which can lead to more variability in teaching quality reports within a class. In contrast, if students with similar response tendencies are concentrated in the same classroom, a lower variability in teaching quality reports could result. In addition, prior research has shown that overuse of one side of the rating scale is problematic because it leads to inflated (disacquiescence: deflated) item scores and may affect the correlations among items or factors. That is, if students consistently agree (or disagree) with items that tap into different aspects of teaching quality, this will increase the shared variance among items (Bolt et al., 2014; Kam & Mayer, 2015; Rammstedt & Farmer, 2013; Weijters et al., 2010).

Even though a large body of research has provided considerable evidence that students’ individual and shared perceptions of teaching quality both hold valuable information (e.g., Aldrup et al., 2018; Downer et al., 2015; Göllner et al., 2018; Schweig, 2016; Wagner et al., 2013), the degree to which acquiescence accounts for variability in student evaluations of teaching quality needs to be clarified. In fact, there is little empirical evidence on the impact of acquiescence on student reports. Moreover, the majority of research examining acquiescence has focused on self-report data on personality traits, personal attitudes or clinical symptoms (e.g., Baumgartner & Steenkamp, 2001; Hinz et al., 2007; Rammstedt & Farmer, 2013; Wetzel et al., 2013), potentially limiting the degree to which existing findings on acquiescence can be transferred to student reports on teaching quality. A notable exception is a study by Spooren et al. (2012), who found no empirical support that acquiescence influences student reports on teaching quality. However, Spooren et al. (2012) examined a sample of university students, who are likely to differ from the much younger students in secondary school, which begins in Grade 5 in Germany. In addition, the study by Spooren et al. (2012) did not take into account the hierarchical structure of student data, which is crucial for accurately evaluating the impact of acquiescence on student reports of teaching quality.

Acquiescence in Multilevel Analyses of Teaching Quality

In general, response tendencies are considered part of students' idiosyncratic perceptions (e.g., Göllner et al., 2018). However, it has been argued that when aggregating individual student data at the classroom level, individual differences in students' response behavior are counterbalanced and thus are less relevant when examining students' shared perceptions of teaching quality (Lüdtke et al., 2009). However, this assumption has not yet been empirically verified. In fact, there are reasons to assume that acquiescent students may not be randomly distributed across classes. For example, individual factors related to acquiescence (e.g., students' cognitive capacities; Soto et al., 2008) have been shown to vary across classes (e.g., Karing et al., 2013; Mullis et al., 2003; Pfost & Artelt, 2013). In the same vein, it is common practice in the German school system to sort students into classes based on their residential area, level of prior achievement, or choice of core and elective subjects, which may in turn contribute to systematic differences between classrooms in students' acquiescence tendencies. To the extent that the number of acquiescent and non-acquiescent students is not equally distributed across classrooms, students' acquiescent response tendencies may affect their shared perceptions of teaching quality and impact analyses at the classroom level. Before using student reports as a reliable and valid method of assessing teaching quality in school, it is critical to evaluate the impact of acquiescence on students' teaching quality reports at both the student and the classroom level.

2 AIMS AND RESEARCH QUESTIONS

From an ecological perspective, the classroom is the most central environment for students' learning and socio-emotional functioning within the broader school context. While students' experiences of teaching quality have been identified as directly relevant for their learning and development, much less is known about dynamics in the classroom that operate across time and between the teacher and the specific students in the classroom. Hence, the present dissertation pursued a twofold aim: First, because the present dissertation's research design strongly relies on student reports on teaching quality and includes students at different grade levels, it was first examined whether students are able to report on teaching quality in a reliable and valid way. Specifically, response tendencies (Section 1.3.3) need to be taken into account when assessing teaching quality with student questionnaires.

The second and main aim of the present dissertation was to explore the contextual and dynamic nature of teaching quality. With respect to the dynamic nature of teaching quality (Section 1.2.1), little is known about how teaching quality develops over time and its longitudinal relations with students development. This research gap is concerning, given that the lower secondary school years coincide with the developmental phase of early adolescence, during which many students experience negative changes and challenges in their academic engagement and their social and emotional adjustment to school (Roeser et al., 2000). In order to foster adolescents' growth and positive development during lower secondary school, it is necessary to know more about the extent to which changes in teaching quality over time are related to changes in both students' academic achievement, academic engagement, and their social and emotional school adjustment during the years spanning early adolescence.

Regarding the contextual nature of teaching quality (Section 1.2.2), it needs to be kept in mind that high-quality teaching is defined and established within the complex social setting of the classroom, to which the teacher and the specific students in the classroom contribute. Relatedly, teaching is a socially mediated process that arises through interactions between the teacher and the specific students in the classroom (Vygotsky, 1978). This implies that students are not only recipients of what happens in the classroom, but also actively contribute to classroom processes - through their agentic behavior and their background characteristics (Eccles & Roeser, 2011). To

illuminate the contextual nature of teaching quality, the present dissertation explored the extent to which the specific students in the classroom contribute to the quality of teaching.

Below, the research questions pursued in the three empirical studies will be described in more detail.

Study 1 (*Ask me, I (Dis)agree! Acquiescence in Student Ratings of Teaching Quality in German Vocational Schools*) examined the extent to which reports on teaching quality provided by students at different grade levels are affected by an acquiescent response style. As outlined earlier, acquiescence ("yay-saying") and its counterpart dis-acquiescence ("nay-saying") represent a well-known type of response error in self-report data that may affect the psychometric properties of student reports on teaching quality. Using a large cross-sectional sample of n = 2,234 fifth-grade and n = 1,832 eighth-grade students from vocational-track schools in Germany, Study 1 evaluated the impact of (dis-)acquiescence on student reports of teaching quality, while additionally controlling for an extreme response style. Taking into account the multilevel structure of student data, the study investigated the degree to which acquiescence differs both across individual students in the same classroom (i.e., student level) and across classrooms (i.e., classroom level). Next, within a structural equation modeling framework, we evaluated the impact of acquiescence on the psychometric properties of teaching quality reports by systematically comparing the factor means and factor inter-correlations before and after statistically controlling for acquiescence. The study design made it possible to comprehensively evaluate the extent to which students' acquiescent responding impacts their teaching quality reports: First, the study's multi-cohort design made it possible to examine the degree to which age-related differences in acquiescence account for age-related differences in student reports on teaching quality. Second, three distinct aspects of teaching quality (i.e., teacher support, clarity of instruction, and content relevance) were included in the study. Importantly, the scales used to assess teaching quality varied with regard to their keying direction, informing about the extent to which acquiescence exhibits comparable effects across scales with positively and negatively worded items.

Study 2 *(Student Development in Early Adolescence: Does Teaching Quality Shape Students' Academic Achievement, Academic Engagement, and Their Social and Emotional School Adjustment?)* is a comprehensive longitudinal study investigating the role of teaching quality for critical domains of students' development (i.e., academic achievement, academic engagement, and their social and emotional school adjustment) during early adolescence. To this end, groups of

students making up specific school classes and their homeroom teachers were followed over three years of lower secondary school (i.e., fifth to eighth grade). In total, the study comprised four measurement points. The pooled sample consisted of N = 3,880 students and N = 126 homeroom teachers from German vocational-track schools. The study design has particular advantages: First, multiple relevant aspects of teaching quality (i.e., disturbances, monitoring, teacher academic and emotional support), academic achievement, students' academic engagement (e.g., academic identification with school, truancy) and students' social (e.g., social relatedness, school belonging) and emotional (i.e., self-esteem, emotional well-being) school adjustment were taken into account simultaneously. Second, the study applied a multi-reporter perspective to examine teaching quality trajectories. Combining information from the student and the teacher perspectives helped to ensure that the findings regarding changes in teaching quality over time are generalizable beyond students' perceptions alone. Third, the main aim of the study was to link the trajectories of teaching quality as rated by both students and their teachers to the trajectories of student development as students moved from fifth to eighth grade. Considering the multilevel structure of the student data, the study comprehensively investigated whether there are systematic differences in adolescent students' academic development and their social and emotional school adjustment within and across classrooms that can be traced back to the quality of teaching provided in the classroom.

Lastly, Study 3 (*Classroom Management: Can it be too Much of a Good Thing?*) focused on the extent to which the specific students in the classroom contribute to the quality of teaching through their background characteristics and their own agentic behavior in the classroom. Using classroom management as an example, the study examined key aspects of classroom management that refer more to student behavior in the classroom (i.e., disruptive behavior) or more to teachers' management actions in the classroom (i.e., monitoring, rule clarity, structure). To examine the extent to which the specific students being taught contribute to the classroom management process, the study examined the associations between students' background characteristics (i.e., gender, SES, prior achievement, school track) and aspects of classroom management. Next, we examined the degree to which students' disruptive behavior in the classroom contributes to the association between teachers' management actions and students' academic achievement. To this end, we examined the predictive value of monitoring, rule clarity and structure for students' academic achievement, when accounting for students' disruptive behavior in the classroom. Importantly, Study 3 used a multi-study design in which the same hypotheses were examined in two independent samples. Study 3.1 used data from a sample of N = 4,645 German tenth-grade students enrolled in

academic- or intermediate-track schools. Study 3.2 used a total sample of N = 6,298 German students from Grades 6 to 10 enrolled in academic-track and different types of vocational-track schools.

3 STUDY 1:

ASK ME, I (DIS)AGREE! ACQUIESCENCE IN STUDENT RATINGS OF TEACHING QUALITY IN GERMAN VOCATIONAL SCHOOLS

Marder, J., Göllner, R., Wagner, W. & Fauth, B. (2021). Ask me, I (Dis)agree! Acquiescence in student ratings of teaching quality in German vocational schools. *Studies in Educational Evaluation, 68*, 100937. https://doi.org/10.1016/j.stueduc.2020.100937

Abstract

Student ratings are frequently used to assess teaching quality even though they might be influenced by an acquiescent response style. Acquiescence is considered particularly common among young respondents and has been recognized as a source of systematic response error. Within a structural equation modeling framework, we examined whether student ratings of teaching quality provided by two age cohorts (fifth- and eighth-grade students) were affected by acquiescence, while additionally controlling for an extreme response style. Across both age cohorts, acquiescence had the highest and most consistent effects on student ratings of teaching quality, whereas extreme responding affected only ratings by fifth-grade students and to a smaller degree. Mean differences in teaching quality ratings between fifth- and eighth- grade students were partly explained by age-related differences in acquiescence. Overall, the effects of acquiescence on student ratings were rather small and primarily affected the ratings provided by younger students (i.e., fifth graders).

Keywords: teaching quality; student ratings; acquiescence; extreme responding

Ask me, I (Dis)agree! Acquiescence in Student Ratings of Teaching Quality in German Vocational Schools

Using questionnaires to ask students about teaching quality has become a well-established approach in educational research and school practice. Students are seen as "expert" judges of teaching quality because they usually experience several teachers in different subjects over an extended period of time, and their perceptions of teaching quality are based on a large number of experiences in their classes (De Jong & Westerhof, 2001; Wubbels & Brekelmans, 2005). Undoubtedly, student ratings are also appealing because they are easy to administer and provide a practical and low-cost way to obtain information about teaching quality from large groups of respondents (Kunter et al., 2008).

At the same time, student ratings may be affected by differences in individual students' response behavior. When asked about teaching quality, students may differ in their tendencies to provide affirmative answers (acquiescence; "yea-saying") or disagree (disacquiescence; "nay-saying") to questions regardless of item content (Baumgartner & Steenkamp 2001; Paulhus, 1991; Winkler et al., 1982). Specifically, there is growing evidence that response formats in which students indicate how strongly they agree or disagree with a given statement are susceptible to the influence of response styles (Podsakoff et al., 2012; Vaaerenbergh & Thomas, 2013; Moors, 2008). When using a rating scale, an acquiescent student would show a preference for the categories reflecting agreement, whereas a disacquiescent student would show a preference for the categories reflecting disagreement.The overuse of one side of the rating scale is problematic because it will lead to inflated (deflated) item scores and has the potential to affect correlations between items and constructs. Moreover, acquiescence might confound the comparison of mean scores, which is of practical relevance when using student ratings of teaching quality in educational research and practice (e.g., Fauth et al., 2020; Eccles & Roeser, 2009; Way et al., 2007). Thus, further research to clarify the effects of acquiescence on student ratings of teaching quality is warranted.

In the present study, we examined whether student ratings of teaching quality provided by two age cohorts (i.e., fifth and eighth graders) were affected by acquiescence. Specifically, we systematically compared factor means, and factor intercorrelations before and after controlling for acquiescence.

Measuring Teaching Quality with Student Ratings

Teaching quality is one key factor that promotes students' achievement, motivation, and subject-related interest (e.g., Hattie, 2009; Kunter et al., 2013; Seidel & Shavelson, 2007) and has therefore become an integral part of international research on educational effectiveness. Student, teacher, or observer ratings are frequently used to provide information about various aspects of teaching practices in the classroom, such as the structure of lessons or whether the teacher is attentive to students' needs and to disruptions in class (e.g., Kunter & Baumert, 2006; Pianta & Hamre, 2009).

When using student ratings to evaluate teaching quality, it has to be considered that students are grouped within classrooms. Therefore, student ratings provide information on two different levels (Lüdtke et al., 2009). First, student ratings can be aggregated into class means, which represent the shared perception of all students within the classroom and provide information about the class as a whole (i.e., classroom level). In contrast, individual student ratings provide information about the idiosyncratic (i.e., non-shared) perceptions of each individual student within a classroom (i.e., student level). Both students' shared and non-shared perceptions have been shown to provide meaningful information about teaching quality and students' learning (Aldrup et al., 2018; Downer et al., 2015; Aldrup et al., 2018; Wagner et al., 2013).

Although student ratings are a well-established approach to assessing teaching quality, empirical research has identified several issues with student ratings of teaching quality. For example, prior studies have found that student ratings and ratings provided by teachers or external observers for the same teaching quality aspects do not seem to correspond very well (e.g., Kunter & Baumert, 2006). Furthermore, differences in teaching quality ratings between students within classroom are mostly more pronounced than differences in students' shared perceptions across classrooms. For example, prior studies (e.g., Kunter et al., 2008; Wagner et al., 2013) indicate that a much larger proportion (73% to 90%) of the variance in student ratings of teaching quality is attributable to differences within classrooms than to differences between classrooms. Moreover, students' ability to distinguish between different aspects of teaching quality has been called into question (Scherer et al., 2016; Wagner et al., 2013). Investigations of the underlying factor structure of student ratings have typically found higher correlations among the proposed factors of teaching quality than observed in ratings provided by teachers or observers or theoretically assumed (e.g., Kunter & Baumert, 2006). Whereas each approach to assessing teaching quality is known to hold

its own strength and weaknesses (e.g., Kunter & Baumert, 2006; Wubbels et al., 1993), it has been suggested that student ratings of teaching quality might be influenced by factors other than the quality of teaching, such as a student's individual response style (Aldrup et al., 2018; Scherer & Gustafsson, 2015; Scherer et al., 2016).

Particularly when evaluating teaching quality, students might show a preference for the response categories reflecting agreement (acquiescence) or the categories reflecting disagreement (disacquiescence). For example, two students might be equally satisfied with their teachers' pace during lessons but differ in their tendency to agree with items on a rating scale (e.g., some students tend to indicate stronger agreement, whereas other students tend to prefer categories around the midpoint of the rating scale). As a consequence, some students systematically assign higher ratings than others, which can lead to more variability in teaching quality ratings within a class or even result in lower variability if students with similar rating tendencies are concentrated in the same classroom. Second, if students consistently agree (or disagree) with teaching quality items that actually tap different aspects of teaching quality, this will naturally increase the shared variance among items and affect the relations between teaching quality factors (Kam & Mayer, 2015; Rammstedt & Farmer, 2013; Weijters et al., 2010b). Consequently, acquiescence may represent an important source of variability in student evaluations of teaching quality. In order to use student ratings as a reliable and valid method of assessing teaching quality in school, it is critical to examine and evaluate the impact of acquiescence on students' teaching quality ratings

Students' Tendencies to Respond in an Acquiescent Way

In general, response styles are conceptualized as a respondent's systematic tendency to use certain categories of a response scale on a basis other than the target construct (Baumgaertner & Steenkamp, 2001; Paulhus, 1991). Specifically, acquiescence (ARS) and disacquiescence (DRS) are defined as a respondent's tendency to consistently agree (disagree) with an item regardless of item content, even though this may lead to self-contradictory responses (Bentler et al., 1971; Weijters et al., 2013). However, when respondents use the rating scale positions expressing extreme agreement (disagreement), acquiescence may resemble an extreme response style (ERS). An extreme response style reflects systematic use of the most extreme positions at the two ends of the rating scale rather than the moderate ones (Paulhus, 1991). The difference between acquiescence and an extreme response style is that respondents with an extreme response style select extreme responses in both directions (Kam & Fan, 2017), whereas respondents with an

acquiescent (disacquiescent) response style consistently prefer the direction of the rating scale expressing agreement (disagreement). Prior research has frequently demonstrated that acquiescence has the potential to distort the measurement properties of data, for example by skewing the distribution toward the endpoint of a scale (Ray, 1985) or affecting scale means by inflating or deflating item scores (Baumgartner & Steenkamp, 2001; Bentler et al., 1971; Bolt, Lu, & Kim, 2014). Furthermore, because respondents high in acquiescence tend to agree with items regardless of item content keying direction, positive correlations between constructs that are keyed in the same direction may be artificially inflated, whereas negative correlations between opposite-keyed constructs may be deflated or even turn positive (e.g., Kam & Meyer, 2015; Rammstedt & Farmer, 2013; Winkler et al., 1982).

A common explanation for the psychological function of acquiescence is to simplify the process of forming a judgment and responding to items (Baumgartner & Steenkamp, 2001). Answering questions on a rating scale is a cognitively demanding process in which respondents are required to accomplish several interrelated tasks. According to prominent models of survey response (e.g., Sudman et al., 1996; Tourangeau et al., 2000), respondents first have to interpret the question to understand what is meant, then retrieve relevant information from memory to form a mental representation of the subject or behavior in question, and then integrate this information into a final judgment. Finally, they need to report their opinion by choosing one of the provided response categories. Because acquiescence (disacquiescence) reflects a respondent's preference for the agreement (disagreement) position on a rating scale, it affects the final task in the response process, in which a response category has to be selected (Weijters et al., 2013).

Acquiescence in Student Ratings of Teaching Quality

Acquiescence has typically been conceptualized as "individual differences in use of the response scale" (e.g., Cronbach, 1950). Respondents who lack a sufficient level of attention or motivation when responding to questions (Huang et al., 2012; Maniaci & Rogge, 2014), or have relatively lower cognitive ability and cognitive skills (Gudjonsson, 1990; Lechner & Rammstedt, 2015) seem to be more prone to acquiescent responding. Specifically, low language and reading literacy and low verbal ability may hamper comprehension and interpretation of the question and lead to a higher chance of being uncertain about its meaning (Messick, 1966). Consequently, respondents provide answers without fully processing the item content before responding (Meade & Craig, 2012). Seen from this point of view, the influence of acquiescence on responses should

be greatest when “vague, ambiguous or difficult to answer items are involved” (Moors, 2008, p. 781). Indeed, teaching quality items often ask about large time intervals (e.g., the whole school year) or require information about the behavior of others (e.g., the teacher/students in this class), which may pose a challenge for information retrieval or the process of forming a judgement, thus increasing the vulnerability of teaching quality items to response styles. Consequently, if students face items they find difficult to answer, they might tend to choose one side of the scale over the other (e.g., Baumgartner & Steenkamp, 2001; Goldberg, 1963).

Considering that students are nested in classes, it has mostly been argued that individual differences in acquiescence are of less relevance in analyses of students’ shared perceptions of teaching quality, because analyses at the classroom level are considered to have the methodological advantage of counterbalancing student idiosyncrasies (Lüdtke et al., 2009). However, acquiescent students might not actually be randomly distributed across classes, not least because individual factors related to acquiescence have been shown to vary across classes (e.g., Karing et al., 2013; Kunter et al., 2007; Mullis et al., 2003; Pfost & Artelt, 2013). Moreover, the common practice of sorting students into homogeneous classes based on their ability and achievement or elective subject may contribute to classroom differences in students’ acquiescence tendencies. If the number of acquiescent and non-acquiescent students (or disacquiescent and non-disacquiescent) is not equally distributed between classrooms, students’ responding tendencies can lead to a false sense of agreement between students from the same class which will consequently bias students’ aggregated ratings at the classroom level. Hence, more research is needed to clarify whether there are systematic differences in acquiescence both between individual students in the same classroom (i.e., student level) and between classrooms (i.e., classroom level) and the degree to which acquiescence affects students’ individual and shared perceptions of teaching quality.

In addition to variability in acquiescence within and across classes, there might also be differences in students’ acquiescence tendencies across grade levels. Previous research has provided extensive evidence that an acquiescent response style is more commonly observed and substantially more pronounced in low and moderately educated groups than in groups with a high level of formal education (e.g., Billiet & McClendon, 2000; Costello & Roodenburg, 2015; Rammstedt et al., 2010; Meisenberg & Williams, 2008). Furthermore, younger children are expected to exhibit greater variability in acquiescence (Soto et al., 2008). From the perspective of developmental psychology, late childhood and adolescence are characterized by fundamental

changes in cognitive capacities, self-concept, and reading and verbal comprehension skills, which affect one's ability to think abstractly and report information about oneself (Eccles et al., 1991; Flavell et al., 1993; Toomela, 2003). Compared with adolescents, younger children have a smaller vocabulary, less capacity to evaluate the logical consistency of statements, and less experience with questionnaires. Even though acquiescence has mostly been studied concerning self-report data assessing a variety of constructs, such as personality traits or personal attitudes (e.g., Baumgartner & Steenkamp, 2001; Rammstedt & Farmer, 2013; Wetzel et al., 2013), it can be assumed that student teaching quality questionnaires, in which students evaluate their own behavior as well as their teacher's actions in the classroom, are comparable to self-report questionnaires (Spooren et al., 2012). Therefore, the cognitively demanding process of interpreting an item, retrieving information, forming a judgment, and providing a response (Tourangeau et al., 2000) involved in evaluating aspects of teachers' teaching quality might be more sensitive to the influence of a systematic response style for younger students at the beginning of secondary school than for older students.

Thus, it can be asked whether acquiescence should be considered in assessments of teaching quality using student ratings across students of different ages (e.g., Aldrup et al., 2018) or when evaluating teaching quality across longer time spans (e.g., several school years), particularly when student ratings are used as the primary source of information on teaching quality. In addition, student ratings are often used for evaluation purposes within the context of teacher evaluation systems, decision-making and school practice (Fauth et al., 2020). Particularly when teaching quality assessments are used to draw inferences about teacher effectiveness (Gitomer & Bell, 2013), it has to be considered that when including ratings by students in different grade levels, observed differences in teaching quality between younger and older students might reflect not only genuine differences in the students' perceptions of teaching quality but also differences in their tendencies to acquiesce (Cheung & Rensvold, 2000).

For example, longitudinal analyses focusing on changes in students' perception of their school environment in secondary school (i.e., from sixth to eighth grade) have shown that on average, students tend to perceive their schools and teachers increasingly negatively over time, as indicated by a consistent decline in their overall evaluation of relevant aspects of the school climate (e.g., teacher support, clarity of school rules, and teacher-student relationship; Reddy et al., 2003; Wang & Dishion, 2012; Way et al., 2007). Similar results were obtained in a study by Booth and

Gerard (2014), who analyzed changes in student ratings of their school environment from seventh to tenth grade. This phenomenon has predominantly been explained in terms of a "misfit" between students' changing needs and what their learning environment provides (Eccles & Roeser, 2009; Midgley et al., 1989). Meanwhile, other factors that might contribute to these declines over time have rarely been considered when interpreting the results. Therefore, it remains unclear whether the reported declines in various aspects of teaching quality over time indeed indicate a deterioration of the teaching quality, or whether these changes might be confounded to some degree with changes in students' response behavior over time.

The Present Study

We investigated students' acquiescent response style in a large cross-sectional sample of German fifth- and eighth-grade students and evaluated the impact of students' acquiescent responding on three distinct aspects of teaching quality (i.e., teacher support, clarity of instruction, and content relevance). The scales used in the present study varied with regard to their keying direction, which allowed us to investigate the degree to which acquiescence exhibits comparable effects across scales with positively and negatively worded items. Even though acquiescence has been found in self-report questionnaires assessing many different constructs (see Baumgartner & Steenkamp, 2001), and there is some evidence that student ratings of teaching quality are affected by an acquiescent response style, a systematic investigation of acquiescence in student ratings is still missing (but see Spooren, 2012).

Specifically, it has to be clarified whether there are systematic differences in acquiescence both across individual students in the same classroom and across classrooms. Considering the possibility that acquiescent students may not necessarily be distributed randomly across classes, we expect the impact of acquiescence on students' teaching quality ratings to differ across both individual students in the same classroom as well as across classrooms (Hypothesis 1).

To provide a direct measure of acquiescence, we calculated an acquiescence index based on an external set of items to avoid a conceptual overlap between the items used to assess acquiescence and those used to assess teaching quality (Billiet & McClendon, 2000; Weijters et al., 2010a, 2010b). Subsequently, within a structural equation modeling framework, we regressed the calculated index values on the teaching quality items to statistically control for acquiescence in student ratings of teaching quality. Specifically, we expected acquiescence to be more pronounced

in ratings provided by fifth-grade students than those provided by eighth-grade students (Hypothesis 2).

We then examined the degree to which acquiescence affected the factor structure of teaching quality (i.e., factorial inter-correlations and factor means) in fifth- and eighth-grade students' ratings of teaching quality. After controlling for acquiescence, we expected the inter-correlations between factors with the same keying to be weakened, whereas we expected the inter-correlations between factors with opposite keying to be strengthened (i.e., more negative; Hypothesis 3).

Finally, we examine the degree to which age-related differences in an acquiescent response style can explain differences between fifth- and eighth-grade students' mean ratings of teaching quality. We expected fifth graders to have a higher acquiescence index score than eighth graders. After statistically controlling for acquiescence, we expected the mean differences between fifth- and eighth-grade students' ratings of teaching quality to decrease (Hypothesis 4).

Method

Sample

We used data from a German large-scale study designed to analyze the academic development of students in vocational secondary school tracks ("Tradition and Innovation in School Systems Study" [TRAIN]; Jonkmann et al., 2013). Participating schools were randomly sampled from two German federal states (Baden-Württemberg and Saxony). The Ministries of Education and Cultural Affairs of the states of Baden-Württemberg and Saxony reviewed the study and approved the instruments and the reassessment of students over time in 2008. Students' participation in the TRAIN study was voluntary and required active parental consent.

In the German school system, students transition to secondary school at the end of fourth grade. Based on their prior achievement, students are allocated to an academic or vocational school track. High achieving students are generally allocated to an academic school track, while the vocational school tracks prepare students to enter vocational training rather than higher education. The vocational school tracks are further divided into an intermediate school track ('Realschule') and a lower school track ('Hauptschule'). Additionally, some German federal states offer comprehensive schools that combine both tracks. The underlying reason for grouping students based on their abilities is to form more homogeneous learning environments designed to meet the

instructional and emotional needs of each specific group of students (Maaz et al., 2008). The dataset used for the present investigation comprised n = 2,234 fifth grade students from 128 school classes (M_{age} = 11.08, SD = 0.54; 46.8% female) and n = 1,832 eighth grade students from 95 school classes (M_{age} = 14.19, SD = 0.63; 48.0% female). The number of students participating from each class ranged from 4 to 35 in fifth grade (M = 24.78, SD = 4.97) and from 9 to 33 in eighth grade (M = 24.92, SD = 4.76). Students were administered questionnaires to assess their background characteristics as well as aspects of the quality of their homeroom teacher's teaching. In Germany, the homeroom teacher teaches his or her class at least one subject and bears special responsibility for this class. It's important to note that within each school track, fifth- and eighth-grade students are taught by the same body of teachers and are considered to have comparable social backgrounds, academic abilities and experiences in the educational environment (cf. Dumont et al., 2013; Schiepe-Tiska, 2019).

Overall, 26.8 % of the students in the present study were enrolled in the lower track ('Hauptschule') and 38.2 % in the intermediate track ('Realschule'), while 35 % of the students were enrolled in the comprehensive track. 21.0 % of the fifth-grade students and 18.2 % of the eighth-grade students had a migration background. Of these, 18.1 % of the fifth grade students and 14.2 % of the eighth-grade students were themselves born in Germany but had at least one parent born outside Germany, whereas 2.9 % of the fifth-grade students and 3.9 % of the eighth-grade students reported that both they themselves and their parents were born outside Germany.

Measures

Teaching Quality

We examined teachers' learning support, clarity of instruction, and content relevance as three well-known measures of teaching quality. The items were adapted from published scales (Baumert et al., 1996; Baumert et al., 2009; Gruehn, 2000). All items were rated on a 4-point Likert scale ranging from 1 = *strongly disagree* to 4 = *strongly agree* and are displayed in Appendix A. Means, standard deviation and intraclass correlation coefficients (ICC) for each item in fifth and eighth grade are presented in Table 1.

Teacher Support. Teachers' interest in students' learning and provision of learning-related assistance was measured with seven items (e.g., "Our homeroom teacher supports us in our

learning"; fifth grade: α = .88; eighth grade: α = .91). Items on the *teacher support* scale were positively worded, so that higher values indicate a higher level of teacher support.

Content Relevance. Five items assessed the extent to which the teacher failed to focus on the relevant subject matter (e.g., "Our homeroom teacher often talks about something completely unrelated to our subject"; fifth grade: α = .79; eighth grade: α = .77). Items on the *content relevance* scale were negatively worded, so that higher values indicate a lower level of the construct.

Clarity of Instruction. Six items measured the extent to which the teacher expressed himself/herself clearly and provided comprehensible instructions. *Clarity of instruction* was assessed with a balanced scale consisting of three positively and three negatively worded items (e.g., "Our homeroom teacher sometimes gives unclear instructions and nobody really knows what to do" [negatively worded]; "Our homeroom teacher knows how to explain complicated things very well" [positively worded]; fifth grade: α = .57; eighth grade: α = .79). The three negatively worded items were presented first.

Acquiescent Response Style Indexes

When measuring acquiescence in a student questionnaire of teaching quality, one challenge consists of separating acquiescence from substantive information (Baumgartner & Steenkamp, 2001), because students' responses may reflect a combination of both a response style and "true" information on students' perception of teaching quality (He et al., 2017; Mottus et al., 2012). For this reason, and consistent with previous research (e.g., Kam & Mayer, 2015; Kam & Zhou, 2015; Weijters et al., 2010a, 2010b), we calculated an acquiescence indicator using an external set of items. Because response styles are considered to be unrelated to item content, students are expected to exhibit similar response styles in their responses to both the external item set and the teaching quality items (He & van de Vijver, 2015; Mottus et al., 2012). Specifically, we calculated overall indicators for an acquiescent response style (ARS) and disacquiescent response style (DRS) on the basis of antithetical item pairs drawn from the Big Five Inventory (BFI; selected by Soto et al., 2008), which students also filled in during the regular assessment. Overall, we used 16 antithetical item pairs that captured the same construct with logically opposite item content (e.g., "I see myself as someone who…is talkative" vs. "…tends to be quiet"). The item pairs were evaluated by Soto et al. (2008) regarding their content and interitem correlations to avoid confounding the measurement of response styles with the measurement of personality. Because the item pairs are

balanced in content and keying (Rammstedt & Farmer, 2013), simultaneous agreement or disagreement to both statements can be attributed to a response style rather than to item content (Weijters et al., 2013; Rammstedt & Farmer, 2013).

The BFI items were answered on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. To calculate mean scores for ARS and DRS for each participant, we followed the procedure outlined by Baumgartner & Steenkamp (2001) and Weijters et al. (2010b) by assigning weights to each response category. For ARS, defined as the tendency to agree with items regardless of content, we assigned 2 points to Category 5 (strongly agree), 1 point to Category 4 (agree), and 0 points to the remaining categories. Similarly, for DRS, the categories reflecting disagreement were weighted by assigning 2 points to Category 1 (strongly disagree) and 1 point to Category 2 (disagree). The method of recoding original item responses to calculate index values reflecting response styles has been applied in numerous previous studies (e.g., Bachman & O' Malley, 1984; Baumgartner & Steenkamp, 2001; Mottus et al., 2012, Weijters et al., 2010a, 2010b). To combine the bidirectional effects of ARS and DRS, we subtracted students' DRS index from their ARS index to obtain an index indicating students' net acquiescent response style (NARS). The NARS index reflects the ratio of weighted agreement and disagreement across the 16 item pairs and was used in all subsequent analyses. The NARS index ranged from -2 to 2. A positive index value indicates a respondent's tendency to agree rather than disagree (e.g., Baumgartner & Steenkamp, 2001; Weijters et al., 2010a, 2010b).

To further distinguish acquiescence (disacquiescence) from an extreme response style (ERS), we additionally calculated an ERS index and used the NARS and ERS indices simultaneously in all analyses. The ERS index was based on the proportion of extreme responses and calculating by weighting the extreme categories 1 (*strongly disagree*) and 5 (*strongly agree*) with 1 point and the remaining categories with 0 points. The ERS index ranged from 0 to 1 and thus reflected the ratio of extreme and non-extreme answers given by each respondent. For each indicator, we averaged the scores obtained using this coding scheme.

Means, standard deviation and ICCs for the response style index are presented in Table 1. The calculated indexes showed good reliabilities (fifth grade: $.89 \leq \alpha \leq .95$; eighth grade: $.81 \leq \alpha \leq .89$). For each response style, fifth-grade students showed significantly higher index values than eighth-grade students (all *p*s > .05) and ICCs. In fifth grade, the response style indexes' average ICC was $M = 0.07$, ranging from 0.03 to 0.10 (eighth grade: $M = 0.03$, ranging from 0.01 to 0.05).

Because the response style indexes showed sufficient between-class variation to make aggregation feasible, we aggregated students' individual response style indexes at the classroom level to form an indicator of a class's average level of acquiescence and extreme responding.

Table 1

Descriptive Statistics for Teaching Quality Items and Response Style Indexes

	Fifth grade		Eighth grade	
	M (SD)	*ICC*	*M (SD)*	*ICC*
Teacher support				
tsup1	3.21 (0.83)	.06	2.90 (0.89)	.17
tsup2	3.45 (0.72)	.09	3.04 (0.86)	.20
tsup3	3.19 (0.85)	.11	2.85 (0.92)	.20
tsup4	3.34 (0.80)	.08	2.94 (0.85)	.17
tsup5	3.31 (0.84)	.08	2.93 (0.91)	.21
tsup6	3.27 (0.83)	.10	2.97 (0.89)	.18
tsup7	3.37 (0.79)	.08	3.10 (0.88)	.15
Clarity of instruction				
clari4	2.39 (1.02)	.02	2.42 (0.88)	.12
clari5	2.07 (1.02)	.02	2.29 (0.94)	.14
clari6	2.08 (1.01)	.05	2.33 (0.95)	.15
clari1	2.97 (1.00)	.02	2.65 (0.91)	.07
clari2	3.16 (0.93)	.04	2.78 (0.90)	.14
clari3	3.32 (0.87)	.05	2.88 (0.93)	.15
Content relevance				
conre1	2.51 (1.03)	.05	2.61 (0.96)	.13
conre3	1.96 (1.05)	.04	2.13 (0.97)	.11
conre4	2.14 (1.01)	.06	1.97 (0.86)	.07
conre5	2.14 (1.12)	.06	2.21 (1.05)	.13
ARS	0.61 (0.44)	.10	0.50 (0.32)	.05
DRS	0.47 (0.32)	.03	0.41 (0.26)	.01
ERS	0.36 (0.28)	.04	0.25 (0.21)	.03
NARS	0.14 (0.63)	.09	0.08 (0.46)	.04

Note. clari1 to clari3 and conre1 to conre5 are negatively worded items.

Statistical Analyses

Measurement Model for Teaching Quality and Testing for Measurement Invariance

To establish the measurement model for teaching quality, we specified a three-factorial model with teacher support, clarity of instruction, and content relevance as seperate teaching quality factors. Each item was constrained to load onto its respective factor without any cross-loadings. The residuals of the three negatively worded items in the balanced *clarity of instruction* scale were allowed to correlate with each other in order to account for their correlated uniqueness. The model was conducted as multigroup confirmatory factor analyses (MG-CFA) with the fifth- and eighth-grade students in separate groups.

A central aim of the present study was to examine fifth- and eighth-grade students' teaching quality ratings before and after controlling for acquiescence. Meaningful comparisons between two groups of respondents with regard to their teaching quality ratings require certain levels of measurement invariance of the teaching quality factors (i.e., configural, metric, and scalar), which we tested for prior to comparing the regression coefficients of the response style indexes, factor intercorrelations and factor means across fifth- and eighth-grade students. Within a multigroup confirmatory factor analysis (MG-CFA), measurement invariance can be examined by increasingly constraining the model parameters to equality across the compared groups (Cheung & Rensvold, 2000; Vandenberg & Lance, 2000) and evaluating changes in the resulting model fit indices. *Configural invariance* is established if the factor structure (i.e., number of factors and loading patterns) is found to be identical across groups. In a configural invariance model, the covariance matrices for both groups can be fitted using the same factor model (Horn & McArdle, 1992). Therefore, the configural invariance model does not include any equality constraints, meaning that factor loadings and item intercepts are freely estimated in each group. *Metric invariance* requires that the factor loadings do not differ across groups, which can be tested by constraining the factor loadings to equality across groups. Metric invariance must be established in order to compare correlation and regression coefficients across groups (Meredith & Teresi, 2006). If the assumption of metric invariance is holds, *scalar invariance* can then be tested by additionally constraining the item intercepts to equality across groups. Scalar invariance should be established to ensure valid comparisons of (latent) factor means across groups (Millsap, 2012; Widaman & Reise, 1997). Each level of measurement invariance is considered to be established if constraining the model

parameters does not lead to a significant decrease in model fit compared to the freely estimated (i.e., configural) model.

Evaluating the Impact of Acquiescence on Students' Teaching Quality Ratings

To account for students' acquiescence tendencies, we included students' individual index values and the aggregated class-mean index values for each response style as predictors of students' teaching quality ratings into the measurement model for teaching quality and regressed each teaching quality item on both the individual and the class-mean index values of acquiescence and extreme responding. The individual index values were centered around the group mean, partialling out the variance in the teaching quality items due to differences in students' response styles within classroom (Cheung & Rensvold, 2000). In contrast, the aggregated class-mean indexes account for the variance in teaching quality due to differences in response styles between classes. The conceptual model of teaching quality and response styles is depicted in Figure 1.

To further evaluate whether individual students' acquiescence is counterbalanced at the classroom level or whether a class's average level of acquiescence reflects an accumulation of the acquiescence tendencies of all students in the classroom, we set students' individual and class-mean aggregated index values for each response style to equality within the groups of fifth- and eighth-grade students and evaluated the resulting changes in the model fit.

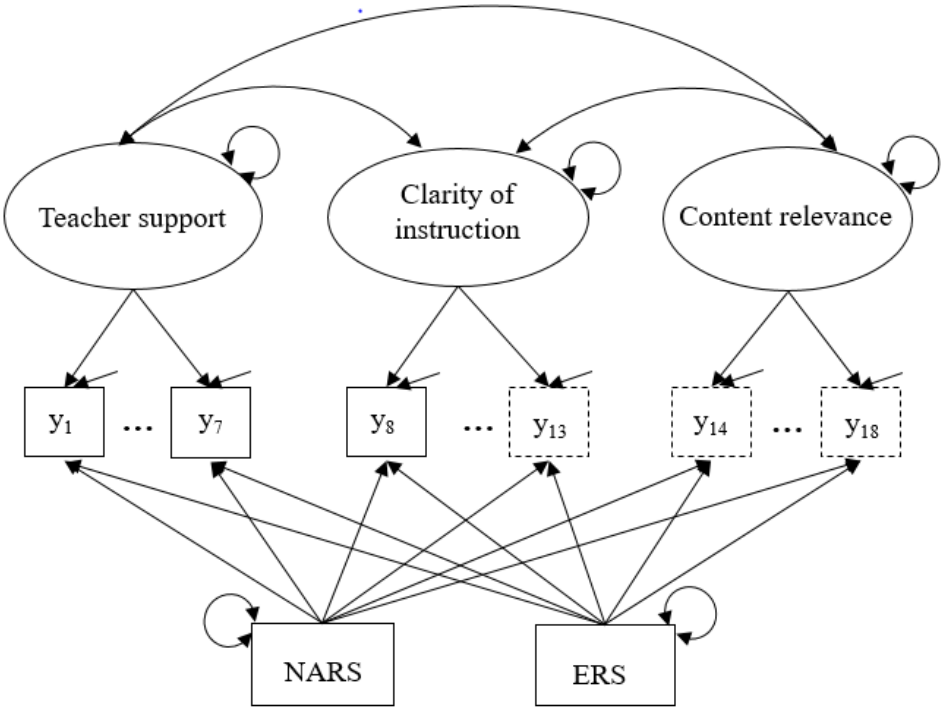

Figure 1. Conceptual model of teaching quality and response style indexes. Negatively worded items are presented in dashed boxes

Age-related Differences in Students' Acquiescent Responding

In order to evaluate whether fifth- and eighth-grade students differ in the strength of their acquiescence tendencies, we compared the regression coefficients for the individual response style indexes and the aggregated class-mean indexes on predicting the teaching quality items. To ensure comparability of the regression coefficients across fifth- and eighth-grade students, we constrained the factor loadings of the teaching quality items in fifth- and eighth-grade students to equality (Metric Response Style Model). We then set the regression coefficients for acquiescence and extreme responding to equality across fifth- and eighth-grade students and compared the resulting model fit to the Metric Response Style Model, in which the regression coefficients were estimated freely for fifth- and eighth-grade students.

Factorial Intercorrelations and Mean-level Differences

Next, we evaluated whether acquiescence affected the correlations between the teaching quality factors. Thus, we compared the intercorrelations between the teaching quality factors before and after we controlled for acquiescence (Metric Teaching Quality Model vs. Metric Response Style Model). Similarly, we compared the latent mean differences in the teaching quality factors between fifth- and eighth-grade students in the scalar invariance models before and after controlling for acquiescence (Scalar Teaching Quality Model vs. Scalar Response Style Model).

Evaluating Model Fit

We evaluated all models on the basis of their goodness-of-fit indices. Good model fit was indicated by root mean square error of approximation (RMSEA) values ≤ .06, confirmatory fit index (CFI) values ≥ .95, and standardized root mean square residual (SRMR) values ≤ .08 (Hu & Bentler, 1999).

For all model comparisons, we evaluated changes in the model fit by considering the chi-square difference test. The combination of the change in the chi-square value and the number of degrees of freedom provides information about the extent to which a model represents an improvement against the previous model. If model fit does not significantly worsen (as indicated by a significant chi-square difference test), this indicates that both models fit the data equally well. However, because the chi-square test strongly depends on sample size and might be overly sensitive to even trivial misfit (Little, 2013), we additionally inspected changes in the goodness-of-fit indices. We considered a decrease in the CFI of less than 0.010 and an increase in the RMSEA and SRMR of less than 0.015 as support for the more constrained model (Chen, 2007; Cheung & Rensvold, 2002).

All analyses were conducted in M*plus* 7 (Muthén & Muthén, 1998-2017) using robust maximum likelihood estimation (MLR; Yuan & Bentler, 2000). Due to the model complexity and the relatively low sample size at the classroom level, all analyses were based on students' individual data, assuming the comparability of factors (i.e., measurement models and factor inter-correlations) at the within- and the between-classroom level (Stapleton, Yang, & Hancock, 2016). To account for the nested data structure of students grouped within classrooms, we applied the well-established type = complex procedure in Mplus. In the type = complex procedure, the standard errors and chi-

square values of the model estimates are adjusted for the clustering of students in order to avoid higher alpha error (Muthén & Satorra, 1995).

On average, we had to deal with 4.2% missing values on single items or scales in the student data. To use all available information, we applied full information maximum likelihood estimation in the structural equation models used in this study (Enders, 2010). All significance testing was performed at the .05 level (two-tailed).

Results

Impact of Acquiescence on Students' Teaching Quality Ratings (Hypothesis 1)

The three-factorial measurement model of teaching quality with fifth- and eighth-grade students in separate groups and without modeling acquiescence (Configural Model of Teaching Quality) exhibited good fit to the data: χ^2 (226) = 1463.94, $p < .001$; RMSEA = .052; CFI = .947; SRMR = .080. The standardized factor loadings of the teaching quality items are displayed in Table 3 (left column). However, the loading pattern for the c*larity of instruction* factor (balanced) warrants some attention. In fifth grade, two out of three negatively worded items exhibited positive and significantly weaker loadings on the c*larity of instruction* factor, whereas the loading pattern for eighth-grade students' ratings was in accordance with the items' keying direction (see Table 3, left column).

Next, we examined the degree to which students' ratings were affected by acquiescence. Specifically, we used students' individual and class-mean aggregated index values of acquiescence and extreme responding as predictors of students' teaching quality ratings and inspected the regression coefficients (Configural Response Style Model, Table 4). The model fit the data well (χ^2 (238) = 1480.33, $p < .001$; RMSEA = .051; CFI = .946; SRMR = .061). Overall, acquiescence affected students' teaching quality ratings most consistently. Across both age cohorts, students' individual and class-mean acquiescence index values predicted their teaching quality ratings. That is, acquiescence accounted for variance in teaching quality ratings between students within the classroom as well as for variance in teaching quality between classrooms. Consequently, individual differences in acquiescence among students within a classroom are not necessarily counterbalanced in students' shared perceptions of teaching quality (i.e., the classroom level). In contrast, extreme responding was only predictive for fifth-grade students' teaching quality ratings and solely

explained variance in teaching quality ratings between individual students within the same classroom.

To clarify whether a class's average level of acquiescence exhibits the same associations with students' teaching quality ratings as students' individual acquiescence tendencies, we set students' individual and class-mean aggregated index values of acquiescence to equality within each group of respondents (i.e., fifth- and eighth-grade students) and evaluated the resulting changes in model fit. The model with equality constraints for students' individual and class-mean aggregated index values of acquiescence and extreme responding also fit the data well (χ^2 (306) = 1554.27, $p < .001$; RMSEA = .045; CFI = .946; SRMR = .067), and the chi-square difference test did not indicate significant differences in the models' fit to the data ($\Delta\chi^2 = 73.9$, $\Delta df = 68$, $p = .290$; ΔRMSEA = -.006; ΔCFI = .000Δ; SRMR = .006). We conclude that the impact of the class-mean aggregated index values is comparable to the influence of the individual index values. For all subsequent analyses, we proceeded with the individual and class-mean aggregated index values of acquiescence and extreme responding set to equality. The standardized regression coefficients for the individual and class-mean aggregated index values of acquiescence when set to equality are presented in Table 5. Across both cohorts, acquiescence was predictive for the majority of items used to assess teaching quality, whereas ratings provided by fifth-grade students were additionally affected by an extreme response style (Table 5). However, the regression coefficients for extreme response style were rather small and clearly lower than the regression coefficients for acquiescence in fifth-grade.

Table 2

Measurement Invariance of Teaching Quality for the Teaching Quality Model (without modeling acquiescence) and the Response Style Model (including the modeling of acquiescence)

	χ^2 (*df*)	RMSEA	CFI	SRMR
Teaching Quality Model				
Model a: Teaching Quality Model (Configural Invariance)	1463.94 (226)	.052	.947	.080
Model b: Metric Teaching Quality Model (Full Metric Invariance)	1614.15 (240)	.053	.941	.092
Model c: Metric Teaching Quality Model (Partial Metric Invariance)	1514.25 (237)	.051	.945	.082
Model d: Scalar Teaching Quality Model (Partial scalar invariance)	1637.14 (248)	.052	.940	.083
Δ Model b – Model a	150.21 (14)	.001	-.006	.012
Δ Model c – Model a	50.31 (11)	-.002	-.002	.002
Δ Model d – Model c	122.89 (11)	.001	-.005	.001
Response Style Model				
Model e: Response Style Model (Configural Invariance)	1554.27 (306)	.045	.946	.067
Model f: Metric Response Style Model (Full Metric Invariance)	1678.92 (320)	.046	.941	.074
Model g: Metric Response Style Model (Partial Metric Invariance)	1594.79 (317)	.045	.945	.068
Model h: Scalar Response Style Model (Partial scalar invariance)	1663.32 (328)	.045	.942	.068
Δ Model f – Model e	124.65 (14)	.001	-.005	.007
Δ Model g – Model e	40.52 (11)	.000	-.001	.001
Δ Model h – Model g	68.53 (11)	.000	.000	.000

Table 3

Standardized Factor Loadings for Fifth- and Eighth-Grade Students

	Fifth grade		Eighth grade	
	No modeling of acquiescence[a] λ (*SE*)	Modeling of acquiescence[b] λ (*SE*)	No modeling of acquiescence[a] λ (*SE*)	Modeling of acquiescence[b] λ (*SE*)
Teacher support				
tsup1	**0.66 (0.02)***	**0.61 (0.02)***	**0.73 (0.02)***	**0.72 (0.02)***
tsup2	**0.76 (0.02)***	**0.72 (0.02)***	**0.83 (0.02)***	**0.81 (0.02)***
tsup3	**0.70 (0.02)***	**0.66 (0.02)***	**0.79 (0.02)***	**0.78 (0.02)***
tsup4	**0.74 (0.02)***	**0.71 (0.02)***	**0.78 (0.02)***	**0.76 (0.02)***
tsup5	**0.75 (0.02)***	**0.72 (0.01)***	**0.79 (0.01)***	**0.78 (0.02)***
tsup6	**0.70 (0.02)***	**0.66 (0.02)***	**0.73 (0.02)***	**0.72 (0.02)***
tsup7	**0.73 (0.02)***	**0.69 (0.02)***	**0.75 (0.02)***	**0.74 (0.02)***
Clarity of instruction				
clari4	**0.68 (0.03)***	**0.66 (0.03)***	**0.74 (0.02)***	**0.73 (0.02)***
clari5	**0.82 (0.02)***	**0.81 (0.02)***	**0.86 (0.01)***	**0.85 (0.01)***
clari6	**0.79 (0.02)***	**0.78 (0.02)***	**0.87 (0.01)***	**0.86 (0.01)***
clari1	**0.21 (0.03)***	**0.16 (0.03)***	**-0.15 (0.04)***	**-0.15 (0.04)***
clari2	0.21 (0.04)	-0.03 (0.03)	**-0.32 (0.05)***	**-0.32 (0.05)***
clari3	-0.01 (0.04)	-0.06 (0.04)	**-0.37 (0.05)***	**-0.39 (0.05)***
Content relevance				
conre1	**0.56 (0.03)***	**0.48 (0.03)***	**0.54 (0.03)***	**0.53 (0.03)***
conre3	**0.72 (0.02)***	**0.73 (0.02)***	**0.81 (0.03)***	**0.81 (0.03)***
conre4	**0.64 (0.03)***	**0.52 (0.02)***	**0.51 (0.04)***	**0.50 (0.04)***
conre5	**0.71 (0.02)***	**0.62 (0.02)***	**0.67 (0.03)***	**0.66 (0.04)***

Note. λ = factor loading; *SE* = standard error; *$p < .05$.

[a]Configural Teaching Quality Model; [b]Configural Response Style Model.

Age-related Differences in Students' Acquiescent Responding (Hypothesis 2)

As a prerequisite for comparing the regression coefficients of acquiescence and extreme responding across fifth- and eighth-grade students, we tested for metric invariance by constraining the factor loadings in eighth grade to be equal to the factor loadings in fifth grade (Metric Response Style Model). Even though changes in the fit indices remained within the acceptable range (Model f in Table 2), we considered metric invariance to be only partially established due to differences in the loading pattern of the negatively worded items on the *clarity of instruction* factor between fifth- and eighth-grade students. With a subset of the three negatively worded items freed, only minor changes in the fit statistics occurred when the partial Metric Response Style Model (Model g in Table 2) was compared against the less restrictive Configural Response Style Model (Δ Model g vs. Model e in Table 2: $\Delta\chi^2 = 40.52$, $\Delta df = 11$, $p < .05$; ΔRMSEA = .000; ΔCFI = -.001; ΔSRMR = .001). Thus, we consider teaching quality to be largely comparable across fifth- and eighth-grade students.

In order to test whether fifth- and eighth-grade students significantly differed in the strength of their acquiescence tendencies, we constrained the regression coefficients of acquiescence and extreme responding to be equal across fifth- and eighth-grade students within the partial Metric Response Tendency Model (Table 5) and compared the fit indices to the metric model in which the response tendencies were freely estimated within each group (Table 4). The chi-square difference test revealed significant differences in the models' fit to the data ($\Delta\chi^2 = 67.95$, $\Delta df = 28$, $p < .05$; ΔRMSEA = -.002; ΔCFI = -.002; ΔSRMR = .002). Hence, the impact of acquiescence on fifth-grade students' teaching quality ratings was more pronounced than for eighth-grade students. However, only minor changes in the models' overall fit which might be due to the fact that overall, the effects of acquiescence on students' teaching quality ratings were rather small.

The widest differences between fifth- and eighth-grade students occurred with regard to the negatively worded items in the *clarity of instruction* and *content relevance* scales. Ratings provided by fifth graders on these items were even more strongly affected by acquiescence than the positively worded items, whereas acquiescence was of less importance in ratings provided by eighth-grade students. In addition, extreme responding changed in accordance with the items' keying direction, indicating that some of the fifth graders expressed extreme agreement on positively worded and extreme disagreement on negatively worded items (Table 5).

Table 4

Standardized Regression Coefficients for Response Styles in Predicting Student Ratings (Configural Response Style Model)

	Fifth Grade				Eighth Grade			
	Within Classroom		Between Classrooms		Within Classroom		Between Classrooms	
	NARS β (SE)	ERS β (SE)	NARS β (SE)	ERS β (SE)	NARS β (SE)	ERS β (SE)	NARS β (SE)	ERS β (SE)
Teacher Support								
tsup1	**0.15 (.03)***	**0.15 (.02)***	**0.16 (.04)***	0.03 (.04)	**0.10 (.03)***	0.01 (.03)	**0.20 (.05)***	-0.03 (.06)
tsup2	**0.11 (.03)***	**0.16 (.02)***	**0.11 (.04)***	0.01 (.04)	**0.08 (.03)***	0.01 (.03)	**0.19 (.06)***	-0.03 (.07)
tsup3	**0.12 (.03)***	**0.13 (.02)***	**0.19 (.04)***	-0.01 (.04)	**0.06 (.03)***	0.00 (.03)	**0.20 (.06)***	-0.03 (.06)
tsup4	**0.17 (.03)***	**0.10 (.02)***	**0.12 (.04)***	0.00 (.04)	**0.11 (.03)***	0.02 (.03)	**0.19 (.05)***	-0.06 (.06)
tsup5	**0.14 (.02)***	**0.12 (.02)***	**0.13 (.04)***	-0.01 (.04)	**0.08 (.03)***	-0.00 (.02)	**0.21 (.06)***	-0.03 (.07)
tsup6	**0.14 (.03)***	**0.13 (.02)***	**0.14 (.05)***	-0.06 (.04)	**0.07 (.03)***	0.01 (.03)	**0.22 (.06)***	-0.03 (.06)
tsup7	**0.13 (.03)***	**0.12 (.02)***	**0.18 (.04)***	-0.03 (.04)	**0.09 (.03)***	0.02 (.03)	**0.19 (.05)***	-0.01 (.06)
Clarity of instruction								
clari4	**0.16 (.03)***	0.01 (.03)	0.04 (.03)	0.00 (.03)	**0.08 (.03)***	-0.03 (.03)	**0.15 (.04)***	-0.03 (.05)
clari5	**0.10 (.03)***	**0.09 (.03)***	0.06 (.03)	-0.02 (.03)	**0.07 (.03)***	0.01 (.03)	**0.20 (.05)***	-0.02 (.05)
clari6	**0.08 (.03)***	**0.10 (.03)***	0.07 (.04)	-0.01 (.04)	**0.07 (.03)***	-0.01 (.03)	**0.22 (.05)***	-0.04 (.05)
clari1	**0.20 (.03)***	-0.03 (.03)	**0.10 (.04)***	0.03 (.03)	**0.04 (.03)**	0.01 (.03)	-0.07 (.04)	0.03 (.04)
clari2	**0.24 (.02)***	**-0.10 (.03)***	**0.09 (.04)***	0.04 (.04)	**0.07 (.03)***	-0.05 (.03)	**-0.10 (.05)***	0.03 (.06)
clari3	**0.25 (.03)***	**-0.11 (.03)***	**0.09 (.04)***	0.01 (.04)	**0.09 (.03)***	-0.04 (.03)	**-0.12 (.05)***	0.02 (.06)
Content relevance								
conre1	**0.22 (.03)***	0.02 (.02)	**0.14 (.04)***	-0.04 (.03)	**0.08 (.03)***	-0.00 (.02)	0.02 (.05)	0.00 (.04)
conre3	**0.28 (.02)***	**-0.06 (.02)***	**0.15 (.04)***	-0.02 (.04)	**0.11 (.03)***	-0.05 (.03)	**-0.12 (.04)***	0.05 (.04)
conre4	**0.26 (.03)***	**-0.05 (.02)***	**0.23 (.03)***	-0.04 (.03)	**0.12 (.03)***	-0.06 (.03)	0.02 (.05)	0.06 (.05)
conre5	**0.25 (.03)***	**0.16 (.04)***	**0.16 (.04)***	-0.03 (.04)	**0.08 (.03)***	-0.03 (.03)	0.03 (.05)	0.03 (.04)

Note. β = standardized regression coefficient; *SE* = standard error; *$p < .05$.

Table 5

Mean Standardized Regression Coefficients for Response Styles in Predicting Student Ratings (Metric Response Style Model)

	Fifth Grade		Eighth Grade	
	NARS β (SE)	ERS β (SE)	NARS β (SE)	ERS β (SE)
Teacher Support				
tsup1	**0.14 (.02)***	**0.09 (.02)***	**0.15 (.03)***	-0.00 (.04)
tsup2	**0.09 (.01)***	**0.09 (.02)***	**0.14 (.04)***	0.00 (.04)
tsup3	**0.13 (.02)***	**0.07 (.02)***	**0.13 (.03)***	-0.01 (.04)
tsup4	**0.14 (.02)***	**0.05 (.02)***	**0.15 (.03)***	-0.02 (.04)
tsup5	**0.12 (.02)***	**0.06 (.02)***	**0.15 (.04)***	-0.01 (.04)
tsup6	**0.12 (.02)***	**0.06 (.02)***	**0.14 (.03)***	-0.00 (.03)
tsup7	**0.13 (.02)***	**0.05 (.02)***	**0.14 (.03)***	0.01 (.03)
Clarity of instruction				
clari4	**0.11 (.02)***	-0.01 (.02)	**0.11 (.02)***	-0.03 (.03)
clari5	**0.08 (.02)***	0.04 (.02)	**0.13 (.02)***	0.01 (.03)
clari6	**0.06 (.02)***	**0.05 (.02)***	**0.14 (.03)***	-0.02 (.03)
clari1	**0.17 (.02)***	**-0.03 (.02)***	-0.01 (.02)	0.01 (.03)
clari2	**0.18 (.02)***	**-0.06 (.02)***	-0.01 (.03)	-0.03 (.03)
clari3	**0.19 (.02)***	**-0.08 (.02)***	-0.01 (.03)	-0.02 (.03)
Content relevance				
conre1	**0.18 (.02)***	-0.01 (.02)	0.05 (.03)	-0.01 (.02)
conre3	**0.23 (.02)***	**-0.05 (.02)***	0.01 (.03)	-0.02 (.03)
conre4	**0.25 (.02)***	**-0.05 (.02)***	**0.08 (.03)***	-0.01 (.03)
conre5	**0.22 (.02)***	-0.03 (.02)	**0.06 (.02)***	-0.01 (.02)

Note. The regression coefficients reflect the average impact of students' individual and the class-mean level of acquiescence on the teaching quality items for fifth- and eighth-grade students.

Changes in Intercorrelations Between Teaching Quality Factors after Controlling for Acquiescence (Hypothesis 3)

We used scales with varying directions of item keying to test whether acquiescence would inflate the correlations between factors that were keyed in the same direction (i.e., *teacher support* and *clarity of instruction*) and deflate the correlations between factors that were keyed in opposite directions (i.e., *teacher support* and *content relevance*; *clarity of instruction* and *content relevance*). We compared correlation coefficients from the partial Metric Teaching Quality Model with correlation coefficients from the partial Metric Response Style Model in which we controlled for acquiescence. Model fits are displayed in Table 2, and correlations between the teaching quality factors are displayed in Table 6.

The largest changes occurred in fifth graders' ratings, whereas the correlations for eighth grade students' ratings remained largely unchanged. As expected, partialling out the variance due to the NARS and ERS indexes strengthened the correlations between oppositely keyed constructs so that they became more negative. In fifth grade, the correlation between *teacher support* and *content relevance* even turned from positive to negative (from r = .08 to r = -.05; eighth grade: from r = -.04 to r = -.07). For *content relevance* and *clarity of instruction,* the correlation increased from r = -.05 to r = -.14 (eigth grade: from r = -.08 to r = -.11). For constructs that were keyed in the same direction (i.e., *teacher support* and *clarity of instruction*), the strength of the positive correlations was slightly attenuated after controlling for acquiescence (fifth grade: from r = .52 to r = .49; eighth grade: from r = .67 to r = .66).

Table 6

Intercorrelations and Mean Ratings of Teaching Quality Factors without the modeling acquiescence (Teaching Quality Model) and including the modeling of acquiescence (Response Style Model)

	without modeling acquiescence	modeling acquiescence	Difference
Factor Intercorrelations[a]			
Fifth grade			
Teacher support	.52* [.46, .58]	.49* [.43, .56]	.03
Clarity of instruction			
Teacher support	.08* [.01, .15]	-.05 [-.12, .02]	.13
Content relevance			
Clarity of instruction	-.05 [-.12, .02]	-.14* [-.20, -.08]	.09
Content relevance			
Eighth grade			
Teacher support	.67* [.61, .73]	.66* [.61, .73]	.01
Clarity of instruction			
Teacher support	-.04 [-.12, .04]	-.07 [-.16, .01]	.03
Content relevance			
Clarity of instruction	-.08* [-.16, -.01]	-.11* [-.19, -.03]	.03
Content relevance			
Factor Mean-Level Differences[b]			
Teacher support	-.33* (0.05)	-.21* (0.04)	.12
Clarity of instruction	.05 (0.04)	.09* (0.04)	.04
Content relevance	-.35* (0.04)	-.27* (0.05)	.08

Note. [a]Factor Intercorrelations were drawn from the Metric Teaching Quality Model and the Metric Response Style Model. Teacher support was positively worded, content relevance was negatively worded, clarity of instruction was balanced (i.e., half positively and half negatively worded); values in square brackets indicate the 95% confidence interval for each correlation coefficient; *$p < .05$; [b]Mean-level differences were drawn from the partial Scalar Teaching Quality Model and the Scalar Response Style Model and were calculated by substracting the factor mean in fifth grade from the factor mean in eighth grade.

Age-related Differences in Students' Mean Ratings of Teaching Quality after Controlling for Acquiescence (Hypothesis 4)

Next, we examined the extent to which mean-level differences in fifth- and eighth-grade students' ratings of teaching quality can be explained by mean differences in students' acquiescent response style. Comparing factor means across two groups of respondents requires scalar invariance, which we tested for by constraining the item intercepts to equality across fifth- and eighth-grade students in both model in which we did not control for acquiescence (Teaching Quality Model) and in the model in which we accounted for acquiescence (Response Style Model). With a subset of the three negatively worded items on the *clarity of instruction* factor freed, partial scalar invariance was established in both models, because the differences in the fit indices between the partial scalar invariance model and the respective partial metric model were small (Teaching Quality Model [Model d vs. Model c in Table 2]: $\Delta\chi^2 = 122.89$, $\Delta df = 11$, $p < .05$; ΔRMSEA = .001; ΔCFI = -.005; ΔSRMR = .001; Response Style Model [Model h vs. Model g in Table 2]: $\Delta\chi^2 = 68.53$, $\Delta df = 11$, $p < .05$; ΔRMSEA = .000; ΔCFI = -.003; ΔSRMR = .000).

Next, we compared mean-level differences between fifth- and eighth-grade students' teaching quality ratings in the model in which we did not control for acquiescence (partial Scalar Teaching Quality Model) to mean-level differences in the model in which we accounted for acquiescence (partial Scalar Response Style Model). In the partial Scalar Teaching Quality Model, there were substantial differences between fifth- and eighth-grade students' mean ratings of teaching quality. Specifically, fifth-grade students evaluated teaching quality more positively than eighth-grade students. The mean difference was $\Delta M = -0.33$ ($SE = 0.05$, $p < .05$) for *teacher support*, and $\Delta M = -0.35$ ($SE = 0.04$, $p < .05$) for *clarity of instruction*. In line with our hypothesis, controlling for students' acquiescent response style led to decreases in the mean differences between fifth- and eighth-grade students' ratings. The reduced mean-level differences in the partial Scalar Response Style Model were $\Delta M = -0.21$ ($SE = 0.04$, $p < .05$) for *teacher support,* and $\Delta M = -0.27$ ($SE = 0.04$, $p < .05$) for *clarity of instruction* (see Table 6). With regard to *content relevance* (negatively worded), the mean-level difference *(*$\Delta M = 0.05$; $SE = 0.04$, $p = .178$) became significant after partialling out the variance due to acquiescence ($\Delta M = 0.09$, $SE = 0.04$, $p < .05$). The results indicate that mean differences in teaching quality ratings between fifth- and eighth-grade students can—at least to some degree—be attributed to differences in students' response styles.

Discussion

Acquiescence has been shown to interfere with the assessment of self-reported data (e.g., Baumgartner & Steenkamp, 2001; Rammstedt & Farmer, 2013). In contrast, rather little is known about the impact of acquiescence on student questionnaires concerning teaching quality, in which students evaluate their own behavior as well as the behavior of others (i.e., teachers' actions in the classroom). Given that student ratings are increasingly used for evaluation purposes within research and practice (Fauth et al., 2020), the impact of acquiescence on student ratings of teaching quality warrants further attention. In the present study, we examined acquiescence in student ratings of three teaching quality factors (teacher support, clarity of instruction, and content relevance) across two age cohorts (fifth- and eighth-grade) of German vocational track students. Specifically, we used an index-based method to measure acquiescence and disacquiescence while additionally controlling for an extreme response style.

Across both age cohorts, acquiescence had the highest and most consistent effects on student ratings of teaching quality. As expected, acquiescence was more pronounced in fifth than in eighth grade. In contrast, extreme responding affected only ratings by fifth-grade students and only to a smaller degree. Comparing students' teaching quality ratings before and after controlling for acquiescence revealed minor changes in the intercorrelations of the teaching quality factors. After controlling for acquiescence, mean differences in teaching quality ratings between fifth- and eighth-grade students were clearly reduced but still evident. This result indicates that differences in teaching quality ratings provided by fifth- and eighth-grade students can—to some degree—be explained by age-related differences in acquiescent responding. Overall, the effects of acquiescence on student ratings were rather small and primarily affected the ratings provided by younger students (i.e., fifth graders).

Students' Acquiescence Tendencies

Most commonly, acquiescence has been conceptualized as individual differences in use of the response scale. However, the results of our study suggest that acquiescence differs across both individual students within the same classroom as well as across classrooms. That is, acquiescence should not only be regarded as a characteristic of individual students, but also as a characteristic of the classroom as a whole. Most importantly, our results demonstrated that acquiescence accounted for variance in students' teaching quality ratings both within and across classrooms. This finding

calls into question the assumption that individual differences in acquiescence are counterbalanced at the classroom level (i.e., students' shared perceptions). In fact, there are few reasons to assume that acquiescent and non-acquiescent students are equally distributed across classrooms or that the response behavior of a particular student is subjected to a reference-group effect (i.e., changes in relation to the response behavior of the other students in the classroom), so that non-acquiescence ultimately offsets acquiescence at the classroom level. Instead, our results suggest that a class's average level of acquiescent responding rather reflects the accumulation of the individual response behavior of all students within the classroom.

In addition, our results showed differences in the predictive effects of acquiescence on teaching quality ratings between fifth- and eighth-grade students. Across both cohorts, acquiescence was predictive for the majority of items used to assess teaching quality, while ratings provided by fifth-grade students were additionally affected by an extreme response style. Specifically, we found that in fifth-grade students' ratings, negatively worded items were affected more strongly by acquiescence than positively worded items and extreme responding varied in accordance with the items' keying direction. These findings are of practical but also theoretical importance:

First, our results indicate that negatively worded items seemed to be problematic for students in early secondary school (i.e., fifth grade). For example, substantial differences in the loading patterns for the balanced *clarity of instruction* scale between fifth- and eighth-grade students emerged, and controlling for acquiescence resulted in only minor corrections of the negatively worded items' loading pattern in the sample of fifth graders. In line with previous research (e.g., Gehlbach, 2015), it appears that negatively worded items are especially difficult for younger students to understand and might be affected by factors other than acquiescence (Weijters et al., 2009; Gehlbach, 2015). For example, it has been argued that answering negatively worded items is cognitively more demanding and exacerbates issues with item interpretation (Swain et al., 2008; Wong et al., 2003). In support of this assumption, we found that in fifth grade, the items' keying direction seemed to interact with students' extreme response tendencies because the loadings of the regression coefficients for the ERS index varied in accordance with the items' keying direction. This finding was surprising because, based on the definitions of response styles given in the literature, we expected to find a consistent tendency to choose extreme positions regardless of item content or keying direction. Instead, assigning ratings consistent with the way

the items are stated may require respondents to pay more attention to the item content than responding in a careless or inattentive way. In this regard, extreme responding might reflect a response strategy other than acquiescence, such as a "confirmatory response strategy", a positive test strategy involving the activation of beliefs to support rather than to reject a statement or question (Weijters et al., 2013), or socially desirable responding (e.g., Miller, 2012).

Second, even though acquiescence exhibited the most consistent effects on all teaching quality items, we found that in fifth grade, acquiescence impacted negatively worded items more strongly than positively worded items, indicating that acquiescence does not necessarily have identical effects on different types of scales (i.e., positively worded, negatively worded or balanced). Instead, it seems that students' response styles reflect an interaction between student (i.e., grade level) and item characteristics (i.e., keying direction), which has also been suggested for other domains (e.g., De Beuckelaer et al., 2010; Kam & Mayer, 2015; Kam & Zhou, 2015).

In summary, our findings advise against using negatively worded items in student evaluations of teaching quality, for example within the context of using balanced scales as a tool to control for acquiescence (e.g., Billiet & McClendon, 2000). In a balanced scale, the same construct is assessed with an equal number of positively and negatively worded items. The underlying assumption is that when the negatively worded items are reverse-coded, acquiescence to positively worded items will be cancelled out by acquiescence to negatively worded items, meaning that the mean score of a scale will not be affected (Billiet & Davidov, 2008; Billiet & McClendon, 2002). However, this applies only if the strength of acquiescence to positively worded items is equal to the strength of acquiescence to negatively worded items (Kam & Mayer, 2015), and our findings demonstrate that this is not necessarily the case. In addition, administering negatively worded items to young students (i.e., fifth grade) introduces several other problems that affect the validity of student teaching quality ratings.

Impact of Acquiescence on Students' Teaching Quality Ratings and Age-Related Differences in Acquiescence

Student ratings have often been criticized for being undifferentiated and influenced by factors other than teaching quality (e.g., response styles or teacher popularity; Fauth et al., 2014) that blur the factor structure and artificially increase the correlations between distinct teaching quality factors. Therefore, a further aim of the present study was to evaluate the impact of

acquiescence on the factor structure (i.e., factor intercorrelations and factor means) of teaching quality in fifth and eighth grade.

With regard to the negatively worded items, differences in the factor structure of the teaching quality model emerged between fifth- and eighth-grade students, resulting in only partial measurement invariance. Nevertheless, the findings of our study support even young students' ability to provide differentiated ratings of teaching quality. Overall, the magnitude of change in the factorial intercorrelations after controlling for acquiescence was rather small, especially when the constructs were keyed in the same direction (i.e., positively correlated) or the ratings were provided by older students (i.e., eighth grade). In fifth grade, however, we found that acquiescence affected negative intercorrelations between oppositely keyed constructs to a greater extent than positive intercorrelations between constructs with the same keying direction. Again, this finding is in line with other research on acquiescence (e.g., Kam & Mayer, 2015; Kam & Zhou, 2015) stating that acquiescence seems to interact with different types of scales (i.e., positively worded, negatively worded or balanced) and does not necessarily have identical effects on positively and negatively worded items. Overall, we consider the effects of acquiescence on the factorial structure of teaching quality as rather unproblematic for practical use. However, when using ratings provided by younger students (i.e., fifth grade), it should be kept in mind that negative correlations between constructs with opposite keying directions are likely to be underestimated, which may be of importance when investigating the internal structure of teaching quality and its relations to external constructs.

Lastly, we examined the degree to which age-related differences in acquiescence contribute to mean-level differences in teaching quality ratings between fifth- and eighth-grade students. Prior research has identified declines in students' perception of their school environment across their secondary school years (i.e., sixth to eighth grade; e.g., Booth & Gerard, 2014; Reddy et al., 2003; Wang & Dishion, 2012; Way et al., 2007). In line with previous research, we found that fifth graders assessed higher levels of teaching quality (i.e., teacher support and clarity of instruction) than eighth-grade students. After controlling for acquiescence, the mean differences in teaching quality between the two student cohorts were clearly reduced but still evident, suggesting that differences in teaching quality ratings provided by fifth- and eighth-grade students can—to some degree—be explained by age-related differences in response behaviour.

Despite the fact that acquiescence only accounted for a small share of the mean-level differences between fifth- and eighth-grade students, we think that differences in acquiescence

should be considered when drawing inferences based on student ratings of teaching quality, for example within the context of teacher or school evaluation and decision-making. Particularly when ratings are provided by students from different age cohorts, (mean-level) differences in ratings may not entirely reflect genuine differences in the level of teaching quality, but may be —at least to a certain degree— due to age-related differences in students' response behaviour.

Strengths and Limitations

In the present study, we used an an index-based method to provide a direct measure of acquiescence and estimated the effects of acquiescence at the item indicator level for three aspects of teaching quality. One advantage of our study was combining scales with different keying directions in order to investigate whether acquiescence has comparable effects on positively and negatively worded items, which allowed for a more detailed analysis of the impact of acquiescence on the factor structure of teaching quality. However, our study was not designed to systematically evaluate the interaction between students' acquiescence tendencies and scales with varying keying directions as a component of survey design. In this regard, an ideal study design would assess each construct with both positively and negatively worded items that are randomly assigned within classrooms. However, many existing teaching quality scales do not contain both positively and negatively worded items and creating positively and negatively worded items with identical content meaning is difficult to achieve (Weijters et al., 2010a).

Also noteworthy is that we calculated the index values on the basis of antithetical item pairs drawn from the BFI (selected by Soto et al., 2008) to avoid a conceptual overlap between the items used to assess acquiescence and those used to assess teaching quality (Weijters et al., 2010a, 2010b). Because response styles are, by definition, independent of specific item content, students' acquiescence tendencies quantified on an external item set are likely to generalize to their teaching quality ratings (Mottus et al., 2012). Nevertheless, it remains unclear whether the assumption of response consistency across both item sets (i.e., students' responded to the BFI and the teaching quality items in the same way) held or was violated in our study. In this regard, anchoring vignettes (King et al., 2004) could be an alternative approach to further validate the findings of the present study. When anchoring vignettes are used to assess response styles, respondents additionally rate several items or short texts representing different levels of the target construct (e.g., high or low level of teacher support; see He et al., 2017 for an example), and systematic differences in responses to the same vignettes across respondents are supposed to mainly reflect differences in response

styles. However, as pointed out by Bolt et al., (2014), even though anchoring vignettes can be a promising approach to increase response consistency, it is still difficult to ensure that respondents use the rating scale in the same way for the vignette questions and the assessment of the target construct.

One additional strengths of the present study was the assessment of acquiescence across two age cohorts. Nonetheless, it should be mentioned that the results of the present study are restricted to ratings provided by students from German vocational school tracks (i.e., lower, intermediate, and combined tracks) because students attending the highest secondary school track ('Gymnasium') did not participate in the study. The proportion of students with a migration background is usually higher in the vocational school tracks than in the academic school tracks, and especially fifth-grade students from vocational school tracks might have lower reading and text comprehension abilities than fifth-grade students from a higher school track. We therefore assume that students in the lower school tracks are more likely to rely on response styles than students in the highest, academic track, even though this has yet to be empirically investigated. However, because students' age was closely related to acquiescent responding, we would expect that acquiescence is relevant to consider in ratings provided by young students across all school tracks. In addition to age-related differences in students' acquiescent responding, teaching evaluations provided by students of different age cohorts might also be affected by response shifts that have been shown to occur in longitudinal assessments (e.g., Oort, 2005). For example, mean-level differences in fifth- and eighth-grade students' evaluations of teaching quality may also indicate that older students have more experience with their teachers and better knowledge of how to differentiate between different levels of teaching quality (e.g., recalibration of response categories, Oort 2005).

Finally, further response styles should be considered in student ratings of teaching quality. For example, taking into account a midpoint response style in the regression analysis was beyond the scope of our study because the 4-point scales did not offer a neutral middle category. Leaving out the middle category forced students to move to either the agreement or to the disagreement side of the rating scale, which could have been consequential in two ways: On the one hand, this might have resulted in an overestimation of the impact of acquiescence on students' teaching quality ratings. On the other hand, the impact of other response styles (e.g., a midpoint response style) might have been underestimated (Bolt et al., 2014).

Similarly, further research is warranted regarding the distinction between acquiescence and an extreme response style. Specifically, when all items were positively worded, it seemed that extreme responding reflected an "extreme acquiescence" tendency, and in fact acquiescence has often been conceptualized as a weak form of extreme responding (Van de Vijver & He, 2014). However, when the items were negatively worded, extreme responding obviously differed from acquiescence and presumably reflected a different underlying cognitive process. Because extreme responding is typically conceptualized on a bi-directional bias (e.g., preference for the extreme categories at both ends of the scale; Kam & Fan, 2017), it might be that the appearance of extreme responding in our data reflected a confirmatory response strategy or social desirability. Using a scale with a larger number of answer categories (as done by Weijters et al., 2010b; 2013) might allow for more differentiated results regarding the influence of acquiescence and better discrimination between acquiescence and an extreme response style.

Conclusion

Our study adds to the growing body of research on the use of student ratings to assess teaching quality. While it has been hypothesized that student ratings might be affected by response styles (e.g., Scherer et al., 2016; Spooren et al., 2012), a direct analysis of acquiescence in student ratings of teaching quality has not been provided so far. Overall, we consider the effects of acquiescence on student ratings of teaching quality as less problematic than one might assume and only relevant to consider when using ratings provided by younger students (e.g., in fifth grade) and when using negatively worded items. To further extend the generalizability of the present study's conclusions, we hope that future research will continue to validate our findings across other school contexts and cultures.

References

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology*, 110(8), 1066–1083. https://doi.org/10.1037/edu0000256

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: black-white differences in response styles. *Public Opinion Quarterly, 48*(2), 491. https://doi.org/10.1086/268845

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., … Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente* [Professional competence of teachers, cognitively activating mathematics instruction, and the development of students' mathematical literacy (COACTIV): Documentation of measurement instruments]. Max Planck Institute for Human Development.

Baumert, J., Roeder, P. M., Gruehn, S., Heyn, S., Koller, O., Rimmele, R., Schnabel, K. & Seipp, B. (1996). Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU) [Educational pathways and psychosocial development in adolescence]. In K.-P. Treumann, G. Neubauer, R. Moeller, & J. Abel (Eds.), *Methoden und Anwendungen empirischer pädagogischer Forschung* [Methods and applications of empirical educational research] (pp. 170–180). Waxmann.

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76*(3), 186–204. https://doi.org/10.1037/h0031474

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research, 36*(4), 542-562. https://doi.org/10.1177/0049124107313901

Billiet, J. B., & McClendon, M. J. (2000). Modeling Acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(4), 608–628. https://doi.org/10.1207/S15328007SEM0704_5

Bolt, D. M., Lu, Y., & Kim, J. S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods, 19*(4), 528–541. https://doi.org/10.1037/met0000016

Booth, M. Z., & Gerard, J. M. (2014). Adolescents' stage-environment fit in middle and high school: The relationship between students' perceptions of their schools and themselves. *Youth & Society, 46*(6), 735–755. https://doi.org/10.1177/0044118X12451276

Cabooter, E. F. K. (2010). The impact of situational and dispositional variables on response styles with respect to attitude measures. Doctoral dissertation, Ghent University.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187–212. https://doi.org/10.1177/0022022100031002003

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Costello, S., & Roodenburg, J. (2015). Acquiescence response bias - Yeasaying and higher education. *Australian Educational and Developmental Psychologist, 32*(2), 105–119. https://doi.org/10.1017/edp.2015.11

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10*(1), 3–31. https://doi.org/10.1177/001316445001000101

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity, 44*, 761–775. https://doi.org/10.1007/s11135-009-9225-z.

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*(1), 51–85. https://doi.org/10.1023/A:1011402608575

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2015). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *The Journal of Early Adolescence, 35*(5–6), 722–758. https://doi.org/10.1177/0272431614564059

Dumont, H., Neumann, M., Maaz, K., and Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. Internationale und nationale Befunde [The effect of student body composition on academic achievement: International and national evidence]. *Psychol. Erziehung Unterricht 60*, 163–183. https://doi.org/10.2378/peu2013.art14d

Eccles, J. S., & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Individual bases of adolescent development* (pp. 404–434). John Wiley & Sons Inc. https://doi.org/10.1002/9780470479193.adlpsy001013

Eccles, J. S., Lord, S., & Midgley, C. (1991). What are we doing to early adolescents? The impact of educational contexts on early adolescents. *American Journal of Education, 99*(4), 521-542. https://doi.org/10.1086/443996

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29,* 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., & Trautwein, U. (2020). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational Psychology, 112*(6), 1284–1302. https://doi.org/10.1037/edu0000416

Flavell, J. H., Miller, P. H., & Miller, S. A.(1993). *Cognitive developments*. Prentice-Hall.

Gehlbach, H. (2015). Seven survey sins. *The Journal of Early Adolescence, 35*(5–6), 883–897. https://doi.org/10.1177/0272431615578276

Gitomer, D. H., & Bell, C. A. (2013). Evaluating teaching and teachers. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology* (pp. 415–444). American Psychological Association.

Goldberg, L. R. (1963). A model of Item Ambiguity in Personality Assessment. Educational and *Psychological Measurement, 23*(3), 467–492. https://doi.org/10.1177/001316446302300305

Gruehn, S. (2000). *Unterricht und schulisches Lernen : Schüler als Quellen der Unterrichtsbeschreibung* [Teaching and learning: Students as raters of teaching quality]. Waxmann.

Gudjonsson, G. H. (1990). The relationship of intellectual skills to suggestibility, compliance and acquiescence. *Personality and Individual Differences, 11*, 227–231. http://dx.doi.org/10.1016/01918869(90)90236-K

Hattie, J. (2009). Visible learning: a synthesis of over 800 meta-analyses relating to achievement. *Routledge.* https://doi.org/10.1080/00071005.2011.584660

He, J., & van de Vijver, F. J. R. (2015). Effects of a general response style on cross-cultural comparisons: Evidence from the teaching and learning international survey. *Public Opinion Quarterly, 79,* 267-290. https://doi.org/10.1093/poq/nfv006

He, J., Buchholz, J., & Klieme, E. (2017). Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures. *Journal of Cross-Cultural Psychology, 48*(3), 319–334. https://doi.org/10.1177/0022022116687395

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3-4), 117–144. https://doi.org/10.1080/03610739208253916

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: *A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*, 99-114. https://doi.org/10.1007/s10869-011-9231-8

Jonkmann, K., Rose, N., & U. Trautwein (Eds.) (2013). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen – Abschlussbericht für die Länder Baden-Württemberg und Sachsen* [Tradition and innovation: Academic and psychosocial development in vocational track schools in the states of Baden-Württemberg and Sachsen]. Projektbericht an die Kultusministerien der Länder.

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The Case of Job Satisfaction. *Organizational Research Methods, 18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kam, C. C. S., & Zhou, M. (2015). Does Acquiescence affect individual items consistently? *Educational and Psychological Measurement, 75*(5), 764–784. https://doi.org/10.1177/0013164414560817

Kam, C. C.S., & Fan, X. (2017). Approaches to handling common response styles and issues in educational surveys. *Oxford Research Encyclopedia of Education*. Retrieved from http://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-90

Karing, C., Pfost, M. & Artelt, C. (2013). Is secondary school teacher judgment accuracy related to the development of students' reading literacy? In M. Pfost, C. Artelt & S. Weinert (Eds.), *The Development of Reading Literacy from Early Childhood to Adolescence. Empirical Findings from the Bamberg BiKS Longitudinal Study*. (pp. 279-310). University of Bamberg Press.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98,* 191-207. https://doi.org/10.1017/S000305540400108X

Knowles, E. S., & Condon, C. A. (1999). Why people say ‘‘yes’’: A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*, 379–386. https://doi.org/10.1037/0022-3514.77.2.379

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. https://doi.org/10.1037/a0032583

Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction, 18(*5), 468–482. https://doi.org/10.1016/J.LEARNINSTRUC.2008.06.008

Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological Assessment, 27*(4), 1301–1311. https://doi.org/10.1037/pas0000151

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*, 120–131. http://dx.doi.org/10.1016/j.cedpsych.2008.12.001

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*, 99–106. http://dx.doi.org/10.1111/j.1750-8606.2008.00048.x

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61-83. https://doi.org/10.1016/j.jrp.2013.09.008

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. Psychological Methods, 17, 437-455. https://doi.org/10.1037/a0028085

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44*(7), 1539–1550. https://doi.org/10.1016/j.paid.2008.01.010

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44* (Suppl 3), S69–S77. https://doi.org/10.1097/01.mlr.0000245438.73837.89

Messick, S. (1966). The psychology of acquiescence: An interpretation of research evidence. *ETS Research Bulletin Series,* i–44. https://doi.org/10.1002/j.2333-8504.1966.tb00357.x.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes toward mathematics before and after the transition to junior high school. *Child Development, 60*(4), 981. https://doi.org/10.2307/1131038

Miller, A. L. (2012). Investigating social desirability bias in student self-report surveys. *Educational Research Quarterly, 36*, 30–47.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge. https://doi.org/10.4324/9780203821961

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity, 42*(6), 779–794. https://doi.org/10.1007/s11135-006-9067-x

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). *PIRLS 2001 International report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Boston College.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267–316. http://dx.doi.org/10.2307/271070

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus user's guide: Statistical analysis with latent variables.* Eighth Edition. Muthén & Muthén.

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research, 14*(3), 587-598. https://doi.org/10.1080/10705511.2012.687671

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press.

Pfost, M., & Artelt, C. (2013). Reading literacy development in secondary school and the effect of differential institutional learning environments. In M. Pfost, C. Artelt & S. Weinert (Eds.), *The development of reading literacy from early childhood to adolescence. Empirical findings from the bamberg BiKS longitudinal study*. (pp. 229-277). University of Bamberg Press

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. https://doi.org/10.3102/0013189X09332374

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569. http://dx.doi.org/10.1146/annurev-psych-120710-100452

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment, 25*(4), 1137–1145. https://doi.org/10.1037/a0033323

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44*(4), 53–61. https://doi.org/10.1016/j.jrp.2009.10.005

Ray, J. J. (1984). Reinventing the wheel: Winkler, Kanouse, and Ware on acquiescent response set. *Journal of Applied Psychology, 69*(2), 353–355. https://doi.org/10.1037/0021-9010.69.2.353

Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology, 15*(1), 119–138. https://doi.org/10.1017/S0954579403000075

Scherer, R., & Gustafsson, J.-E. (2015). Student assessment of teaching as a source of information about aspects of teaching quality in multiple subject domains: an application of multilevel bifactor structural equation modeling. *Frontiers in Psychology, 6,* 1550. https://doi.org/10.3389/fpsyg.2015.01550

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7*, 1–16. https://doi.org/10.3389/fpsyg.2016.00110

Schiepe-Tiska, A. (2019). School tracks as differential learning environments moderate the relationship between teaching quality and multidimensional learning goals in mathematics. *Frontiers in Education, 4*(4). https://doi.org/10.3389/feduc.2019.00004

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from Ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718

Spooren, P., Mortelmans, D., & Thijssen, P. (2012). “Content” versus “style”: Acquiescence in student evaluation of teaching? *British Educational Research Journal, 38*(1), 3–21. https://doi.org/10.1080/01411926.2010.523453

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics, 41*(5), 481–520. https://doi.org/10.3102/1076998616646200

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. Jossey-Bass.

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research, 45*, 116-131. https://doi.org/10.1509/jmkr.45.1.116

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology, 85*(4), 723–735. https://doi.org/10.1037/0022-3514.85.4.723

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). T*he psychology of survey response.* Cambridge University Press.

Van de Vijver, F. and J. He (2014), Report on social desirability, midpoint and extreme responding in TALIS 2013. *OECD Education Working Papers*, No. 107, OECD Publishing. https://doi.org/10.1787/5jxswcfwt76h-en.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25,* 195–217. https://doi.org/10.1093/ijpor/eds021

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and

generalizability of domain-independent assessments. *Learning and Instruction, 28,* 1–11. https://doi.org/10.1016/j.learninstruc.2013.03.003

Wang, M., & Dishion, T. (2012). The Trajectories of adolescents' perceptions of school climate, deviant peer affiliation, and behavioral problems during the middle school years. *Journal of research on adolescence, 22*(1), 40-53. https://doi.org/10.1111/j.1532-7795.2011.00763.x

Way, N., Reddy, R., & Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological and behavioral adjustment. *American Journal of Community Psychology, 40*(3–4), 194–213. https://doi.org/10.1007/s10464-007-9143-y

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320–334. https://doi.org/10.1037/a0032121

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement, 34*(2), 105–121. https://doi.org/10.1177/0146621609338593

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The Stability of individual response styles. *Psychological Methods, 15*(1), 96–110. https://doi.org/10.1037/a0018721

Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence Over 8 Years. *Assessment, 23*(3), 279–291. https://doi.org/10.1177/1073191115583714

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research*. (pp. 281–324). American Psychological Association. https://doi.org/10.1037/10222-009

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67*(5), 555–561. https://doi.org/10.1037/0021-9010.67.5.555

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research, 30*, 72–91. https://doi.org/10.1086/374697

Wubbels, T., & Brekelmans, M. (2005). Two decades of research on teacher-student relationships in class. *International Journal of Educational Research, 43*(1–2), 6–24. https://doi.org/10.1016/j.ijer.2006.03.003

Wubbels, T., Brekelmans, M., & Hooymayers, H. (1993). Comparison of teachers' and students' perceptions of interpersonal teacher behavior. In T. Wubbels & J. Levy (Eds.), *Do you know what you look like? Interpersonal relationships in education* (pp. 64 – 80). Falmer Press

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. *Sociological Methodology, 30*(1), 165–200. https://doi.org/10.1111/0081-1750.00078

4 STUDY 2:

STUDENT DEVELOPMENT IN EARLY ADOLESCENCE: DOES TEACHING QUALITY SHAPE STUDENTS' ACADEMIC ACHIEVEMENT, ACADEMIC ENGAGEMENT, AND THEIR SOCIAL AND EMOTIONAL SCHOOL ADJUSTMENT?

Marder, J., Trautwein, U., Aldrup, K., Fauth., B. & Göllner, R. (2022). Student Development in Early Adolescence: Does Teaching Quality Shape Students' Academic Achievement, Academic Engagement, and Their Social and Emotional School Adjustment? *Manuscript submitted for publication.*

Abstract

How important is teaching quality for student development? Whereas there is sound evidence for associations with student academic achievement, much less is known about how teaching quality is associated with other important domains of student development (i.e., academic engagement, social and emotional school adjustment). Using a sample of $N = 3{,}880$ German lower secondary school students and their homeroom teachers, we provide a comprehensive analysis of how the trajectories of central aspects of students' academic engagement (e.g., academic identification) and their social (e.g., school belonging) and emotional (e.g., self-esteem) school adjustment relate to the trajectory of teaching quality from fifth to eighth grade. Multilevel longitudinal growth curve models performed at the student and classroom levels revealed significant mean-level changes in students' development. Overall, both teachers and students reported declines in teaching quality. Students experienced declines in their academic engagement, but adjusted to school quite well. Moreover, teaching quality buffered the downward trend in most aspects of students' academic engagement and promoted students' academic achievement at the classroom level. In contrast, students' individual perceptions of teaching quality (i.e., the student level) were most important for their social and emotional school adjustment, indicating that students' social and emotional school adjustment during adolescence is heavily driven by individual processes.

Keywords: teaching quality, academic achievement, academic engagement, school adjustment, multilevel longitudinal modeling

Student Development in the Years of Early Adolescence: Does Teaching Quality Shape Students' Academic Achievement, Academic Engagement, and Their Social and Emotional School Adjustment?

The adolescent years are a period of significant developmental change, during which many students experience struggles in their academic development and their social and emotional adjustment to school (Roeser et al., 2000). Given that these domains are important predictors of students' academic success (Furrer & Skinner, 2003; Gillen-O'Neel & Fuligni, 2013; Hughes et al., 2015; Morinaj & Hascher, 2019, professional development, and lifelong learning (National Research Council, 2012), schools face the challenge of promoting positive youth development.

Several theoretical accounts (e.g., Bronfenbrenner, 1979; Eccles & Midgley, 1989; Eccles et al., 1993; Ryan & Deci, 2000) claim that environmental contexts, such as the learning environment at school, affect student development across multiple domains, including their academic engagement and their social and emotional school adjustment. Arguably, the way teachers shape the classroom setting through the quality of their teaching can be expected to exert a strong influence on student development (e.g., Eccles & Midgley, 1989; Eccles et al., 1993). But how much of a difference does teaching quality really make for students' academic achievement and engagement as well as their social and emotional school adjustment? To date, longitudinal research investigating the role of teaching quality in student development during early adolescence is still scarce. In particular, comprehensive studies examining multiple relevant aspects of teaching quality (i.e., monitoring, teacher support) and student development (i.e., academic achievement, academic engagement, social and emotional school adjustment) at the same time are lacking. Moreover, the assessment of teaching quality was often restricted to individual student perceptions, rather than using classroom aggregates or teacher self-reports as additional indicators of teaching quality.

To gain a deeper understanding of the extent to which teaching quality matters for students' development, the present longitudinal study followed school classes made up of specific groups of students and their homeroom teachers over 3 years of lower secondary school (i.e., fifth to eighth grade). By considering the hierarchical data structure of students nested in classes and combining data from students and their homeroom teachers, the study provides a comprehensive examination

of the extent to which teaching quality shapes students' academic development and school adjustment over the years spanning early adolescence.

Student Development in Early Adolescence

The lower secondary school years coincide with the years spanning early adolescence, which is considered as a vulnerable phase for students' academic engagement as well as their social and emotional school adjustment (Eccles & Midgley, 1989). Extensive research has documented declines in students' academic engagement during the early adolescent years. For example, students have been shown to perceive school as less interesting and valuable during adolescence (Dotterer et al., 2009; Engels et al., 2017; Wang et al., 2015) which can be accompanied by increases in problem behaviors such as school truancy and misconduct (Archambault et al., 2009; Virtanen et al., 2021; Wang & Dishion, 2012). Moreover, adolescent students are at risk for experiencing struggles in their social school adjustment (e.g., the extent to which they feel socially connected to school; Gillen-O'Neel & Fuligni, 2013; Witherspoon & Ennett, 2011) and in their emotional school adjustment (e.g., their well-being and self-esteem; Morin et al., 2017; Way et al., 2007). These negative trajectories, in turn, constitute risk factors for students' long-term development because research has found that students' social and emotional functioning prevents withdrawal from school and is important for students' academic success, particularly during adolescence (e.g., Furrer & Skinner, 2003; Gillen-O'Neel & Fuligni, 2013; Hughes et al., 2015).

The obvious relevance of positive student development prompts two questions that are of high theoretical and practical interest (e.g., Bronfenbrenner, 1979; Eccles et al., 1993; Trautwein et al., 2015). First, how much does students' development differ as a consequence of their experiences of their learning environment, rather than being largely determined by factors outside the realm of schools? Second, if learning environments matter, what are the key characteristics that foster positive development? On a more general level, the available evidence indicates that the quality of learning environments indeed matters, with school environments that systematically align opportunities with adolescent students' developmental needs being more successful in supporting positive student development over time (Eccles et al., 1991; Roeser et al., 2000). For instance, research on concurrent or short-term longitudinal associations around school transitions (Benner, 2011; Evans et al., 2018; Longobardi et al., 2019; Roeser & Eccles, 1998) found adolescent students' academic and socio-emotional functioning in school to be associated with their experiences of the broadly defined school climate. In addition to that, a number of notable

longitudinal studies have documented declines in several aspects of the school climate (e.g., academic and social support from teachers and peers, school behavior management, opportunity for student autonomy; Reddy et al., 2003; Wang & Dishion, 2012; Wang & Eccles, 2012; Way et al., 2007) and in students' perceptions of teacher-student relationships (Engels et al., 2016; Hughes & Cao, 2018) across the lower secondary school years, with negative consequences for adolescent students' academic success, school engagement, and socio-emotional functioning.

Interestingly, when assessing the extent to which students' experiences of their learning environments matter for their development, a large body of research has assessed characteristics at the school level (e.g., in terms of students' overall experiences with their teachers or school climate aspects), rather than focusing on their experiences with specific teachers or classroom processes. From an environmental perspective (Bronfenbrenner, 1979), students' classrooms are the most immediate learning environment within the school context and, therefore, might be particularly important for understanding student development. More specifically, the processes happening therein—such as the extent to which a specific teacher keeps the classroom well organized and supports students' learning and their social and emotional functioning in the classroom—should be expected to be among the most proximal factors in the school context that impact student development (e.g., Eccles & Roeser, 2011). Thus, focusing on teaching quality (e.g., classroom management, monitoring, emotional and academic support) may help to more fully understand the processes related to academic development and their social and emotional school adjustment during early adolescence. By longitudinally linking multiple aspects of teaching quality to both students' academic development and socio-emotional school adjustment within the same sample, the present study offers a comprehensive investigation of whether there are systematic differences in student development during early adolescence that can be related to students' classroom experiences.

Teaching Quality

There is a large number of conceptual models that describe key aspects of successful teaching (e.g., Creemers & Kyriakides, 2006; Ferguson, 2014; Hamre & Pianta, 2010; Klieme et al., 2009), including an organizational and a supportive component. In the present research, we focus on classroom management and teacher support as two overarching dimensions featuring prominently in several modern conceptions of teaching quality (see Hamre & Pianta, 2010; Klieme et al., 2009) and which arguably should be associated with students' academic achievement, academic engagement, and their social and emotional school adjustment. Each dimension covers a

variety of specific behavioral indicators that have been shown to be relevant for students' academic achievement and learning-related outcomes (e.g., Kunter et al., 2013; Praetorius et al., 2018; Wagner et al., 2013), but can also be expected to affect students' social and emotional school adjustment because they respond to adolescents' needs for competence, autonomy, and relatedness (for an overview, see Emmer & Gerwels, 2006; Léon et al., 2017).

Classroom management broadly refers to the organization of the classroom and the management of time, attention, and students' behavior. The amount of disturbances in the classroom and the extent to which the teachers monitor what is happening in the classroom are frequently used to assess classroom management. Classroom management aims at establishing the conditions for teaching and learning in the classroom and at maximizing the productive use of learning time and has most consistently been associated with students' achievement gains (Aldrup et al., 2018; Scherer et al., 2016; Wagner et al., 2016). Moreover, classroom management facilitates the development of cognitive and behavioral self-control (Emmer & Strough, 2001; McCaslin et al., 2006; Rimm-Kaufman et al., 2015), which are seen as key competences for success in all areas of schooling (Hamre & Pianta, 2010; Raver, 2004).

Teacher support refers to the nature and the quality of the teachers' and students' interactions with one another, regarding emotional and instructional matters. Teacher emotional support addresses the overall emotional tone in the classroom that results from interactions between teachers and students and includes a positive teacher-student relationship, respectful and caring teacher behavior, and teachers' regard for students. Since adolescent students are particularly sensitive to their social acceptance, positive relationships with teachers and the feeling of being cared for are key elements of students' feelings of social relatedness and competence (Furrer & Skinner, 2003). Teacher academic support refers to teachers' care for students with respect to instructional matters, such as providing constructive feedback and helping students with content-related challenges. Support from the teacher in overcoming learning-related difficulties supports students' cognitive development, enhancing their feelings of competence and autonomy (Brophy, 2006; Niemiec & Ryan, 2009).

Despite this strong theoretical rationale, the current understanding of how teaching quality is associated with the development of students' academic achievement, academic engagement, and their social and emotional school adjustment during early adolescence remains limited for two reasons: (a) the paucity of longitudinal studies in teaching quality research and (b) the neglect to

fully consider the multilevel nature of student ratings of teaching quality in prior research. In the following section, we address these two points in more detail.

Longitudinal Development of Teaching Quality in Lower Secondary School

There is initial evidence that teaching quality is variable over time, for example when teaching students at different grade levels (Fauth et al., 2020; Gaertner & Brunner, 2018). In multi-cohort studies relying on student or observer ratings, teaching quality was systematically lower in higher grade levels (e.g., eighth grade) compared to lower grade levels (e.g., fifth grade; Aldrup et al., 2018; Grossman et al., 2014; Marder et al., 2021; Mihaly & McCaffrey, 2014). In addition, the literature on mean-level changes in teacher-student relationships reports decreases in relationship quality and teacher support over the first year of lower secondary school (Lazarides et al., 2019; Opdendakker et al., 2012) and over the course of lower secondary school, from both the student (Wu & Hughes, 2014) and the teacher perspective (Hughes & Cao, 2018). However, little research has addressed the question of whether such changes in teaching quality aspects are related to students' academic development and school adjustment across early adolescence. Given that adolescence is a time during which students are particularly in need of supportive relationships, experience challenges related to identity development and desire more autonomy and less control (Roeser et al., 2000; Wentzel, 2003), shifts in the extent to which students feel socially connected, competent and supported by their teachers in the classroom can be expected to have a strong influence on their academic development and school adjustment (Pianta & Allen, 2008).

When examining teaching quality in lower secondary school, existing research has mostly focused on how students perceive their classroom environment. Student ratings are particularly appealing, because students have extensive experience with their teacher and can provide valuable information that may not be accessible from an outside perspective (Feldlaufer et al., 1988; Fraser & O'Brien, 1985). That is, students' own experiences and perceptions of their teachers are expected to be more strongly related to their feelings, beliefs and behavior than more objective accounts of the classroom and thus may become important predictors of students' academic and psychological development over time (Eccles & Roeser, 2011; Maehr & Midgley, 1991; Roeser et al., 1996). Prior research has suggested using information by multiple reporters (e.g., student ratings and teacher self-reports) to ensure that findings are generalizable beyond students' perceptions alone (Aldrup et al., 2018; Kane & Staiger, 2012; Kunter & Baumert, 2006; Marder et al., 2021). To date, there is no longitudinal multi-reporter research on teaching quality across lower secondary school.

When combining two different sources of information (e.g., student ratings and teacher self-reports), one perspective can add valuable information missing from the other perspective, making it possible to examine whether changes in teaching quality across lower secondary school are reflected in ratings provided by both students and teachers. Hence, there is a need for a longitudinal investigation of teaching quality that combines information from multiple reporters to better understand the processes related to student development in the school context during early adolescence.

Multilevel Issues in Teaching Quality Data

Examining the extent to which students' experiences of their classroom environment account for differences in students' development is related to the methodological issue of appropriately modeling such data (e.g., Lüdtke et al., 2009; Raudenbush & Bryk, 2002). Because students are nested within classes, student reports of teaching quality inform about both the average level of teaching quality experienced by all students in the classroom (classroom level) as well as the idiosyncratic (i.e., non-shared) experiences of individual students (student level). This distinction is particularly important because students' individual perceptions of their teacher might diverge from the overall level of perceived teaching quality in the classroom due to heterogeneity among students. For example, a particular student might have a distant relationship with the teacher and report a lower level of perceived support, despite an overall high level of support in the classroom (or vice versa). From this conceptual point of view, student ratings of teaching quality reflect information about both the extent to which students within the same classroom differ in their perceptions of teaching quality (student level) and the extent to which the average level of teaching quality differs across classrooms (classroom level). Importantly, previous research has shown that both students' individual and shared perceptions of teaching quality are related to student learning outcomes (Aldrup et al., 2018; Downer et al., 2015; Göllner et al., 2018). Thus, applying a multi-reporter and multilevel examination can help to provide a comprehensive investigation of whether there are systematic differences in the development of adolescent students' academic achievement, academic engagement, and their social and emotional school adjustment – within and across classrooms – that can be traced back to their classroom experiences.

The Present Study

A large part of research on teaching quality research has relied on cross-sectional or short-term longitudinal examinations of teaching quality. However, such designs do not allow to fully understand the dynamic nature of students' experiences with their teachers within the classroom across longer time spans and the consequences for student development, particularly the years spanning early adolescence. Addressing this research gap, the present study provides a comprehensive longitudinal multilevel examination of critical domains of student development (i.e., academic achievement, academic engagement, social and emotional school adjustment) and teaching quality and combines information from the student and teacher perspectives.

Using a large sample of German lower secondary school students and their homeroom teachers, we examine the co-development of student development and teaching quality, from both the student and the teacher perspective. In the German school system, each class is assigned a homeroom teacher who serves both an educational and an advisory function. That is, the homeroom teacher teaches at least one subject to the class, but also spends additional time with his/her class for counseling, dealing with class organizational matters, and addressing students' overall learning development and personal needs or concerns (Aldrup et al., 2018). In addition, homeroom teachers typically retain their classes for several school years. Thus, the analyses allow for important insights into the extent to which teaching quality matters for student development.

Our first research question addresses the trajectories of critical domains of students' development (i.e., academic achievement, academic engagement, social and emotional school adjustment) across lower secondary school. We expected students' academic achievement as measured via a standardized achievement test to increase over time. In contrast, we expected decreases in students' academic engagement as well as social and emotional school adjustment over time.

Our second research question focuses on the trajectories of teaching quality (i.e., classroom management and teacher support) across lower secondary school, from both the student and the teacher perspective. Based on prior research on school climate trajectories and initial evidence from teaching quality research, we expected both academic and emotional support by teachers to decline during lower secondary school. However, due to the lack of prior studies on changes in teaching quality, we made no predictions regarding the direction of change in classroom management.

Third, our most central research question addressed the co-development of student development (i.e., academic achievement, academic engagement, social and emotional school adjustment) and teaching quality over time, at both the student and classroom levels and from both the student and teacher perspectives. We expected positive associations between the trajectories of student development and teaching quality. Moreover, we expected students' academic achievement, academic engagement, and their social and emotional school adjustment to be more strongly related to students' perceptions of teaching quality than to teacher-reported teaching quality (Wagner et al., 2016). We made no predictions about whether the associations would be more pronounced at the student or the classroom level.

Method

Sample

The present study used data from the large-scale longitudinal study Tradition and Innovation in School Systems Study (TRAIN; Jonkmann et al., 2013) which investigated students' developmental pathways in lower secondary school during the period of early adolescence, specifically in vocational school tracks. In the German school system, students transition to lower secondary school at the end of fourth grade. Based on their prior achievement, students are sorted into either an academic- or vocational-track school. High-achieving students generally enroll in an academic-track school, whereas vocational-track schools typically prepare students to enter vocational training rather than higher education. The vocational track can be further divided into intermediate-track schools ('Realschule') and lower-track schools ('Hauptschule'). Additionally, some German federal states offer comprehensive schools combining the intermediate and vocational tracks. The underlying rationale for grouping students based on their abilities is to form more homogeneous learning environments tailored to meet the instructional and emotional needs of different groups of students (Maaz et al., 2008).

The TRAIN study encompasses four waves of data collection that took place once during the first six weeks of one school year. Each wave of data collection was spread over two days. The same cohort of students and their homeroom teachers were first assessed in fifth grade, after students transitioned from primary to lower secondary school, and then followed through eighth grade. The pooled sample consisted of N = 3,880 students (45.2% female) from 136 secondary school classes and contained all individuals who provided information at a minimum of one

measurement wave. In fifth grade, $N = 2{,}894$ students (130 classes) participated, while in 6th grade $N = 2{,}936$ students (131 classes), in 7th grade $N = 2{,}993$ students (132 classes), and in eighth grade $N = 3{,}060$ students (135 classes) participated. The average class size in fifth grade was $M = 22.26$ students per class ($SD = 4.7$; 6th: $M = 23.95$, $SD = 5.4$; 7th: $M = 26.67$, $SD = 5.5$; 8th: $M = 28.10$, $SD = 6.1$). Students' mean age in fifth grade was $M = 10.71$ years ($SD = 0.62$). 27.2% of the students reported having a migrant background, meaning that at least one parent or the students themselves had been born outside Germany.

The students' current homeroom teacher also participated at each measurement wave and provided information about their homeroom class and their teaching. Homeroom teachers typically retain their classes for at least two years, although it is possible for some homeroom teachers to accompany their class for a shorter or longer period of time. A total of $N = 126$ homeroom teachers participated in the study. On average, homeroom teachers spent $M = 11.13$ hours ($SD = 5.33$) per week within their homeroom class, and each class experienced an average of $M = 1.2$ homeroom teacher changes as they moved from fifth to eighth grade ($SD = 0.97$).

The Ministries of Education and Cultural Affairs of the German federal states of Baden-Württemberg and Saxony reviewed the study in 2008 and approved the instruments and reassessment of students over time. Students' participation in the TRAIN study was voluntary and required active parental consent.

The TRAIN data set has been used in a number of prior studies, including a study that examined the longitudinal development of students' self-esteem (Grade 5 to Grade 8; Wagner et al., 2017) and a study on students' math interest and perceived teacher support (Grade 5 to Grade 6; Lazarides et al., 2019). However, the present study is clearly different from prior research with the TRAIN data because none of the prior studies provided a comprehensive analysis of the co-development of students' academic achievement, academic engagement, their social and emotional school adjustment, and teaching quality during lower secondary school.

Measures

In the TRAIN study, students provided a variety of information describing their academic learning and socio-emotional functioning in lower secondary school. To provide a comprehensive analysis of the development of students' academic engagement as well as their social and emotional school adjustment, we used a range of variables that are considered key aspects of each domain. Descriptive results and Cronbach's α at each measurement wave are presented in Table 1. All items are included in Appendix A.

Academic Achievement

To examine the students' academic achievement, we used standardized achievement tests for math and German. The German test included several short texts and related questions and addressed students' reading comprehension. The mathematics test covered grade-specific content such as arithmetic rules, the metric system, or linear equations. The tests showed good reliability (Cronbach's $\alpha > .70$) in both domains. For the statistical analyses, we used weighted likelihood estimates of students' achievement scores (Aldrup et al., 2018).

To investigate the relationship between students' math or German achievement and teaching quality in that subject, we formed an achievement variable whose value depended on whether students' homeroom teacher taught their class math or German. If the homeroom teacher taught math, we used students' math achievement score. If the homeroom teacher taught German, we used students' German achievement score. If the homeroom teacher taught both subjects, we averaged both achievement scores. All analyses including students' academic achievement were based on the subsample of students whose homeroom teacher taught math, German or both (n = 2649 students from 90 classes and their corresponding homeroom teachers).

Academic Engagement

To assess students' academic engagement, we included variables that correspond to two critical subcomponents of academic engagement (Fredricks et al., 2004): students' behavioral engagement (i.e., truancy) and emotional engagement (i.e., value beliefs in math and German, academic identification, school satisfaction).

Subject Value Beliefs (Math/German). Three items assessed the extent to which students enjoyed and valued doing well in math/German (e.g., 'It's important to me to be good at math/German'). The 4-point scale ranged from 1 = *strongly disagree* to 4 = *strongly agree*. Across the four measurement waves, scale reliability ranged from $\alpha_{student}$ = .67 to .74. Students' value beliefs were calculated depending on whether their homeroom teacher taught them math or German. If the homeroom teacher taught both subjects, we averaged both values.

Academic Identification. Six items assessed the extent to which students found it important to be someone who does well at school (e.g., 'It's particularly important to me to do well at school'). Students answered on a 4-point scale ranging from 1 = *not at all important* to 4 = *really important*; $\alpha_{student}$ = .87 to .89. The scale was newly developed for the TRAIN study.

School Satisfaction. Four items developed by Baumert et al. (1997) measured the extent to which students liked being at school and going to school (e.g., 'I enjoy doing my tasks at school'). The 4-point scale ranged from 1 = *strongly disagree* to 4 = *strongly agree;* $\alpha_{student}$ = .68 to .73.

Truancy. Six types of truancy per school year were assessed on a scale ranging from 1 = *never,* 2 = *two or three times*, 3 = *three or four times*, 4 = *five times or more*; $\alpha_{student}$ = .93 to .95. The items were adapted from Kittl et al., (2005).

Social School Adjustment

To assess students' social school adjustment, we focused on the extent to which students feel socially connected to school.

Social Relatedness. Four items assessed students' feelings of being socially connected to others in school (e.g., 'My classmates quite like me'). The 4-point scale ranged from 1 = *strongly disagree* to 4 = *strongly agree;* $\alpha_{student}$ = .70 to .83. This scale has been shown to be both reliable and valid in prior research (e.g., Köller et al., 2010).

School Belonging. Six items assessed the extent to which students feel that they are part of the school community, including their feelings of being liked and accepted by peers within the school context (e.g., 'My school is a place where I can make friends easily'). The 4-point scale ranged from 1 = *strongly disagree* to 4 = *strongly agree;* $\alpha_{student}$ = .70 to .83. This scale was adapted from the PISA 2012 cycle (Mang et al., 2012).

Emotional School Adjustment

To assess students' emotional school adjustment, we used students' emotional well-being and self-esteem.

Self-esteem. Students' self-esteem was assessed with the KINDL-R self-esteem subscale (Ravens-Sieberer et al., 2013). The scale consists of four items (e.g., 'In the last week, I was proud of myself') and ranges from 1 = *never* to 5 = *always*; $\alpha_{student}$ = .71 to .78).

Emotional Well-being. Students' emotional well-being during the last week was assessed with the KINDL-R emotional well-being subscale (Ravens-Sieberer et al., 2013), consisting of four items (e.g., 'In the last week, I felt lonely'). The scale ranged from 1 = *never* to 5 = *always;* $\alpha_{student}$ = .66 to .72.

Teaching Quality

Teaching quality was assessed from the student and teacher perspectives using items with parallel content. Students were asked to rate their homeroom teacher's lessons, while teachers were asked to self-evaluate their teaching in their respective homeroom class. All items were rated on a 4-point scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. Descriptive results (i.e., means, standard deviations, range, scale reliability and ICCs for the student and teacher data are presented in Table 1 and Table 2, respectively). The scales were adapted from Baumert et al. (2009).

Classroom Management. Two scales were used to assess classroom management. First, *disturbances* (four items) refers to students' disruptive behavior in class (e.g., student perspective: 'In this class, we rarely chatter loudly'; $\alpha_{student}$ = .72 to .87; teacher perspective: e.g., 'In this class, students rarely chatter loudly'*;* $\alpha_{teacher}$ = .85 to .91).

Second, *monitoring* (four items) assessed the extent to which the homeroom teacher notices and prevents disruptions and students being inattentive in class (e.g., 'Our homeroom teacher always knows what's going on in class'; $\alpha_{student}$ = .80 to .87/'I always know what's going on in class'; $\alpha_{teacher}$ = .64 to .85).

Teacher Support. Two scales were used to assess teacher support. First, *teacher academic support* (five items) assessed the homeroom teacher's sensitivity to students' instructional needs and assistance with content-related issues was assessed with five items (e.g., 'Our homeroom

teacher provides additional assistance if we need help'; $\alpha_{student}$ = .84 to .91/'I support my students if they need any extra help'; $\alpha_{teacher}$ = .66 to .74).

Second, t*eacher emotional support* (seven items) measured the extent to which the homeroom teacher acts as a confidant and believes/has confidence in his/her students (e.g., 'Our homeroom is someone we can trust'; $\alpha_{student}$ = .70 to .83/'I try to build my trust with my students; $\alpha_{teacher}$ = .71 to .81).

Table 1
Descriptive Results for Student Ratings

	M (SD)	Min	Max	α	ICC (1)
Achievement					
T1	0.50 (.55)	-0.63	1.92		.27
T2	0.74 (.62)	-1.09	2.16		.25
T3	1.14 (.74)	-0.40	2.59		.31
T4	1.58 (.87)	-0.56	3.07		.28
Academic Engagement					
Academic identification					
T1	3.37 (.60)	1	4	.87	.02
T2	3.35 (.63)	1	4	.89	.02
T3	3.23 (.65)	1	4	.89	.02
T4	3.19 (.63)	1	4	.88	.02
School satisfaction					
T1	2.84 (.79)	1	4	.71	.08
T2	2.64 (.79)	1	4	.73	.08
T3	2.55 (.74)	1	4	.69	.05
T4	2.47 (.72)	1	4	.68	.05
Value beliefs (Math/German)					
T1	3.88	1	4	.67	.04
T2	2.87	1	4	.72	.07
T3	2.77	1	4	.74	.07
T4	2.65	1	4	.72	.08
Truancy					
T1	1.14 (.45)	1	4	.93	.02
T2	1.14 (.46)	1	4	.94	.05
T3	1.16 (.44)	1	4	.95	.08
T4	1.24 (.61)	1	4	.95	.06
Social School Adjustment					
School belonging					
T1	3.20 (.57)	1	4	.70	.02
T2	3.21 (.60)	1	4	.78	.03
T3	3.28 (.57)	1	4	.79	.01
T4	3.28 (.58)	1	4	.83	.03
Social relatedness					
T1	3.14 (.70)	1	4	.70	.03
T2	3.16 (.73)	1	4	.79	.02
T3	3.24 (.69)	1	4	.80	.02
T4	3.23 (.68)	1	4	.83	.02

Emotional School Adjustment					
Self-esteem					
T1	3.46 (1.0)	1	5	.71	.02
T2	3.48 (.91)	1	5	.74	.03
T3	3.55 (.87)	1	5	.74	.01
T4	3.49 (.86)	1	5	.78	.03
Emotional well-being					
T1	4.14 (.73)	1	5	.66	.02
T2	4.00 (.79)	1	5	.69	.03
T3	4.10 (.72)	1	5	.67	.04
T4	4.00 (.75)	1	5	.72	.03
Teaching Quality					
Disturbances					
T1	2.44 (.67)	1	4	.72	.10
T2	2.35 (.71)	1	4	.80	.11
T3	2.46 (.74)	1	4	.82	.17
T4	2.47 (.78)	1	4	.87	.22
Monitoring					
T1	3.23 (.65)	1	4	.80	.07
T2	3.10 (.70)	1	4	.83	.13
T3	3.12 (.72)	1	4	.86	.14
T4	2.99 (.76)	1	4	.87	.17
Teacher academic support					
T1	3.29 (.63)	1	4	.84	.13
T2	3.11 (.74)	1	4	.88	.18
T3	3.13 (.73)	1	4	.90	.20
T4	3.00 (.79)	1	4	.91	.19
Teacher emotional support					
T1	3.36 (.62)	1	4	.81	.12
T2	3.17 (.75)	1	4	.84	.17
T3	3.17 (.72)	1	4	.98	.20
T4	3.06 (.78)	1	4	.90	.21

Table 2

Descriptive Results for Teacher Self-reports

	M (SD)	Min	Max	α
Teaching Quality				
Disturbances				
T1	2.53 (.79)	1.00	4.00	.91
T2	2.36 (.73)	1.00	4.00	.88
T3	2.37 (70)	1.00	4.00	.86
T4	2.34 (70)	1.00	3.75	.85
Monitoring				
T1	3.39 (.41)	2.00	4.00	.85
T2	3.32 (.39)	2.25	4.00	.73
T3	3.30 (.34)	2.67	4.00	.64
T4	3.24 (.42)	1.25	4.00	.73
Teacher academic support				
T1	3.57 (.03)	2.67	4.00	.71
T2	3.47 (.35)	2.67	4.00	.74
T3	3.40 (.36)	2.20	4.00	.66
T4	3.36 (.40)	2.20	4.00	.73
Teacher emotional support				
T1	3.66 (.03)	2.75	4.00	.81
T2	3.54 (.36)	2.75	4.00	.75
T3	3.52 (.38)	2.50	4.00	.71
T4	3.42 (.50)	1.00	4.00	.79

Covariates

Prior research has shown that students' background characteristics are related to teaching quality and should be considered in evaluations of teaching quality (Campbell & Ronfeldt, 2018; Fauth et al., 2021; Göllner et al., 2020). Thus, we included students' *gender* (0 = female, 1 = male), *migration background* (0 = no migration background, 1 = migration background), and *socioeconomic background* (SES) at both the student and the class level. Students' SES was derived from students' responses regarding their parents' occupation and was calculated based on the International Socioeconomic Index of Occupational Status (Ganzeboom & Treiman, 2003). The parent with the highest score was used in the analyses (M = 45.55, SD = 12.6).

At the class level, we additionally controlled for *school track* using two dummy-coded variables. The lower school track served as the reference category (intermediate track: 1 = intermediate, 0 = other; comprehensive track: 1 = comprehensive, 0 = other). To account for homeroom teacher changes from one school year to the next, we calculated a dummy-coded variable for each measurement wave that was included as a time-varying covariate at the class level (1= different teacher from previous year, 0 = same teacher as previous year).

Statistical Analyses

Univariate Latent Growth Curve Models

To model the trajectories of students' academic achievement, academic engagement, and their social and emotional school adjustment, and teaching quality from fifth to eighth grade (Research Questions 1 and 2), we applied multilevel latent growth curve modeling (LGCM), a method that is frequently used to examine how constructs develop over time and how changes in one construct are related to changes in others. In a LGCM, the construct of interest is conceptualized as a function of time, and the goal is to explain mean-level differences from one time point to the next. Specifically, two latent factors are estimated that represent the average initial status in fifth grade (intercept) and average change from fifth to eighth grade (slope). To take into account the multilevel structure of student data (i.e., students clustered within classes), we conducted the latent growth curve models at both the student and the classroom level by aggregating the individual student data at the classroom level to represent a class's mean value. Here, the intercept at the student level represents within-classroom differences in the average level across the entire period from fifth to eighth grade, while the slope represents within-classroom

differences in how individual students' values change over time. At the classroom level, the intercept represents the classroom's average level during the period from fifth to eighth grade, while the slope represents the classrooms' average change from fifth to eighth grade as well as between-classroom differences in fifth grade. Moreover, the variance of the intercept and the slope indicates whether there are considerable differences between individual students (student level) or entire classrooms (classroom level) in the average level of teaching quality over the entire period from fifth to eighth grade and in the pattern of change over time.

One important question when conducting latent growth curve models concerns the most parsimonious form of the estimated time function. To determine this, we compared a model assuming a linear time trend with a model allowing deviation from a linear function. This was achieved by relaxing the factor loadings of the slope factor to be freely estimated (Bollen & Curran, 2006). Comparing the different solutions at the student allowed us to select the most appropriate model for all further analyses.

Bivariate Latent Growth Curve Models

The central aim of the study was to explore the co-development of student development and teaching quality during early adolescence (Research Question 3). To this end, we ran a series of bivariate latent growth curves in which the trajectories of student development (i.e., academic achievement, academic engagement, and their social and emotional school adjustment) were set in relation to the trajectories of teaching quality. To this end, we examined the extent to which the intercept (initial level) and slope (rate of change) of students' academic achievement, and each aspect of students' academic engagement, and their social and emotional school adjustment were associated with the intercept and slope of their homeroom teachers' classroom management, monitoring, academic and social support, as rated by both students and teachers. The correlation between the intercepts reflects the correlation between the average levels of the two constructs during the period from fifth to eighth grade, whereas the correlation between slope factors indicates how changes in student' academic achievement, academic engagement, and their social and emotional school adjustment are associated with changes in teaching quality. Each bivariate latent growth curve model was calculated as an unconditional model in which we did not account for any of the covariates and as a covariate model in which we controlled for homeroom teacher changes and students' background characteristics.

Covariate Models

Covariates were included at the individual (i.e., gender, migration background, SES) as well as at the classroom level (i.e., school track, homeroom teacher changes, proportion of male/female students, proportion of students with migration backgrounds, average level of SES) in both the univariate and bivariate growth curve models. With the exception of homeroom teacher changes, the covariates were modeled as time-invariant covariates that directly predicted variability in initial status in fifth grade and changes from fifth to eighth grade. Homeroom teacher changes from one year to the next were modeled as a single time-variant covariate. A graphic representation of the growth curve model including covariates is shown in Figure 1.

We conducted all analyses with the Mplus 7.3 software (Muthén & Muthén, 1998-2012). Model fit was evaluated by means of the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI values above .90, RMSEA values below .06, and SRMR values below .08 are considered indicative of satisfactory to good model fit (Hu & Bentler, 1995). All significance testing was performed at the $\alpha =$. 05 level.

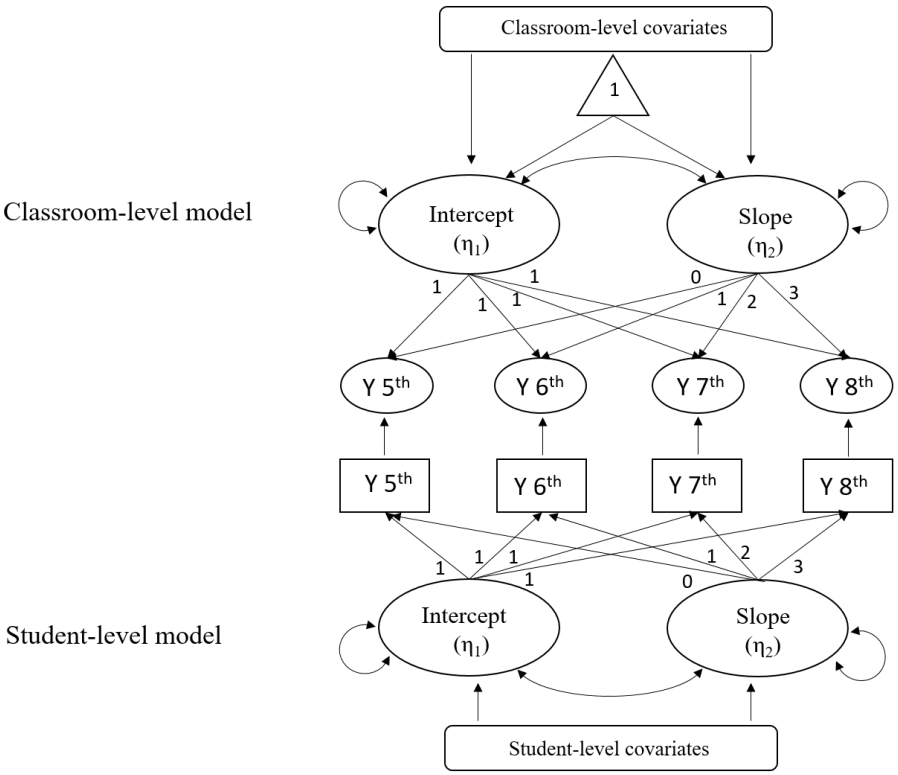

Figure 1. Graphic representation of the multilevel longitudinal growth curve model

Missing Data

One challenge in empirical longitudinal studies are cases with missing data. Missing information for students may have occurred if students were held back a grade, were unavailable at one or more measurement points, or did not consent to participate in a given measurement point. Overall, $N = 2208$ students (46 % female) and $N = 99$ teachers participated at all four measurement points. Students who participated at all measurement points were less likely to have a migration background ($\chi^2 = 97.76$, $p < .05$; 82 % without migration background vs 63 % with migration background), and were more likely to be from comprehensive-track or intermediate-track schools ($\chi^2 = 168.55$, $p < .05$; 67 % comprehensive-track, 64% intermediate-track, 45 % lower track). However, no significant differences in students' SES emerged ($T = 1.81$, $SE = 0.45$, $p = 0.70$). Although some significant differences between both groups existet, the revealed differences were relatively small.

Results

Preliminary Analysis

Our first step was to test whether a model with a linear rate of change adequately represents the trajectories of students' academic achievement, academic engagement, and their social and emotional school adjustment and teaching quality from fifth to eighth grade. For this, we compared a linear trend model to a model with a freely-estimated rate of change (Table S1 in the supplement). The unconditional models without covariates with a linear rate of change fit the data well, and exhibited satisfactory model fit overall (Table S2), whereas the unconditional models with a freely estimated growth factor did not converge in 5 out of 17 models (see Table S1). The fit indices for the remaining models showed that a linear time trend was appropriate, as relaxing the loading constraints did not result in a superior model fit and the differences in model fit indices were rather small. Given these results, we continued the analysis using a linear time trend for all variables and subsequent analytical steps.

Each trajectory was then calculated as an unconditional model (i.e., without controlling for the covariates, Table S2) and as a covariate model (Table 3), to examine whether the trajectories of student development remained unchanged when additionally controlling for students' background characteristics. Overall, the covariates did not show many systematic associations with the intercepts (i.e., average level from 5^{th} to 8^{th} grade) and the slopes (i.e., average amount of

change from 5th to 8th grade) of students' development and teaching quality (see Tables S3, S4, S5) and there were only minor changes in the trajectories after having controlled for the covariates. For this reason, we report the results for the covariate model.

Trajectories of Student Development

Our first research question addresses the trajectories of change in student development (i.e., academic achievement, academic engagement, social and emotional school adjustment) from fifth to eighth grade of lower secondary school, with a special focus on the extent to which the trajectories differ at the classroom level.

Students' Academic Achievement and Academic Engagement

The estimated growth trajectories showed significant gains in students' academic achievement (slope = 0.30, $p < .05$). In contrast, students' emotional engagement largely exhibited negative changes over time (school satisfaction: slope = -0.15; subject value beliefs: slope = - 0.12; all *p*s < .05, but academic identification: slope = - 0.02, $p = .433$). Students' behavioral engagement (truancy: slope = 0.03, $p = .086$) exhibited no significant change over time.

In addition to that, we inspected the extent to which the trajectories differed across individual students and across classrooms. At the student level, there was a significant amount of variability in the intercepts and the slopes (intercept variance: $\sigma^2 = 0.06$ to $\sigma^2 = 0.51$; slope variance: $\sigma^2 = 0.01$ to $\sigma^2 = 0.03$, all *p*s < .05), indicating that within the same classroom, the trajectories of students' academic achievement and academic engagement differed considerably across individual students. At the class level, systematic differences in the trajectories between classrooms were found for students' academic achievement (intercept variance: $\sigma^2 = 0.07$, slope variance: $\sigma^2 = 0.01$, all *p*s < .05) and school satisfaction (intercept variance: $\sigma^2 = 0.02$, slope variance: $\sigma^2 = 0.01$, all *p*s < .05). In addition to this, the slope ICC indicated that 8-21% of the variance in individual students' academic achievement and engagement was attributable to differences between classrooms (Table 3), which points to differences in the development of students' academic achievement and engagement due to their classroom environment.

Students' Social and Emotional School Adjustment

The trajectories indicated that on average, adolescent students adjust to school quite well. The trajectories of students' social school adjustment indicated a positive development over time (school belonging: slope = 0.06, social relatedness: slope = 0.07, all *p*s < .05). For students' emotional school adjustment, self-esteem (slope = 0.06, $p < .05$) slightly increased from fifth to eighth grade, whereas students' emotional well-being exhibited no significant changes over time (slope = - 0.04, $p = .083$).

At the student level, there was a significant amount of variability in the intercepts and the slopes (intercept variance: $\sigma^2 = 0.06$ to $\sigma^2 = 0.51$; slope variance: $\sigma^2 = 0.01$ to $\sigma^2 = 0.03$, all *p*s < .05), indicating that within the same classroom, the trajectories of students' academic achievement, academic engagement, and their social and emotional school adjustment differed considerably. In contrast, there were no considerable differences in the trajectories of students' social and emotional school adjustment across classes (classroom level, see Table 3). Neither the covariates exhibited any systematic effects on the trajectories at the class level (Table S4, S5), nor did the slope ICC indicate systematic differences in the trajectories between classes (0-4 %, Table 3). In summary, the findings suggest that students' learning environment may play a less crucial role in students' social and emotional school adjustment during lower secondary school than in other domains of student development.

Trajectories of Teaching Quality

Our second research question addressed the development of teaching quality from fifth to eighth grade, as assessed by both the students and their homeroom teachers. From the student perspective (covariate model), there was a significant amount of change in teaching quality from fifth to eighth grade. Specifically, monitoring: slope = - 0.09; academic support: slope = - 0.09; social support: slope = - 0.10 (all *p*s < .05, Table 3) declined over time. No changes over time were found for the amount of disturbances in the classroom (slope = - 0.05, $p = 0.164$). The average level and changes in teaching quality from 5^{th} to 8^{th} grade were not systematically affected by any of the covariates at the classroom level (Table S4, Table S5).

However, the trajectories showed significant variance at both the student level (intercept variance: $\sigma^2 = 0.10$ to $\sigma^2 = 0.13$; slope variance: $\sigma^2 = 0.02$ to $\sigma^2 = 0.03$, all *p*s < .05) and the classroom level (intercept variance: $\sigma^2 = 0.03$ to $\sigma^2 = 0.05$; slope variance: $\sigma^2 = 0.01$, all *p*s < .05;

Table 2), indicating substantial differences in the average level and patterns of change in students' perceptions of teaching quality over time within and across classrooms. That is, although teaching quality decreased on average from fifth to eighth grade, not all students followed the same developmental trend. Specifically, the slope ICC indicated that 21% to 42% of the variance in students' individual perceptions of changes in teaching quality over time was attributable to differences between classrooms (Table 3).

From the teacher perspective, the models indicated decreases in academic support as students moved from fifth to eighth grade (slope = 0.08, p >.05). Classroom management, monitoring and social support exhibited negative, but non-significant changes over time (classroom management: slope = - 0.08, p = .205; monitoring: slope = - 0.08, p = .098; social support: slope = - 0.06, p = .116; Table 3). In the unconditional models without covariates, there were significant declines in all aspects of teacher-reported teaching quality over time (classroom management: slope = - 0.06, monitoring: slope = - 0.06, academic support: slope = - 0.07; social support: slope = - 0.08; all ps < .05; Table S2). In addition, there were significant differences in teaching quality trajectories between teachers, with the largest variance in classroom management trajectories (Table 3). The average level and changes in teacher-reported teaching quality from fifth to eighth grade were largely unrelated to the covariates (i.e., students' background characteristics, homeroom teacher changes and the school track, see Table S4, Table S5).

In summary, both students and teachers reported decreases in perceived teaching quality from fifth to eighth grade of lower secondary school. However, the decline was more pronounced from the student perspective, and there was significant variability in the trajectories across classes and teachers.

Table 3

Model Fit and Coefficients for the Covariate Univariate Models

Model	χ^2 *(df)*	RMSEA	CFI/TLI	$SRMR_{s/c}$	Intercept M_i (σ_i^2)	Slope M_s (σ_s^2)	ICC (Intercept)	ICC (Slope)
Student Ratings								
Achievement[a]	51.52* (30)	0.02	.99/.97	.01/.03	.68* (.07)*	.30* (.01)*	0.12	0.21
Academic Engagement								
Academic identification	57.44* (27)	0.02	.98/.94	.02/.04	3.32* (.01)*	-.02 (.00)	0.05	0.10
School satisfaction[b]	58.49* (28)	0.02	.99/.96	.02/.03	2.66* (.02)*	-.15* (.00)*	0.07	0.09
Subject value beliefs[a]	51.85* (28)	0.02	.98/.94	.02/.06	2.89* (.01)*	-.12* (.00)	0.07	0.00
Truancy[a]	34.47* (28)	0.01	.99/.98	.02/.04	1.22* (.00)*	.03 (.00)	0.06	0.08
Social School Adjustment								
School belonging[b]	108.13* (28)	0.03	.97/.90	.02/.07	3.21* (.00)	.06* (.00)	0.01	0.00
Social Relatedness[b]	82.94* (28)	0.02	.98/.93	.02/.06	3.19* (.00)	.07* (.00)	0.01	0.04
Emotional School Adjustment								
Self-esteem	46.72* (27)	0.01	.99/.96	.01/.06	3.55* (.01)	.06* (.00)	0.03	0.00
Emotional well-being[b]	92.51* (28)	0.02	.95/.85	.02/.07	4.05* (.01)*	-.04 (.00)	0.04	0.04
Teaching Quality								
Disturbances[a]	72.72* (28)	0.02	.95/.85	.02/.04	2.41* (.04*)	-.05 (.01*)	0.33	0.42
Monitoring[a]	40.83* (28)	0.01	.99/.97	.02/.03	3.24* (.02*)	-.09* (.01*)	0.19	0.21
Teacher academic support[a]	41.36* (28)	0.01	.99/.96	.01/.04	3.27* (.05*)	-.09* (.01*)	0.26	0.36
Teacher emotional support[a]	56.44* (28)	0.02	.98/.93	.02/.04	3.28* (.05*)	-.10* (.01*)	0.26	0.40
Teacher Ratings								
Disturbances	27.67* (18)	.01	.98/.92	.00/.03	2.31* (.21*)	-.08 (.05*)		
Monitoring	24.06* (18)	.01	.98/.93	.00/.04	3.35* (.03*)	-.08 (.01*)		
Teacher support academic	29.79* (18)	.01	.97/.89	.00/.04	3.48* (.04*)	-.08* (.01*)		
Teacher support social	15.78* (18)	.00	1.0/1.0	.00/.03	3.69* (.05*)	-.06 (.02*)		

Note. *$p < .05$.

[a]For model identification, the residual variance at the first measurement time point at the class level had to be fixed to zero.

[b]For model identification, the residual variance at the last measurement time point at the class level had to be fixed to zero.

The Co-development of Student Development and Teaching Quality

Finally, we examined the extent to which students' learning environments account for systematic differences in student development. To this end, we investigated whether the trajectories of student development (i.e., academic achievement, academic engagement, their social and emotional school adjustment) were associated with the trajectories of teaching quality from fifth to eighth grade of lower secondary school. As there were no substantial differences in the estimated associations between the unconditional (Table S6, S7) and the covariate model (Table 4), we present the results for the covariate model, beginning with the associations at the student level.

Students' Academic Achievement, Academic Engagement, and Teaching Quality

The average levels of students' academic achievement and academic engagement at the student level were strongly related to student-reported teaching quality. That is, students who reported a higher average level of teaching quality during the period from fifth to eighth grade also reported higher average levels of academic achievement and academic engagement (see Table 4). Turning to the associations between the slopes, students' academic achievement was unrelated to teaching quality (slopes: $r = 0.02$ to $r = 0.15$; all ps $> .05$) from fifth to eighth grade. Besides this, the development of students' academic engagement was strongly associated with student-perceived teaching quality. That is, students who experienced positive change in their academic identification (slopes: $r = 0.49$ to $r = 0.69$; all ps $< .05$), school satisfaction (slopes: $r = 0.37$ to $r = 0.52$; all ps $< .05$) and value beliefs (slopes: $r = 0.55$ to $r = 0.62$; all ps $< .05$) also experienced more positive change in teaching quality over time. However, changes in truancy over time were largely unrelated to teaching quality at the student level (but disturbances: slope: $r = -0.42$, $p < .05$).

At the classroom level, classes that were provided with higher average levels of teaching quality also indicated higher average levels of academic engagement from fifth to eighth grade. However, the average level of students' academic achievement was unrelated to the average level of teaching quality over time (Table 4). Turning to the associations between the slopes at the classroom level, the development of students' academic achievement and academic engagement was positively associated with the development of student-perceived teaching quality over time. That is, positive change in students' academic achievement (slopes: $r = 0.58$ to $r = 0.70$, all ps $< .05$; except for classroom management: $r = -0.19$, $p > .05$), academic identification (slopes: $r = 0.51$ to $r = 0.92$, all ps $< .05$), and school satisfaction (slopes: $r = 0.47$ to $r = 0.55$, all ps $< .05$) was

related to positive changes over time in teaching quality, indicating classes with more positive development of academic achievement and academic engagement were systematically provided with high teaching quality over time.

In summary, the findings suggest that the extent to which adolescent students' are provided with high teaching quality from fifth to eighth grade seems to contribute to (systematic differences in) the development of their academic achievement and academic engagement from a longitudinal perspective.

Lastly, we examined whether the trajectories of students' development were related to the trajectories of teaching quality from the teacher's perspective (see Table S7 presenting results for the models without covariates and Table 5 for the covariate models). In the covariate model, teachers' self-reports of teaching quality was largely unrelated to student development, with associations only emerging among the intercepts. However, there were no additional associations between change over time in students' academic achievement, academic engagement and teacher-reported teaching quality from fifth to eighth grade (Table 5).

Students' Social and Emotional School Adjustment and Teaching Quality

Overall, the development of students' social and emotional school adjustment was largely associated with teaching quality at the student-level. That is, students who reported a higher average level of teaching quality during the period from fifth to eighth grade also reported higher average levels of their social (i.e., school belonging, perceived social relatedness), and emotional (i.e., self-esteem and emotional wellbeing) school adjustment (see Table 4).

With regard to the associations between the slopes, changes over time in school belonging, self-esteem and emotional well-being were related to changes in students' individual perceptions of teaching quality. Students who experienced increases in school belonging (slopes: $r = 0.18$ to $r = 0.33$; all ps $< .05$), self-esteem (slopes: $r = 0.23$ to $r = 0.42$; all ps $< .05$) and emotional well-being (slopes: $r = 0.23$ to $r = 0.25$; all ps $< .05$) perceived more positive changes in teaching quality over time compared to their classmates. However, changes in student-perceived disturbances were largely unrelated to change over time in students' social and emotional school adjustment. Moreover, it is important to note that changes over time in students' social relatedness were largely unrelated to changes in their perceptions of teaching quality ($r = 0.05$ to $r = 0.14$; ps $> .05$, but academic support: $r = 0.17$, $p < .05$; Table 4).

In contrast, the development of students' social and emotional school adjustment was unrelated to teaching quality at the classroom level (Table 4), which suggests that the development of students' social and emotional school adjustment is driven more strongly by individual processes, rather than by systematic differences in students' learning environments.

From the teachers' perspective, teaching quality was unrelated with the development of students' social and emotional school adjustment from fifth to eighth grade.

Table 4
Associations between Student Development and Teaching Quality in the Covariate Models (Student Ratings)

	Model Fit				Associations			
					Student level		Classroom level	
Model	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$	$r_{intercept}$	r_{slope}	$r_{intercept}$	r_{slope}
Achievement + Disturbances	143.22* (82)	.02	.98/.95	.03/.05	.04 (.04)	.02 (.13)	.19 (.18)	.19 (.25)
Achievement + Monitoring	127.53* (82)	.02	.98/.96	.02/.05	**.13 (.03)***	.15 (.11)	.02 (.18)	**.70* (.19)**
Achievement + Teacher ac. support	128.46* (82)	.02	.98/.96	.02/.05	**.14 (.03)***	.02 (.10)	-.02 (.13)	**.58* (.17)**
Achievement + Teacher emo. support	130.49* (82)	.02	.98/.96	.02/.05	**.15 (.03)***	.06 (.11)	.03 (.17)	**.62* (.18)**
Academic Engagement								
Academic Ident. + Disturbances[a]	148.52* (71)	.02	.97/.93	.02/.05	**.31 (.04)***	**.43 (.15)***	**.45 (.13)***	.37 (.26)
Academic Ident. + Monitoring[a]	103.20* (71)	.01	.99/.98	.02/.04	**.59 (.03)***	**.69 (.11)***	**.51 (.13)***	**.92 (.33)***
Academic Ident. + Teacher ac. support[a]	117.37* (71)	.01	.99/.96	.02/.05	**.53 (.03)***	**.63 (.13)***	**.36 (.13)***	**.51 (.24)***
Academic Ident. + Teacher emo. support[a]	136.66* (71)	.02	.98/.95	.02/.05	**.54 (.03)***	**.66 (.14)***	**.40 (.14)***	**.59 (.24)***
School satisfaction + Disturbances	200.79* (80)	.02	.96/.91	.02/.05	**.43 (.03)***	**.37 (.09)***	**.53 (.12)***	**.47 (.14)***
School satisfaction + Monitoring	155.99* (80)	.02	.98/.95	.02/.04	**.46 (.03)***	**.38 (.07)***	**.57 (.10)***	**.47 (.19)***
School satisfaction + Teacher ac. support	157.10* (80)	.02	.97/.95	.02/.05	**.53 (.03)***	**.42 (.09)***	**.46 (.11)***	**.55 (.16)***
School satisfaction + Teacher emo. support	183.00* (80)	.02	.97/.94	.02/.05	**.51 (.03)***	**.52 (.10)***	**.49 (.11)***	**.53 (.16)***
Value Beliefs + Disturbances	176.62* (80)	.02	.95/.88	.03/.06	**.43 (.04)***	**.54 (.15)***	**.67 (.11)***	**.**76 (1.12)
Value Beliefs + Monitoring	140.29* (80)	.02	.97/.94	.02/.06	**.51 (.03)***	**.60 (.13)***	**.70 (.12)***	.33 (.69)
Value Beliefs + Teacher ac. support	160.05* (80)	.02	.97/.92	.02/.07	**.51 (.04)***	**.61 (.13)***	**.72 (.12)***	.70 (.59)
Value Beliefs + Teacher emo. support	167.55* (80)	.02	.96/.91	.02/.07	**.57 (.05)**	**.60 (.14)**	**.62 (.17)**	.68 (.77)
Truancy + Disturbances[a]	123.38* (72)	.01	.97/.92	.02/.05	**-.12 (.05)***	**-.42 (.17)***	**-.39 (.16)***	-.20 (.25)
Truancy + Monitoring[a]	96.67* (72)	.01	.99/.97	.02/.04	**-.28 (.03)***	-.17 (.13)	**-.45 (.16)***	-.06 (.31)
Truancy + Teacher ac. support[a]	90.77* (72)	.01	.99/.98	.02/.05	**-.31 (.03)***	-.22 (.12)	-.31 (.16)	. 17 (.29)

Truancy + Teacher emo. support[a]	104.80* (72)	.01	.98/.96	.02/.05	**-.31 (.04)***	-.15 (.14)	-.30 (.16)	.11 (.27)
Social School Adjustment								
Schoolbelonging + Disturbances	232.95* (80)	.02	.95/.89	.03/.07	**.28 (.03)***	.16 (.10)	.21 (.25)	.86 (.59)
Schoolbelonging + Monitoring	214.72* (80)	.02	.96/.91	.02/.08	**.27 (.03)***	**.18 (.08)***	.36 (.23)	.58 (.54)
Schoolbelonging + Teacher ac. support	191.54* (80)	.02	.97/.93	.02/.07	**.29 (.03)***	**.29 (.08)***	.22 (.22)	.74 (.46)
Schoolbelonging + Teacher emo. support	221.14* (80)	.02	.96/.91	.02/.07	**.30 (.03)***	**.33 (.09)***	.31 (.22)	.77 (.41)
Social relatedness + Disturbances	185.42* (80)	.02	.97/.93	.03/.07	**.27 (.03)***	.05 (.09)	.05 (.30)	.35 (.24)
Social relatedness + Monitoring	163.88* (80)	.02	.98/.95	.02/.07	**.20 (.03)***	.13 (.07)	.23 (.32)	.23 (.26)
Social relatedness + Teacher ac. support	163.94* (80)	.02	.97/.95	.02/.07	**.23 (.03)***	**.17* (.07)**	.24 (.31)	.31 (.24)
Social relatedness + Teacher emo. support	194.05* (80)	.02	.97/.93	.02/.07	**.22 (.03)***	.14 (.08)	.28 (.28)	.42 (.23)
Emotional School Adjustment								
Self-esteem + Disturbances	164.64* (79)	.02	.96/.91	.02/.07	**.29 (.04)***	**.42 (.13)***	.11 (.20)	.54 (.95)
Self-esteem + Monitoring	116.95* (79)	.01	.99/.97	.02/.06	**.37 (.03)***	**.23 (.11)***	.04 (.18)	.61 (1.13)
Self-esteem + Teacher ac. support	105.91* (79)	.01	.99/.98	.02/.07	**.42 (.03)***	**.40 (.12)***	.21 (.17)	.08 (.36)
Self-esteem + Teacher emo. support	127.16* (79)	.01	.98/.96	.02/.08	**.46 (.03)***	**.35 (.12)***	.22 (.27)	.46 (.53)
Emo. Well-being + Disturbances	198.34* (80)	.02	.94/.87	.02/.07	**.32 (.03)***	.07 (.11)	**.50 (.16)***	.68 (.47)
Emo. Well-being + Monitoring	152.69* (80)	.02	.97/.94	.02/.07	**.32* (.05)**	**.25* (.10)**	**.55 (.17)***	.75 (.59)
Emo. Well-being + Teacher ac. support	160.24* (80)	.02	.97/.93	.02/.07	**.43 (.03)***	**.28 (.10)***	**.47 (.17)***	.76 (.58)
Emo. Well-being + Teacher emo. support	184.72* (80)	.02	.96/.91	.02/.07	**.43 (.04)***	**.23 (.10)***	**.54 (.17)***	.85 (.51)

Note. *$p < .05$. Correlation coefficients are shown in their standardized form.

[a]We allowed for time specific correlations between the residuals of Academic Identification/Truancy and Teaching quality at each measurement time point, because the assessment of these aspects took place on the same day of assessment.

Table 5

Associations between Student Development and Teaching Quality in the Covariate Models (Teacher Self-reports)

	Model Fit				Associations Classroom level	
Model	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$	$r_{intercept}$	r_{slope}
Achievement + Disturbances	108.26* (60)	.02	.98/.95	.01/.04	.19 (.12)	.12 (.21)
Achievement + Monitoring	93.52* (60)	.02	.98/.96	.01/.05	-.16 (.17)	.15 (.35)
Achievement + Teacher ac. support	95.64* (60)	.02	.98/.96	.01/.05	.03 (.14)	-.04 (.26)
Achievement + Teacher emo. support	86.46* (60)	.01	.99/.97	.01/.05	.20 (.16)	-.12 (.24)
Academic Engagement						
Academic Ident. + Disturbances	102.71* (57)	.01	.98/.94	.02/.05	**.40* (.13)**	.30 (.21)
Academic Ident. + Monitoring	92.19* (57)	.01	.98/.95	.02/.05	.15 (.17)	.00 (.28)
Academic Ident. + Teacher ac. support	94.16* (57)	.01	.98/.95	.02/.05	-.08 (.15)	-.12 (.25)
Academic Ident. + Teacher emo. support	95.13* (57)	.01	.98/.95	.02/.04	-.17 (.20)	-.52 (.28)
School satisfaction + Disturbances	106.00* (58)	.02	.98/.95	.02/.05	**.44 (.12)***	.29 (.19)
School satisfaction + Monitoring	98.00* (58)	.01	.98/.96	.02/.04	.20 (.16)	-.09 (.25)
School satisfaction + Teacher ac. support	99.87* (58)	.01	.98/.96	.02/.04	.24 (.15)	.35 (.27)
School satisfaction + Teacher emo. support	85.92* (58)	.01	.99/.97	.02/.04	**.32 (.14)***	.08 (.20)
Value Beliefs + Disturbances	115.10* (58)	.02	.96/.89	.02/.07	**.41 (.17)***	.39 (.66)
Value Beliefs + Monitoring	85.01* (58)	.01	.98/.94	.02/.06	.08 (.21)	-.22 (1.2)
Value Beliefs + Teacher ac. support	97.83* (58)	.01	.97/.93	.02/.06	.20 (.19)	1.1 (1.9)
Value Beliefs + Teacher emo. support	84.90* (58)	.01	.98/.94	.02/.05	.28 (.18)	.68 (1.7)
Truancy + Disturbances	102.95* (58)	.01	.96/.91	.02/.06	**-.47 (.13)***	-.41 (.31)
Truancy + Monitoring	77.90* (58)	.01	.98/.95	.02/.06	.16 (.19)	-.08 (.40)

Truancy + Teacher ac. support	76.83* (58)	.01	.98/.96	.02/.05	**.41 (.14)***	-.03 (.37)
Truancy + Teacher emo. support	69.47* (58)	.01	.99/.97	.02/.05	.18 (.15)	.59 (.36)
Social School Adjustment						
Schoolbelonging + Disturbances	144.45* (58)	.02	.97/.92	.03/.07	.35 (.25)	.85 (.48)
Schoolbelonging + Monitoring	152.80* (58)	.02	.96/.90	.02/.07	**.76 (.25)***	.04 (.49)
Schoolbelonging + Teacher ac. support	141.02* (58)	.02	.97/.92	.02/.07	.23 (.24)	.09 (.59)
Schoolbelonging + Teacher emo. support	128.34* (58)	.02	.97/.93	.02/.07	-.01 (.23)	-.41 (.62)
Social relatedness + Disturbances	122.93* (58)	.02	.98/.94	.02/.06	.29 (.43)	.27 (.28)
Social relatedness + Monitoring	118.43* (58)	.02	.98/.94	.02/.06	.74 (.42)	-.61 (.43)
Social relatedness + Teacher ac. support	144.08* (58)	.02	.97/.92	.02/.08	.47 (.39)	-.39 (.42)
Social relatedness + Teacher emo. support	120.10* (58)	.02	.98/.94	.02/.07	.39 (.33)	-.62 (.38)
Emotional School Adjustment						
Self-esteem + Disturbances	86.00* (57)	.01	.98/.96	.01/.06	.27 (.17)	.79 (1.6)
Self-esteem + Monitoring	83.44* (57)	.01	.98/.96	.01/.06	.34 (.21)	-.63 (1.3)
Self-esteem + Teacher ac. support	93.73* (57)	.01	.98/.95	.01/.07	.04 (.18)	-1.2 (2.6)
Self-esteem + Teacher emo. support	69.96* (57)	.01	.99/.98	.01/.06	-.05 (.19)	-2.0 (7.9)
Emo. Well-being + Disturbances	127.98* (58)	.02	.96/.89	.02/.06	.32 (.15)	.36 (.42)
Emo. Well-being + Monitoring	121.31* (58)	.02	.96/.89	.02/.07	.28 (.18)	-.92 (.85)
Emo. Well-being + Teacher ac. support	133.42* (58)	.02	.95/.88	.02/.07	.18 (.17)	-.24 (.50)
Emo. Well-being + Teacher emo. support	115.99* (58)	.02	.96/.90	.02/.06	.13 (.17)	-.77 (.63)

Note. *$p < .05$. Correlation coefficients are shown in their standardized form.

Discussion

A critical issue in educational research concerns the factors that foster or impede positive student development during adolescence. The present study focused on the classroom as the most proximal learning environment and applied a teaching quality framework to describe the central processes that happen within it. Combining student and teacher data from a longitudinal perspective, we investigated the co-development of students' academic achievement, academic engagement and their social and emotional school adjustment during the early adolescent years at the individual student and classroom levels. The multilevel longitudinal growth curve analyses revealed significant changes in both, student development and teaching quality over time, with the results were more pronounced for student-rated teaching quality than for teacher-rated teaching quality. In addition, teaching quality contributed to student development in a longitudinal perspective, whereas the specific patterns of association differed between the levels of analyses: Whereas change in teaching quality over time contributed to differences in the development of students' academic achievement and engagement at the student- and at the classroom level, the development of students' social and emotional school adjustment was more strongly driven by changes in students' individual perceptions of teaching quality at the student level. Overall, the revealed associations were more pronounced for student reports than for teacher self-reports of teaching quality.

Student Development in early Adolescence

The results of our study make an important contribution to research on adolescent development by providing a differentiated view on critical domains of adolescent student development across lower secondary school: In line with prior research (e.g., Eccles et al., 1993), students experienced declines in various aspects of their academic engagement during lower secondary school. In contrast, the trajectories for students' social and emotional school adjustment revealed rather positive changes over time. On average, students' perceived school belonging and social relatedness and their self-esteem slightly increased over time, even after controlling for the covariates (i.e., SES, gender, immigrant background, teacher changes and school track). Consequently, it seems that students were able to adjust to the secondary school environment in a rather favorable way, in their own view.

Lastly, it should be mentioned that there was little difference in the average development of adolescent students' social and emotional school adjustment between classes (i.e., the classroom level; see also Witherspoon & Ennett, 2011), whereas the trajectories significantly differed between students within the same classroom (i.e., the student level). This might indicate that the development of students' social and emotional school adjustment during adolescence reflects a more individual process, and may explain why aggregated class means have been used relatively infrequently in analyses of student development during adolescence.

Development of Teaching Quality

The present study provided a comprehensive overview of the development of teaching quality as students progressed through lower secondary school and combined ratings from both the teacher and the student perspective. Students in our sample reported declines in teaching quality from fifth to eighth grade (with the exception of classroom disturbances), which matches previous findings on negative changes of school climate during lower secondary school (e.g., Reddy et al., 2003; Way et al., 2007; Wang & Dishion, 2012). Moreover, our study's results extend existing research on changes in teaching quality over time by indicating that the previously reported declines in teacher support over the first year of secondary school (e.g., teacher-student relationship, Gehlbach et al., 2012; Maulana et al., 2014; Opdenakker et al., 2012) reflect a more continuous process across the secondary school years. In line with stage-environment fit theory (Eccles & Midgley, 1989), the declines in student-reported teaching quality over time may indeed indicate a misfit between students' developmental needs (e.g., competence, autonomy, relatedness; Ryan & Deci, 2020) and the teaching they experience in their classrooms (Roeser et al., 2000).

However, there was a significant amount of variance in the trajectory of teaching quality over and above students' background characteristics, suggesting that on average, teaching quality decreased across lower secondary school, but some classes may have experienced steeper decreases, whereas other classes may even have experienced an increase in teaching quality. Consequently, it seems that teachers clearly differ in their ability to continuously adapt their teaching to the developmental needs of adolescent students. On the other hand, it might be argued that adolescent students grow increasingly dissatisfied with their schools in general. Consequently, the declines in teaching quality may represent an overarching developmental trend during adolescence, irrespective of the actual teaching quality students receive (e.g., Way et al., 2007). The present study contributes to this question by combining student and teacher reports of teaching

quality. Indeed, declines in teaching quality were also reported from the teacher perspective, suggesting that students are indeed sensitive to changes in their teachers' organizational, social and instructional behavior and are able to accurately report changes in teaching quality in a student questionnaire.

The Co-development of Student Development and Teaching Quality

The final aim of our study was to examine the extent to which teaching quality in the classroom contributes to systematic differences in student development during early adolescence. The use of multilevel modeling extends previous research, which has predominantly conducted analyses at the student level and provides more detailed information on the associations between the trajectories of student development and teaching quality.

In line with previous research, our results showed that teaching quality seems to be particularly important for students' academic achievement and academic engagement. At both the student and classroom levels, positive changes in teaching quality buffered the downward trend in most aspects of students' academic engagement and fostered students' academic achievement at the classroom level. It is noteworthy that in our study, monitoring, academic and social support were positively related to students' academic achievement over time, whereas most cross-sectional studies report positive relations with the dimension of classroom management in particular (e.g., Fauth et al., 2014; Kunter et al., 2006). From this longitudinal perspective, it seems that the development of students' achievement and academic engagement benefit not only from a high average level of teaching quality, but also and particularly from the extent to which a teacher adapts his/her teaching behavior to students' changing developmental needs (e.g., Olivier et al., 2021; Patall et al., 2010; Stroet et al., 2013). Given that adolescence is a time of significant developmental change, during which students become increasingly self-aware, require appropriate challenge and competence-building experiences, and are in need of supportive relationships apart from their parents (Roeser et al., 2000; Wentzel, 2003), teachers' engagement to address students' developmental needs by providing a supportive learning environment and establishing positive teacher-student relationships seem to be important mechanisms by which to support students' academic development during the phase of early adolescence (Emmer & Gerwels, 2006).

Regarding students' social and emotional school adjustment, changes in teaching quality were related to changes in individual students' feelings of school belonging, self-esteem and

emotional well-being. In line with previous research (e.g., Furrer & Skinner, 2003; Rucinski et al., 2018), our study shows that teachers can significantly contribute to students' positive socio-emotional functioning by helping students experience school as a positive, supportive and respectful environment. This finding is compelling, considering that these associations emerged at the student level and classes did not systematically differ in their social and emotional development. Consequently, it appears that students' social and emotional school adjustment reflects a more individual process that is shaped by students' unique experiences with their teacher: for example, the amount of support a student personally perceives, or whether the teacher meets an individual student's needs. In contrast, it seems to be less important for students' social and emotional school adjustment whether the classroom is well organized or whether the teacher is, on average, more supportive than other teachers. Even though teachers represent an important attachment figure apart from parents during adolescence (Furman & Buhrmester, 1992; Way & Greene, 2006), students' social relatedness was unrelated to all aspects of teaching quality apart from academic support in this study. It might be that other factors besides teaching quality, such as peer relations, are also highly relevant for students' feelings of social relatedness. In sum, the results of our study suggest that supportive teachers play a particularly important role in fostering students' feelings of school belonging, self-esteem and emotional well-being by maintaining good-quality relationships with individual students and adapting their levels of support to individual students' needs. However, our findings raise the question of whether items should be formulated on the individual student vs. classroom level when examining the impact of students' learning environment on student development. That is, scales addressing the individual student's socio-emotional functioning (e.g., "In the last week, *I* felt lonely") may be less suited to capturing the effects of teaching quality on students' social and emotional adjustment in the classroom. Referring to the specific context (e.g., "In the last week, *I* felt lonely *in my class*") may be a more suitable alternative approach to assessing student development in the classroom environment.

Overall, the associations with student development were more pronounced for student-reported teaching quality than for teacher self-reports. Prior research has repeatedly shown only moderate concordance between student- and teacher-reported teaching quality (Aldrup et al., 2018; Kunter & Baumert, 2006; Wagner et al., 2016), most likely because a student's idiosyncratic perceptions of his/her teacher might be different from the teacher's perspective on the level of teaching quality provided to all students in the class. This might additionally support our study's assumption that students' individual perceptions of their teacher's organizational, social and

instructional behavior are a uniquely important predictor of their social and emotional school adjustment during lower secondary school.

Limitations

Although we believe that our study makes a significant contribution to understanding the complex dynamics of young adolescent students' development in lower secondary school and to teaching quality research, some limitations must be kept in mind. First, only students from vocational-track schools participated in our study. Therefore, the trajectories of students' academic development, socio-emotional school adjustment and teaching quality described in our study may not generalize to the entire population of German secondary school students, particularly students in academic-track schools. For example, in academic-track schools, there is typically a stronger emphasis on academic performance and testing procedures, including stronger between-student comparisons (Maaz et al., 2011). It remains to be seen whether the trajectories of student development and their relations with teaching quality differ in German academic-track schools and in other school systems.

Second, our study did not include data on students' transition from primary to secondary school and did not cover the entire span of secondary school. In the German system, students leave school to enter vocational training after 9th grade in the lower track and after 10th grade in the intermediate track. Thus, examining whether changes in student development and teaching quality are particularly related to school transitions or accelerate or deaccelerate during the peak of adolescence was beyond the scope of our study.

Third, with respect to the direction of the relation between student development and teaching quality, it should be kept in mind that the associations reported in our study are correlational. It could be argued that well-adjusted students perceived their classrooms more favorably. Although our study cannot answer this question, prior studies (e.g., Way et al., 2007; Reddy et al., 2003) found that aspects of the school climate were uni-directionally related to student development, implying that changes in student development might indeed be attributable to changes in teaching quality over time rather than vice versa.

Conclusion

In sum, the results suggest that students are sensitive to their learning environment not only during school transitions (e.g., Eccles et al., 1996), but also during secondary school itself. The fact that student development was closely linked to students' perceptions of teaching quality over time highlights teachers' potential to foster the development of young adolescent students. Maintaining high teaching quality over time promotes the academic achievement and engagement of all students within a class (i.e., the classroom level), and the social and emotional school adjustment of individual students (i.e., student level).

Moreover, the results of our study demonstrate that teaching quality, in terms of teachers' organizational, social and instructional interactions with students, is variable over time, thus highlighting the role of teaching quality as a critical aspect of students' learning environment. Potential reasons for these changes are manifold, and further research should address why students and teachers tend to perceive declines in teaching quality over time and what accounts for variation in their perceptions. Thus, our study's results underscore the importance of not only examining the associations between teaching quality and student outcome variables at single time points, but also considering the dynamic nature of teaching quality and examining the effects of changes in teaching quality over time on student development, specifically students' academic achievement. We encourage future research to further examine the longitudinal relations between supportive teacher behavior and students' academic achievement.

References

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology, 110*(8), 1066–1083. https://doi.org/10.1037/edu0000256

Archambault, I., Janosz, M., Morizot, J., & Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: relationship to dropout. *The Journal of school health*, *79*(9), 408–415. https://doi.org/10.1111/j.1746-1561.2009.00428.x

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., … Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente* [Professional competence of teachers, cognitively activating mathematics instruction, and the development of students' mathematical literacy (COACTIV): Documentation of measurement instruments]. Max Planck Institute for Human Development.

Baumert, J., Gruehn, S., Heyn, S., Köller, O. & Schnabel, K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation- Band 1* [Educational pathways and psychosocial development in adolescence. Dokumentation-Volume 1]. Max-Planck-Institute for Human Development.

Benner, A. D. (2011). The transition to high school: Current knowledge, future directions. *Educational Psychology Review, 23*(3), 299–328. https://doi.org/10.1007/s10648-011-9152-0

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press

Brophy, J. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 17–43). Lawrence Erlbaum Associates Publishers.

Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal, 55*(6), 1233 – 1267.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366. https://doi.org/10.1080/09243450600697242

Dotterer, A. M., McHale, S. M., & Crouter, A. C. (2009). The development and correlates of academic interests from childhood through adolescence. *Journal of Educational Psychology, 101*(2), 509–519. https://doi.org/10.1037/a0013987

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2015). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *The Journal of Early Adolescence, 35*(5-6), 722–758. https://doi.org/10.1177/0272431614564059

Eccles, J. S., & Roeser, R. W. (2011). School and community influences on human development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 571–643). Psychology Press.

Eccles, J. S., Lord, S., & Buchanan, C. M. (1996). School transitions in early adolescence: What are we doing to our young people? In J. A. Graber, J. Brooks-Gunn, & A. C. Petersen (Eds.), *Transitions through adolescence: Interpersonal domains and context* (pp. 251–284). Lawrence Erlbaum Associates, Inc.

Eccles, J. S., Lord, S., & Midgley, C. (1991). What are we doing to early adolescents? The impact of educational contexts on early adolescents. *American Journal of Education, 99*(4), 521–542. https://doi.org/10.1086/443996

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*(2), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

Eccles, J. S., & Midgley, C. (1989). Stage/environment fit: Developmentally appropriate classrooms for early adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (pp. 139–181). Academic Press.

Emmer, E. T. & Strough, L. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*(2), 103–112. https://doi.org/10.1207/S15326985EP3602_5

Emmer, E. T., & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 407–437). Lawrence Erlbaum Associates Publishers.

Engels, M. C., Colpin, H., Van Leeuwen, K., Bijttebier, P., Den Noortgate, W. V., Claes, S., Goossens, L., & Verschueren, K. (2017). School engagement trajectories in adolescence: The role of peer likeability and popularity. *Journal of School Psychology*, 64, 61–75. https://doi.org/10.1016/j.jsp.2017.04.006

Engels, M. C., Colpin, H., Van Leeuwen, K., Bijttebier, P., Van Den Noortgate, W., Claes, S., Goossens, L., & Verschueren, K. (2016). Behavioral engagement, peer status, and teacher-student relationships in adolescence: A longitudinal study on reciprocal influences. *Journal of youth and adolescence*, *45*(6), 1192–1207. https://doi.org/10.1007/s10964-016-0414-5

Evans, D., Borriello, G. A., & Field, A. P. (2018). A review of the academic and psychological impact of the transition to secondary education. *Frontiers in psychology*, *9*, 1482. https://doi.org/10.3389/fpsyg.2018.01482

Fauth, B., Atlay, C., Dumont, H., & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, 101355. https://doi.org/10.1016/j.learninstruc.2020.101355

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29,* 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., & Trautwein, U. (2020). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational* Psychology, 112(6), 1284–1302. https://doi.org/10.1037/edu0000416

Feldlaufer, H., Midgley, C., & Eccles, J. S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high school. *The Journal of Early Adolescence, 8*(2), 133–156. https://doi.org/10.1177/0272431688082003

Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 98–143). John Wiley & Sons.

Fraser, B. J., & O'Brien, P. (1985). Student and teacher perceptions of the environment of elementary school classrooms. *The Elementary School Journal, 85*(5), 567–580. https://doi.org/10.1086/461422

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). Academic engagement : Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59- 109. https://doi.org/10.3102/00346543074001059

Furman, W., & Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child Development*, *63*(1), 103–115. https://doi.org/10.2307/1130905

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95*(1), 148–162.

https://doi.org/10.1037/0022-0663.95.1.148

Gaertner, H., & Brunner, M. (2018). Once good teaching, always good teaching? The differential stability of student perceptions of teaching quality. *Educational Assessment, Evaluation and Accountability*, *30*(2), 159–182. https://doi.org/10.1007/s11092-018-9277-5

Ganzeboom H.B.G., & Treiman D.J. (2003). Three iternationally standardised measures for comparative research on occupational status. In Hoffmeyer-Zlotnik J.H.P., Wolf C. (Eds.), *Advances in cross-national comparison*. Springer. https://doi.org/10.1007/978-1-4419-9186-7_9

Gehlbach, H., Brinkworth, M. E., & Harris, A. D. (2012). Changes in teacher–student relationships. *British Journal of Educational Psychology, 82*(4), 690–704.

https://doi.org/10.1111/j.2044-8279.2011.02058.x

Gillen-O'Neel, C., & Fuligni, A. (2013). A longitudinal study of school belonging and academic motivation across high school. *Child Development*, 84(2), 678- 692. https://doi.org/10.1111/j.1467-8624.2012.01862.x

Göllner, R., Fauth, B., Lenske, G., Praetorius, A.K., Wagner, W. (2020). Do student ratings of classroom management tell us more about teachers or classroom composition? *Zeitschrift für Pädagogik, 66,* 156–172.

Göllner, R., Wagner, W., Eccles, J. S., & Trautwein, U. (2018). Students' idiosyncratic perceptions of teaching quality in mathematics: A result of rater tendency alone or an expression of dyadic effects between students and teachers? *Journal of Educational Psychology, 110*(5), 709–725. https://doi.org/10.1037/edu0000236

Grossman, P., Cohen, J., & Brown, L. (2014). Understanding instructional quality in English Language Arts: Variations in the relationship between PLATO and value-added by content and context. In K. Kerr, R. Pianta, & T. Kane (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 303–331). Jossey-Bass

Hamre, B. K. & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization and measurement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). Routledge.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hughes, J. N., & Cao, Q. (2018). Trajectories of teacher-student warmth and conflict at the transition to middle school: Effects on academic engagement and achievement. *Journal of School Psychology, 67,* 148–162. https://doi.org/10.1016/j.jsp.2017.10.003

Hughes, J. N., Im, M. H., & Allee, P. J. (2015). Effect of school belonging trajectories in grades 6-8 on achievement: Gender and ethnic differences. *Journal of School Psychology*, *53*(6), 493–507. https://doi.org/10.1016/j.jsp.2015.08.001

Jonkmann, K., Rose, N., & Trautwein, U. (2013). Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und

Mittelschulen in Sachsen [Tradition and innovation: Developmental pathways in the lower and intermediate track in Baden-Württemberg and at comprehensive schools in Saxony]. Abschlussbericht für die Länder Baden-Württemberg und Sachsen.

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

Kittl, H., Mayr, A., & Schiffer, B. (2005). *Early school leaving and drop out. Interim report 2005*. Graz Institute of Education.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Köller, O., Baumert, J., Cortina, K. S. & Trautwein, U. (2010). Bildungsverläufe und psychosoziale Entwicklung im Jugendalter und jungen Erwachsenenalter [Educational pathways and psychosocial development in early and late adolescenc]. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Eds.), *Bildungspsychologie* [Educational psychology] (pp. 245 – 252). Hogrefe.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. http://dx.doi.org/10.1037/a0032583

Lazarides, R., Gaspard, H., & Dicke, A.-L. (2019). Dynamics of classroom motivation: Teacher enthusiasm and the development of math interest and teacher support. *Learning and Instruction, 60,* 126–137. https://doi.org/10.1016/j.learninstruc.2018.01.012

Léon, J., Medina-Garrido, E., & Núñez, J. L. (2017). Teaching quality in math class: The development of a scale and the analysis of its relationship with engagement and achievement. *Frontiers in Psychology, 8,* 895. https://doi.org/10.3389/fpsyg.2017.00895

Longobardi, C., Settanni, M., Prino, L. E., Fabris, M. A., & Marengo, D. (2019). Students' psychological adjustment in normative school transitions from kindergarten to high school: Investigating the role of teacher-student relationship quality. *Frontiers in Psychology, 10,* Article 1238. https://doi.org/10.3389/fpsyg.2019.01238

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106. https://doi.org/10.1111/j.1750-8606.2008.00048.x

Maehr, M. L., & Midgley, C. (1991). Enhancing student motivation: A schoolwide approach. *Educational Psychologist, 26*(3-4), 399–427. https://doi.org/10.1207/s15326985ep2603&4_9

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K. & Gonzáles Rodríguez, E. (2018). *PISA 2012 Skalenhandbuch – Dokumentation der Erhebungsinstrumente* [Documentation of measurement instruments]. Waxmann.

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K. & Gonzáles Rodríguez, E. (2018). *PISA 2012 Skalenhandbuch – Dokumentation der Erhebungsinstrumente* [Documentation of measurement instruments]. Waxmann.

Marder, J., Göllner, R., Wagner, W. & Fauth, B. (2021). Ask me, I (Dis)agree! Acquiescence in student ratings of teaching quality in German vocational schools. Studies in Educational Evaluation, 68, 100937. https://doi.org/10.1016/j.stueduc.2020.100937

Maulana, R., Opdenakker, M. C., & Bosker, R. (2014). Teacher-student interpersonal relationships do change and affect academic motivation: a multilevel growth curve

modelling. *The British journal of educational psychology*, *84*(Pt 3), 459–482. https://doi.org/10.1111/bjep.12031

McCaslin, M., Bozack, A. R., Napoleon, L., Thomas, A., Vasquez, V., Wayman, V., & Zhang, J. (2006). Self-regulated learning and classroom management: Theory, research, and considerations for classroom practice. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 223–252). Lawrence Erlbaum Associates Publishers.

Mihaly, K., & McCaffrey, D. F. (2014). Grade level variation in observational measures of teacher effectiveness. In K. Kerr, R. Pianta, & T. Kane (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 9–49) Jossey-Bass.

Morin, A., Arens, A. K., Tracey, D., Parker, P. D., Ciarrochi, J., Craven, R. G., & Maïano, C. (2017). Self-Esteem trajectories and their social determinants in adolescents with different levels of cognitive ability. *American journal on intellectual and developmental disabilities*, *122*(6), 539–560. https://doi.org/10.1352/1944-7558-122.6.539

Morinaj, J., & Hascher, T. (2019). School alienation and student well-being: A cross-lagged longitudinal analysis. *European Journal of Psychology of Education, 34*(2), 273–294. https://doi.org/10.1007/s10212-018-0381-1

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide: Statistical Analysis with Latent Variables.* Eighth Edition. Los Angeles, CA: Muthén & Muthén.

National Research Council. 2012. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. The National Academies Press. https://doi.org/10.17226/13398.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education, 7,* 133–144. https://doi.org/10.1177/1477878509104318

Olivier, E., Galand, B., Morin, A. J. S., & Hospel, V. (2021). Need-supportive teaching and student engagement in the classroom: Comparing the additive, synergistic, and global

contributions. *Learning and Instruction, 71*, Article 101389. https://doi.org/10.1016/j.learninstruc.2020.101389

Opdenakker, M.-C., Maulana, R., & den Brok, P. (2012). Teacher—Student interpersonal relationships and academic motivation within one school year: Developmental changes and linkage. *School Effectiveness and School Improvement, 23*(1), 95–119. https://doi.org/10.1080/09243453.2011.619198

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102(4), 896. https://doi.org/10.1037/a0019545

Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. In M. Shinn & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 21–39). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195327892.003.0002

Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications Inc.

Ravens-Sieberer, U., Klasen, F., Bichmann, H., Otto, C., Quitmann, J., & Bullinger, M. (2013). Erfassung der gesundheitsbezogenen Lebensqualität von Kindern und Jugendlichen [Assessment of Health-Related Quality in Children and Adolescents]. Das Gesundheitswesen, 75(10), 667-678. https://doi.org/10.1055/s-0033-1349555

Raver, C.C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75*, 346-353. https://doi.org/10.1111/j.1467-8624.2004.00676.x

Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A Latent Growth Curve Study. *Development and Psychopathology*, 15, 119-138. https://doi.org/10.1017/S0954579403000075

Rimm-Kaufman, S. E., Baroody, A. E., Larsen, R. A. A., Curby, T. W., & Abry, T. (2015). To what extent do teacher–student interaction quality and student gender contribute to fifth graders' engagement in mathematics learning? *Journal of Educational Psychology, 107*(1), 170–185. https://doi.org/10.1037/a0037252

Roeser, R. W., & Eccles, J. S. (1998). Adolescents' perceptions of middle school: Relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence, 8*(1), 123–158. https://doi.org/10.1207/s15327795jra0801_6

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (2000). School as a context of early adolescents' academic and social-emotional development: A summary of research findings. *The Elementary School Journal, 100*(5), 443–471. https://doi.org/10.1086/499650

Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology, 88*(3), 408–422. https://doi.org/10.1037/0022-0663.88.3.408

Rucinski, C. L., Brown, J. L., & Downer, J. T. (2018). Teacher–child relationships, classroom climate, and children's social-emotional and academic development. *Journal of Educational Psychology*, *110*(7), 992-1004. https://doi.org/10.1037/edu0000240

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7,* Article 110. http://dx.doi.org/10.3389/fpsyg.2016.00110

Stroet, K., Opdendakker, M.C., & Minnaert, Al. (2013). Effects of need supportive teaching on early adolescents‘ motivation and engagement: a review of the literature. *Educational Research Review, 9*, 65-87. https://doi.org/10.1016/j.edurev.2012.11.003

Trautwein, U., Dumont, H., & Dicke, A. (2015). Schooling: Impact on cognitive and motivational development. *International Encyclopedia of the Social & Behavioral Sciences, 21*, 219-124. https://doi.org/10.1016/B978-0-08-097086-8.26056-X

Virtanen, T.E., Räikkönen, E., Engels, M.C., Vasalampi, K., & Lerkkanen, M.-K. (2021). Student engagement, truancy, and cynicism: a longitudinal study from primary school to upper secondary education. *Learning and Individual Differences*, *86*, Article 101972. https://doi.org/10.1016/j.lindif.2021.101972

Wagner, J., Lüdtke, O., Robitzsch, A., Göllner, R., & Trautwein, U. (2017). Self-esteem development in the school context: The roles of intrapersonal and interpersonal social predictors. *Journal of Personality, 86,* 481– 497. https://doi.org/10.1111/jopy.12330

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction, 28,* 1–11. https://doi.org/10.1016/j.learninstruc.2013.03.003

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108,* 705–721. http://dx.doi.org/10.1037/edu0000075

Wang, M. T., & Dishion, T. J. (2012). The Trajectories of Adolescents' Perceptions of School Climate, Deviant Peer Affiliation, and Behavioral Problems During the Middle School Years. *Journal of research on adolescence: the official journal of the Society for Research on Adolescence*, *22*(1), 40–53. https://doi.org/10.1111/j.1532-7795.2011.00763.x

Wang, M.-T., & Eccles, J. S. (2012). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of Research on Adolescence, 22*(1), 31-39. https://doi.org/10.1111/j.1532-7795.2011.00753.x

Wang, M.-T., Chow, A., Hofkens, T., & Salmela-Aro, K. (2015). The trajectories of student emotional engagement and school burnout with academic and psychological development:

Findings from Finnish adolescents. *Learning and Instruction*, 36, 57–65. https://doi.org/10.1016/j.learninstruc.2014.11.004

Way, N., & Greene, M. L. (2006). Trajectories of perceived friendship quality during adolescence: The patterns and contextual predictors. *Journal of Research on Adolescence, 16*(2), 293- 320. doi:10.1111/j.1532-7795.2006.00133.x

Way, N., Reddy, R., & Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: associations with trajectories of psychological and behavioral adjustment. *American journal of community psychology*, *40*(3-4), 194–213. https://doi.org/10.1007/s10464-007-9143-y

Wentzel, K. R. (2003). Sociometric status and adjustment in middle school: A longitudinal study. *The Journal of Early Adolescence, 23*(1), 5–28. https://doi.org/10.1177/0272431602239128

Widlund, A., Tuominen, H., & Korhonen, J. (2021). Development of school engagement and burnout across lower and upper secondary education: Trajectory profiles and educational outcomes. *Contemporary Educational Psychology*, *66*(Complete). https://doi.org/10.1016/j.cedpsych.2021.101997

Witherspoon, D., & Ennett, S. (2011). Stability and change in rural youths' educational outcomes through the middle and high school years. *Journal of youth and adolescence*, *40*(9), 1077–1090. https://doi.org/10.1007/s10964-010-9614-6

Wu, J., & Hughes, J.N. (2014). Teacher Network of Relationships Inventory: Measurement Invariance Across Ages 6 to 15. *School Psychology Quarterly*, 30, 23–36. https://doi.org/10.1037/spq0000063

Supplement

Table S1

Fit Indices for the Freely Estimated Univariate and Unconditional Models

Model	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$	Slope Loadings$_c$ β(0), β(1), β(2), β(3)	Slope Loadings$_b$ β(0), β(1), β(2), β(3)	Intercept M_i (σ_i^2)	Slope M_s (σ_s^2)
Student Reports								
Student Development								
Achievementa	8.76* (7)	.01	.99/.99	.00/.02	0, 0.3, 0.9, 3.0	0, 0.9, 1.8, 3.0	.56* (.27*)	.37* (.00*)
Academic Identification	34.48* (4)	.05	.96/.89	.02/.05	0, 1.8, 2.4, 3.0	0, 0.5, 2.4, 3.0	3.36* (.02*)	-.06* (.00)
School satisfactionc	19.82* (6)	.03	.99/.98	.02/.03	0, 2.3, 2.9, 3.0	0, 1.6, 2.4, 3.0	2.82* (.05*)	-.12* (.01*)
Value Beliefsa (Math/German)	6.73* (5)	.01	.99/.99	.02/.03	0, 2.5, 3.0, 3.0	0, 1.5, 2.2, 3.0	3.08* (.02)	-.15* (.00)
Truancy	10.65* (4)	.02	.98/.93	.02/.08	0, 1.3, 3.7, 3.0	0, 0.3, 1.1, 3.0	1.14* (.00*)	.04* (.00)
Schoolbelongingb	48.52* (5)	.05	.97/.93	.02/.05	0, 2.4, 2.9, 3.0	0, 0.4, 3.0, 3.0	3.20* (.01*)	.02* (.00)
Social Relatedness	no convergence							
Self-esteem	no convergence							
Emotional wellbeing	no convergence							
Teaching Quality								
Disturbances	no convergence							
Monitoringa	10.11* (5)	.02	.99/.98	.01/.06	0, 1.8, 2.5, 3.0	0, 1.3, 1.7, 3.0	3.23* (.03*)	-.08* (.01*)
Teacher ac. supporta	11.98* (5)	.02	.99/.97	.00/.07	0, 2.2, 2.8, 3.0	0, 1.4, 2.1, 3.0	3.29* (.05*)	-.09* (.01*)
Teacher emo. support	No convergence							
Teacher Reports								
Disturbances	3.27* (3)	.00	.99/.99	.00/.03	--	0, 1.6, 2.9, 3.0	2.53* (.61*)	-.06* (.05*)
Monitoringa	7.75* (4)	.02	.80/.70	.00/.06	--	0, 2.0, 3.0, 3.0	3.42* (.16*)	-.04* (.02*)
Teacher ac. supporta	4.75* (4)	.01	.97/.94	.00/.04	--	0, 1.1, 2.6, 3.0	3.61* (.07*)	-.07* (.01*)
Teacher emo. support	4.99* (3)	.02	.95/.89	.00/.08	--	0, 0.9, 2.0, 3.0	3.65* (.06*)	-.07* (.02*)

Note. $*p < .05$.

[a]For model identification, the residual variance at the first measurement time point at the between class level had to be fixed to zero.

[b]For model identification, the residual variance at the last measurement time point at the between class level had to be fixed to zero.

[c]For model identification, the residual variance at the first measurement time point at both levels had to be fixed to zero.

Table S2

Fit Indices and Coefficients for the Unconditional Univariate Models

Model	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$	Intercept M_i (σ_i^2)	Slope M_s (σ_s^2)
Student Reports						
Student Developmemt						
Achievement[a]	13.49* (11)	0.01	.99/.99	.01/.02	.55* (.27*)	.35* (.01*)
Academic Identification	32.65* (8)	0.03	.97/.96	.02/.05	3.37* (.02*)	-.06* (.00)
School satisfaction	50.26* (8)	0.04	.97/.96	.02/.04	2.77* (.04*)	-.11* (.00)
Value Beliefs[a] (Math/German)	27.41* (9)	0.03	.97/.96	.03/.05	3.06* (.02*)	-.14* (.00)
Truancy[a]	13.34* (9)	0.01	.99/.98	.03/.09	1.13* (.00*)	.03* (.00)
Schoolbelonging[b]	64.08* (9)	0.04	.96/.95	.03/.03	3.20* (.01*)	.03* (.00)
Social Relatedness[b]	63.62* (9)	0.04	.97/.96	.03/.05	3.13* (.01*)	.03* (.00)
Self-esteem	32.81* (8)	0.03	.98/.96	.02/.11	3.46* (.02)	.02* (.00)
Emotional wellbeing[b]	99.13* (9)	0.05	.90/.97	.04/.08	4.10* (.01)	-.04* (.00)
Teaching Quality						
Classroom Management[a]	65.47* (9)	0.04	.88/.85	.04/.06	2.42* (.05*)	.01 (.01*)
Monitoring[a]	19.13* (9)	0.02	.99/.98	.02/.03	3.23* (.03*)	-.08* (.01*)
A. teacher support[a]	21.96* (9)	0.02	.98/.97	.02/.06	3.28* (.05*)	-.09* (.01*)
S. teacher support[a]	39.05* (9)	0.03	.95/.94	.03/.07	3.34* (.05*)	-.10* (.01*)
Teacher Reports						
Classroom Management	8.28* (5)	.01	.97/.96	.00/.05	2.51* (.47*)	-.06* (.05*)
Monitoring	7.36* (5)	.01	.88/.85	.00/.06	3.40* (.08*)	-.06* (.01*)
Teacher ac. support	7.34* (5)	.01	.96/.95	.00/.05	3.56* (.06*)	-.07* (.01*)
Teacher emo. support	4.82* (5)	.01	.99/.99	.00/.08	3.65* (.07*)	-.08* (.02*)

Note. *$p < .05$.

[a]For model identification, the residual variance at the first measurement time point at the classroom level had to be fixed to zero.

[b]For model identification, the residual variance at the last measurement time point at the classroom level had to be fixed to zero.

Table S3

Impact of the Covariates at the Student Level (Student Reports)

	Gender		Migra		SES	
Model	To Interc.	To Slope	To Interc.	To Slope	To Interc.	To Slope
Student Development						
Achievement	.02 (.03)	**-.15 (.07)***	**-.12 (.03)***	-.03 (.08)	**.06 (.02)***	-.03 (.06)
Academic Ident.	-.04 (.03)	.04 (.04)	**.14 (.03)***	.01 (.04)	**-.05 (.03)***	.01 (.03)
School satisfaction	**-.20 (.03)***	.06 (.04)	**.11 (.03)***	-.02 (.03)	-.02 (.03)	.01 (.04)
Value Beliefs	.01 (.03)	.07 (.05)	**.08 (.03)***	**-.15 (.06)***	-.01 (.02)	.01 (.04)
Truancy	**.17 (.03)***	.08 (.04)	.01 (.03)	.01 (.05)	.00 (.03)	-.03 (.04)
Schoolbelonging	.02 (.02)	.06 (.04)	**.08 (.02)***	**.08 (.04)***	-.02 (.02)	**-.10 (.03)***
Social Relatedness	.02 (.02)	.03 (.03)	**.09 (.03)***	.02 (.03)	-.01 (.02)	**-.10 (.04)***
Self-esteem	**.18 (.02)***	**.14 (.05)***	.03 (.03)	.03 (.06)	**-.05 (.02)***	.02 (.04)
Emotional wellbeing	.01 (.03)	**.10 (.04)***	-.01 (.03)	**-.10 (.05)***	-.05 (.03)	**-.09 (.04)***
Teaching Quality						
Disturbances	.03 (.03)	.06 (.05)	.01 (.03)	**-.12 (.07)***	**-.06 (.03)***	.06 (.05)
Monitoring	**-.10 (.03)***	.01 (.04)	**.08 (.03)***	-.02 (.05)	**-.05 (.03)***	-.03 (.04)
Teacher ac. support	**-.12 (.03)***	.02 (.04)	.03 (.03)	**-.08 (.04)***	**-.05 (.03)***	-.01 (.04)
Teacher emo. support	**-.13 (.03)***	.04 (.04)	.00 (.02)	**-.09 (.04)***	-.03 (.02)	.01 (.04)

Note. *$p < .05$. Regression weights are shown in their standardized form.

Table S4

Impact of the Covariates at the Classroom Level (Student Reports)

	Mittel_D		Real_D		Gender		Migra		SES	
Model	To Interc.	To Slope	To Interc.	To Slope	To Interc.	To Slope	To Interc.	To Slope	To Interc.	To Slope
Student Reports										
Student Development										
Achievement	**.44 (.11)***	.24 (.27)	**.50 (.09)***	.13 (.25)	**.08 (.06)***	-.15 (.14)	**-.21 (.09)***	-.28 (.29)	**.22 (.08)***	.15 (.21)
Acad. Identification	**-.37 (.14)***	**-.81 (.29)***	.04 (.15)	**-.42 (.20)***	-.13 (.10)	-.06 (.13)	**.45 (.12)***	**-.74 (.27)***	-.03 (.11)	-.06 (.19)
School satisfaction	**-.43 (.15)***	.29 (.22)	-.20 (.13)	-.12 (.17)	**-.26 (.09)***	.03 (.13)	**.33 (.14)***	-.30 (.22)	.19 (.14)	-.11 (.17)
Value Beliefs	**-.42 (.18)***	**-1.4 (.52)***	**-.40 (.16)***	**-.66 (.32)***	-.03 (.13)	.38 (.23)	**.32 (.15)***	**-.92 (.45)***	.06 (.16)	-.50 (.33)
Truancy	**-.50 (.15)***	-.08 (.36)	-.25 (.19)	-.05 (.26)	.03 (.15)	-.01 (.18)	.18 (.16)	-.19 (.31)	-.14 (.14)	-.29 (.29)
Schoolbelonging	.04 (.21)	**-.95 (.38)***	**.63 (.20)***	-.56 (.27)	.11 (.12)	.12 (.16)	.15 (.25)	-.05 (.32)	**.38 (.16)***	.08 (.22)
Social Relatedness	**-.48 (.20)***	-.32 (.31)	**.55 (.20)***	-.44 (.23)	.21 (.16)	-.21 (.17)	-.05 (.21)	.33 (.36)	.10 (.15)	.33 (.29)
Self-esteem	**-.63 (.21)***	-.85 (.71)	.16 (.18)	.16 (.42)	.04 (.11)	.34 (.34)	-.01 (.21)	-.85 (.78)	.09 (.16)	-.50 (.53)
Emotional wellbeing	-.37 (.24)	-.21 (.45)	.06 (.16)	-.49 (.39)	-.05 (.12)	-.04 (.26)	-.07 (.23)	-.38 (.46)	.27 (.15)	.08 (.33)
Teaching Quality										
Disturbances	.18 (.17)	.19 (.19)	.09 (.14)	-.13 (.15)	-.08 (.11)	-.04 (.12)	.18 (.16)	-.06 (.18)	.04 (.15)	.13 (.13)
Monitoring	-.12 (.15)	-.04 (.22)	-.09 (.13)	-.13 (.17)	-.10 (.09)	.03 (.13)	**.28 (.13)***	-.14 (.24)	-.03 (.12)	.25 (.16)
A. teacher support	-.15 (.16)	-.20 (.18)	-.10 (.15)	-.24 (.15)	-.16 (.09)	-.10 (.12)	.12 (.18)	-.11 (.19)	-.01 (.14)	**.24 (.12)***
S. teacher support	.02 (.18)	-.24 (.19)	-.02 (.13)	**-.32 (.15)***	-.15 (.09)	-.12 (.11)	.16 (.17)	-.16 (.20)	.05 (.14)	**.26 (.12)***
Teacher Reports										
Disturbances	**.38 (.14)***	.03 (.19)	.04 (.14)	.02 (.19)	-.17 (.09)	-.03 (.12)	-.16 (.14)	.08 (.21)	.03 (.12)	-.01 (.17)
Monitoring	-.05 (.19)	.18 (.26)	.07 (.15)	.16 (.21)	-.11 (.12)	-.22 (.13)	-.18 (.19)	.10 (.25)	-.05 (.17)	.21 (.20)
Teacher ac. support	-.15 (.16)	.01 (.25)	-.13 (.14)	-.05 (.19)	.10 (.09)	**-.31 (.14)***	-.20 (.16)	-.05 (.15)	-.02 (.14)	-.00 (.18)
Teacher emo. support	-.25 (.15)	-.21 (.18)	-.02 (.15)	.03 (.17)	.-03 (.10)	-.19 (.11)	-.25 (.17)	-.00 (.18)	.05 (.16)	.11 (.16)

Note. *$p < .05$. Regression weights are shown in their standardized form.

Table S5

Impact of Homeroom Teacher Changes at the Classroom Level (Student Reports)

Model	Homeroom Teacher Change		
	T2	T3	T4
Student Reports			
Student Development			
Achievement	.10 (.07)	.12 (.06)	-.00 (.07)
Acad. Identification	.12 (.13)	-.03 (.09)	-.04 (.12)
School satisfaction	.03 (.05)	.08 (.09)	**.38 (.12)***
Value Beliefs	-.03 (.15)	.25 (.14)	.28 (.16)
Truancy	-.03 (.16)	-.15 (.13)	.11 (.16)
Schoolbelonging	.09 (.12)	-.02 (.13)	-.22 (.18)
Social Relatedness	-.04 (.14)	.02 (.12)	**-.35 (.17)***
Self-esteem	-.11 (.15)	-.02 (.15)	**-.46 (.17)***
Emotional wellbeing	**-.30 (.13)***	**.28 (.12)***	.24 (.24)
Teaching Quality			
Class. Management	**-.16* (.07)**	**.24* (.08)**	**.27* (.09)**
Monitoring	-.03 (.13)	**.30* (.09)**	.16 (.10)
Teacher ac. support	.03 (.11)	**.24* (.09)**	**.19* (.09)**
Teacher emo. support	.03 (.11)	**.25* (.09)**	**.24* (.10)**
Teacher Reports			
Disturbances	-.11 (.07)	-.05 (.08)	.02 (.10)
Monitoring	**-.29* (.13)**	-.07 (.14)	-.08 (.16)
Teacher ac. support	**-.19* (.09)**	.04 (.09)	.08 (.11)
Teacher emo. support	-.16 (.12)	.02 (.09)	-.02 (.09)

Note. *$p < .05$. Homeroom teacher changes from one school year to the next was included time-varying covariate. Student values at a given measurement time point were regressed on homeroom teacher change at that measurement time point.

Table S6

Associations between Student Development and Teaching Quality in the Unconditional Models (Student Reports)

	Model Fit				Association			
					Student level		Classroom level	
Model	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$	$r_{intercept}$	r_{slope}	$r_{intercept}$	r_{slope}
Achievement + Disturbances	93.98* (44)	.02	.97/.96	.03/.07	**-.14* (.05)**	.02 (.13)	-.08 (.13)	.16 (.23)
Achievement + Monitoring	56.60* (44)	.01	.99/.99	.02/.05	.04 (.04)	.15 (.11)	**-.45* (.12)**	**.56* (.18)**
Achievement + Teacher ac. support	64.74* (44)	.01	.99/.99	.02/.07	.01 (.04)	.03 (.10)	-.20 (.10)	**.41* (.16)**
Achievement + Teacher emo. support	74.46* (44)	.02	.99/.98	.02/.07	.05 (.04)	.05 (.11)	-.06 (.11)	**.42* (.18)**
Academic Engagement								
Academic Ident. + Disturbances[a]	124.11* (33)	.03	.95/.92	.03/.05	**.30* (.07)**	**.43* (.15)**	**.46* (.13)**	.34 (.24)
Academic Ident. + Monitoring[a]	59.41* (33)	.02	.99/.98	.02/.05	**.60* (.06)**	**.69* (.11)**	**.78* (.14)**	**.88* (.29)**
Academic Ident. + Teacher ac. support[a]	78.88* (33)	.02	.98/.97	.02/.06	**.56* (.07)**	**.62* (.13)**	**.48* (.14)**	**.58* (.22)**
Academic Ident. + Teacher emo. support[a]	98.94* (33)	.02	.97/.95	.02/.07	**.55* (.06)**	**.66* (.14)**	**.44* (.14)**	**.69* (.22)**
School satisfaction + Disturbances	168.16* (41)	.03	.95/.93	.03/.07	**.48* (.05)**	**.38* (.09)**	**.52* (.12)**	**.55* (.21)**
School satisfaction + Monitoring	116.69* (41)	.02	.97/.96	.02/.05	**.51* (.04)**	**.39* (.07)**	**.80* (.10)**	.50 (.29)
School satisfaction + Teacher ac. support	120.52* (41)	.02	.97/.96	.02/.07	**.56* (.05)**	**.43* (.08)**	**.58* (.09)**	**.48* (.23)**
School satisfaction + Teacher emo. support	148.68* (41)	.03	.96/.94	.03/.07	**.57* (.04)**	**.53* (.10)**	**.50* (.09)**	**.45* (.22)**
Value Beliefs + Disturbances	143.09* (42)	.03	.92/.90	.04/.08	**.45* (.07)**	**.53* (.14)**	**.57* (.15)**	.36 (.31)
Value Beliefs + Monitoring	95.23* (42)	.02	.97/.95	.03/.08	**.58* (.05)**	**.58* (.12)**	**.59* (.14)**	.29 (.30)
Value Beliefs + Teacher ac. support	91.47* (42)	.03	.97/.97	.02/.13	**.32* (.06)**	.17 (.11)	.16 (.24)	.31 (.20)
Value Beliefs + Teacher emo. support	139.30* (42)	.03	.94/.92	.03/.10	**.57* (.05)**	**.62* (.13)**	**.59* (.14)**	**.52* (.25)**
Truancy + Disturbances[a]	86.69* (34)	.02	.94/.90	.03/.08	**-.29* (.11)**	**-.41* (.17)**	-.16 (.27)	-.23 (.28)
Truancy + Monitoring[a]	52.28* (34)	.01	.98/.97	.02/.08	**-.28* (.10)**	-.16 (.13)	.29 (.36)	.05 (.42)
Truancy + Teacher ac. support[a]	52.68* (34)	.01	.98/.97	.02/.08	**-.33* (.10)**	-.22 (.12)	.30 (.28)	.30 (.41)
Truancy + Teacher emo. support[a]	66.14* (34)	.02	.97/.96	.02/.08	**-.32* (.09)**	-.14 (.14)	.13 (.28)	.24 (.38)

Social School Adjustment								
Schoolbelonging + Disturbances	157.83* (42)	.03	.95/.93	.03/.08	**.30* (.05)**	.12 (.10)	**.46* (.17)**	.76 (.50)
Schoolbelonging + Monitoring	124.04* (42)	.02	.97/.96	.03/.09	**.22* (.05)**	**.16* (.08)**	-.06 (.19)	.46 (.37)
Schoolbelonging + Teacher ac. support	114.82* (42)	.02	.97/.96	.03/.07	**.28* (.04)**	**.27* (.08)**	.21 (.18)	**.79* (.40)**
Schoolbelonging + Teacher emo. support[a]	143.74* (42)	.03	.96/.95	.03/.08	**.32* (.04)**	**.31* (.09)**	**.40* (.17)**	**.86* (.37)**
Social relatedness + Disturbances	140.65* (42)	.03	.96/.94	.03/.06	**.26* (.04)**	.04 (.09)	**.48* (.20)**	.25 (.22)
Social relatedness + Monitoring	109.54* (42)	.02	.98/.97	.02/.09	**.17* (.04)**	.13 (.07)	.15 (.21)	.12 (.24)
Social relatedness + Teacher ac. support	108.05* (42)	.02	.97/.97	.02/.08	**.21* (.04)**	**.16* (.07)**	.28 (.19)	.34 (.20)
Social relatedness + Teacher emo. support	137.74* (42)	.03	.96/.95	.03/.08	**.20* (.04)**	.13 (.08)	**.40* (.17)**	**.44* (.20)**
Emotional School Adjustment								
Self-esteem + Disturbances	132.29* (41)	.03	.94/.93	.03/.10	**.32* (.06)**	**.42* (.13)**	.29 (.20)	.62 (.86)
Self-esteem + Monitoring	72.19* (41)	.01	.98/.98	.02/.10	**.31* (.05)**	**.21* (.10)**	.27 (.21)	.76 (1.0)
Self-esteem + Teacher ac. support	75.15* (41)	.02	.98/.98	.02/.10	**.38* (.05)**	**.39* (.11)**	.26 (.18)	.21 (.39)
Self-esteem + Teacher emo. support	99.54* (41)	.02	.97/.96	.02/.11	**.41* (.05)**	**.34* (.12)**	**.34* (.16)**	.56 (.45)
Emo. Well-being + Disturbances	188.60* (42)	.03	.91/.88	.03/.08	**.25* (.06)**	.06 (.11)	**.62* (.27)**	.77 (.97)
Emo. Well-being + Monitoring	119.63* (42)	.02	.96/.94	.02/.09	**.32* (.05)**	**.24* (.09)**	.26 (.22)	.89 (1.3)
Emo. Well-being + Teacher ac. support	132.76* (42)	.03	.95/.93	.03/.09	**.33* (.05)**	**.28* (.10)**	.24 (.22)	.85 (1.1)
Emo. Well-being + Teacher emo. support	154.04* (42)	.03	.94/.92	.03/.10	**.34* (.05)**	**.23* (.10)**	.33 (.21)	.91 (.86)

Note. *$p < .05$. Correlation coefficients are shown in their standardized form.

[a]We allowed for time specific correlations between the residuals of Academic Identification/Truancy and Teaching Quality at each measurement time point, because the assessment of these aspects took place on the same day of assessment.

Table S7

Associations between Student Development and Teaching Quality in the Unconditional Models (Teacher Reports)

	Model Fit				Association	
					Classroom level	
Model	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$	$r_{intercept}$	r_{slope}
Achievement + Disturbances	41.34* (28)	.03	.99/.99	.01/.04	.19 (.13)	.01 (.13)
Achievement + Monitoring	36.43* (28)	.01	.99/.99	.01/.06	-.05 (.15)	.21 (.31)
Achievement + Teacher ac. support	34.83* (28)	.01	.99/.99	.01/.05	.10 (.13)	-.08 (.26)
Achievement + Teacher emo. support	34.76* (28)	.01	.99/.99	.01/.08	.13 (.10)	-.19 (.22)
Academic Engagement						
Academic Ident. + Disturbances	61.25* (25)	.02	.97/.96	.02/.08	.07 (.16)	.32 (.20)
Academic Ident. + Monitoring	51.92* (25)	.02	.97/.96	.02/.06	.19 (.17)	.00 (.29)
Academic Ident. + Teacher ac. support	45.55* (25)	.02	.98/.97	.02/.05	-.07 (.17)	-.05 (.25)
Academic Ident. + Teacher emo. support	53.85* (25)	.02	.97/.97	.02/.05	-.24 (.14)	-.42 (.24)
School satisfaction + Disturbances	76.55* (25)	.02	.97/.96	.02/.08	.03 (.14)	.30 (.26)
School satisfaction + Monitoring	73.42* (25)	.02	.97/.96	.02/.06	.14 (.16)	-.07 (.29)
School satisfaction + Teacher ac. support	66.81* (25)	.02	.97/.97	.02/.04	.10 (.16)	.34 (.33)
School satisfaction + Teacher emo. support	70.15* (25)	.02	.97/.96	.02/.05	.11 (.13)	-.08 (.28)
Value Beliefs + Disturbances	51.75* (26)	.02	.97/.96	.03/.10	.08 (.17)	.02 (.30)
Value Beliefs + Monitoring	47.13* (26)	.02	.97/.96	.03/.09	.06 (.21)	-.36 (.42)
Value Beliefs + Teacher ac. support	41.70* (26)	.02	.98/.97	.03/.06	.09 (.22)	.31 (.45)
Value Beliefs + Teacher emo. support	47.30* (26)	.02	.97/.96	.03/.07	.15 (.15)	.39 (.36)
Truancy + Disturbances	53.24* (26)	.02	.95/.94	.03/.10	**-.77* (.21)**	-.50 (.50)
Truancy + Monitoring	32.96* (26)	.01	.98/.98	.03/.08	.33 (.23)	-.17 (.48)

Truancy + Teacher ac. support	32.54* (26)	.01	.99/.98	.03/.06	.21 (.22)	.08 (.42)
Truancy + Teacher emo. support	33.89* (26)	.01	.98/.98	.03/.08	.08 (.20)	.78 (.57)
Social School Adjustment						
Schoolbelonging + Disturbances	85.36* (26)	.02	.97/.96	.03/.08	**.46* (.15)**	**.69* (.27)**
Schoolbelonging + Monitoring	83.43* (26)	.02	.97/.96	.03/.09	.24 (.18)	-.26 (.41)
Schoolbelonging + Teacher ac. support	82.66* (26)	.02	.97/.96	.03/.06	.19 (.17)	.08 (.45)
Schoolbelonging + Teacher emo. support	79.75* (26)	.02	.97/.96	.03/.08	.15 (.17)	-.20 (.37)
Social relatedness + Disturbances	82.28* (26)	.02	.97/.96	.03/.06	.16 (.18)	.25 (.23)
Social relatedness + Monitoring	77.72* (26)	.02	.97/.96	.03/.07	-.00 (.25)	-.57 (.34)
Social relatedness + Teacher ac. support	100.52* (26)	.03	.96/.95	.03/.11	.17 (.21)	-.25 (.32)
Social relatedness + Teacher emo. support	90.42* (26)	.03	.97/.95	.03/.11	.03 (.21)	-.43 (.26)
Emotional School Adjustment						
Self-esteem + Disturbances	47.98* (25)	.02	.98/.97	.02/.09	.02 (.19)	.66 (.90)
Self-esteem + Monitoring	46.91* (25)	.02	.98/.97	.02/.09	.04 (.23)	-.62 (.95)
Self-esteem + Teacher ac. support	53.65* (25)	.02	.98/.97	.02/.10	-.24 (.23)	-.94 (1.5)
Self-esteem + Teacher emo. support	42.45* (25)	.01	.98/.98	.02/.08	-.37 (.22)	-1.5 (3.9)
Emo. Well-being + Disturbances	112.30* (26)	.03	.92/.90	.04/.08	.27 (.21)	.39 (.80)
Emo. Well-being + Monitoring	107.02* (26)	.03	.92/.90	.04/.08	-.01 (.24)	-1.3 (1.5)
Emo. Well-being + Teacher ac. support	112.11* (26)	.03	.92/.89	.04/.08	-.10 (.20)	-.50 (.79)
Emo. Well-being + Teacher emo. support	105.66* (26)	.03	.92/.90	.04/.08	-.20 (.19)	-1.06 (1.2)

Note. *$p < .05$. Correlation coefficients are shown in their standardized form.

5 STUDY 3:

CLASSROOM MANAGEMENT: CAN IT BE TOO MUCH OF A GOOD THING?

Marder, J., Thiel, F., & Göllner, R. (2021). Classroom management: Can it be too much of a good thing? *Manuscript submitted for publication.*

Abstract

Establishing a well-organized classroom conducive to learning is a key element of high-quality teaching. However, less is known about the degree to which the classroom management process is affected by the specific students in the classroom. Using two large-scale datasets of German secondary school students, the present two studies examined the extent to which students' disruptive behavior in the classroom contributes to the effectiveness of teachers' classroom management actions (i.e., monitoring, rule clarity and structure) for students' pre-adjusted math achievement. When accounting for the average level of disruptions in the classroom, both students' disruptive behavior and teachers' monitoring activity were negatively associated with students' math achievement, whereas no associations occurred for teachers' structure and rule clarity. The results suggest that monitoring needs to be adjusted to specific students' behavior and the situational demands of each class, highlighting the complexity of classroom management.

Keywords: Classroom Management, Mathematic Achievement, Student Reports

Classroom Management: Can it be too Much of a Good Thing?

Classroom management, or a teacher's ability to maintain discipline and order in the classroom, has often been identified as a central aspect of good teaching and as a basic condition for students' learning and development in the classroom (Emmer and Stough 2001, Hattie, 2009; Seidel & Shavelson, 2007; Wang et al., 1993). Yet simultaneously managing students' behavior and providing instruction is an area of great concern for both new and experienced teachers (Hamre & Pianta, 2010). Teachers report disciplinary problems in the classroom as a critical factor affecting their mental health and teaching enjoyment (Aldrup et al., 2018b; Aloe et al., 2014; Emmer & Stough, 2001). In addition, teachers often report that they lack practical training in classroom management and have to learn (painfully) through experience in the field (Jones, 2006; Stough & Montague, 2015). Consequently, there is an urgent need for educational research to identify conditions affecting successful classroom management.

In general, classroom management refers to the actions the teacher takes to establish appropriate conditions for teaching and learning in the classroom (Kounin, 1970; Doyle, 2006; Ophardt & Thiel, 2013; 2017). That is, the broad range of processes happening in the classroom need to be aligned to ensure that the program of action can be followed and that allocated learning time can be used efficiently. In this respect, classroom management is particularly challenging because it requires a broad set of management actions (e.g., Creemers et al., 2013; Hamre & Pianta, 2010; Klieme, Pauli & Reusser, 2009), as well as practical knowledge about when and how to react in order to adequately respond to what happens in the classroom (Doyle, 2006; Ophardt & Thiel, 2017).

However, as anyone who has stood in front of a classroom can attest, successful teaching is not solely the result of teachers' actions alone, but needs to be considered in light of the classroom context. From an ecological perspective, the classroom functions as a social system that includes both the teacher and the students. Consequently, the degree to which discipline and order are established in the classroom needs to be understood as an interplay between the teacher and students in the classroom (Doyle, 2006).

Thus, in the present two studies, we distinguish between teachers' management actions (i.e., monitoring) and student actions in the classroom (i.e., disruptive behavior).). Specifically, we examine the extent to which the specific students in the classroom contribute to a) degree to which

discipline and order are established in the classroom and b) to the effectiveness of teachers' managing actions regarding students' academic achievement. To this end, we examine the predictive value of teachers' management actions (i.e., monitoring) for students' academic achievement, accounting for students' disruptive behavior in the classroom, in two different samples of German secondary school students (Study 1 and Study 2). In Study 2, we seek to validate Study 1's findings and assess two additional management actions teachers can take (i.e., rule clarity and structure).

Theoretical Perspective on Classroom Management

Classroom management is a basic condition for the effective use of learning time and has concerned researchers and practitioners for decades. Since the early 20th century, classroom management concepts have evolved from "practitioners' wisdom" on disciplining students (see Brophy, 2006) to more elaborate, evidence-based management strategies teachers take to create, facilitate, and maintain an effective learning environment. Classroom management has consistently been associated with students' learning and development (for an overview, see Seidel & Shavelson, 2007; Wang et al., 1993) and is a core element in most teaching quality frameworks (e.g., Creemers et al., 2013; Hamre & Pianta, 2010; Klieme et al., 2009).

Based on early work by Kounin (1970), today's understanding of classroom management has shifted from "reacting/responding" to student misbehavior to taking proactive actions to promote positive student behavior and prevent disturbances. To encourage positive student behavior and ensure smooth transitions between classroom activities, teachers must communicate clear expectations and establish clear rules and stable routines. To prevent disruptions and ensure effective use of time, teachers must continuously monitor what is happening in the classroom ("withitness"; Kounin, 1970) and redirect misbehavior if necessary (Kunter & Voss, 2013). On a more profound level, classroom management competency comprises both the general organization of the classroom (e.g., expectations, rules and routines), and guiding the program of action (e.g., establishing smooth transitions, monitoring, dealing with disturbances; for a discussion, see Doyle, 2006; Ophardt & Thiel, 2017).

Whereas these aspects, which clearly refer to management actions by the teacher, are included in most teaching quality frameworks, ecological conceptualizations of teaching and learning in the classroom present a more systematic view on classroom management involving

additional aspects: the teacher and the students being taught, as well as the classroom environment (see Emmer & Evertson, 2017; Eccles & Roeser, 2011). This theoretical perspective is consequential for understanding and assessing the complex nature of classroom management: First, order or chaos in the classroom are less a result of the teacher's management actions alone, but are defined and achieved within the specific context of each class, to which the teacher, the students and their specific needs contribute. Relatedly, through their own behavior in the classroom (i.e., disruptive behavior), students may contribute to the effectiveness of teachers' management actions regarding student achievement. For example, Burns et al., (2021) showed that disruptions in the classroom attenuated the positive association between teaching quality and student achievement. Second, classroom management requires not only a set of management actions, but also more complex knowledge about when and how to act in order to sustain classroom order and on-task behavior (Doyle, 2006; Wolff et al., 2021). For example, a defining feature of classrooms is the immediacy of events (Doyle, 2006): Whereas some teaching activities and management strategies can be planned in advance, students' behavior and the flow of the lesson are difficult to anticipate. Accordingly, effective classroom management requires constant awareness of classroom situations (i.e., perceiving and interpreting classroom events) and active decision-making in order to respond to situational demands (i.e., deciding whether and how to act in order to sustain on-task behavior; Ophardt & Thiel, 2017, Wolff et al., 2016). With respect to students' misbehavior specifically, efficient classroom management requires the teacher to anticipate the consequences of both student misbehavior and their reaction to it on the program of action and flow of the lesson. That is, attempts to redirect misbehavior inherently run the risk of calling even more attention to disturbances, meaning that both student misbehavior and teachers' reaction to it may draw attention away from on-task behavior (Doyle, 2006). From this perspective, teachers need to adapt their classroom management strategies to the given situation and the specific students in the classroom. This critical skill also needs to be more strongly considered in empirical research on classroom management.

Classroom Management in Educational Research

A well-organized classroom in which time can be used efficiently for learning is consistently seen as an indication of high-quality teaching (Kounin, 1970; Doyle, 2006). Across various theoretical frameworks, classroom management is conceptualized as an overarching domain comprising a broad range of teachers' management actions that have been shown to be important for organizing the classroom and guiding the program of action (e.g., communicating clear rules, establishing stable routines, monitoring, dealing with disturbances; e.g., Creemers et al., 2013; Hamre & Pianta, 2010; Klieme et al., 2009). Additionally, given that successful classroom management is defined and achieved within the complex social structure of the classroom (Doyle, 2006), aspects referring more to students' actions, (e.g., the amount of disruptions in the classroom), are considered to be indicators of effective classroom management as well. The underlying assumption is that students' disruptive behavior reflects the consequences of teachers' management actions. If there are no indications of classroom disciplinary issues, it is assumed that effective management strategies are in place. Hence, a lack of disciplinary issues and disruptions in the classroom can be seen as an indicator of efficient classroom management (Pianta et al., 2008; Göllner et al., 2020).

Interestingly, more recent work on teaching quality has explicitly distinguished between aspects of classroom management referring more to teachers' management actions (e.g. monitoring, teacher awareness), aspects referring more to students' actions (e.g., disruptive behavior), and aspects involving both (e.g., inefficient time use; for an overview, see Fauth et al., 2020b; Göllner et al., 2020). On the one hand, distinguishing between students' disruptive behavior and teachers' management actions may be a first attempt to consider that order in the classroom is jointly accomplished by the teacher and the students. On the other hand, students' disruptive behavior is among the most frequently used aspects to assess classroom management (Fauth et al., 2020b; Lipowsky et al., 2009), and some instruments (e.g., The Tripod Classroom Environment Survey; Ferguson, 2010) and empirical studies assess classroom management exclusively via student behavior (e.g., Aldrup et al., 2018a; Kane et al., 2012; Fauth et al., 2014; Kunter et al., 2013; Wagner et al., 2016). This might be problematic, because interpreting the level of discipline and order in the classroom solely as the result of teachers' management actions does not appropriately capture the complexity of a classroom in which student and teacher behavior are closely intertwined (Doyle, 1986, 2006). In support of this notion, prior research has shown that

the specific students being taught have an impact on evaluations teaching quality, with the most consistent effects observed for classroom management. For instance, Fauth et al. (2020a) showed that when the same students reported on teaching quality at several measurement time points, their reports revealed a certain amount of temporal consistency, even when the students were taught by different teachers at each measurement time point. Most important for the present study, the highest stability was found for measures of classroom management (r = .40, p < .05), compared to emotional support (r = .09, p < .05) or clarity of instruction (r = .11, p < .05). In addition to that, other studies found students' background characteristics (e.g., gender, SES, migration background, prior interest and achievement) to be associated with evaluations of classroom management (Fauth et al., 2021; Campbell & Ronfeldt, 2018; Göllner et al., 2020; Hochweber et al., 2014). For example, in a study by Göllner et al. (2020), classrooms with a higher proportion of male students as well as those with lower math performance received lower scores on classroom management factors referring more to students than to the teacher. In summary, these findings further underline the notion that teachers' management actions may be more effective in classes with a favorable student composition. Thus, rather than focusing on students' disruptive behavior as an indication of effective classroom management, the appropriateness of teachers' management actions needs to be evaluated in light of the class in which they are applied.

The Present Studies

^ Despite the theoretical agreement that successful classroom management involves both student and teacher actions, little is known empirically about the extent to which the classroom management process is affected by the specific students in the classroom. Addressing this research gap, the present two studies examined the degree to which students' disruptive behavior in the classroom contributes to the association between teachers' management actions and students' academic achievement (i.e., as an indicator of the effectiveness of teachers' management actions). To this end, we used student reports of several key aspects of classroom management from two large-scale datasets (Study 1 and Study 2). Because classroom management is commonly used as an umbrella term for a variety of teaching functions (Brophy, 1999; Emmer & Gerwels, 2006), the studies refer to a variety of conceptually distinct classroom management aspects that refer more to teachers' management actions or students' actions in the classroom, respectively. With regard to the former, we use measures referring to teachers' general organization of the classroom (i.e., rule clarity) and the degree to which the teacher guides the program of action during the lesson (i.e.,

monitoring and structure). To account for students' disruptive behavior in the classroom, we used measures referring to student actions (i.e., disturbances and ineffective time use).

Based on student reports on classroom management, we investigate whether aspects of classroom management referring to teacher and student actions, respectively, can be empirically distinguished. Next, we address the main research questions: First, we examine the degree to which students' background characteristics are associated with classroom management aspects. We expected that students' disruptive behavior would exhibit more pronounced associations with students' background characteristics than teachers' management strategies (Research Question 1). Second, we examine how students' disruptive behavior and teachers' management strategies are associated with student achievement (Research Question 2). To examine the degree to which the students' disruptive behavior contributes to the effectiveness of teachers' management actions, we examine the predictive value of monitoring (Study 1 and Study 2; rule clarity and structure, Study 2) on students' academic achievement while simultaneously taking into account students' disruptive behavior.

Study 1

Method

Sample

The dataset used in the present study stemmed from the German extension of the 2003 cycle of the Programme for International Student Assessment (PISA; OECD, 2004[1]). In this national extension (PISA-E), a nationally representative subsample of 15-year-old PISA students in Grade 9 and their math teachers took part in a follow-up assessment at the end of Grade 10 (Prenzel et al., 2006). Participation in the study was voluntary. Students in the PISA classes were administered achievement tests as well as questionnaires assessing background data and aspects of teaching quality in math class at the end of Grades 9 and 10. The sample consisted of N = 4,645 students (57 % female) from K = 259 secondary school classes. Students were enrolled in intermediate or academic-track schools. Due to time constraints, items in PISA are administered using a multi-matrix design, in which classes are split into halves, and each half is administered a different set of

[1]We thank the German PISA consortium (Prenzel et al., 2007; Prenzel et al., 2013) and the Research Data Centre (FDZ) at the IQB in Berlin for their approval and support in conducting the secondary analysis.

items. A matrix design has the advantage of reducing student burden while maintaining sufficient item coverage across content and student responses. N = 2,917 students (on average 12 per class) provided information on teaching quality in their math class.

Instruments

Classroom Management

We used three well-known measures of classroom management: monitoring, classroom disturbances and inefficient time use. The monitoring scale referred to teachers' management actions, whereas the disturbances and inefficient time use scales referred to students' actions in the classroom or to the interplay between students and the teacher (see Table 1 for items). These measures were assessed at the measurement point in Grade 10. Students gave responses on a 4-point Likert scale ranging from 1 = *strongly disagree* to 4 = *strongly agree.* Descriptive statistics for the classroom management indicators are shown in Table 1.

Disturbances. Two items assessed disruptions in the classroom (e.g., "In math, the lesson is often disturbed"; $\alpha = 0.81$).

Ineffective Time Use. Two items assessed the amount of time lost due to disruptions in the classroom, (e.g., "In math, it is long after the beginning of the lesson by the time the students get quiet and start working"; $\alpha = 0.74$).

Monitoring. Four items assessed the extent to which the math teacher is aware of students' actions in the classroom and of students getting distracted (e.g., "The teacher always knows exactly what is going on in class"; $\alpha = 0.73$).

Table 1

Descriptive Statistics for Student Reports on Classroom Management Items

Classroom Management Aspect		M	SD	ICC
Disturbances (Cronbach's α = 0.81)				
Ds 1	In math class, the lesson is often disturbed.	2.43	0.99	0.32
Ds 2	In math, a lot of nonsense is going on all the time.	2.63	1.02	0.31
Ineffective time use (Cronbach's α = 0.74)				
Tu 1	In math, it is long after the beginning of the lesson by the time the students get quiet and start working.	2.62	1.00	0.29
Tu 2	In math, a lot of time in class is wasted.	2.66	1.01	0.26
Monitoring (Cronbach's α = 0.73)				
Mo 1	My teacher always knows exactly what is going on in class.	2.52	0.97	0.26
Mo 2	My teacher always checks our homework very accurately.	2.23	0.98	0.29
Mo 3	My teacher makes sure that we pay attention.	2.83	0.92	0.23
Mo 4	My teacher immediately notices when students start doing something else.	2.61	0.96	0.20

Students' Math Achievement

To measure students' math achievement, we used the standardized math achievement test scores from the PISA dataset. Students' math achievement was assessed at the end of Grade 10. The test covered standard content from the German federal states' curricula for Grade 10 mathematics. To account for students' prior math achievement, we included students' mathematical literacy scores measured in Grade 9 as part of the international PISA assessment. All test items had a closed response format, and different subsets of test items were administered using the multi-matrix design. The tests were scaled using Rasch analysis and the weighted likelihood estimates were used as person parameters.

Students' Background Characteristics

Prior research has suggested that students' background characteristics are related to teaching quality and should be considered in evaluations of teaching quality (Campbell & Ronfeldt, 2018; Cohen & Goldhaber, 2016; Fauth et al., 2021; Göllner et al., 2020). Thus, we took into account students' *prior math achievement* (Grade 9), *gender* (0 = female, 1 = male) and

socioeconomic background (SES) at both the within-class and the between-class level. In PISA, students' socioeconomic background was estimated using the PISA index of economic, social and cultural status (ESCS). This composite score is derived from a range of variables assessing parents' education and occupations as well as material wealth and cultural capital (see PISA 2003 Technical Report for a detailed description). At the between-class level, we additionally controlled for *school track* (1= academic track, 0 = intermediate track).

Statistical analysis

Factor Analyses

First-order Factor Model

To investigate whether students are able to differentiate between theoretically distinct aspects of classroom management, we started by examining the data's factor structure. In a two-level model, the aspects of monitoring, disturbances and ineffective time use were modeled as correlated first-order factors at the within- and between-class level.

Nested factor Model

In order to examine teachers' monitoring while simultaneously accounting for the average level of disruptions in the classroom, we examined a nested factor model. A nested factor model (or bifactor model) assumes a general factor accounting for the communality of all items in the model. Thus, all items in the model should load on to the general factor, and the remaining "specific" factors then explain the residual variation not covered by the general factor (Chen, 2006; Reise, 2013; see Shernoff et al., 2017 and Scherer & Gustafsson, 2015 for an application with student reports). We used all of the indicators of disturbances/effective time use to identify the general factor (see Figure 1). Thus, the general factor in our model should explain the variance resulting from the level of disruptions in the classroom. The specific monitoring factor should then reflect the amount of monitoring unrelated to the level of disruptions in the classroom.

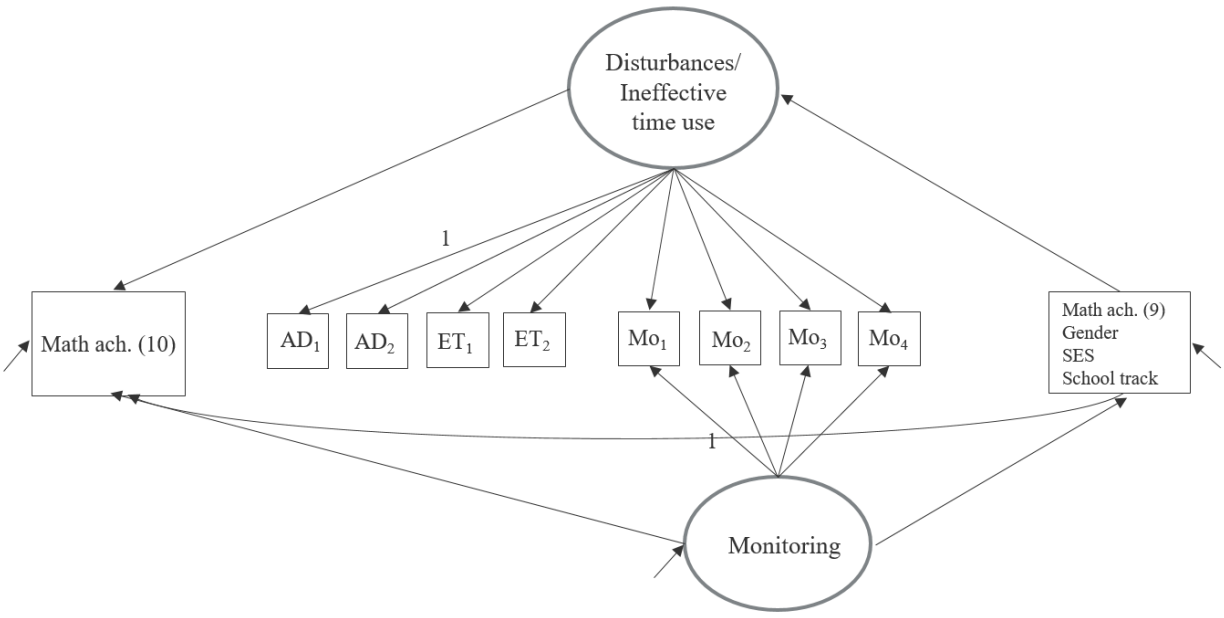

Figure 1. Graphic representation of the nested factor model representing a general (Disturbances/Ineffective time use) and a specific (i.e., Monitoring) classroom management factor. The models were estimated simultaneously at the student and classroom level. The covariate school track was used only at the classroom level.

Students' Background Characteristics and Classroom Management

To examine whether a teacher's classroom management is associated with students' background characteristics, we included students' prior math achievement, gender, and SES as predictors of the general (i.e., level of disruptions) and specific (i.e., monitoring) classroom management factor (see Figure 1). At the between-class level, we additionally controlled for school track (i.e., academic vs. intermediate track).

To avoid a violation of the assumption of uncorrelated general and specific factors when examining the associations with students' background characteristics in the nested model, we applied the residual approach proposed by Koch et al. (2017). Thus, within the nested factor model, the background characteristics were used to predict the general factor and were regressed on the specific factor (see Koch et al., 2017 for more details).

Classroom Management and Students' Math Achievement

Finally, we examined the associations between classroom management measures and students' pre-adjusted math achievement. Specifically, the nested model allowed us to examine the effects of teachers' classroom management strategies (i.e., monitoring) on students' academic achievement accounting for the level of disruptions in the classroom. To control for students' prior math achievement, we used students' math achievement score at the first measurement point (i.e., Grade 9) as a covariate predicting students' math achievement at the second measurement point (i.e., Grade 10).

We evaluated the model fit of the first-order and nested factor models based on their goodness-of-fit indices. Following the guidelines developed for evaluating conventional single-level structural equation models (Browne & Cudeck, 1993; Hu & Bentler, 1999), a good model fit is indicated by root mean square error of approximation (RMSEA) values below .05, comparative fit index (CFI) and Tucker-Lewis index (TLI) values below .95, and standardized root mean square residual (SRMR) values below .08. In addition, a satisfactory model fit was indicated by RMSEA values below .08, CFI/TLI values above .90 and an SRMR below .10 (Muijselar et al., 2017).

We conducted all analyses using Mplus 7.3 (Muthén & Muthén, 1998 –2012). We used robust maximum likelihood estimation (MLR) to obtain reliable standard errors and fit tests for continuous variables and non-normally distributed data. To deal with missing data, we used full information maximum likelihood (FIML) estimation, in which model variables are used to predict missing data. At the within-class level, we centered the covariates around their group means. All continuous variables were z-standardized ($M = 0$, $SD = 1$) before analysis.

Results

First-order Factor Model

The three-factorial model showed a good model fit: $\chi^2(34) = 195.62$, $p < .001$; RMSEA = .04; CFI = 0.97; TLI = 0.95; $SRMR_{student\text{-}level} = .04$; $SRMR_{class\text{-}level} = .04$. However, two aspects addressing the level of disruptions in the classroom (i.e., disturbances and ineffective time use) were nearly perfectly correlated both at the within- ($r = .94$, $p < 0.05$) and the between-class level ($r = .97$, $p < 0.05$). In contrast, these two factors' correlations with monitoring were substantially smaller (within-class level: $-0.31 \leq r \leq -0.32$; between-class level: $-0.81 \leq r \leq -0.85$; all *ps* < 0.05). For this reason, we tested a second model in which we combined disturbances and ineffective time use into a single factor. The resulting two-factor model's fit was comparable to that of the three-factor model: χ^2 (38) = 212.45, $p < .001$; RMSEA = .04; CFI = .97; TLI = .95; $SRMR_s = .04$; $SRMR_c = .04$. For reasons of parsimony, we elected to use the two-factor model in all subsequent analyses. In summary, the revealed factorial structure supports students' ability to differentiate between their own and their teachers' actions in the classroom.

Nested Factor Model

To address our research questions, we applied a nested factor model with one general factor representing the level of disruptions in the classroom and one specific factor with factor loadings on monitoring (see Figure 1). The fit indices indicated a good model fit, χ^2 (32) = 174.26, $p < .001$; RMSEA = 0.04; CFI = 0.97; TLI = 0.95; $SRMR_s = 0.03$; $SRMR_c = 0.04$. The factor loadings are shown in Table S1 in the supplement.

The model fit the data comparably well when constraining the factor loadings for disturbances/ineffective time use (i.e., general factor) and for monitoring (i.e., specific factor) to equality at the within- and between-class level, indicating measurement equivalence across levels; $\Delta\ \chi^2$: -11.19; Δ df: 6; Δ RMSEA: 0.00; Δ CFI: 0.00 and Δ TLI: 0.01; $\Delta\ SRMR_s$: - 0.01 and $\Delta\ SRMR_c$: 0.00.

Students' Background Characteristics and Classroom Management

We examined the association between students' background characteristics (i.e., gender, SES and prior math achievement in Grade 9) and classroom management at the within- and between-class level, while additionally controlling for school track (i.e., academic vs. vocational track) at the between-class level (Research Question 1). Results from the multilevel regression analyses are shown in Table 2.

At the within-class level, students' background characteristics did not significantly account for variation in students' classroom management reports. At the between-class level, however, gender was significantly associated with disturbances/ineffective time use in class. That is, classrooms with a higher proportion of male students reported slightly more disturbances and ineffective time use ($\beta = 0.06$, $SE = 0.02$, $p < 0.05$). SES, prior math achievement and school track were unrelated to disturbances/effective time use, and none of the covariates were associated with teachers' monitoring in the classroom.

Table 2

Associations between Students' Background Characteristics and Classroom Management

	General factor Disturbances /Ineffective time use		Specific factor Monitoring	
Predictor variable (Covariate)	Within-class β (*SE*)	Between-class β (*SE*)	Within-class β (*SE*)	Between-class β (*SE*)
Math Ach.T1	-.03 (.03)	-.15 (.09)	-.01 (.01)	-.07 (.04)
SES	-.02 (.03)	-.03 (.06)	-.01 (.01)	.04 (.06)
Gender	-.01 (.02)	**.06 (.02)***	-.04 (.03)	.15 (.12)
School track	-	-.02 (.06)	-	-.04 (.06)

Note. *$p < .05$. Model fit indices: $\chi^2(83) = 300.35$, $p < .001$; RMSEA = .02; CFI/TLI = .97/.95; $SRMR_c = .03$; $SRMR_s = .05$. Coefficients are shown in their unstandardized form.

Classroom Management and Students' Math Achievement

To examine the associations between classroom management and students' pre-adjusted math achievement (Research Question 2), we regressed students' math achievement in Grade 10 on the classroom management factors and additionally controlled for students' gender, SES, school track, and prior math achievement in Grade 9.

At the within-class level, classroom management was unrelated to individual students' math achievement (disturbances/ineffective time use: $\beta = 0.01$, $SE = 0.02$, $p = 0.937$; monitoring: $\beta = 0.03$, $SE = 0.03$, $p = 0.439$). In contrast, both classroom management factors were associated with students' math achievement at the between-class level. The average level of disruptions was negatively related to the average level of math achievement in a class ($\beta = -0.07$, $SE = 0.03$, $p < 0.05$). In addition, the average level of monitoring in a class was negatively associated with students' math achievement in the nested model ($\beta = -0.24$, $SE = 0.12$, $p < 0.05$). That is, accounting for the level of disruptions in the classroom (i.e., the general factor), monitoring (i.e., the specific factor) was negatively associated with students' math achievement. In sum, the findings suggest that the level of disruptions in the classroom is not only an important predictor for students' academic achievement, but also contributes to the effectiveness of teachers' monitoring in the classroom.

Discussion

Classroom management is usually assessed via a wide range of aspects referring to teacher or student actions in the classroom. However, the results of Study 1 point to empirical and conceptual differences between these classroom management aspects: First, classroom management aspects referring to students' actions (i.e., the level of disruptions) were related to the proportion of male and female students in the classroom, whereas no such relations were found for teachers' actions (i.e., monitoring). Second, the considered classroom management aspects differed in their associations with students' math achievement: A high level of disruptions in the classroom was negatively associated with students' mathematic achievement. However, results from the nested model suggested that monitoring was negatively associated with students' mathematic achievement after accounting for the level of disruptions in the classroom.

Study 2

Study 1 suggested that students' background characteristics are more closely related to classroom management aspects referring to students' than to teachers' actions in the classroom. Specifically, the level of disruptions in the classroom seems to be an important predictor for students' academic achievement, and contributed to the effectiveness of teachers' monitoring.

In Study 2, we examined the same research questions as in Study 1 in order to replicate Study 1's findings. In addition, Study 2 extends Study 1 in the following ways: We used a more diverse dataset in terms of grade level and school track. Specifically, the sample in Study 2 comprises one academic and two vocational school tracks and includes secondary school students from Grade 6 to Grade 10. Most importantly, we aimed to examine a broader range of teachers' management strategies in Study 2. Therefore, Study 2 included additional aspects of classroom management referring to teachers' general organization of the classroom (i.e., rule clarity) and the degree to which the teacher guides the program of action during the lesson (i.e., monitoring and structure).

Methods

The Ministry of Education and Cultural Affairs of the German federal state of Baden-Württemberg approved the study and the data collection (date of approval: February 12, 2018, file number: 31-6600.0/279). The ethics committee of [Institution; blinded for review] confirmed that the procedures were in line with ethical standards for research with human subjects (date of approval: May 4, 2018, file number: A2.5.4-074_aa).

Sample

The present study uses data from the longitudinal large-scale study "Teaching Quality from the Students' Perspective (UNITAS)", which was conducted in the federal state of Baden-Württemberg, Germany. The UNITAS study was designed to examine teaching quality in secondary school via student reports in two different subjects, mathematics and German language arts. To this end, students (and teachers) were administered achievement tests as well as questionnaires assessing background data and aspects of teaching quality in math and German classes. The dataset comprises two measurement points: The first assessment wave took place in

spring 2018. Participants were reassessed a year later, in spring 2019. Participation in the study was voluntary. In the present study, we use data from the second measurement point assessing teaching quality in mathematics.

A total of $N = 6298$ students (51.8% female) from Grades 6 to 10 from 391 classes in 26 schools in Baden-Wuerttemberg participated in the study at the second measurement point. On average, 16 students per class participated in the study. The UNITAS study included students from each of the main school tracks in Baden-Württemberg. The sample comprises $n = 3610$ (57.3%) students from academic-track schools, $n = 2058$ (32.7%) students from intermediate-track schools, $n = 555$ (8.8%) students from multitrack schools, and $n = 75$ (1.2%) students from lower-track schools.

Instruments

Classroom Management

To measure classroom management in Study 2, we used a similar set of items as in Study 1, adding the aspects of rule clarity and structure (see Table 3 for complete set of items and descriptive statistics for the classroom management indicators). The scale *disturbances* referred to students' actions in the classroom or to the interplay between students and the teacher, whereas the scales *monitoring, rule clarity* and *structure* referred to teachers' management actions.

Disturbances. Six items assessed disruptions in the classroom and the degree to which the beginning of the class is delayed ($\alpha = 0.85$). Thus, the aspect of inefficient time use (Study 1) was covered by the disturbances in the classroom scale in this study.

Monitoring. Four items assessed the extent to which the math teacher is aware of students' actions in the classroom and of students getting distracted ($\alpha = 0.76$)

Rule clarity. Three items assessed the extent to which students exactly know what rules apply in the classroom (e.g., "In class, it is clear for us what we are allowed to do and what not"; $\alpha = 0.70$).

Structure. Five items addressed the extent to which the lesson is well-structured and students clearly know what the lesson will be about (e.g., "In our teacher's lessons, we always know what to do next"; $\alpha = 0.69$).

Table 3

Descriptive Statistics for Students Reports on Classroom Management Items

Classroom Management Aspect		***M***	***SD***	**ICC**
Disturbances (Cronbach's α = 0.85)				
Ds 1	In class, sometimes it is loud and chaotic.	2.63	0.94	0.26
Ds 2	The lessons are often disturbed.	2.51	0.86	0.28
Ds 3	At the beginning of a lesson, our teacher sometimes has to wait for a long time until the class calms down.	2.45	0.99	0.23
Ds 4	In class, there is constant loud talking.	2.26	0.80	0.26
Ds 5	Sometimes time is wasted in class.	2.54	0.90	0.29
Ds 6	In class, it usually takes a very long time until everyone is ready to work.	2.28	0.79	0.18
Monitoring (Cronbach's α = 0.76)				
Mo 1	Our teacher knows exactly what is going on in the classroom.	3.02	0.65	0.27
Mo 2	Our teacher intervenes before disturbances and disruptions arise.	2.84	0.75	0.21
Mo 3	Our teacher makes sure that we pay attention.	3.01	0.68	0.17
Mo 4	Our teacher realizes immediately when students are engaged in something else.	2.76	0.81	0.16
Rule Clarity (Cronbach's α = 0.70)				
Rc 1	In class, we have agreed on common rules of behaviour.	2.63	1.15	0.18
Rc 2	We know the rules we have to follow in class.	3.40	0.62	0.07
Rc 3	In class, it is clear for us what we are allowed to do and what not.	3.43	0.56	0.09
Rc 4	We know what happens when we break the rules.	3.21	0.84	0.12
Structure (Cronbach's α = 0.69)				
St 1	Our teacher summarizes the most important information at the end of the lesson.	2.05	0.99	0.12
St 2	In our class, everything that we do is thoroughly planned by our teacher.	3.26	0.65	0.16
St 3	In our teacher's lessons, we always know what to do next.	2.94	0.69	0.12
St 4	At the beginning of the lesson, our teacher summarizes what we did last.	2.74	0.93	0.12
St 5	At the beginning of the lesson, our teacher informs us what we will do today.	2.53	0.95	0.13

Students' Math achievement

To measure students' math achievement, we used a standardized math achievement test (MBK 5-12+; Ennemoser et al., 2011; Krajewski & Ennemoser, 2013). The MBK 5-12+ consists of several timed tasks, such as marking numbers on a number line or number dictation. In the present study, the test revealed satisfactory internal consistency ($\alpha = .83$). To account for students' prior math achievement, we included students' prior math test scores as measured at the first measurement point.

Covariates

To account for students' background characteristics, we included students' *prior math achievement, gender* (0 = female, 1 = male), and *socioeconomic background* (SES) at both the within-class and the between-class level. Students' socioeconomic status was assessed based on the number of books in the students' home (adapted from Kunter et al., 2002). 3.4% of students had no or only a few books at home. 6.4% of students reported enough books to fill a single shelf of a bookshelf, and 13.7% reported enough books to fill a whole bookshelf. 19.4% of students reported having enough books to fill three full bookshelves and 26.4% reported more than 200 books at home (more than three full bookshelves).

At the between-class level, we additionally controlled for *grade level* (i.e., Grade 6 to 10) and *school track*. We combined the school track information into a singly dummy-coded variable (1= academic track, 0 = other).

Statistical Analyses

The modeling approach and statistical analysis were identical to Study 1. Because we included a greater variety of classroom management aspects in Study 2, we evaluated the association between each classroom management aspect (i.e., disturbances, monitoring, rule clarity, structure) and students' math achievement while accounting for the level of disruptions in the classroom. Following the procedures outlined in Study 1, we used all the indicators for disturbances to identify the general factor, whereas the specific factor was based on the indicators of monitoring, rule clarity, or structure, respectively. Due to the complexity of the model, monitoring, rule clarity and structure were entered separately into the nested model.

We expected a negative associations between disturbances, monitoring and students' math achievement, respectively. However, we made no prediction about the association between rule clarity, structure and students' math achievement, respectively.

Results

First-order factor model

To investigate the factor structure of students' classroom management reports, we examined a two-factor model with disturbances and each classroom management aspect (i.e., monitoring, rule clarity and structure) as correlated first-order factors at the within- and between-class level. Across all models, the fit indices showed at least satisfactory fit to the data (see Table S2 in the supplement).

Nested Factor Model

Next, we applied a nested factor model with one general factor (i.e., disturbances) representing the level of disruptions in the classroom and one specific factor with factor loadings on monitoring, rule clarity or structure, respectively. Overall, the fit indices showed a good fit to the data across all models. Across all models, RMSA values were 0.03, further indices ranged between CFI: 0.97 to 0.98 and TLI: 0.93 to 0.95; $SRMR_w$: 0.02 and $SRMR_b$: 0.04 to 0.05 (Table S3). The factor loadings in each model are presented in Tables S4 to S6 in the supplement).

Across all models, constraining the factor loadings for the general and the specific factor to equality across levels led to only minor differences in model fit: Δ RMSEA: 0.00; Δ CFI: 0.00 and Δ TLI: 0.01; Δ $SRMR_w$: 0.00 and Δ $SRMR_b$: 0.00 to 0.03 (Table S3). For each model, we thus assumed measurement equivalence across levels.

Students' Background Characteristics and Classroom Management

We examined the association between students' background characteristics (i.e., gender, SES and prior math achievement in Grade 9) and classroom management at the within- and between-class level, while additionally controlling for school track (i.e., academic vs. vocational track) at the between-class level (Research Question 1). Model fit indices and results from the multilevel regression analyses are shown in Table 4.

At the within-class level, students' background characteristics affected students' reports on classroom management only to a minor degree. At the between-class level, there were more systematic effects of students' background characteristics on the general factor. Across all models, classes with a higher proportion of higher-SES students and higher prior math achievement reported less disruptions in the classroom (prior math achievement: β = - 0.10 to - 0.11, *SE* = 0.02, all *p*s < 0.001; SES: β = - 0.53 to – 0.59, *SE* = 0.08 to 0.09; all *p*s < 0.001. In addition, the level of disruptions (i.e., the general factor) was affected by grade level and school track. On average, students in academic-track schools and in higher grade levels reported a lower level of disruptions (school track: β = - 0.27 to - 0.31, *SE* = 0.05 to 0.06, all *p*s < 0.001; grade level: β = - 0.63 to - 0.68, *SE* = 0.18 to 0.19, all *p*s < 0.001).

However, students' background characteristics (i.e., prior math achievement, SES, gender) were unrelated to the specific factors in all models.

Table 4

Associations between Students' Background Characteristics and Classroom Management

Model	Fit indices				Predictor Variable	Dependent Variable Within-class		Dependent Variable Between-class	
	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$		General factor β (*SE*)	Specific factor β (*SE*)	General factor β (*SE*)	Specific factor β (*SE*)
Monitoring									
	708.66 (153)	.02	.97/.95	.02/.05	Math Ach.T1	-.01 (.00)	.01 (.09)	**-.10 (.02)***	.07 (.29)
					SES	-.01 (.03)	-.01 (.01)	**-.53 (.08)***	.00 (.04)
					Gender	-.01 (.02)	**-.08 (.03)***	.04 (.05)	.10 (.06)
					School track	-	-	**-.27 (.05)***	**-.21 (.07)***
					Grade level	-	-	**-.68 (.18)***	**-.09 (.02)***
Rule Clarity									
	965.69 (153)	.03	.95/.93	.02/.09	Math Ach.T1	-.00 (.00)	.07 (.07)	**-.11 (.02)***	-.10 (.19)
					SES	-.01 (.03)	-.01 (.01)	**-.56 (.08)***	.02 (.03)
					Gender	-.01 (.02)	**-.06 (.03)***	.05 (.05)	.02 (.03)
					School track	-	-	**-.30 (.06)***	**-.09 (.04)***
					Grade level	-	-	**-.68 (.19)***	**-.04 (.01)***
Structure									
	1080.10 (178)	.03	.95/.92	.02/.07	Math Ach.T1	-.01 (.00)	.17 (.10)	**-.11 (.02)***	-.34 (.34)
					SES	-.02 (.03)	-.01 (.01)	**-.59 (.09)***	.02 (.04)
					Gender	-.01 (.02)	-.05 (.04)	.04 (.05)	-.01 (.06)
					School track	-	-	**-.31 (.06)***	-.11 (.07)
					Grade level	-	-	**-.63 (.19)***	.01 (.02)

Note. **p* < .05. The general factor is identified by the indicators of Disturbances. Coefficients are shown in their unstandardized form.

Classroom Management and Students' Math Achievement

To examine the associations between classroom management and students' pre-adjusted math achievement (Research Question 2), we regressed students' math achievement in Grade 10 on the classroom factors and additionally controlled for students' gender, SES, prior math achievement in Grade 9, school track, and grade level. The fit indices and results for the multilevel regression analyses are shown in Table 5.

At both the within- and between-class levels, classroom management was significantly associated with students' math achievement, with the most consistent results for level of disruptions (i.e., general factor). At the within-class level, the amount of disruptions perceived by an individual student was negatively related to individual student achievement across all models (β = - 0.09 to - 0.10, *SE* = 0.02, all *p*s < 0.05). In addition, students who reported more monitoring and more rule clarity (i.e., specific factors) compared to their classmates exhibited higher math achievement (monitoring: β = 0.07, *SE* = 0.03, $p < 0.05$; rule clarity: β = 0.11, *SE* = 0.03, $p < 0.001$). However, the amount of structure perceived by an individual student was not related to his or her math achievement (β = 0.03, *SE* = 0.03, $p = 0.193$).

At the between-class level, the average level of disruptions in class (i.e., general factor) was negatively related to students' average level of math achievement across all models (β = - 0.09, *SE* = 0.05, all *p*s < 0.05). Importantly, when accounting for the amount of disruptions in the classroom, the average level of monitoring (i.e., specific factor) was negatively related to the average level of student achievement (β = - 0.19, *SE* = 0.11, $p < 0.05$). In contrast, the average level of rule clarity (β = - 0.02, *SE* = 0.24, $p = 0.468$) and structure (β = - 0.15, *SE* = 0.11, $p = 0.081$) did not affect student achievement in the nested model. In sum, the findings suggest that when taking into account the level of disruptions in the classroom, teachers' monitoring negatively impacted student achievement. In contrast, no such relation was found for clarity of classroom rules and structure of lessons.

Table 5

Association between Classroom Management on Students' Math Achievement

Model	Fit Indices				Predictor	Dependent Variable: Math Achievement	
	χ^2 (df)	RMSEA	CFI/TLI	$SRMR_{s/c}$		Within-class β (*SE*)	Between-class β (*SE*)
Monitoring							
	764.21 (173)	.02	.97/.95	.02/.05	General factor	**-.09 (.02)***	**-.09 (.05)*** [a]
					Specific factor	**.07 (.03)***	**-.19 (.11)*** [a]
Rule Clarity							
	1004.41 (173)	.03	.97/.93	.02/.09	General factor	**-.10 (.02)***	**-.09 (.05)*** [a]
					Specific factor	**.11 (.03)***	-.02 (.24)
Structure							
	1160.78 (200)	.03	.95/.92	.02/.07	General factor	**-.10 (.02)***	**-.09 (.05)*** [a]
					Specific factor	.03 (.03)	-.15 (.11)

Note. *$p < .05$. The general factor is identified by the indicators of Disturbances. Coefficients are shown in their unstandardized form.
[a]This *p* value is one-tailed because the hypotheses were directional.

Discussion

In Study 2, we tested the same research questions as in Study 1 in order to replicate Study 1's findings in a more diverse sample. Moreover, we included two more classroom management aspects reflecting teachers' management actions. Thus, we used *rule clarity* as an indicator of teachers' general classroom organization, while *monitoring* and *structure* were used as indicators of teachers' guiding the program of action.

As in Study 1, the level of disruptions in the classroom was most consistently related to students' background characteristics. In contrast, teachers' management actions were less systematically associated with students' background characteristics. In addition, the classroom management aspects differed in their association with students' math achievement. As in Study 1, the level of disruptions in the classroom was negatively associated with students' math achievement. Moreover, results from the nested model in Study 2 replicated the negative association between teachers' monitoring and students' math achievement at the between- class level. In contrast, no such negative associations occurred for the other aspects of teachers' classroom management (i.e., rule clarity or structure). In sum, these findings point to empirical and conceptual differences between measures of classroom management referring to students' and teachers' actions in the classroom, respectively.

General Discussion

In educational research, classroom management is a central element in most frameworks of good teaching and has most consistently been related to students' learning and academic development. Even though teaching is inherently a highly interactive process between students and teachers (Doyle, 2006), the interplay between students and teachers is not explicitly considered in the evaluation of classroom management. Using two large samples of German secondary school students who provided classroom management reports, we examine the extent to which the specific students in the classroom shape successful classroom management. The results of the present studies highlight that students significantly contribute to classroom management processes, which is of high importance for educational research and practice: First, classroom management aspects referring more to teachers' actions or students' actions in the classroom, respectively, reflect theoretically and empirically distinct aspects of classroom management. More specifically, in both studies, classroom management aspects referring to students' behavior in the classroom (i.e.,

disruptive behavior) were closely related to students' background characteristics and turned out to be most predictive for students' achievement. Moreover, the results suggest that students' disruptive behavior contributes to the effectiveness of teachers' management actions in the classroom. Thus, the specific classroom context needs to be considered when evaluating teachers' management actions.

Classroom Management and Student Background Characteristics

Classroom management is typically used as an umbrella term for a broad range of aspects describing the extent to which a learning environment conducive to instruction is created and maintained in the classroom (Emmer & Gerwels, 2006). One distinguishing feature of these aspects is the extent to which they refer more to students' or to teachers' actions in the classroom.

The results of the present studies highlight conceptual differences between these two perspectives: First, the factor analyses identified aspects of classroom management referring more to student or to teacher actions as clearly distinct factors. Second, classroom management aspects referring more to student actions in the classroom (i.e., students' disruptive behavior) were more closely related to students' background characteristics than aspects referring to teacher management actions (e.g., monitoring). Consequently, it seems that the level of discipline and order in a classroom is not the result of teachers' actions alone, but is essentially co-constructed by the students in the classroom. This finding is highly consequential for the assessment and evaluation of classroom management, in which students' disruptive behavior is a frequent – and often the only -- indicator of classroom management (e.g., Aldrup et al., 2018a; Kunter et al., 2013; Wagner et al., 2016; Wallace et al., 2016). First, asking about student behavior may not fully allow inferences to be drawn about a teacher's classroom management skills. Instead, students must be considered as an equally important part of the classroom management process. Second, classroom management aspects addressing teacher or student behavior should not be used as co-equal or interchangeable indicators in the evaluation of classroom management, but rather tap into different aspects of the classroom management process.

Classroom Management and Student Achievement

The assumption that students co-construct the classroom management process (e.g., Doyle, 2006; Praetorius et al., 2018) is also consequential with regard to the effectiveness of teachers' classroom management actions. Thus, evaluating a teacher's management actions against the

background of the specific class in which they are applied may yield important insights into the complexity of classroom management. Consequently, we evaluated several aspects of teachers' classroom management in terms of teachers' general organization of the classroom (i.e., rule clarity) and guiding the program of action during the lesson (i.e., monitoring and structure) in light of students' disruptive behavior in class. In line with previous studies, we found that students' disruptive behavior was most consistently related to students' achievement, which may explain why prior research has frequently used students' disruptive behavior as an indicator of classroom management. However, it should be kept in mind that the scales referring to students' disruptive behavior reflect information about both teachers and the students taught, rather than a teacher's classroom management skills per se.

Regarding teachers' management actions, it is highly interesting that both studies revealed a negative association between teachers' monitoring and students' achievement. While this finding was unexpected, it might be explained in terms of the interplay between students and teachers in the classroom. To some degree, monitoring can be considered both a reaction to students' disruptive behavior as well as a proactive strategy to prevent the (further) spread of off-task behavior (Brophy, 1986; Kounin, 1970). Because the nested model accounted for the shared amount of disruptions inherent in all classroom management items, the remaining variance in monitoring may be interpreted as "too much" monitoring in relation to the level of disruptions in a given classroom. There are at least two explanations for why monitoring may have been "too much" in relation to the level of disruptions in the classroom.

First, it needs to be considered that adolescent students are known to be sensitive to experiences of autonomy and control (Eccles & Midgley, 1989; Eccles & Roeser, 2009). In fact, some of the items on the monitoring scale tap into classroom management processes that might have been perceived as controlling behavior by teachers (e.g., *"Our teacher realizes immediately when students are engaged in something else"*). Thus, when adolescent students feel overly surveilled by their teacher, their motivation to engage in learning activities may be compromised (e.g., Patall et al., 2010). This interpretation is supported by Yue's (2021) finding of a negative relationship between teachers' monitoring and students' enjoyment in class.

Second, to the degree that monitoring refers to intervening actions by the teacher (i.e*., "Our teacher intervenes before disturbances and disruptions arise"*), a high amount of monitoring may

itself interfere with the lesson flow or, in the worst-case scenario, call even more attention to misbehavior and thus further compromise learning time (Brophy, 2006; Emmer & Gerwels, 2006). Taken together, it seems that teachers' monitoring activity needs to be adjusted to students' specific needs and to the situational demands of the classroom environment in order to be effective (Olivier et al., 2021; Stroet et al., 2013). This critical skill needs to be more strongly considered in evaluations of teaching quality.

It's important to note that no such negative associations with students' math achievement were found for other aspects of teachers' classroom management (i.e., rule clarity and structure). One might argue that from a theoretical perspective, teachers' general organization of the classroom (i.e., rule clarity) and the degree to which the lesson is well-structured (i.e., structure) refer more to preventive actions the teacher takes to foster desirable student behavior and strengthen the focus on the program of action. In contrast, monitoring describes a more active process that is both proactive and reactive to students' behavior in the classroom.

In sum, the results of our studies demonstrate that classroom management is a complex construct that needs to be distinguished more precisely in empirical research. On the one hand, it is necessary to distinguish between student behavior and teachers' management actions in the classroom. On the other hand, aspects referring more to the teacher also seem to differ greatly with regard to the underlying processes (i.e., proactive or reactive) and conceptual aspect of classroom management (i.e., general classroom organization or guiding the program of action) they address.

Implications and Further Directions

The present studies support the notion that classroom management is a complex endeavor, which can be measured by a range of conceptually distinct aspects. Although they all fall under the broad dimension of classroom management, these aspects tap into different parts of the classroom management process and thus cannot be used as co-equal or interchangeable indicators of classroom management. Specifically, classroom management aspects referring more to students' actions in the classroom do not solely reflect the teacher's ability to provide a well-organized classroom, but rather provide information about the interplay between students and the teacher in the classroom.

With regard to teachers' management actions in the classroom, the results indicate that high-quality classroom management needs to meet the needs of the specific students and the

situational demands of a classroom (Olivier et al., 2021. That is, effective classroom management requires the teacher to determine which events and situations require what kind of actions (Ophardt & Thiel, 2017; Wolff et al., 2021). Specifically with respect to classroom management strategies that are reactive in nature (e.g., interventions), there seems to be a fine line between dealing with inappropriate behavior promptly when it occurs and teachers' management actions themselves interfering with students' feelings of autonomy and the flow of the lesson (Emmer & Gerwels, 2006). In this respect, Ophardt and Thiel (2017) suggest a stepwise approach for effectively dealing with disturbances in the classroom that might provide some orientation. First, the teacher reacts to inappropriate behavior by strengthening the primary vector of action, for example by maintaining eye contact or lifting his or her voice when talking. If necessary, the teacher reminds misbehaving students verbally or nonverbally to comply with the rules. Eventually, the teacher may sanction inappropriate behavior. Thus, when it comes to dealing with disturbances in the classroom, the primary emphasis is on the extent to which the teacher keeps the flow of the lesson going rather than ending misbehavior immediately.

However, the complexity of the classroom management process needs to be more strongly considered in both teacher training and in the assessment and evaluation of classroom management. In order to evaluate teachers' classroom management competencies, items are needed that more strongly capture the adaptivity of teachers' management behavior and more strongly focus on teachers' "withitness" (Kounin, 1970) rather than teacher control.

Limitations

The consideration of a broad range of classroom management aspects referring more to student or to teacher behavior, respectively, is a central advantage of the present study. Nonetheless, some limitations need to be discussed.

First, the present studies relied solely on student reports on classroom management. Because students report on teaching quality from their own idiosyncratic perspectives, combining classroom management aspects referring to student and teacher actions, respectively, may have caused systematic differences in the assessment of these classroom management aspects. That is, providing information about someone else's behavior (i.e., teachers' monitoring) can be expected to impose higher cognitive demands than evaluating one's own behavior (i.e., students' disruptive behavior; Tourangeau et al., 2000), which might have caused differences in evaluations of the

classroom management aspects referring to student and teacher behavior, respectively. To follow up on the suggested explanation that monitoring may have been perceived as controlling teacher behavior, it needs to be investigated whether information provided by other sources, such as observer reports or teacher self-reports, would lead to similar results.

A second limitation refers to the relationship between students' disruptive behavior on the one hand and teachers' management actions on the other hand. Even though we found the level of disruptions in the classroom to be closely related to students' background characteristics, this finding does not mean that teachers' management actions are unimportant for students' disruptive behavior in the classroom. Consequently, the nested model may have overestimated the association between monitoring and students' achievement, because students' disruptive behavior already reflects the consequences of teachers' monitoring to some degree. In future research, assessing students' disruptive behavior and teachers' management actions at several points within a lesson or over the course of the school year would offer a promising opportunity to gain unique insights into the dynamics of student behavior and teachers' management actions.

Lastly, the items addressing ineffective use of time in the classroom loaded onto the disturbances factor, which was interpreted in terms of students' disruptive behavior in the classroom. Nonetheless, it should be noted that ineffective use of time in the classroom is not necessarily the result of students' behavior alone, but rather taps into the interplay between students' and the teacher's actions in the classroom. To disentangle item content from item wording, an ideal study design would need items that distinguish more clearly between teacher and student actions that contribute to (in-)effective use of time in the classroom.

References

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018a). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology, 110*(8), 1066–1083. https://doi.org/10.1037/edu0000256

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018b). Student misbehavior and teacher well-being: Testing the mediating role of the teacher-student relationship. *Learning and Instruction, 58,* 126–136. https://doi.org/10.1016/j.learninstruc.2018.05.006

Aloe, A.M., Shisler, S.M., Norris, B.D., Nickerson, A.B., & Rinker, T.W. (2014). A multivariate meta-analysis of student misbehavior and teacher burnout. *Educational Research Review, 12*, 30-44. https://doi.org/10.1016/j.edurev.2014.05.003

Brophy, J. (1999). Perspectives of classroom management: yesterday, today and tomorrow. In H. J. Freiberg, & J. E. Brophy (Eds.), *Beyond behaviourism: changing the classroom management paradigm*. Allyn & Bacon.

Brophy, J. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 17–43). Lawrence Erlbaum Associates Publishers.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Sage Publishing Inc.

Burns, E., Martin, A.J., Collie, R.J., & Mainhard, T. (2021). Perceived classroom disruption determines the positive educational effects of perceived need-supportive teaching in science. *Learning and Instruction, 75*, Article 101498. https://doi.org/10.1016/j.learninstruc.2021.101498

Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring morethan we bargained for? *American Educational Research Journal, 55*(6), 1233 – 1267.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A Comparison of bifactor and second-order models of quality of life. *Multivariate behavioral research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher, 45*(6), 378–387. https://doi.org/10.3102/0013189X16659442

Creemers, B. P. M., L. Kyriakides, and P. Antoniou. (2013). *Teacher professional development for improving quality of teaching*. Springer.

Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 97–125). Lawrence Erlbaum Associates Publishers.

Eccles, J. S., & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Individual bases of adolescent development* (pp. 404–434). John Wiley & Sons Inc. https://doi.org/10.1002/9780470479193.adlpsy001013

Eccles, J. S., & Roeser, R. W. (2011). School and community influences on human development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 571–643). Psychology Press

Eccles, J. S., & Midgley, C. (1989). Stage/environment fit: Developmentally appropriate classrooms for early adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (pp. 139–181). Academic Press.

Emmer, E. T. & Strough, L. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*(2), 103–112. https://doi.org/10.1207/S15326985EP3602_5

Emmer, E. T., & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 407–437). Lawrence Erlbaum Associates Publishers.

Emmer, E., & Evertson, C. (2017). *Classroom management for middle and high school teachers* (10th ed.). Pearson.

Ennemoser, M., Krajewski, K., & Schmidt, S. (2011). Entwicklung und Bedeutung von Mengen-Zahlen-Kompetenzen und eines basalen Konventions- und Regelwissens in den Klassen 5 bis 9 [Development and importance of sets and numbers competencies and of a basic knowledge of conventions and rules in classes 5 through 9]. *Zeitschrift Für Entwicklungspsychologie und Pädagogische Psychologie, 43*(4), 228–242. https://doi.org/10.1026/0049-8637/a000055

Fauth, B.*, Atlay, C.*, Dumont, H. & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, 101355. https://doi.org/10.1016/j.learninstruc.2020.101355

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29,* 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fauth, B., Göllner, R., Lenske, G., Praetorius, A.K., & Wagner, W. (2020b). Who sees what?: conceptual considerations on the measurement of teaching quality from different perspectives. Zeitschrift für Pädagogik. Beiheft, 66(1):138-155.

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., & Trautwein, U. (2020a). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational Psychology, 112*(6), 1284–1302. https://doi.org/10.1037/edu0000416

Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 98–143). John Wiley & Sons.

Göllner, R., Fauth, B., Lenske, G., Praetorius, A.K., Wagner, W. (2020). Do student ratings of classroom management tell us more about teachers or classroom composition? *Zeitschrift für Pädagogik, 66,* 156–172.

Hamre, B. K. & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization and measurement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). Routledge.

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

Hochweber, J., Hosenfeld, I., & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology, 106*(1), 289–300. https://doi.org/10.1037/a0033829

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jones, V. (2006). How do teachers learn to be effective classroom managers? In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 887–907). Lawrence Erlbaum Associates Publishers.

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

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Koch, T., Holtmann, J., Bohn, J., & Eid, M. (2018). Explaining general and specific factors in longitudinal, multimethod, and bifactor models: Some caveats and recommendations. *Psychological methods*, *23*(3), 505–523. https://doi.org/10.1037/met0000146

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Rinehart & Winston.

Krajewski, K., & Ennemoser, M. (2013). Entwicklung und Diagnostik der Zahl-Größen-Verknüpfung zwischen 3 und 8 Jahren [Development and diagnostics of the linkage of numbers and quantities between ages 3 and 8]. In M. Hasselhorn, A. Heinze, W. Schneider, & U. Trautwein (Eds.), *Diagnostik mathematischer Kompetenzen. Jahrbuch der pädagogisch-psychologischen Diagnostik. Tests und Trends* [Diagnosis of mathematical competencies. Yearbook of pedagogical-psychological diagnosis. Tests and trends] (pp. 41–65). Hogrefe.

Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: a multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 97-124). Springer.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. http://dx.doi.org/10.1037/a0032583

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M.,… Stanat, P. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente* [PISA 2000: Documentation of Scales]. Heenemann GmbH & Co.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Muijselaar, M. M. L., Swart, N. M., Steenbeek-Planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). The dimensions of reading comprehension in Dutch children: Is differentiation by text and question type necessary? *Journal of Educational Psychology, 109*(1), 70–83. https://doi.org/10.1037/edu0000120

Muthén, L. K., & Muthén, B. O. (1998 – 2012). *Mplus user's guide. Seventh Edition.* Muthén & Muthén.

OECD (2004). *Learning for tomorrow's world: First results from PISA 2003.* OECD Publishing.

Olivier, E., Galand, B., Morin, A. J. S., & Hospel, V. (2021). Need-supportive teaching and student engagement in the classroom: Comparing the additive, synergistic, and global contributions. *Learning and Instruction*, *71*, Article 101389. https://doi.org/10.1016/j.learninstruc.2020.101389

Ophardt D., & Thiel, F. (2017). Klassenmanagement als Basisdimension der Unterrichtsqualität [Classroom management as a dimension of teaching quality]. In Schweer M. (Ed.), *Lehrer-Schüler-Interaktion* [Teacher-student interaction]. Springer VS. https://doi.org/10.1007/978-3-658-15083-9_11

Ophardt, D. & Thiel, F. (2013). *Klassenmanagement – Ein Handbuch für Studium und Praxis* [Clasroom management – a handbook for study and practice]. Kohlhammer.

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology, 102*(4), 896. https://doi.org/10.1037/a0019545

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manuel, Pre- K*. Brookes Publishing.

Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (2007). *PISA 2003 [Version 1, Code book and dataset]*. Institute for Educational Quality Improvement.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (2013). *PISA-I-Plus 2003, 2004 [Version 1, Code book and dataset]*. Institute for Educational Quality Improvement.

Prenzel, M., Carstensen, C.H., Schöps, K. & C. Maurischat (2006). Die Anlange des Längsschnitts bei PISA 2003. In PISA-Konsortium Deutschland (Eds.), PISA 2003 - Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres (pp. 29-62). Münster.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Scherer, R., & Gustafsson, J.-E. (2015). Student assessment of teaching as a source of information about aspects of teaching quality in multiple subject domains: An application of multilevel bifactor structural equation modeling. *Frontiers in Psychology, 6,* Article 1550. https://doi.org/10.3389/fpsyg.2015.01550

Seidel, T., Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454-499. https://doi.org/10.3102/0034654307310317.

Shernof, D. J., Ruzek, E. A., Sannella, A. J., Schorr, R. Y., Sanchez-Wall, L., & Bressler, D. M. (2017). Student engagement as a general factor of classroom experience: associations with student practices and educational outcomes in a university gateway course. *Frontiers in psychology*, *8*, 994. https://doi.org/10.3389/fpsyg.2017.00994

Stough, L. M., & Montague, M. L. (2015). How teachers learn to be classroom managers. In E. T. Emmer, & E. J. Sabronie (Eds.), Handbook of classroom management: Research and contemporary issues (pp. 446-458). Routlegde.

Stroet, K., Opdendakker, M.C., & Minnaert, Al. (2013). Effects of need supportive teaching on early adolescents‘ motivation and engagement: a review of the literature. *Educational Research Review, 9*, 65-87. https://doi.org/10.1016/j.edurev.2012.11.003

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge University Press.

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108,* 705–721. http://dx.doi.org/10.1037/edu0000075

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the tripod student perception survey. *American Educational Research Journal, 53*(6), 1834–1868. https://doi.org/10.3102/0002831216671864.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*, 249-294. https://doi.org/10.3102/00346543063003249

Wolff, C.E., Jarodzka, H. & Boshuizen, H.P.A. (2021). Classroom management scripts: a theoretical model contrasting expert and novice teachers' knowledge and awareness of classroom events. *Educational Psychology Review 33*, 131–148. https://doi.org/10.1007/s10648-020-09542-0

Yue, D. (2021). *Classroom management, class related achievement emotions, and disruptive classroom behavior*. Dissertation, 3 Mikrofiches (249 pages), Freie Universität Berlin.

Supplement

Study 1

Table S1

Factor Loadings in the Nested Factor Model (Monitoring)

	Within-class		Between-class	
Indicators	General factor Est. *(SE)*	Specific factor Est. *(SE)*	Indicators	General factor Est. *(SE)*
Disturbances/ Ineffective time use				
Ds1	.73 (.02)*	-	.98 (.01)*	-
Ds2	.75 (.02)*	-	.99 (.01)*	-
Tu1	.71 (.02)*	-	.96 (.01)*	-
Tu2	.63 (.02)*	-	.89 (.02)*	-
Monitoring				
Mo1	-.24 (.03)*	.53 (.02)*	-.71 (.05)*	.51 (.07)*
Mo2	-.15 (.03)*	.40 (.03)*	-.45 (.05)*	.12 (.10)
Mo3	-.24 (.03)*	.62 (.03)*	-.81 (.05)*	.33 (.09)*
Mo4	-.16 (.03)*	.68 (.02)*	-.79 (.04)*	.63 (.07)*

Note. *$p < .05$. The general factor was identified by the indicators of Disturbances/Ineffective time use. The specific factor was identified by the indicators of Monitoring. Factor loadings are shown in their standardized form.

Study 2

Table S2

Fit Indices of the First-Order Factor Models

Model	χ^2	Df	RMSEA	CFI	TLI	$SRMR_w$	$SRMR_b$
Ds + Mo	579.99	68	.04	.97	.94	.02	.07
Ds + Rc	555.97	68	.03	.97	.94	.02	.09
Ds + St	868.32	86	.04	.95	.91	.04	.09

Note. $*p < .05$. Models included Disturbances and each classroom management aspect (i.e., Monitoring, Rule clarity and Structure) as correlated first-order factors. Ds = Disturbances, Mo = Monitoring, Rc = Rule Clarity, St = Structure.

Table S3

Fit Indices of the Nested Factor Models

Model	χ^2	Df	RMSEA	CFI	TLI	$SRMR_w$	$SRMR_b$
Freely estimated							
Monitoring	433.93	62	.03	.98	.95	.02	.04
Rule Clarity	473.32	62	.03	.98	.95	.02	.05
Structure	646.54	78	.03	.97	.95	.02	.05
Constrained							
Monitoring	475.98	70	.03	.98	.95	.02	.04
Rule Clarity	581.30	70	.03	.97	.94	.02	.08
Structure	666.76	87	.03	.97	.94	.02	.05

Note. $*p < .05$. The general factor was identified by the indicators of Disturbances, while the specific factor was identified by the indicators of Monitoring, Rule Clarity and Structure, respectively.

Table S4

Factor Loadings in the Nested Factor Model (Monitoring)

	Within-class		Between-class	
	General factor	Specific factor		General factor
Indicators	Est. *(SE)*	Est. *(SE)*	Indicators	Est. *(SE)*
Disturbances/ Ineffective time use				
Ds1	.66 (.01)*	-	.98 (.01)*	-
Ds2	.66 (.01)*	-	.95 (.01)*	-
Ds3	.53 (.01)*	-	.86 (.02)*	-
Ds4	.68 (.01)*	-	.97 (.01)*	-
Ds5	.60 (.01)*	-	.91 (.02)*	-
Ds6	.57 (.01)*	-	.93 (.02)*	-
Monitoring				
Mo1	-.27 (.02)*	.47 (.02)*	-.59 (.05)*	.68 (.04)*
Mo2	-.34 (.02)*	.44 (.02)*	-.82 (.03)*	.53 (.04)*
Mo3	-.28 (.02)*	.53 (.02)*	-.56 (.05)*	.75 (.04)*
Mo4	-.25 (.02)*	.55 (.02)*	-.58 (.05)*	.81 (.04)*

Note. *$p < .05$. The general factor was identified by the indicators of Disturbances/Ineffective time use. The specific factor was identified by the indicators of Monitoring. Factor loadings are shown in their standardized form.

Table S5

Factor Loadings in the Nested Factor Model (Rule Clarity)

	Within-class		Between-class	
	General factor	Specific factor		General factor
Indicators	Est. *(SE)*	Est. *(SE)*	Indicators	Est. *(SE)*
Disturbances/				
Ineffective time use				
Ds1	.65 (.01)*	-	.98 (.01)*	-
Ds2	.66 (.01)*	-	.95 (.01)*	-
Ds3	.53 (.01)*	-	.86 (.02)*	-
Ds4	.68 (.01)*	-	.97 (.01)*	-
Ds5	.59 (.01)*	-	.92 (.02)*	-
Ds6	.58 (.01)*	-	.94 (.01)*	-
Rule Clarity				
Rc1	-.17 (.02)*	.42 (.01)*	-.19 (.06)*	.80 (.04)*
Rc2	-.19 (.02)*	.71 (.02)*	-.57 (.06)*	.83 (.05)*
Rc3	-.22 (.02)*	.64 (.02)*	-.66 (.05)*	.68 (.05)*
Rc4	-.15 (.02)*	.47 (.02)*	-.21 (.07)*	.77 (.04)*

Note. *$p < .05$. The general factor was identified by the indicators of Disturbances/Ineffective time use. The specific factor was identified by the indicators of Rule Clarity. Factor loadings are shown in their standardized form.

Table S6

Factor Loadings in the Nested Factor Model (Structure)

	Within-class		Between-class	
Indicators	General factor Est. *(SE)*	Specific factor Est. *(SE)*	Indicators	General factor Est. *(SE)*
Disturbances/ Ineffective time use				
Ds1	.65 (.01)*	-	.98 (.01)*	-
Ds2	.66 (.01)*	-	.95 (.01)*	-
Ds3	.53 (.01)*	-	.86 (.02)*	-
Ds4	.67 (.01)*	-	.97 (.01)*	-
Ds5	.60 (.01)*	-	.92 (.02)*	-
Ds6	.58 (.01)*	-	.94 (.01)*	-
Structure				
Rc1	-.15 (.02)*	.53 (.02)*	-.19 (.07)*	.88 (.07)*
Rc2	-.25 (.02)*	.40 (.02)*	-.56 (.05)*	.36 (.12)*
Rc3	-.24 (.02)*	.40 (.02)*	-.73 (.04)*	.43 (.10)*
Rc4	-.10 (.02)*	.59 (.01)*	-.38 (.06)*	.63 (.07)*
Rc5	-.11 (.02)*	.55 (.02)*	-.19 (.07)*	.68 (.07)*

Note. *$p < .05$. The general factor was identified by the indicators of Disturbances/Ineffective time use. The specific factor was identified by the indicators of Structure. Factor loadings are shown in their standardized form.

6 GENERAL DISCUSSION

There is growing recognition that schools are crucial socio-cultural contexts that shape students' learning and socio-emotional functioning (Benner et al., 2008; Eccles & Roeser, 2009; 2011). In particular, the quality of teaching provided has been identified as one of the most decisive factors shaping the school context. Most importantly for the present dissertation, schools are considered to be "dynamic systems", in which students and teachers interact with each other on a daily basis and undergo developmental processes over time (Eccles & Roeser, 2011). Guided by this perspective, the overarching aim of the present dissertation was to explore the dynamic and contextual nature of teaching quality by considering changes in teaching quality over time and the extent to which students "co-construct" the quality of teaching.

In the present dissertation, teaching quality was examined along two central dimensions: classroom management and teacher support. These dimensions refer to teachers' organizational, social and instructional actions with students in the classroom and are key elements in most well-established teaching quality frameworks (Klieme et al., 2009; Pianta et al., 2008). Because how students themselves perceive and interpret what happens in the classroom is expected to be most important for their learning and development (e.g., Eccles et al., 2011), the present dissertation places particular focus on students as informants about teaching quality. In the present dissertation, reports provided by students at different grade levels were used to assess teaching quality. Therefore, the present dissertation began by examining whether students at different grade levels are able to report on teaching quality in a reliable and valid way. Study 1 of the present dissertation showed that response tendencies (i.e., acquiescence) primarily affected information provided by younger (i.e., fifth grade) students; however, the impact of acquiescence on the student data was considered unproblematic for practical use. The empirical findings of Study 2 and Study 3 further support the assessment of assessing teaching quality via student reports. Overall, it was shown that students' individual and shared perceptions contain meaningful information on teaching quality and are important predictors for students' learning and socio-emotional functioning in secondary school.

Next, the dynamic and contextual nature of teaching quality was examined. With respect to the former, teaching quality was found to be variable over time (Study 2). In addition, the development of teaching quality was longitudinally related to the development of students' academic achievement, academic engagement, and their social and emotional school adjustment across the lower secondary school years. Regarding the contextual nature of teaching quality (Study 3), the specific students in the classroom were found to contribute to the classroom management process through their background characteristics and their own agentic behavior (i.e., disruptive behavior).

In the following section, the central findings of the three empirical studies will be summarized and discussed. In addition, strengths and limitations of the present dissertation will be taken into account (Section 6.1). Furthermore, based on the advances made by the empirical studies regarding the dynamic and contextual nature of teaching quality, opportunities for further research will be identified and implications for conceptualizing and measuring teaching quality in educational research will be derived (Section 6.2). Subsequently, implications for educational practice (Section 6.3) will be discussed. The chapter closes with an overall conclusion for the present dissertation (Section 6.4).

6.1 Discussion of the central empirical findings

Teaching is a socially mediated process (e.g., Vygotsky, 1978) that emerges through the interactions between the teacher and the students in the classroom (Doyle, 2006; Fauth et al., 2020). This notion is highly important for the assessment and understanding of teaching quality in empirical educational research. First, students' own perceptions of teaching quality are assumed to be stronger determinants of their learning and socio-emotional functioning than an objective or third-perspective account of the same (Eccles & Roeser, 2009). Second, rather than a "static, baseline predictor" of student outcomes (Way et al., 2007, p. 195), teaching quality is a dynamic process happening in the context of the classroom. The present dissertation made significant contributions to a) the use of student reports, and b) exploring the dynamic, and c) contextual nature of teaching quality. First, this dissertation's findings on assessing teaching quality via student ratings will be discussed with regard to the extent to which students at different grade levels are able to provide reliable and valid information on teaching quality. Second, the dynamic nature of teaching quality will be discussed with regard to the longitudinal development of teaching quality during the early adolescent years. Third, the contextual nature of teaching quality will be discussed with regard to the extent to which the specific students in the classroom contribute to the classroom management process.

6.1.1 Assessing Teaching Quality with Student Reports: Students' Ability to Provide Valid and Reliable Information on Teaching Quality

Students' perceptions of what happens in the classroom are an important intermediary between the quality of teaching provided and their learning and socio-emotional functioning. Thus, assessing teaching quality from the student perspective allows for valuable insights that may not be captured by other sources (e.g., teacher self-reports, observers; Feldlaufer, et al., 1988; Fraser & O'Brien, 1985; Kunter & Baumert, 2006). Thus, before using student reports to examine the dynamic and contextual nature of teaching quality, the present dissertation began by examining whether students at different grade levels are able to report on teaching quality in a reliable and valid way. Studies 1, 2 and 3 within the present dissertation make a significant contribution to the discussion of using student reports to assess teaching quality.

Study 1 addressed a major concern that has been raised when assessing teaching quality with student questionnaires: The impact of response tendencies, defined as individual differences

in using the response scale, on student data on teaching quality. In particular, Study 1 investigated the extent to which student ratings provided by students in different grades (i.e., fifth grade and eighth grade) were affected by an acquiescent response tendency (i.e., acquiescence: "yay-saying"; dis-acquiescence: "nay-saying"), while additionally controlling for an extreme response style. Analyses were conducted at the student and at the classroom level. Overall, the results of Study 1 showed that acquiescence was more pronounced in ratings provided by younger (i.e., fifth grade) than in older (i.e., eighth grade) students and that student reports provided by fifth graders were additionally impacted by an extreme response style. More specifically, acquiescence (and extreme responding) had a greater impact on negatively worded items than on positively worded items in the teaching quality reports provided by fifth grade students. Finally, acquiescence was found to differ both across individual students within the same classroom as well as across classrooms, contradicting the assumption that individual differences in acquiescence are counterbalanced at the classroom level (i.e., students' shared perceptions). Whereas these findings may seem concerning at first glance, comparing students' teaching quality ratings before and after controlling for acquiescence revealed only minor differences in the psychometric properties (i.e., factor intercorrelations, factor means) of student data on teaching quality. Overall, the insights gained from Study 1 suggest that acquiescence primarily affects ratings provided by younger students (i.e., fifth grade) and when negatively worded items are used. Nonetheless, we consider the impact of acquiescence on student data on teaching quality to be rather small and unproblematic for practical use.

The empirical findings from Study 2 and Study 3 provided further support for students' ability to provide reliable and valid information on teaching quality.

Factorial analyses were conducted as a key methodological component of Studies 1 and 3 to explore the factorial structure of student reports on teaching quality. Study 1 examined *teacher support*, c*ontent relevance*, and c*larity of instruction*, three well-known measures of teaching quality. Study 3, which examined two different large-scale data sets of secondary school students, included various measures of classroom management. Some of these measures referred more to students' behavior or to the interaction between students and the teacher in the classroom (i.e., *disturbances*), while other measures addressed teachers' managing actions in the classroom (i.e., *monitoring, rule clarity, structure*). In both studies, a clear factor structure emerged and fit the data well. Hence, both studies support the ability of students at all investigated grade levels to

distinguish between different aspects of teaching quality. More specifically, not only were students able to distinguish between the overarching dimensions of teaching quality (Study 1), they were also able to differentiate between measures belonging to the same overarching dimension (i.e., classroom management, Study 3). Moreover, students were able to distinguish between their own vs. the teacher's actions in the classroom (Study 3). In summary, the results support the discriminant validity of student reports on teaching quality.

In addition to that, Study 2 examined the co-development of multiple aspects of teaching quality and critical domains of student development (i.e., academic achievement, academic engagement, their social and emotional school adjustment). The teaching quality trajectories were modeled from both the student and the teacher perspective. With the exception of students' disruptive behavior, students reported declines in all aspects of teaching quality over time. The same declining trend was also found from the teacher perspective, albeit to a lesser extent. Taken together, the findings underscore students' sensitivity to changes in their teachers' organizational, social and instructional behavior, confirming students' ability to report on changes in teaching quality over time in a student questionnaire.

Finally, the results of Study 2 and Study 3 support the predictive validity of student reports on teaching quality. Study 2 provided insights into which aspects of teaching quality are concurrently and longitudinally associated with critical domains of students' development at both the student and the classroom level. Overall, the revealed associations were more pronounced for student reports than for teacher self-reports of teaching quality. In addition, the results of Study 3 showed that students' disruptive behavior and teachers' monitoring were associated with students' math achievement, over and above their individual background characteristics.

In summary, both students' individual and shared perceptions of teaching quality have been shown to be valid and uniquely important predictors of students' learning and socio-emotional functioning in secondary school.

6.1.2 The Dynamic Nature of Teaching Quality: Teaching Quality During Early Adolescence

One possibility to explore the dynamic nature of teaching quality is to examine teaching quality from a longitudinal perspective. Focusing on the lower secondary school years (i.e., Grade 5 to Grade 8) is a particularly worthwhile endeavor. During lower secondary school, adolescent students experience critical developmental processes related to the onset of puberty and are at risk for experiencing struggles in their academic engagement and their social and emotional school adjustment (Engels et al., 2017; Roeser et al., 2000; Virtanen et al., 2021). Given that teaching quality has been identified as one of the most decisive factors within the school context (e.g., Eccles & Roeser, 2011) and corresponds to adolescent students' developmental needs for autonomy, competence and relatedness (Ryan and Deci, 2017; Connell & Wellborn, 1991; Eccles & Midgley, 1989), it is surprising that relatively little is known about the extent to which changes in teaching quality during secondary school are associated with changes in students' learning and socio-emotional functioning. To address this research gap, Study 2 followed the same group of students and their homeroom teachers from fifth to eighth grade of secondary school and linked the trajectories of multiple aspects of teaching quality to the trajectories of student development (i.e., academic achievement, academic engagement, their social and emotional school adjustment).

Overall, multilevel longitudinal growth curve models revealed significant changes in teaching quality and student development as students moved from fifth to eighth grade, underscoring the dynamic nature of teaching quality. Most consistently with prior research (e.g., Eccles et al., 1993; Engels et al., 2017; Way et al., 2007), students reported mostly negative changes in teaching quality and their academic engagement. In contrast, students adjusted to the secondary school environment in a rather favorable way, contributing to the more mixed literature on adolescent students' social and emotional school adjustment during secondary school (Widlund et al., 2021).

Another central finding of Study 2 were significant associations between the trajectories of students' development (i.e., academic achievement, academic engagement, social and emotional school adjustment) and teaching quality, with the results more pronounced for student compared to teacher reports on teaching quality. In line with prior research (e.g., Hughes, 2011; Opdenakker et al., 2012; Spilt et al., 2012), students' academic achievement and academic engagement were most consistently related to teaching quality over time. That is, more favorable changes in teaching

quality buffered the downward trend in students' academic engagement and promoted students' academic achievement at the classroom level. Importantly, the teaching quality dimensions of monitoring, teacher academic and emotional support were all positively related to students' academic achievement over time, whereas most cross-sectional studies have identified classroom disturbances in particular as associated with students' achievement (e.g., Fauth et al., 2014; Kunter et al., 2013; see also Evertson & Weinstein, 2006). Thus, from a longitudinal perspective, it appears that students' academic achievement benefits not only from a high average level of teaching quality, but also and particularly from the extent to which a teacher adjusts his or her instructional and interpersonal interactions with students to their changing developmental needs (e.g., Patall et al., 2010; Stroet et al., 2013). Accordingly, providing a supportive learning environment and maintaining a positive teacher-student relationship seem to be valuable resources for fostering students' academic development during early adolescence (Emmer & Gerwels, 2006).

In addition, students' social and emotional school adjustment was solely related to teaching quality at the student level. Hence, the results of Study 2 suggest that students' social and emotional school adjustment is a rather individual process that is shaped by students' individual experiences with their teachers. Specifically, the amount of support a student personally perceives and/or the extent to which a teacher meets individual students' needs appear to be important mechanisms for supporting students' social and emotional school adjustment during early adolescence. In contrast, it seems to be less important for students' social and emotional school adjustment whether the classroom is well organized or whether the teacher is, on average, more supportive than other teachers.

Taken together, the results of Study 2 showed that teaching quality is a rather dynamic process that is variable over time, underscoring its dynamic nature and supporting the conceptualization of schools and classrooms as complex and dynamic social environments (e.g., Eccles & Roeser, 2011). Not only do students develop over time (i.e., in terms of their developmental needs, academic and socio-emotional development), so too do characteristics of the classroom environment (i.e., teaching quality). In particular, the results suggest that teaching quality needs to be adjusted to adolescent students' developmental needs in order to create "developmentally appropriate" learning environments (Eccles & Midgley, 1989; Eccles et al., 1993) that support positive adolescent development.

6.1.3 The Contextual Nature of Teaching Quality: Classroom Management in the Classroom Context

In the quest to understand what constitutes high-quality teaching, it needs to be considered that high-quality teaching is defined and established within the classroom context, to which the teacher and the specific students in the class contribute (e.g., Doyle, 2006). This has several consequences for the evaluation of teaching quality: First, it implies that students are not only recipients of teaching in the classroom, but actively contribute to the processes happening in the classroom through interactions with their peers and with the teacher. As Kennedy (2010, p. 595) notes: "*Students influence teaching practices when they are restless, gregarious, or frustrated, and even when they are happy. And they interrupt the learning of other students as well*". Second, it implies that situational factors, such as students' behavior, needs and demands, may strongly affect how the teacher organizes and conducts the lesson (e.g., Doyle, 2006; Emmer & Stough, 2001; Shavelson & Dempsey-Atwood, 1976). However, the interplay between students and the teacher in the classroom has not been explicitly considered when evaluating teaching quality. Study 3 addressed this research gap. Using classroom management as an example, Study 3 investigated the extent to which the specific students in the classroom contribute to the classroom management process through a) their background characteristics and b) their own agentic behavior (i.e., disruptive behavior) across two different large-scale data sets. To this end, Study 3 included various key aspects of classroom management and differentiated between aspects referring more to student behavior (i.e., disruptive behavior) or more to teachers' management actions in the classroom (i.e., monitoring, rule clarity, structure).

Overall, the results of Study 3 highlight the importance of considering the specific classroom context in which classroom management is applied. Across both data sets, it was found that measures of classroom management referring more to student actions in the classroom (i.e., disruptive behavior) were more closely related to students' background characteristics than measures referring to teacher actions, which is in line with other research in the field (Göllner et al., 2018). Consequently, it seems that the degree to which discipline and order are established in a classroom is less a result of teachers' managing actions alone, but is co-constructed by the students in the classroom.

To investigate the extent to which the specific students in the classroom contribute to the effectiveness of teachers' classroom management strategies, we calculated a nested model

examining the associations between several management strategies and students' pre-adjusted mathematic achievement, while taking into account students' disruptive behavior in the classroom. Across both data sets, students' disruptive behavior and teachers' monitoring activity were negatively associated with students' pre-adjusted mathematic achievement. Because the nested model accounted for the shared amount of disturbances inherent in all of the classroom management items, the remaining variance in monitoring may be interpreted as an indicator of "too much" monitoring in relation to the level of disturbances in a given classroom. Thus, the results suggested that teachers' managing activities (i.e., monitoring) needed to match the level of disturbances in the class. In contrast, no associations were found for the other management strategies (i.e., rule clarity and structure). From a theoretical perspective, it can be argued that rule clarity and structure refer more to teachers' preventive actions to maintain an orderly classroom environment, whereas monitoring describes a more active process that can be considered both preventive and reactive to students' disruptive behavior to a certain degree.

Taken together, the results of Study 3 highlight the role of students in the classroom management process and underscore the complexity of classroom management. On the one hand, the results support the notion that discipline and order are less a result of teachers' managing actions alone, but need to be jointly accomplished by the teacher and the students in the classroom (Doyle, 2006). When conceptualizing teaching quality as an offer-use situation (e.g., Helmke, 2003; Fend, 2008), it can be argued that students' disruptive behavior reflects the degree to which the teacher's offer (i.e., turning the allocated lesson time into learning time) is actually used by the students. On the other hand, students also seem to affect the classroom management processes located on the offer side. That is, efficient classroom management requires the teacher to maintain "situational awareness" (Ophardt & Thiel, 2017; Wolff et al., 2021) in order to adjust his or her classroom management strategies (i.e., monitoring) to the requirements of the specific students and the situational demands of each class.

The notion that the specific students in the classroom affect both the offer and the use side of teaching quality underscores its contextual nature. Thus, the findings corroborate and extend prior research on the variability of teaching quality, which has shown that teaching quality can differ substantially across classes, even when the same teacher teaches the same content to the same grade level (e.g., Fauth et al., 2020).

6.1.4 Strengths and Limitations

When considering the contribution of the present dissertation to understanding effective teaching from the student perspective, it is important to keep some general strengths and limitations in mind.

First, the present dissertation had the advantage of using several large-scale data sets including students at different grade levels of secondary school. However, students from the highest, academic track in Germany (i.e., Gymnasium) were not included in the TRAIN data set, which was used in Study 1 and Study 2. This was because the TRAIN study focused on the academic and social development of students from vocational-track schools. This may limit the generalizability of the empirical findings to the entire population of German secondary school students. To some extent, the different school tracks represent differential learning environments (i.e., Baumert et al., 2006; Scharenberg, 2014; Schiepe-Tiska, 2019), in part because students in academic- and vocational-track schools typically differ with regard to their social background, prior academic achievement and future academic development (Maaz et al., 2008; Dumont et al., 2013). Consequently, it is possible that the strength of students' acquiescence tendencies differs between the academic and vocational tracks. Relatedly, future research is needed to examine whether and to what degree the trajectories of teaching quality, students' academic achievement, academic engagement, their social and emotional school adjustment differ in academic-track schools. Nonetheless, focusing on students from vocational-track schools seems particularly worthwhile, given that these students often struggle with educational attainment and academic achievement (for more detail, see Baumert et al., 2003; Köller & Baumert, 2001, Pekrun et al., 2006).

Second, in order to particularly focus on students' perceptions of their learning environment, the present dissertation drew on student reports of teaching quality as its primary data source. The nested structure of the student data was adequately addressed by conducting multilevel analyses. Nevertheless, using the same data source (i.e., student reports) increases the risk of shared method bias (Podsakoff et al., 2003). Study 2 addressed these issues and examined longitudinal changes in teaching quality as assessed from both the student and teacher perspectives. However, even though both students and teachers mostly reported negative changes in teaching quality over time, longitudinal associations between teaching quality and student development were only found when the former was measured from the student perspective. Including information from an outside

perspective on students' development (e.g., parent information) would help to reduce the possible impact of shared method bias.

Third, the present dissertation assessed teaching quality along central dimensions included in most well-established teaching quality frameworks (Klieme et al., 2009; Pianta et al., 2008). As such, Study 2 focused on disturbances, monitoring, teacher academic and emotional support, while Study 3 exclusively focused on classroom management. It should be mentioned here that most prominent theoretical models of teaching quality (e.g., Klieme et al., 2009; Pianta et al., 2008) include a third dimension of high-quality teaching, usually referred to as "cognitive activation". In general, cognitive activation refers to teaching practices that help students actively engage with the learning content in order to develop a deep understanding of it. For example, the teacher may provide students with complex tasks, encourage discussion among students, and activate prior knowledge (Lipowsky et al., 2009; Pianta et al., 2008). However, given that students generally lack professional knowledge of teaching, their ability to evaluate aspects concerning the teaching methodology of the lesson is questionable. Nonetheless, future research should examine whether and to what degree the present dissertation's findings regarding students' ability to report on teaching quality and the dynamic and contextual nature of teaching quality can be generalized across all teaching quality dimensions.

Lastly, the present dissertation included a number of scales assessing central dimensions of teaching quality that have been validated in prior large-scale studies (e.g., Baumert et al., 2009). However, the scales' specific wording should be addressed. In Study 1, scales with varying keying directions were used to examine whether acquiescence exhibits a comparable impact on positively and negatively worded items. In Study 3, a broad range of classroom management measures were used referring either to teachers' management actions, students' behavior or the interplay between students and the teacher. Given that the empirical studies were not designed to systematically evaluate the impact of item wording (i.e., keying direction, referring to students vs. the teacher), it cannot be ruled out that the item wording interacted with the item content (i.e., aspect of teaching quality) to some degree. Albeit hard to achieve, an ideal study design would assess each aspect of teaching quality with types of item wordings, making it possible to disentangle item content from item wording.

6.2 Theoretical Implications and Future Directions

The present dissertation explored the dynamics that operate across time and between the teacher and the specific students in the classroom. Specifically, the present dissertation investigated teaching quality from a longitudinal perspective and examined the extent to which the specific classroom environment contributes to teaching quality. The results highlight promising opportunities for further research on teaching quality and are highly relevant for the conceptualization and measurement of teaching quality in empirical educational research.

6.2.1 Exploring the Dynamic Nature of Teaching Quality

By examining teaching quality across three consecutive school years within the same school setting (i.e., Grade 5 to Grade 8 of secondary school), Study 2 of the present dissertation shed light on the dynamic nature of teaching quality and extended previous findings on teaching quality. Within this longitudinal perspective, a highly interesting finding was that teacher support was associated with students' academic achievement, whereas classroom management was not. Given that the majority of empirical research on teaching quality relies on concurrent or short-term longitudinal analyses, more research examining the extent to which concurrent associations between teaching quality and student outcomes differ from longitudinal associations is necessary to corroborate our results and more fully understand the complex, dynamic nature of teaching quality.

However, there was also a significant amount of variance in the trajectories of teaching quality over and above students' background characteristics, suggesting that some students or entire classes may have experienced changes over time that diverged from the average downward trend (e.g., a weaker or stronger decrease, or even an increase in teaching quality). Identifying subgroups of students or classes that diverge from the average trend (e.g., latent trajectory classes, Spilt et al., 2012, Widlund et al., 2021) would be a promising starting point for further research. Furthermore, examining what accounts for variation in students' trajectories could be helpful in order to provide them with specialized support and interventions at an early stage of secondary school.

6.2.2 Exploring the Contextual Nature of Teaching Quality

The results of Study 3 highlight the importance of considering the specific classroom context in which teaching quality is applied. However, assessing teaching quality merely from the student perspective provides insights into only one of "multiple realities of classroom life" (Hoy & Weinstein, 2006, p. 190). In this regard, it would be highly interesting to investigate whether the key finding of Study 3 (i.e., the negative association between monitoring and student achievement when accounting for the level of disturbances in the classroom) can be replicated when using teacher self-reports or external observations. On the one hand, the extent to which students feel excessively monitored by their teacher might be best evaluated by the students themselves. On the other hand, external observers could provide helpful guidance in the context of trainings for teachers, helping them better tailor their teaching to the specific students in the classroom. Enriching the present dissertation's findings with data from other sources (e.g., teacher self-reports, observers) would provide a deeper understanding of the contextual nature of teaching quality. Relatedly, it would be highly interesting to examine the degree to which teachers themselves experience differences in the quality of their teaching, and particularly in the way they organize and manage the classroom and interact with students when teaching different classes. Moreover, assessing student and teacher behavior at several measurement points (e.g., within the same lesson, within the same school year) might allow for more insight into how teacher and student actions in the classroom mutually impact one another over time.

6.2.3 Conceptualizing and Measuring Teaching Quality

Broadly speaking, high-quality teaching provides students with appropriate conditions that allow them to engage in meaningful learning processes (Kunter & Voss, 2013). While teaching is a multifaceted and complex endeavor, teaching quality needs to be conceptualized and defined more narrowly in order to assess and investigate high-quality teaching in educational research and practice. Integrating the present dissertation's findings on the dynamic and contextual nature of teaching quality may strengthen the understanding of teaching quality in educational research and practice, with consequences for how teaching quality is conceptualized and measured in empirical educational research.

Conceptualizing Teaching Quality

Broadly speaking, teaching can be understood as an inherently *instructional* activity that is initiated and driven by the teacher. In theoretical models of teaching quality (e.g., Klieme et al., 2009; Pianta et al., 2008), teaching is largely described in terms of teachers' organizational, social and instructional actions in the classrooms, which are considered important for students' learning and socio-emotional functioning. However, focusing on teacher actions alone may not fully address the complexity of effective teaching in the classroom setting. Drawing on the findings from the present dissertation, high-quality teaching should also be understood as a "*co-constructive*" process. That is, it needs to be considered that teaching takes place within the complex social structure of the classroom and through interactions between the teacher and the students. This implies that teaching quality is not completely determined by teachers' actions in the classroom alone, a notion which is largely supported by the empirical findings of the present dissertation. In particular, Study 2 showed that teaching quality varied when teaching students at different grade levels. Study 3 showed that students significantly contribute to classroom management processes through their background characteristics and their disruptive behavior. Consequently, the quality of teaching can to at least some degree be understood as "co-constructed" by the specific students in the classroom (see also Fauth et al., 2020; Praetorius et al., 2018).

Finally, teaching quality can be understood as an "*adaptive*" process. Teaching in the social classroom setting involves a constant awareness of classroom situations (i.e., perceiving and interpreting classroom events) as well as active-decision making to respond to situational demands (i.e., deciding whether and how to act in order to sustain and pursue teaching goals; Doyle, 2006; Ophardt & Thiel, 2017). As Doyle (1977, p.169) notes: *"adaptations to momentary classroom conditions may be the most important teacher behaviors from the perspective of the student. There is little reason to presume on a priori grounds that behaviors which are either stable or generalizable across settings are necessarily those that are the most powerful correlates of achievement in a given classroom situation".* In support of this notion, Study 2 suggested that teaching quality needs to meet students' developmental needs, whereas Study 3 suggested that teachers' monitoring activity needs to be adjusted to the level of disturbances in a given classroom. From this perspective, high-quality teaching can also be understood as an "adaptive" process. That is, a teacher's competence to adjust his/her organizational, social and instructional actions to a

diverse set of rapidly changing classroom situations and to the specific students in a given classroom appears to be a critical, yet overlooked characteristic of high-quality teaching.

Measuring Teaching Quality

Understanding teaching quality as a "co-constructive" or "adaptive" process is also consequential for how teaching quality is measured in empirical educational research. As noted earlier, teaching quality is conceptualized as a hierarchical structure consisting of several broad dimensions (e.g., classroom management, teacher support), which in turn encompass several subordinate aspects describing teachers' actions in the classroom (e.g., monitoring, clarity of instruction, caring teacher behavior, etc.). In general, empirical studies greatly vary with regard to the range and specific aspects they use to assess teaching quality. To some extent, this suggests that within each dimension of teaching quality, the subordinate aspects are considered interchangeable with one another. However, the results of the present dissertation (Study 2) point to empirical and conceptual differences between the subordinate aspects of teaching quality within a given quality dimension.

For the dimension of classroom management, it was shown that the subordinate aspects differ in the extent to which they refer to teachers' actions (e.g., monitoring) or students' actions (e.g., disruptive behavior) in the classroom. Specifically, classroom management aspects referring to students' behavior in the classroom were more closely related to students' background characteristics than aspects referring to teachers' managing actions. Thus, it can be argued that assessing teaching quality via student behavior (e.g., "In this class, students rarely chatter loudly") describes teaching quality as a co-constructive process to which both the teacher and the students in the classroom contribute. This notion is highly consequential: Asking about student behavior in the classroom may not fully allow inferences to be made about a given teacher's ability to provide high-quality teaching, but rather provides information about the interplay between the students and the teacher in the classroom.

In contrast, when teaching quality assessments specifically seek to evaluate the quality of the teachers' organizational, social and instructional actions in the classroom, items should be used that explicitly refer to teachers' actions in the classroom (e.g., "Our teacher makes sure that we pay attention"). Importantly, the present dissertation suggested that providing high-quality teaching requires the teacher to adjust his or her teaching to the specific students being taught, to their

developmental needs and to the given situation in the classroom (Stroet et al., 2013; Wolff et al., 2021). However, this critical skill has yet to be more strongly considered in teacher training as well as in the assessment and evaluation of teaching quality. In particular, items are needed that tap more strongly into the extent to which teachers adjust their actions to a given classroom situation.

Taken together, teaching quality can be measured and interpreted quite differently in educational research and practice depending on the conceptual understanding of teaching quality being employed (e.g., instruction, co-construction, adaptation). Moreover, even within the same quality dimension, teaching quality aspects referring to student or to teacher actions tap into different teaching processes and can not necessarily be considered equivalent or interchangeable measures of teaching quality.

6.3 Practical Implications

Although further research is needed to substantiate the findings of the present dissertation, some conclusions and implications can be drawn that are of relevance for educational policy and practice.

First, some implications regarding the use of student reports on teaching quality can be stated. Overall, the results showed that students across several grade levels (i.e., Grade 5 to Grade 8) are able to provide meaningful information regarding their individual and shared perceptions of teaching quality. However, when assessing teaching quality with student ratings, several aspects need to be considered. First, the findings on response tendencies suggest that negatively worded items are difficult to understand and exacerbate issues with item interpretation, particularly when administered to younger students (i.e., fifth grade). Some literature recommends including both positively and negatively worded items to keep students' level of attention high (for a discussion, see Gehlbach, 2015). However, the results of the present dissertation suggest using positively worded items only and avoiding presenting positively and negatively items in alternating order (at least to fifth grade students; see also Swain et al.,2008; Wong et al., 2003; Weijters et al., 2013).

Second, the results of the present dissertation add to a growing body of literature showing that to some extent, the specific students being taught have an impact on the quality of teaching, and that a given teacher's teaching quality may vary substantially across classes (e.g., Bell et al., 2012; Fauth et al., 2019; Gitomer & Bell, 2013; Göllner et al., 2020; Hochweber et al., 2014; Kennedy, 2010). Consequently, teaching quality needs to be considered within the context of the specific class in which it takes place and is being measured, which is particularly important with respect to teacher evaluations. Accordingly, it would be helpful to conduct multiple assessments (e.g., across time and classes) to gain a more comprehensive impression of a teacher's ability to provide high-quality teaching.

Third, the findings underscore the critical role of teachers for both students' long-term academic development, and their social and emotional school adjustment. Whereas students' academic engagement was related to the overall level of teaching quality perceived by all students within a classroom (i.e., the classroom level), students' individual perceptions of teaching quality (i.e., the student level) were most relevant for their social and emotional school adjustment. Given the importance of positive student development, teachers should be encouraged to foster a

supportive learning environment by maintaining good relationships with each individual student and addressing each individual student's developmental needs. This becomes particularly important when considering that adolescent students' experiences with a single teacher can "make a difference" (Pajares & Urdan, 2008; van Ryzin, 2010) for their general development, even though students may have many teachers at the same time in secondary school.

Finally, the results of the present dissertation shed some light on the complexity of effective teaching. That is, teaching is "far more complex than establishing rules" (Evertson & Weinstein, 2006, p. 5) and dealing with the subject matter (Hattie, 2009). Instead, teaching takes place within the social structures of the classroom setting, to which both the teacher and the specific students contribute. Consequently, classroom dynamics (e.g., teacher-student interactions, interactions among peers, rapidly changing situations and unforeseen events) have a strong bearing on what happens in the classroom and require the teacher to plan and react appropriately (Doyle, 2006). However, adjusting one's teaching to meet both the needs of the specific students in the classroom and situational affordances is not easy to achieve and can clearly be considered "expert teaching" (Hattie, 2009; 2012). In fact, several experimental studies have shown that expert teachers monitor and manage their classroom differently from more inexperienced, novice teachers (e.g., Cortina et al., 2015; van den Bogert et al., 2014). This suggests that both pre- and in-service teacher training programs are needed to support and guide teachers' development of critical teaching competences and thus promote teachers' professional development. On the one hand, the present dissertation highlights the need to provide appropriate opportunities for students to feel competent, autonomous and relate positively to others in the classroom. To this end, secondary school teachers could be equipped with knowledge about developmental processes during the phases of early and late adolescence in order to increase their awareness of their students' developmental needs. On the other hand, teachers' capability to "see learning through the eyes of the students" (Hattie, 2009, p. 110) needs to be raised. According to Hattie, high-quality teaching requires the teacher to have "a better understanding of what learning looks and feels like for the students" (Hattie, 2009, p. 116). That is, it is critical for teachers to reflect on and evaluate their organizational, social and instructional actions in the classroom in order to gauge their effects on students' classroom experiences (see also Emmer & Stough, 2001). Presumably, only then will teachers be able to adjust their teaching to students' developmental needs and react flexibly and appropriately to any and all classroom situations. In this regard, student reports on teaching quality are a promising and

powerful way to provide feedback to teachers on what their teaching looks and feels like from the students' perspective, which could further enhance the dialogue between students and the teacher in the classroom.

6.4 Conclusion

To more fully understand the complexity of effective teaching from the student perspective, it is necessary to take into account the dynamics at play that operate across time and between the teacher and the specific students in the classroom. The present dissertation made an important contribution to this issue, by examining the dynamic and contextual nature of teaching quality from the student perspective. It has thereby generated important insights into students' ability to report on teaching quality, the longitudinal development of teaching quality during lower secondary school (i.e., dynamic nature), the critical role of teachers for students' long-term academic, social and emotional development, and the extent to which students contribute to teaching quality (i.e., contextual nature). The findings are particularly important for assessing teaching quality via student reports and for conceptualizing and measuring teaching quality in educational research and practice.

7 REFERENCES

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation.* Lawrence Erlbaum.

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology, 110*(8), 1066–1083. https://doi.org/10.1037/edu0000256

Anderson, C. S. (1982). The Search for school climate: A review of the research. *Review of Educational Research*, *52*(3), 368–420. https://doi.org/10.3102/00346543052003368

Anderson, J. O. (1989). Evaluation of student achievement: Teacher practices and educational measurement. *Alberta Journal of Educational Research, 35*(2), 123–133.

Archambault, I., Janosz, M., Morizot, J., & Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: Relationship to dropout. *Journal of School Health*, *79*, 408–415. https://doi.org/10.1111/j.1746-1561.2009.00428.x

Assor, A., Kaplan, H., and Roth, G. (2002). Choice is good, but relevance is excellent: autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology, 72*(2), 261–278. https://doi.org/10.1348/000709902158883

Bäckman, O. (2017). High school dropout, resource attainment, and criminal convictions. *Journal of Research in Crime and Delinquency*, *54*(5), 715-749. https://doi.org/10.1177/0022427817697441

Bandura, A., & National Inst of Mental Health. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bardach, L., Yanagida, T., Morin, A. J. S., & Lüftenegger, M. (2021). Is everyone in class in agreement and why (not)? Using student and teacher reports to predict within-class consensus on goal structures. *Learning and Instruction.* Advance Online Publication. http://dx.doi.org/10.1016/j.learninstruc.2020.101400

Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology*, *92*(3), 568–582. https://doi.org/10.1037/0022-0663.92.3.568

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Baumert, J., & Köller, O. (1998). Interest research in secondary level I: An overview. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Eds.), *Interest and learning: Proceedings of the Seeon conference on interest and gender* (pp. 241-256). IPN.

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., … Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente* [Professional competence of teachers, cognitively activating mathematics instruction, and the development of students' mathematical literacy (COACTIV): Documentation of measurement instruments]. Max Planck Institute for Human Development.

Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differentieller Lern- und Entwicklungsmilieus [School structure and the emergence of differential learning environments]. In: Baumert, J., Stanat, P., Watermann, R. (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 95–188). Verlag für Sozialwissenschaften.

Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens [Educational environments- institutional conditions for teaching and learning]. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K. J. Tillmann & M. Weiß (Eds.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* [PISA 2000. A differentiated view on the federal states of Germany] (pp. 261–331). Leske + Budrich.

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Becker, M., Cortina, K. S., Tsai, Y.-M., & Eccles, J. S. (2014). Sexual orientation, psychological well-being, and mental health: A longitudinal analysis from adolescence to young adulthood. *Psychology of Sexual Orientation and Gender Diversity, 1*(2), 132–145. https://doi.org/10.1037/sgd0000038

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2-3), 62–87. https://doi.org/10.1080/10627197.2012.715014

Benner, A. D., Graham, S., & Mistry, R. S. (2008). Discerning direct and mediated effects of ecological structures and processes on adolescents' educational outcomes. *Developmental Psychology, 44*(3), 840–854. https://doi.org/10.1037/0012-1649.44.3.840

Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education, 56*, 205-213. https://dx.doi.org/10.1177/0022487105275904

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61-80.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). John Wiley & Sons.

Bolt, D. M., Lu, Y., & Kim, J. S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, *19*(4), 528–541. https://doi.org/10.1037/met0000016

Bowlby, J. (1969). *Attachment and Loss: Attachment* (Vol. 1). Basic Books.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Bronfenbrenner, U. (1989). Ecological systems theory. *Annals of Child Development*, *6*(1), 187–249.

Bronfenbrenner, U., & Morris, P. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (pp. 993–1028). Wiley.

Brophy, J. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 17–43). Lawrence Erlbaum Associates Publishers.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). McMillan.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction, 21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology, 40*(5), 415–436. https://doi.org/10.1016/S0022-4405(02)00107-3

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, F. (2008). Classroom problem behavior and teacher-child relationships in kindergarten: the moderating role of classroom climate. *Journal of school psychology*, *46*(4), 367–391. https://doi.org/10.1016/j.jsp.2007.06.009

Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring morethan we bargained for? *American Educational Research Journal, 55*(6), 1233 – 1267.

Chernyshenko, O., M. Kankaraš and F. Drasgow (2018), "Social and emotional skills for student success and well-being: Conceptual framework for the OECD study on social and emotional skills", *OECD Education Working Papers*, No. 173, OECD Publishing, Paris. https://doi.org/10.1787/db1d8e59-en.

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* [Instructional quality: a question of the perspective?]. Waxmann.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Self processes and development* (pp. 43–77). Lawrence Erlbaum Associates, Inc.

Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A Meta-analysis. *Review of Educational Research*, *77*(1), 113–143. https://doi.org/10.3102/003465430298563

Cortina, K. S., Miller, K. F., McKenzie, R., Epstein, A., & Feng, G. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with teachers. *International Journal of Science and Mathematics Education, 13(2),* 389-403. https://dx.doi.org/10.1007/s10763-014-9610-5

Crosnoe, R., Johnson, M. K., & Elder, G. H., Jr. (2004). School size and the interpersonal side of education: An examination of race/ethnicity and organizational context. *Social Science Quarterly, 85*(5), 1259–1274. https://doi.org/10.1111/j.0038-4941.2004.00275.x

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*(1), 51- 85. https://doi.org/10.1023/A:1011402608575

den Brok, P., Fisher, D., Rickards, T., & Bull, E. (2006). Californian science students' perceptions of their classroom learning environments. *Educational Research and Evaluation, 12*(1), 3–25. https://doi.org/10.1080/13803610500392053

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2015). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *The Journal of Early Adolescence, 35*(5-6), 722–758. https://doi.org/10.1177/0272431614564059

Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of Research in Education, 5,* 163-198. https://doi.org/10.2307/1167174

Doyle, W. (1985). Recent research on classroom management: Implications for teacher education. *Journal of Teacher Education*, *36*(3), 31–35. https://doi.org/10.1177/002248718503600307

Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 97–125). Lawrence Erlbaum Associates Publishers.

Dumont, H., Neumann, M., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. Internationale und nationale Befunde [Student composition as a factor influencing student achievement. International and national results]. *Psychologie in Erziehung und Unterricht, 60*(3), 163–183. https://doi.org/10.2378/peu2013.art14d

Eccles, J. S., & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Individual bases of adolescent development* (pp. 404–434). John Wiley & Sons Inc. https://doi.org/10.1002/9780470479193.adlpsy001013

Eccles, J. S., & Roeser, R. W. (2011). School and community influences on human development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 571–643). Psychology Press.

Eccles, J. S., Lord, S., & Midgley, C. (1991). What are we doing to early adolescents? The impact of educational contexts on early adolescents. *American Journal of Education, 99*(4), 521–542. https://doi.org/10.1086/443996

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*(2), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

Eccles, J. S., & Midgley, C. (1989). Stage/environment fit: Developmentally appropriate classrooms for early adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (pp. 139–181). Academic Press.

Emmer, E. T. & Strough, L. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103–112. https://doi.org/10.1207/S15326985EP3602_5

Emmer, E. T., & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 407–437). Lawrence Erlbaum Associates Publishers.

Engels, M. C., Colpin, H., Van Leeuwen, K., Bijttebier, P., Den Noortgate, W. V., Claes, S., Goossens, L., & Verschueren, K. (2017). School engagement trajectories in adolescence: The role of peer likeability and popularity. *Journal of School Psychology*, 64, 61–75. https://doi.org/10.1016/j.jsp.2017.04.006

Evans, D., Borriello, G. A., & Field, A. P. (2018). A review of the academic and psychological impact of the transition to secondary education. *Frontiers in psychology*, *9*, 1482. https://doi.org/10.3389/fpsyg.2018.01482

Evertson, C. M., & Weinstein, C. S. (2006). Classroom management as a field of inquiry. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 3–15). Lawrence Erlbaum Associates Publishers.

Fauth, B.*, Atlay, C.*, Dumont, H. & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, 101355. https://doi.org/10.1016/j.learninstruc.2020.101355

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29,* 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fauth, B., Göllner, R., Lenske, G., Praetorius, A.K., & Wagner, W. (2020a). Who sees what?: conceptual considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik. Beiheft, 66*(1), 138-155.

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., & Trautwein, U. (2020b). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational Psychology, 112*(6), 1284–1302. https://doi.org/10.1037/edu0000416

Feldlaufer, H., Midgley, C., & Eccles, J. S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high school. *The Journal of Early Adolescence, 8*(2), 133–156. https://doi.org/10.1177/0272431688082003

Fend, H. (1998): *Qualität im Bildungswesen*. Juventa.

Fend, H. (2008). *Schule gestalten: Systemsteuerung, Schulentwicklung, Unterrichtsqualität* [*Designing school: System control, school development, teaching quality*]. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90867-0

Fraser, B. J. (1991). Validity and use of classroom environment instruments. *Journal of Classroom Interaction, 26*(2), 5–11.

Fraser, B. J., & O'Brien, P. (1985). Student and teacher perceptions of the environment of elementary school classrooms. *The Elementary School Journal, 85*(5), 567–580. https://doi.org/10.1086/461422

Fraser, B. J., & Walberg, H. J. (Eds.). (1991). *Educational environments: Evaluation, antecedents and consequences.* Pergamon Press.

Fraser, B.J., Walberg, H.J. (1981). Psychosocial learning environment in science classrooms: A review of research. *Studies in Science Education*, *8*, 67–92.

Furman, W., & Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child Development*, *63*(1), 103–115. https://doi.org/10.2307/1130905

Gazelle, H. (2006). Class climate moderates peer relations and emotional adjustment in children with an early history of anxious solitude: A child × environment model. *Developmental Psychology, 42*(6), 1179–1192. https://doi.org/10.1037/0012-1649.42.6.1179

Gehlbach, H. (2015). Seven Survey Sins. *The Journal of Early Adolescence*, *35*(5–6), 883–897. https://doi.org/10.1177/0272431615578276

Gitomer, D. H., & Bell, C. A. (2013). Evaluating teaching and teachers. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 3. Testing and*

assessment in school psychology and education (pp. 415–444). American Psychological Association. https://doi.org/10.1037/14049-020

Goldstein, H. (2010). *Multilevel statistical models*. Fourth Edition. London, Edward Arnold. . https://doi.org/10.1002/9780470973394

Göllner, R., Fauth, B., Lenske, G., Praetorius, A.K., Wagner, W. (2020). Do student ratings of classroom management tell us more about teachers or classroom composition? *Zeitschrift für Pädagogik, 66,* 156–172.

Göllner, R., Wagner, W., Eccles, J. S., & Trautwein, U. (2018). Students' idiosyncratic perceptions of teaching quality in mathematics: A result of rater tendency alone or an expression of dyadic effects between students and teachers? *Journal of Educational Psychology, 110*(5), 709–725. https://doi.org/10.1037/edu0000236

Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung* [Teaching and learning: Students as raters of teaching quality]. Waxmann.

Gutman, L. M., & Eccles, J. S. (2007). Stage-environment fit during adolescence: Trajectories of family relations and adolescent outcomes. *Developmental Psychology, 43*(2), 522–537. https://doi.org/10.1037/0012-1649.43.2.522

Hall-Lande, J. A., Eisenberg, M. E., Christenson, S. L., & Neumark-Sztainer, D. (2007). Social isolation, psychological health, and protective factors in adolescence. *Adolescence*, *42*, 265-286.

Hamre, B. K. & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization and measurement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). Routledge.

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

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning.* Routledge/Taylor & Francis Group.

Helmke, A. (2003): Unterrichtsqualität erfassen, bewerten, verbessern [Assessing, evaluating, improving teaching quality]. Klett Kallmeyer

Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (7th ed.). Klett-Kallmeyer.

Hidi, S., & Ainley, M. (2002). Interest and adolescence. In F. Pajares & T. Urdan (Eds.), *Academic motivation of adolescents* (pp. 247-275). IAP.

Hinz, A., Michalski, D., Schwarz, R., & Herzberg, P. Y. (2007). The acquiescence effect in responding to a questionnaire. *Psycho-social medicine*, *4*, Doc07.

Hochweber, J., Hosenfeld, I., & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology, 106*(1), 289–300. https://doi.org/10.1037/a0033829

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge/Taylor & Francis Group.

Hoy, A. W., & Weinstein, C. S. (2006). Student and teacher perspectives on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 181–219). Lawrence Erlbaum Associates Publishers.

Hughes, J. N. (2011). Longitudinal effects of teacher and student perceptions of teacher-student relationship qualities on academic adjustment. *The Elementary School Journal, 112*(1), 38–60. https://doi.org/10.1086/660686

Jonkmann, K., Rose, N., & Trautwein, U. (2013). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen* [Tradition and innovation: Developmental pathways in the lower and intermediate track in Baden-Württemberg and at comprehensive schools in Saxony]. Abschlussbericht für die Länder Baden-Württemberg und Sachsen.

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Karing, C., Pfost, M. & Artelt, C. (2013). Is secondary school teacher judgement accuracy related to the development of students' reading literacy? In M. Pfost, C. Artelt & S. Weinert (Eds.), *The development of reading literacy from early childhood to adolescence. empirical findings from the Bamberg BiKS Longitudinal Study* (pp. 279-309). Bamberg: University of Bamberg Press.

Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher, 39*(8), 591–598. https://doi.org/10.3102/0013189X10390804

Klafki, W. (1963). *Studien zur Bildungstheorie und Didaktik* [Studies concerning educational theory and didactics]. Beltz.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* [On the Development of National Educational Standards. An Expertise]. BMBF.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur" und Unterrichtsgestaltung [Mathematical instruction in secondary school: "Task culture" and instructional settings]. In E. Klieme, & J. Baumert (Eds.), *TIMSS—Impulse für Schule und Unterricht* [TIMSS—Impulses for school and classroom instruction] (pp. 43–57). Federal Ministry of Education and Research.

Köller, O., & Baumert, J. (2001). Leistungsgruppierungen in der Sekundarstufe I. Ihre Konsequenzen für die Mathematikleistung und das mathematische Selbstkonzept der Begabung [Ability-grouping at secondary level 1. Consequences for mathematics achievement and the self-concept of mathematical ability]. *Zeitschrift für Pädagogische Psychologie, 15*(2), 99–110. https://doi.org/10.1024/1010-0652.15.2.99

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Rinehart & Winston.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: a multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 97-124). Springer.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. http://dx.doi.org/10.1037/a0032583

Lenske, G. (2016). *Schülerfeedback in der Grundschule* [Students' feedback in primary school]. Waxmann.

Léon, J., Medina-Garrido, E., & Núñez, J. L. (2017). Teaching quality in math class: The development of a scale and the analysis of its relationship with engagement and achievement. *Frontiers in psychology*, *8*, 895. https://doi.org/10.3389/fpsyg.2017.00895

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Longobardi, C., Settanni, M., Prino, L. E., Fabris, M. A., & Marengo, D. (2019). Students' psychological adjustment in normative school transitions from kindergarten to high school: Investigating the role of teacher-student relationship quality. *Frontiers in Psychology, 10,* Article 1238. https://doi.org/10.3389/fpsyg.2019.01238

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106. https://doi.org/10.1111/j.1750-8606.2008.00048.x

Maehr, M. L., & Midgley, C. (1991). Enhancing student motivation: A schoolwide approach. *Educational Psychologist, 26*(3-4), 399–427. https://doi.org/10.1207/s15326985ep2603&4_9

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential baises, and utility. *Journal of Educational Psychology, 76*(5), 707–754. https://doi.org/10.1037/0022-0663.76.5.707

Marsh, H. W., Ginns, P., Morin, A. J. S., Nagengast, B., & Martin, A. J. (2011). Use of student ratings to benchmark universities: Multilevel modeling of responses to the Australian Course Experience Questionnaire (CEQ). *Journal of Educational Psychology, 103*(3), 733–748. https://doi.org/10.1037/a0024221

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47,* 106–124. http://dx.doi.org/10.1080/00461520.2012.670488

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child development*, *79*(3), 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*(1), 110–133. https://doi.org/10.2307/1130295

McCaslin, M., Bozack, A. R., Napoleon, L., Thomas, A., Vasquez, V., Wayman, V., & Zhang, J. (2006). Self-Regulated learning and classroom management: Theory, research, and considerations for classroom practice. In C. M. Evertson & C. S. Weinstein

(Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 223–252). Lawrence Erlbaum Associates Publishers.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107–122. https://doi.org/10.1007/BF02294746

Meyer, H. (2004). *Was ist guter Unterricht?* Cornelsen.

Midgley, C. (Ed.). (2002). *Goals, goal structures, and patterns of adaptive learning*. Lawrence Erlbaum Associates Publishers.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity, 42*(6), 779–794. https://doi.org/10.1007/s11135-006-9067-x

Moos, R. H. (1979). *Evaluating educational environments: Procedures, measures, findings and policy implications*. Jossey-Bass.

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education, 82*(2), 143–167. https://doi.org/10.1080/00220973.2013.769412

Morinaj, J., & Hascher, T. (2019). School alienation and student well-being: A cross-lagged longitudinal analysis. *European Journal of Psychology of Education, 34*(2), 273–294. https://doi.org/10.1007/s10212-018-0381-1

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J, and O'Connor, K.M. (2003), *TIMSS assessment frameworks and specifications 2003*. Boston College.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st Century*. The National Academies Press. https://doi.org/10.17226/13398.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education, 7,* 133–144. https://doi.org/10.1177/1477878509104318

Nurmi, J.-E., & Kiuru, N. (2015). Students' evocative impact on teacher instruction and teacher-child relationships: Theoretical background and an overview of previous research. *International Journal of Behavioral Development, 39,* 445– 457. https://dx.doi.org/10.1177/01650254155 92514

OECD (2021). *Beyond academic learning: First results from the survey of social and emotional skills*. OECD Publishing. https://doi.org/10.1787/92a11084-en.

Opdenakker, M.-C., Maulana, R., & den Brok, P. (2012). Teacher—Student interpersonal relationships and academic motivation within one school year: Developmental changes and linkage. *School Effectiveness and School Improvement, 23*(1), 95–119. https://doi.org/10.1080/09243453.2011.619198

Ophardt D., & Thiel, F. (2017). Klassenmanagement als Basisdimension der Unterrichtsqualität [Classroom management as a dimension of teaching quality]. In Schweer M. (Ed.), *Lehrer-Schüler-Interaktion* [Teacher-student interaction]. Springer VS. https://doi.org/10.1007/978-3-658-15083-9_11

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology, 95*(3), 695–708. https://doi.org/10.1037/0022-3514.95.3.695

Pajares, F., & Urdan, T. C. (2008). *The ones we remember: Scholars reflect on teachers who made a difference*. Information Age.

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102(4), 896. https://doi.org/10.1037/a0019545

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99*(1), 83–98. https://doi.org/10.1037/0022-0663.99.1.83

Patrick, H., Turner, J. C., Meyer, D. K., & Midgley, C. (2003). How teachers establish psychological environments during the first days of school: Associations with avoidance in mathematics. *Teachers College Record, 105*(8), 1521–1558. https://doi.org/10.1111/1467-9620.00299

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press.

Pekrun, R., vom Hofe, R., Blum, W., Götz, T.,Wartha, S., Frenzel, A. & Jullien, S. (2006). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA). Entwicklungsverläufe, Schülervoraussetzungen und Kontextbedingungen von Mathematikleistungen in der Sekundarstufe [Project for analyzing the development of students‘ mathematic achievement in secondary school]. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* [Investigations of the educational quality of schooling: Final report of the DFG-program] (pp. 21–53). Waxmann.

Pfost, M. & Artelt, C. (2013). Reading literacy development in secondary school and the effect of differential institutional learning environments. In M. Pfost, C. Artelt & S. Weinert (Eds.), *The Development of Reading Literacy from Early Childhood to Adolescence. Empirical Findings from the Bamberg BiKS Longitudinal Study* (pp. 229-277). University of Bamberg Press.

Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. In M. Shinn & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 21–39). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195327892.003.0002

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manuel, Pre- K*. Brookes Publishing.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539-569. http://dx.doi.org/10.1146/annurev-psych-120710-100452

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended

remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145. https://doi.org/10.1037/a0033323

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications Inc.

Raver, C.C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75*, 346-353. https://doi.org/10.1111/j.1467-8624.2004.00676.x

Reusser, K. (2006). Konstruktivismus — vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur [Constructivism— Moving from a guiding principle towards renewing the culture of teaching methodology]. In M. Baer, M. Fuchs, P. Füglister, K. Reusser, & H. Wyss (Eds.), *Didaktik auf psychologischer Grundlage. Von Hans Aeblis kognitionspsychologischer Didaktik zur modernen Lehr-und Lernforschung* (pp. 151–168.) Hep-Verlag.

Reusser, K. (2008). Empirisch fundierte Didaktik – didaktisch fundierte Unterrichtsforschung. Eine Perspektive zur Neuorientierung der Allgemeinen Didaktik [Teaching methodology on the basis of empiricism- educational research on the basis of teaching methodology]. In M. A. Meyer, M. Prenzel & S. Hellekamps (Eds.), *Perspektiven der Didaktik. Zeitschrift für Erziehungwissenschaft*, *Sonderheft 9*, 219–237. https://dx.doi.org/10.1007/978-3-531-91775-7_15

Reusser, K., Pauli, C., & Waldis, M. (Eds.), (2010). *Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und einer schweizerischen Videostudie zum Mathematikunterricht* [Designing teaching and teaching quality – results from an international video-based study on math lessons in Switzerland]. Waxmann.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*(4), 958–972. https://doi.org/10.1037/a0015861

Roeser, R. W., & Eccles, J. S. (1998). Adolescents' perceptions of middle school: Relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence, 8*(1), 123–158. https://doi.org/10.1207/s15327795jra0801_6

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (2000). School as a context of early adolescents' academic and social-emotional development: A summary of research findings. *The Elementary School Journal, 100*(5), 443–471. https://doi.org/10.1086/499650

Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology, 88*(3), 408–422. https://doi.org/10.1037/0022-0663.88.3.408

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The Influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*(4), 493–529. https://doi.org/10.3102/0034654311421793

Rucinski, C. L., Brown, J. L., & Downer, J. T. (2018). Teacher–child relationships, classroom climate, and children's social-emotional and academic development. *Journal of Educational Psychology*, 110(7), 992-1004. https://doi.org/10.1037/edu0000240

Rumberger, R. W. (2001). *Why students drop out and what can be done*. Harvard Civil Rights Project.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and instruction*, *42*, 95–103. https://doi.org/10.1016/j.learninstruc.2016.01.004

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom: Self-report and projective assessments of individual differences in children's perceptions. *Journal of Personality and Social Psychology, 50*(3), 550–558. https://doi.org/10.1037/0022-3514.50.3.550

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179–200). American Psychological Association. https://doi.org/10.1037/10409-006

Scharenberg, K. (2014). Schule und Schulklasse als soziale Kontexte der Entwicklung im Jugendalter [Schools and classrooms as social environments for adolescent development]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *66* (Supplement 54), 317-348. https://doi.org/10.1007/s11577-014-0265-6

Schenke, K., Ruzek, E., Lam, A. C., Karabenick, S. A., & Eccles, J. S. (2018). To the means and beyond: Understanding variation in students' perceptions of teacher emotional support. *Learning and Instruction, 55,* 13–21. https://doi.org/10.1016/j.learninstruc.2018.02.003

Schenke, K., Ruzek, E., Lam, A. C., Karabenick, S. A., & Eccles, J. S. (2017). Heterogeneity of student perceptions of the classroom climate: a latent profile approach. *Learning Environments Research, 20*, 289–306. http://dx.doi.org/10.1007/s10984-017-9235-z

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7,* Article 110. http://dx.doi.org/10.3389/fpsyg.2016.00110

Schiepe-Tiska, A. (2019). School Tracks as Differential Learning Environments Moderate the Relationship Between Teaching Quality and Multidimensional Learning Goals in Mathematics. *Frontiers In Education*, *4*(4). https://doi.org/10.3389/feduc.2019.00004

Schweig, J. D. (2016). Moving beyond means: Revealing features of the learning environment by investigating the consensus among student ratings. *Learning Environments Research, 19*(3), 441–462. https://doi.org/10.1007/s10984-016-9216-7

Seidel, T. (2015). Performance assessment and professional development in university teaching. In: Welpe, I. M., Wollersheim, J., Ringelhan, S., & Osterloh; M. (Ed.), *Incentives and Performance* (pp. 465-477). Springer. https://doi.org/10.1007/978-3-319-09785-5

Seidel, T., Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454-499. https://doi.org/10.3102/0034654307310317.

Shavelson, R., & Dempsey-Atwood, N. (1976). Generalizability of measures of teaching behavior. *Review of Educational Research, 46*(4), 553–611. https://doi.org/10.3102/00346543046004553

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*(4), 765–781. https://doi.org/10.1037/a0012840.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718‑737. https://doi.org/10.1037/0022-3514.94.4.718.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of personality and social psychology*, *100*(2), 330–348. https://doi.org/10.1037/a0021717

Spilt, J. L., Hughes, J. N., Wu, J. Y., & Kwok, O. M. (2012). Dynamics of teacher–student relationships: Stability and change across elementary school and the influence on children's academic success. *Child Development, 83*(4), 1180–1195. https://doi.org/10.1111/j.1467-8624.2012.01761.x

Spooren, P., Mortelmans, D., & Thijssen, P. (2012). “Content” versus “style”: Acquiescence in student evaluation of teaching? *British Educational Research Journal*, *38*(1), 3–21. https://doi.org/10.1080/01411926.2010.523453

Steinberg, L., & Silk, J. S. (2002). Parenting adolescents. In M. H. Bornstein (Ed.), *Handbook of parenting: Children and parenting* (pp. 103–133). Lawrence Erlbaum Associates Publishers.

Stroet, K., Opdendakker, M.C., & Minnaert, Al. (2013). Effects of need supportive teaching on early adolescents‘ motivation and engagement: a review of the literature. Educational Research Review, 9, 65-87. https://doi.org/10.1016/j.edurev.2012.11.003

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. Journal of Marketing Research, 45(1), 116–131. https://doi.org/10.1509/jmkr.45.1.116

Trautwein, U., Dumont, H., & Dicke, A. (2015). Schooling: Impact on cognitive and motivational development. International Encyclopedia of the Social & Behavioral Sciences, 21, 219-124. https://doi.org/10.1016/B978-0-08-097086-8.26056-X

Turner, J. C., & Meyer, D. K. (2000). Studying and understanding the instructional contexts of classrooms: Using our past to forge our future. *Educational Psychologist, 35*(2), 69–85. https://doi.org/10.1207/S15326985EP3502_2

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

van Ryzin, M. (2010). Secondary school advisors as mentors and secondary attachment figures. *Journal of Community Psychology, 38*, 131–154. https://dx.doi.org/10.1002/jcop.20356

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*, 195–217. https://doi.org/10.1093/ijpor/eds021

Virtanen, T.E., Räikkönen, E., Engels, M.C., Vasalampi, K., & Lerkkanen, M.-K. (2021). Student engagement, truancy, and cynicism: a longitudinal study from primary school to upper

secondary education. *Learning and Individual Differences*, *86*, Article 101972. https://doi.org/10.1016/j.lindif.2021.101972

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction, 28,* 1–11. https://doi.org/10.1016/j.learninstruc.2013.03.003

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108,* 705–721. http://dx.doi.org/10.1037/edu0000075

Walberg, H. J., Singh, R., & Rasher, S. P. (1977). Predictive validity of student perception: A cross-cultural replication. *American Educational Research Journal, 14*(1), 45–49. https://doi.org/10.2307/1162518

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*, 249-294. https://doi.org/10.3102/00346543063003249

Wang, M. T., & Dishion, T. J. (2012). The Trajectories of adolescents' perceptions of school climate, deviant peer affiliation, and behavioral problems during the middle school years. *Journal of research on adolescence: the official journal of the Society for Research on Adolescence*, *22*(1), 40–53. https://doi.org/10.1111/j.1532-7795.2011.00763.x

Wang, Y., King, R.B., Wang, F., & Leung, S.O. (2021). Need-supportive teaching is positively associated with students' well-being: A cross-cultural study. *Learning and Individual Differences, 92.* https://doi.org/10.1016/j.lindif.2021.102051

Waxman, H. C. (1991). Investigating classroom and school learning environments: A review of recent research and developments in the field. *Journal of Classroom Interaction, 26*(2), 1–4.

Way, N., & Greene, M. L. (2006). Trajectories of perceived friendship quality during adolescence: The patterns and contextual predictors. *Journal of Research on Adolescence, 16*(2), 293- 320. doi:10.1111/j.1532-7795.2006.00133.x

Way, N., Reddy, R., & Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: associations with trajectories of psychological and behavioral adjustment. *American journal of community psychology*, *40*(3-4), 194–213. https://doi.org/10.1007/s10464-007-9143-y

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. Psychological Met*hods*, *18*(3), 320–334. https://doi.org/10.1037/a0032121

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*(2), 105–121. https://doi.org/10.1177/0146621609338593

Weinert, F. E., Schrader, F.-W. & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research, 13*, 895-914.

Weinstein, R. S. (1983). Student perceptions of schooling. *The Elementary School Journal, 83*(4), 287–312. https://doi.org/10.1086/461319

Weinstein, R. S., Marshall, H. H., Brattesani, K. A., & Middlestadt, S. E. (1982). Student perceptions of differential teacher treatment in open and traditional classrooms. *Journal of Educational Psychology, 74*(5), 678–692. https://doi.org/10.1037/0022-0663.74.5.678

Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 Years. *Assessment*, *23*(3), 279–291. https://doi.org/10.1177/1073191115583714

Widlund, A., Tuominen, H., & Korhonen, J. (2021). Development of school engagement and burnout across lower and upper secondary education: Trajectory profiles and educational outcomes. *Contemporary Educational Psychology*, *66*. https://doi.org/10.1016/j.cedpsych.2021.101997

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R., Davis-Kean, P. (2006). Motivation. In N. Eisenberg (Ed.), *Handbook of child psychology* (pp. 933–1002). Wiley.

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67*(5), 555–561. https://doi.org/10.1037/0021-9010.67.5.555

Wolff, C.E., Jarodzka, H. & Boshuizen, H.P.A. (2021). Classroom Management Scripts: a Theoretical Model Contrasting Expert and Novice Teachers' Knowledge and Awareness of Classroom Events. *Educational Psychology Review 33*, 131–148. https://doi.org/10.1007/s10648-020-09542-0

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research*, *30*(1), 72–91. https://doi.org/10.1086/374697

Wubbels, T., Brekelmans, M., & Hooymayers, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching and Teacher Education*, *8*(1), 47–58. https://doi.org/10.1016/0742-051X(92)90039-6