

Technische Universität Dortmund

Fakultät Erziehungswissenschaft, Psychologie und Bildungsforschung

**Wahrnehmung von Unterrichtsqualität in der Primarstufe**

–

**Übereinstimmungen sowie Prädiktoren der Wahrnehmungen von  
Lernenden, Lehrkräften und externen Urteilenden**

Kumulative Dissertation zur Erlangung des akademischen Grades

Doktorin der Philosophie

(Dr.in phil.)

vorgelegt von

Jennifer Iglar, M. A.

Erstgutachterin: Prof.in Dr.in Nele McElvany

Zweitgutachterin: Prof.in Dr.in Bernadette Gold

Dortmund, 2024

---

Dissertation in der Fakultät für Erziehungswissenschaft, Psychologie und Bildungsforschung  
an der Technischen Universität Dortmund

---

## Danksagung

Die vorliegende kumulative Dissertation ist im Rahmen meiner Tätigkeit am Institut für Schulentwicklungsforschung der TU Dortmund entstanden und die Umsetzung meines Dissertationsvorhabens war nur mit Unterstützung möglich. Demzufolge gilt in erster Linie mein tiefster Dank Prof.in Dr.in Nele McElvany, die mich in den letzten Jahren effektiv unterstützt hat und mir stets mit Rat und Tat zur Seite stand. Danke, dass ich immer auf eine schnelle Rückmeldung zählen konnte und ich jede Idee und Frage mit Dir diskutieren konnte. Außerdem möchte ich Dr.in Annika Ohle-Peters für ihre wertvolle Unterstützung und ihre kompetenten Rückmeldungen zu Manuskripten bedanken. Frau Prof.in Dr.in Bernadette Gold danke ich für ihre Bereitschaft zur Begutachtung dieser Arbeit.

Zudem bedanke ich mich bei meinen Kolleg\*innen des Instituts für Schulentwicklungsforschung. Durch verschiedene Austauschformate erhielt ich stets effektive Rückmeldungen zu meinen Forschungsvorhaben. Ein besonderer Dank gilt Theresa Schlitter, Dr. Andreas Sander, Dr.in Svenja Hartwig, Nicole Kaufmann und Dr.in Annika Teerling, die mich auf dem Weg zur Promotion begleitet haben und stets moralisch unterstützt haben.

Des Weiteren bedanke ich mich insbesondere bei meinen Eltern. Sie haben mir den Werdegang zur Promotion ermöglicht, indem sie mich immer in allen Phasen meines Lebens unterstützt haben (moralisch und finanziell) und mich stets motiviert haben.

Abschließend möchte ich mich bei meiner Familie bedanken. Danke für euer Verständnis, eure Geduld und eure Unterstützung.

---

---

## Inhaltsverzeichnis

Zusammenfassung .....	10
Abstract .....	13
Einleitung .....	15
1 Unterrichtsqualität .....	19
1.1 Theoretische Konzeptualisierung der Unterrichtsqualität .....	19
1.2 Empirische Relevanz der Unterrichtsqualität .....	26
1.3 Unterrichtsqualitätsdimensionen .....	30
1.3.1 Kognitive Aktivierung.....	31
1.3.2 Klassenführung.....	32
1.3.3 Unterstützendes Unterrichtsklima .....	33
2 Wahrnehmung .....	35
2.1 Begriffsbestimmung der Wahrnehmung.....	35
2.2 Informationsverarbeitungsmodelle .....	37
3 Wahrnehmung von Unterrichtsqualität .....	42
3.1 Methoden zur Erfassung von Unterrichtsqualität .....	43
3.1.1 Inferenzgrad.....	43
3.1.2 Beurteilungszeitraum.....	44
3.1.3 Itembezug .....	46
3.2 Perspektiven zur Erfassung von Unterrichtsqualität.....	48
3.2.1 Wahrnehmungen von Lernenden .....	50
3.2.2 Wahrnehmungen von Lehrkräften.....	54
3.2.3 Wahrnehmungen von externen Urteilenden.....	59
4 Forschungsanliegen und Forschungsfragen .....	62
5 Zusammenfassungen der Einzelbeiträge .....	65
5.1 Beitrag I. Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität .....	66
5.2 Beitrag II. Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht .....	67
5.3 Beitrag III. Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt .....	68
5.4 Beitrag IV (weiterführende Analysen). Students' perceived motivational support in elementary school – How accurate are teachers' judgments?.....	69

---

Literaturverzeichnis I .....	71
6 Beiträge der kumulativen Dissertation .....	104
6.1 Beitrag I. Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität .....	104
6.2 Beitrag II. Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht .....	137
6.3 Beitrag III. Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt .....	167
7 Weiterführende Analysen (Beitrag IV) .....	200
7.1 Students' perceived motivational support in elementary school – How accurate are teachers' judgments? .....	200
8 Gesamtdiskussion .....	246
8.1 Zusammenfassung und Diskussion der zentralen Ergebnisse .....	246
8.1.1 Übereinstimmungen von Wahrnehmungen .....	246
8.1.2 Individuelle und methodische Prädiktoren für die Wahrnehmungen verschiedener Perspektiven .....	249
8.2 Stärken und Grenzen der vorliegenden Arbeit .....	255
8.2.1 Stärken .....	255
8.2.2 Grenzen .....	258
8.3 Implikationen .....	261
8.3.1 Implikationen für die Forschung .....	262
8.3.2 Implikationen für die Praxis .....	264
8.4 Ausblick .....	266
Literaturverzeichnis II .....	268
9 Anhang .....	279
9.1 Liste der Einzelbeiträge .....	279
9.2 Eigenanteile der Doktorandin bei den Einzelarbeiten .....	279
9.3 Eidesstattliche Erklärung .....	284

---

## Tabellenverzeichnis

Tabelle 1 Unterrichtsqualitätsmerkmale nach Helmke (2010, 2022), Meyer (2004, 2017) und Brophy (2000) .....	21
Tabelle 2 Zuordnung ausgewählter Hattie-Merkmale zu den Basisdimensionen guten Unterrichts nach Lotz und Lipowsky (2016) .....	28
Tabelle 3 Beispielitems mit unterschiedlichem Itembezug aus Fauth, Göllner, Lenske, Praetorius & Wagner (2018). .....	46

### Beitrag I

Tabelle 1 Interkorrelation zwischen den Schülereinschätzungen der Unterrichtsqualität und den Schülermerkmalen auf Within-Level .....	118
Tabelle 2 Individuelle Schülermerkmale als Prädiktoren für die Einschätzung von Unterrichtsqualität ( $\beta$ -Koeffizienten und Standardfehlern) .....	121

### Beitrag II

Tabelle 1 Anzahl der befragten Lehrkräfte nach Bundesländern (Analysesample).....	147
Tabelle 2 Deskriptiva, Interkorrelation zwischen den Variablen.....	151
Tabelle 3 Prädiktoren von Beteiligung an innovativen Programmen für die Merkmale des Leseunterrichts (Angaben zu standardisierten $\beta$ -Koeffizienten und Standardfehlern) .....	153

### Beitrag III

Tabelle 1 Deskriptiva der Gesamt- und Teilstichproben.....	176
Tabelle 2 Durchführung der Studie .....	178
Tabelle 3 Mittelwerte und Standardabweichungen der Unterrichtsvideos .....	180
Tabelle 4 Prädiktoren (Experimentelle Bedingung A und C, Erfahrungszeit und Interaktionsterme Bedingung A oder C x Erfahrungszeit) für die Urteile der Unterrichtsqualitätsmerkmale .....	186
Tabelle 5 Prädiktoren (Experimentelle Bedingung A und C, Müdigkeit und Interaktionsterme Bedingung A oder C x Müdigkeit) für die Urteile der Unterrichtsqualitätsmerkmale .....	187

### Weiterführende Analysen (Beitrag IV)

Table 1 Descriptive statistics and intercorrelations of students' perceived motivational support and teachers' judgments on the class level and individual level.....	214
Table 2 The level component and global deviation measure on the class level.....	216
Table 3 The three components of judgment accuracy and the global deviation measure on the individual level.....	219

---

Table 4 Predictors of students' intrinsic reading motivation .....	220
Table 5 Predictors of students' intrinsic reading motivation development.....	222

---

## Abbildungsverzeichnis

Abbildung 1. Linsenmodell nach Brunswik (1943) .....	38
Abbildung 2. Sequenzen der Informationsverarbeitung nach Greifeneder et al. (2018) .....	39
Abbildung 3. Perspektiven zur Erfassung von Unterrichtsqualität .....	50
Abbildung 4. Untersuchungsmodell der vorliegenden kumulativen Dissertation .....	63
Abbildung 5. Individuelle Prädiktoren für die Wahrnehmungen von Lernenden.....	250
Abbildung 6. Individuelle Prädiktoren für die Wahrnehmungen von Lehrkräften.....	251
Abbildung 7. Individuelle Prädiktoren für die Wahrnehmung von externen Urteilenden...	252
Abbildung 8. Methodische Prädiktoren für die Wahrnehmung von externen Urteilenden..	253

### Beitrag I

Abbildung 1. Prädiktoren für die Schülereinschätzungen von Unterricht (Angaben zu standardisierten $\beta$ -Koeffizienten und Standardfehlern; dargestellt ohne Interkorrelationen zwischen den Prädiktoren). .....	123
---	-----

### Beitrag II

Abbildung 1. Mediationsanalyse zur Vorhersagekraft motivationaler Lehrkraftmerkmale über die Beteiligung an innovativen Programmen für die Qualität des Leseunterrichts.....	154
--	-----

### Beitrag III

Abbildung 1. Durchschnittliche Abweichungen der Urteile des Videos mit mittlerem Qualitätsniveau der experimentellen Bedingungen A (niedrig/mittel) und C (hoch/mittel) und signifikanten Pfaden unterteilt nach den Unterrichtsqualitätsmerkmalen.....	182
---	-----

### Weiterführende Analysen (Beitrag IV)

Figure 1. One-factor global deviation measure predicting students' intrinsic reading motivation.....	230
Figure 2. One-factor level component predicting students' intrinsic reading motivation....	230
Figure 3. One-factor global deviation measure predicting students' intrinsic reading motivation development .....	231
Figure 4. One-factor level component predicting students' intrinsic reading motivation development .....	232

---

## Zusammenfassung

Unterrichtsqualität ist bedeutsam für den schulischen Lernerfolg (Allen et al., 2013; Kunter & Baumert, 2006). Allerdings kann Unterrichtsqualität nicht direkt erfasst werden und Aussagen über die Qualität des unterrichtlichen Angebots können nur mittels der Wahrnehmungen von Lernenden, Lehrkräften oder externen Urteilenden gewonnen werden (Clausen, 2002). Hierbei weist jede Perspektive ihre Vor- und Nachteile auf. Generell kann aber gesagt werden, dass Studien gezeigt haben, dass Wahrnehmungen von Unterrichtsqualität über den gleichen Unterricht divergieren zwischen Lernenden, Lehrkräften und externen Urteilenden und die Perspektiven keine große Übereinstimmung aufweisen (Fauth, Decristan, Rieser, Klieme & Büttner, 2014b). In meinem Promotionsvorhaben wurde die Perspektivenabhängigkeit der Wahrnehmung von Unterrichtsqualität näher beleuchtet. Ziel der vorliegenden Arbeit war es, individuelle Prädiktoren für die Wahrnehmungen von Lernenden, Lehrkräften und externen Urteilenden aufzudecken und die Wahrnehmungen von Lernenden und Lehrkräften abzugleichen. Allgemein lässt sich Unterrichtsqualität in drei Dimensionen unterteilen: *Kognitive Aktivierung*, *Klassenführung* und *Unterstützendes Unterrichtsklima*. Im Rahmen des Promotionsprojektes werden einzelne Unterrichtsmerkmale aus diesen Dimensionen betrachtet.

Die Datengrundlage der in dieser Arbeit zugrunde liegenden empirischen Beiträge bildeten Erhebungen aus den Projekten „Entwicklung und Überprüfung von Kompetenzmodellen zur integrativen Verarbeitung von Texten und Bildern (BiTe)“ (vgl. McElvany et al., 2012; Ohle et al., 2017) und „Evaluation von Konzepten und Maßnahmen der fächerübergreifenden Leseförderung im Primarbereich (BiSS-EvalLesen)“ (McElvany et al., 2018). Alle vier Beiträge beziehen sich auf Unterricht in der vierten Grundschulklasse.

Beitrag I sollte Aufschluss darüber geben, wie sehr die Wahrnehmungen von Lernenden der Grundschule über ein und denselben Unterricht divergieren und welche individuellen Prädiktoren die Varianz der Wahrnehmungen vorhersagen. Hierfür wurden  $N = 647$  Lernende aus 33 vierten Grundschulklassen zu ihren Wahrnehmungen zu den Unterrichtsqualitätsmerkmalen *Herausfordernde Aufgaben*, *Störungen im Klassenraum* und *Motivierung durch die Lehrkraft* sowie zu demografischen, kognitiven, motivationalen und emotionalen Merkmalen im Rahmen des Projekts BiTe befragt. Die Analysen deuteten auf geringe Übereinstimmungen von Wahrnehmungen innerhalb einer Klasse hin. Vorwissen, Kognitive Fähigkeiten, sozialer Hintergrund, motivationale und emotionale Merkmale

erwiesen sich als Prädiktoren für die Wahrnehmungen von Grundschulklässler\*innen bezüglich der Unterrichtsqualität.

In Beitrag II wurde der Zusammenhang zwischen der intrinsischen Motivation und Selbstwirksamkeitserwartung von Lehrkräften und ihren Wahrnehmungen von Unterrichtsqualität untersucht. Zusätzlich wurde noch die Rolle der Motivation für die Beteiligung an innovativen Programmen in den Blick genommen. Datengrundlage waren  $N = 40$  Lehrkräfte aus sieben Bundesländern, welche sich aktiv an der zweiten Erhebung des Projekts BiSS-EvalLesen beteiligten und welche neben den Prädiktionsvariablen zu den Unterrichtsqualitätsmerkmalen Differenzierung im Leseunterricht und Motivierung von Lernenden befragt wurden. Die empirischen Befunde zeigten gerichtete Zusammenhänge zwischen den individuellen Lehrkraftmerkmalen und den Wahrnehmungen von Differenzierung im Leseunterricht und Motivierung von Lernenden, welche über die Beteiligung an innovativen Programmen mediiert wurden.

Der Beitrag III fokussierte die Perspektive der externen Urteilenden. Ziel dieses Beitrags war zu untersuchen, ob sich Unterschiede in den Wahrnehmungen von externen Urteilenden bezüglich der Unterrichtsqualitätsmerkmale von methodischen Faktoren, hier der Reihenfolge der Unterrichtsszenen, und von individuellen Prädiktoren, Stimmung (Müdigkeit) und die Vorerfahrung in der Schule, vorhersagen lassen. Zusätzlich wurde der Frage nachgegangen, ob die individuellen Prädiktoren das Auftreten des Reihenfolgeeffekts beeinflussen. Für die Überprüfung der Forschungsfragen wurden  $N = 69$  Studierende auf drei Versuchsgruppen aufgeteilt, welche 10-minütige Unterrichtsszenen aus dem Projekt BiTe mit unterschiedlich hohen Unterrichtsqualitäten präsentiert bekamen: Gruppe A (niedrig/mittel), B (mittel/hoch) und C (hoch/mittel). Ergebnisse des Experiments wiesen auf einen Reihenfolgeeffekt hin und so unterschieden sich die Gruppen B und C signifikant in ihren Wahrnehmungen des Unterrichtsvideos mittlerer Qualität. Zusätzlich erwiesen sich die individuellen Merkmale als Prädiktor für die Wahrnehmungen der Motivierungsqualität und verstärkten den Reihenfolgeeffekt.

In den weiterführenden Analysen (Beitrag IV) wurde der Frage nachgegangen, wie akkurat Lehrkräfte die wahrgenommene Motivierungsqualität von ihren Lernenden in der Grundschulklasse einschätzen und welche Bedeutung die Urteilsgenauigkeit der wahrgenommenen Motivierungsqualität für die (Entwicklung der) Lesemotivation hat. Im Rahmen des Projekts BiSS-EvalLesen wurden zum vierten Messzeitpunkt Lehrkräfte befragt, wie ihre Schüler\*innen die Unterstützung von Autonomieerleben, Kompetenzerleben und

soziale Eingebundenheit auf Klassenebene ( $N = 50$  Lehrkräfte) und für 10 Lernende auf individueller Ebene ( $N = 49$  Lehrkräfte) wahrnahmen. Zusätzlich wurden die Lernenden zu ihrer wahrgenommenen Motivationsunterstützung und zu ihrer Lesemotivation befragt. Erwartungskonform zeigte sich, dass Lehrkräfte die wahrgenommene Motivierungsqualität ihrer Lernenden eher ungenau einschätzten. Die Urteilsgenauigkeit erwies sich als prädiktiv für die (Entwicklung der) Lesemotivation.

Aufgrund der vier empirischen Beiträge zur Wahrnehmung von Unterrichtsqualität können neue Erkenntnisse zu Übereinstimmungen von Wahrnehmungen von Unterrichtsqualitätsmerkmalen in Grundschulklassen innerhalb einer Gruppe von Urteilenden (Lernende, Lehrkräfte oder externe Urteilende), zu Übereinstimmungen von Wahrnehmungen zwischen den einzelnen Gruppen von Urteilenden (Lernende und Lehrkräfte) und zu individuellen sowie auch methodischen Prädiktoren für divergierende Wahrnehmungen der einzelnen Perspektiven gewonnen werden. Die Befunde liefern einen Beitrag für die Erweiterung von theoretischen Modellen in der Bildungsforschung (z. B. Angebots-Nutzungs-Modelle) und Implikationen für weitere empirische Forschung. Außerdem können Empfehlungen für die schulische Praxis abgeleitet werden.

## **Abstract**

Teaching quality is important for learning success in school (Allen et al., 2013; Kunter & Baumert, 2006). However, teaching quality cannot be assessed directly and statements about the teaching quality can only be obtained from the perceptions of learners, teachers or external raters (Clausen, 2002). Each perspective has its advantages and disadvantages. In general, however, it can be said that studies have shown that perceptions of teaching quality differ across the same lessons and that perceptions from the different perspectives do not show much agreement (Fauth et al. 2014b). In my dissertation, the perspective dependency of the perception of teaching quality was examined in more detail. The aim of the present work was to examine individual predictors for the perceptions of learners, teachers and externals and to correlate the perceptions of learners and teachers. In general, teaching quality can be divided into three dimensions: cognitive activation, class management and supportive teaching climate. As part of the dissertation, individual teaching characteristics from these dimensions are considered.

The data basis for the empirical contributions in this work are surveys from the projects "Development and testing of competence models for the integrative processing of texts and images (BiTe)" (McElvany et al., 2012; Ohle et al., 2017) and "Evaluation of concepts and measures for interdisciplinary reading promotion in primary education (BiSS-EvalLesen)" (McElvany et al., 2018). All four studies refer to lessons in the fourth grade of elementary school.

With the first Article, I tested how much the perceptions of primary school learners diverge regarding one and the same lesson and which individual predictors predict the variance in perceptions. For this question,  $N = 647$  learners from 33 fourth grades of elementary school were asked about their perceptions of teaching quality characteristics of challenging tasks, disturbances in the classroom and motivation by the teacher as well as demographic, cognitive, motivational and emotional characteristics as part of the BiTe project. The analyses indicated low agreement of perceptions within a class. Previous knowledge, cognitive abilities, social background, motivational and emotional characteristics proved to be predictors for students' perceptions of elementary school regarding teaching quality.

In article II, the relationship between teachers' intrinsic motivation and self-efficacy expectations and their perceptions of teaching quality was examined. In addition, the role of motivation for participation in innovative programs was also considered. The data was based on  $N = 40$  teachers from seven federal states, who actively participated in the second survey

of the BiSS-EvalLesen project and who, in addition to the prediction variables, were asked about the teaching quality characteristics of differentiation in reading lessons and motivation of learners. The empirical findings showed directed correlations between individual teacher characteristics and perceptions of differentiation in reading instruction and motivation, which were mediated through participation in innovative programs.

Article III focused on the external perspective. The aim of this article was to investigate whether differences in the perceptions of external raters can be predicted from methodical predictors, here the order of the teaching scenes, and from individual predictors, mood (tiredness) and previous experiences at school. In addition, the question of whether the individual predictors influence the occurrence of the order effect was investigated. To review the research questions,  $N = 69$  students from university were divided into three test groups, which were presented ten-minute lesson scenes from the BiTe project with different levels of teaching quality: Group A (low/medium), B (medium/high) and C (high/medium). Experimental results indicated an order effect, and so Groups B and C differed significantly in their perceptions of the medium-quality video. In addition, the individual characteristics proved to be a predictor for the perceptions of the motivational quality and increased the order effect.

In further analyses (article IV) it is tested how accurately teachers judge the perceived motivational quality of their learners in the elementary school classes and what significance the accuracy of their judgments of the perceived motivational quality has for the (development of) reading motivation. As part of the BiSS-EvalLesen project, teachers were asked at the fourth point in time how their students perceived the support of experienced autonomy, experienced competence and experienced social inclusion at class level ( $N = 50$  teachers) and for ten learners at individual level ( $N = 49$  teachers). In addition, the learners were asked about their perceived motivational support and their reading motivation. As expected, it was shown that teachers judged the perceived motivational quality of their learners rather imprecisely. Accuracy of judgment proved to be predictive for (development of) reading motivation.

Based on the four empirical articles about perception of teaching quality, new results of correlation perceptions of teaching quality characteristics in elementary school within a group of judges (students, teachers or external judges), to the correspondence of perceptions between the individual groups of judges (students and teachers) and to individual as well as methodological predictors for diverging perceptions of the individual perspectives can be gained. The findings provide a contribution to the expansion of theoretical models in

educational science and implications for further empirical research. In addition, recommendations for school practice can be derived.

### **Einleitung**

Eine Frage, mit der sich die Bildungsforschung beschäftigt ist, wie das Lehren und Lernen gestaltet werden kann, damit Lernende positive Lernerfolge erzielen. Es bleibt der Bedarf bestehen, Erklärungsfaktoren für unterschiedliche Lernerfolge zu finden. Für den Lernerfolg von Lernenden, werden neben ihren Merkmalen wie das Vorwissen oder das häusliche Umfeld als mögliche Erklärung, auch Merkmale der Lehrkräfte und des unterrichtlichen Angebotes herangezogen. In der empirischen Unterrichtsforschung wurde mit dem Ziel die Frage zu beantworten, wie Unterricht zu möglichst positiven Wirkungen führen kann, Angebots-Nutzungs-Modelle von Unterricht entwickelt (Rakoczy, Frick, Weiß-Wittstadt, Tallon & Wagner, 2022). Basis dieser Modelle ist die Konzeption von Fend (1981), welche die Analyse und Darstellung von Voraussetzungen, Prozessen und Ergebnissen des Lehrens und Lernens sowie die Verbindungen vieler Faktoren beim Lehren und Lernen beschreibt (ebd.). Eine Weiterentwicklung dieses Konzepts ist das Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts von Helmke (2022), in dem er Merkmale der Lehrkraft, des Lernenden sowie des Unterrichts selbst berücksichtigt. Demnach wirken Kompetenzen und das Wissen der Lehrkraft auf das unterrichtliche Angebot. Auch personale Lehrkraftmerkmale wie zum Beispiel Engagement und Geduld, aber auch unterrichtsrelevante Werte und Ziele, beeinflussen den Unterricht und die Unterrichtsqualität, welche sich durch die Prozessqualität (fächerübergreifende und fachspezifische Faktoren) und Qualität des Unterrichtsmaterials auszeichnet. Dieses Lernangebot muss von den Lernenden wahrgenommen und interpretiert werden, bevor es zur Nutzung des Unterrichts (Lernaktivitäten) führen kann. Hierunter fällt auch die aktive Lernzeit. Diese wirkt sich dann positiv auf den Lernerfolg aus. Einen Einfluss auf die Nutzung des unterrichtlichen Angebots hat das Lernpotential der Lernenden. Hierunter zählen zum Beispiel das Vorwissen, kognitive und sprachliche Fähigkeiten, Lernmotivation und Lernstrategien. Förderlich für ein gutes Lernpotential ist das häusliche Umfeld und die damit verbundene Förderung und Bildungsnähe der Familie. Darüber hinaus spielt auch der Kontext, hierunter können zum Beispiel die Schulart und die Klassenzusammensetzung gefasst werden, eine Rolle für die Gestaltung des Unterrichts sowie für den Lernerfolg. In Bezug auf das Angebot-Nutzungs-Modell wird deutlich, dass verschiedene Ursachen einen Einfluss auf den Lernerfolg von Lernenden haben. Im Zentrum des Modells stehen aber die

drei Felder „Unterricht (Angebot)“, „Lernaktivität (Nutzung)“ und „Wirkung (Ertrag)“. Somit wird dem unterrichtlichen Angebot verbunden mit der tatsächlichen Nutzung eine starke Bedeutung für den Lernerfolg von Lernenden zugeschrieben (Helmke, 2022).

Im Vergleich zu anderen Versionen des Angebots-Nutzungs-Modells, welche Unterricht als ein Zusammenspiel von Angebot und Nutzung beschreiben (z. B. Lipowsky, 2015; Reusser & Pauli, 2010), betrachten Helmke (2022) und auch zum Beispiel Seidel (2014) Unterricht nur als Angebot. Auch im Hinblick auf die „Nutzung“ gibt es verschiedene Definitionen in den Modellen. So werden entweder kognitive Lernerfolge (Helmke, 2022), zusätzlich auch emotionale motivationale Faktoren (Klieme, Lipowsky, Rakoczy & Ratzka, 2006) oder auch das soziale Miteinander (Kunter & Trautwein, 2013) beschrieben. Im integrativen Modell von Vieluf und Kolleg\*innen (2020) definieren sie Unterricht bezogen auf einen Unterrichtsgegenstand als Ko-Konstruktion des Handelns von Lehrkraft und Lernenden und Unterricht stellt in diesem Modell das Angebot dar. Die Nutzung meint das Zusammenspiel von Kognitionen, Emotionen und Motivationen in konkreten Unterrichtssituationen. Das Angebot und die Nutzung stehen im wechselseitigen Verhältnis. Trotz der Unterschiede der Modelle kann zusammenfassend für alle festgehalten werden, dass Unterricht und die Qualität des Unterrichts im Zusammenspiel mit der Nutzung des Angebots, mediiert über die Wahrnehmung, entscheidend für den Lernerfolg von Lernenden sind (Praetorius & Kleickmann, 2022). Auch in den ersten internationalen Modellen zu Lehr-Lern-Prozessen steht das unterrichtliche Angebot und Lehrkraftverhalten im Zentrum, die mit Faktoren auf Schul-, Lehrkraft- und Lernendenseite verknüpft sind und nur über Interaktion mit dem Lernenden und die Nutzung von Lernenden zum Lernerfolg führen können (schon Cruickshank, 1985; Proctor, 1984). Im Model of teaching-learning process von Huitt (1995) wird das unterrichtliche Angebot, welches als Lehrkraftverhalten (Planung, Management und Instruktionen) im Model betitelt wird, von der Schulpolitik (z. B. Schulklima) und staatlichen Politik (z. B. Leistungstests und Richtlinien) sowie von Lehrkraftmerkmalen (z. B. Wissen und Fähigkeiten) und Lernendenmerkmalen (z. B. Vorwissen und Motivation) beeinflusst. Außerdem steht das unterrichtliche Angebot im Wechselspiel mit dem Lernendenverhalten, der Nutzung des Angebots. Nur über diese Mediation führt das Lehrkraftverhalten zu den Leistungserfolgen der Lernenden (McIlrath & Huitt, 1995).

Im Dynamic Model of Educational Effectiveness wird nach dem Kontext-, dem Schul-, dem Klassen- und dem Lernendenniveau differenziert. Auf jedem Niveau werden relevante Faktoren definiert, welche einen direkten oder indirekten Einfluss auf den Lernerfolg der

Lernenden haben. Im Modell werden das Lehren und das Lernen als zentrale Prozesse der Schule und Lehrkräfte sowie Lernende als wichtigste Akteure in der Schule dargestellt. Lehrkraftmerkmale werden in diesem Modell aber nicht miteinbezogen. Die Qualität des unterrichtlichen Angebots (z. B. Zeitmanagement, Orientierung, Strukturierung) steht im Wechselspiel mit der Variablen auf dem Lernendenniveau. Auf dem Lernendenniveau wird auch die Nutzung des Angebots verortet (z. B. die Eignung, die Gelegenheit zum Lernen und die Ausdauer von Lernenden), welche von individuellen Merkmalen von Lernenden und den Mitschüler\*innen beeinflusst werden (Creemers & Kyriakides, 2008; Kyriakides, Creemers, Panayiotou & Charalambous, 2020).

Grundsätzlich kann Unterrichtsqualität fächerübergreifend durch drei Unterrichtsdimensionen abgebildet werden: kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima (Klieme, Pauli & Reusser, 2009; Pianta & Hamre, 2009; Praetorius & Charalambous, 2018). Demzufolge sollte der Unterricht für die Lernenden kognitiv ansprechend, klar strukturiert und unterstützend sowie motivierend sein. Da Unterrichtsqualität nicht direkt erfassbar ist, müssen Aussagen über Teilnehmende oder Beobachter\*innen eingeholt werden, um Informationen über die Qualität des unterrichtlichen Angebots zu bekommen (Decristan, Hess, Holzberger & Praetorius, 2020). Hierbei werden häufig die Wahrnehmungen von Lernenden, Lehrkräften oder externen Urteilenden berücksichtigt. Jede dieser Perspektiven weist ihre Vor- und Nachteile auf (Clausen, 2002). Aus psychologischen Theorien ist bekannt, dass Wahrnehmungen von verschiedenen Merkmalen beeinflusst werden und dementsprechend auch unterschiedlich ausfallen können. In diesen Modellen werden insbesondere kognitive Merkmale als Einflussvariablen betrachtet. Hierunter fallen beispielsweise das Vorwissen oder die Vorerfahrung (Greifeneder, Bless & Fiedler, 2018). Auch in Bezug auf die Wahrnehmung von Unterrichtsqualität konnte gezeigt werden, dass die Wahrnehmungen über ein und denselben Unterricht zwischen den Perspektiven divergieren (Wagner et al., 2016) und die Wahrnehmungen der einzelnen Perspektiven nur gering übereinstimmen (Fauth et al., 2014b). Wenig beleuchtet wurden bisher individuelle sowie methodische Merkmale, welche die Wahrnehmungen beeinflussen. Ebenfalls bezogen sich viele Studien auf die Wahrnehmungen von Unterrichtsqualität in der Sekundarstufe (Rakoczy et al., 2022; Schatz, 2021; Willems, 2022).

Ziel der vorliegenden Dissertation ist es, mittels psychologischer Theorien zur Wahrnehmung, Modelle zur Lehr-Lern-Forschung, wie beispielsweise das Angebots-Nutzungs-Modell von Helmke (2022), und empirischen Befunden zur Thematik

*Wahrnehmung von Unterrichtsqualität* individuelle sowie Prädiktoren für die Wahrnehmung von Unterrichtsqualität für die drei verschiedenen Perspektiven (Lernenden, Lehrkräften und externe Urteilenden) zu identifizieren. Ebenfalls werden Zusammenhänge der Perspektiven der Lernenden und der Lehrkräfte näher beleuchtet. In Kapitel 1 wird auf die theoretische Konzeptualisierung und empirische Relevanz der Unterrichtsqualität eingegangen. Am Ende des Kapitels werden die fächerübergreifenden Unterrichtsqualitätsdimensionen kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima im Detail vorgestellt. In Kapitel 2 wird näher auf den Wahrnehmungsprozess und mögliche Einflussfaktoren auf diesen eingegangen. Anschließend wird im dritten Kapitel das Thema Wahrnehmung von Unterrichtsqualität präsentiert. Dabei werden unter anderem die drei Perspektiven (Lernende, Lehrkräfte und externe Urteilende) und die Vor- und Nachteile bei deren Wahrnehmung von Unterrichtsqualität vorgestellt. In Kapitel 4 wird das Forschungsanliegen skizziert sowie ein Untersuchungsmodell der vorliegenden Dissertation vorgestellt. Darauf folgend werden in Kapitel 5 die in dieser Dissertation zugrunde liegenden empirischen Beiträge zusammengefasst vorgestellt. Kapitel 6 umfasst alle vier Beiträge der Dissertation. In Kapitel 7 werden alle bedeutsamen Resultate diskutiert. Ebenfalls werden in diesem Kapitel die Stärken und Grenzen der empirischen Beiträge aufgezeigt sowie Implikationen für Forschung und Praxis. Abschließend wird noch ein Ausblick für weitere Untersuchungen im Bereich Wahrnehmung von Unterrichtsqualität gegeben.

# 1 Unterrichtsqualität

Generell kann Unterricht als Angebot angesehen werden eine „Sache“ zu verstehen und ihr für die eigene Person einen Sinn zu geben, somit sich an der Sache als Individuum zu bilden (Biesta & Stengel, 2016; Klieme, 2020). Praktischer beschrieben ist Unterricht eine komplexe Aktivität und eine sich wiederholende Situation, welche durch die Interaktion von Verhalten der Lehrkräfte und Lernenden gestaltet wird (Helmke, Piskol, Pikowsky & Wagner, 2009; Clausen, 2002). Terhart (1994) definierte Unterricht als organisierte Abfolge von Lehr- und Lernsituationen in oftmals vorgesehenen Institutionen unter regelhaften Bedingungen, welche von ausgebildeten Lehrkräften absichtsvoll geplant und initiiert werden. Hierbei steht der Aufbau von Wissen und der Erwerb von Fertigkeiten sowie Fähigkeiten von Lernenden im Vordergrund. Ziele des Lehrens sind die Ermöglichung, das Anregen, die Aufrechterhaltung und das Optimieren von individuellen Lernprozessen (Helmke, 2022; Hasselhorn & Gold, 2017). Wiederholt beschreiben Definitionen von Unterricht eine bestimmte Situation zwischen Lehrkräften und Lernenden, mit dem Ziel von Wissensvermittlung und -erwerb. Nur wenn Unterricht von Lernenden genutzt wird und auch Unterrichtsziele erreicht werden, kann Unterricht einen Beitrag zur Qualifikation, Sozialisation und persönlichen Entwicklung von jungen Menschen leisten. Eine Voraussetzung hierfür ist aber, dass das unterrichtliche Angebot von hoher Qualität ist (Klieme, 2020).

## 1.1 Theoretische Konzeptualisierung der Unterrichtsqualität

Allgemein kann auch in der Bildung Qualität als Summe von Merkmalen, welche sich auf ein Produkt oder eine Dienstleistung auswirken, beschrieben werden, die von der Bedürfnisbefriedigung abhängig ist (Cheng & Cheung, 2004). In der empirischen Bildungsforschung wurden Unterrichtsmerkmale identifiziert, welche Informationen über das Niveau der Kompetenz, des Wertes und Kontrollierbarkeit von Lernaktivitäten liefern (Lazarides & Raufelder, 2021). Zu diesen Merkmalen können beispielsweise die Wahl der Unterrichtsmethode, kognitive Qualität der Aufgaben, Autonomieunterstützung, Bestimmen der Klassenregeln und die Feedbackkultur gezählt werden (z. B. Bieg et al., 2017; Huber, 2021). Seit den 1960er Jahren wird versucht Merkmale guten Unterrichts zu identifizieren. So wurden zu Beginn Studien durchgeführt, welche Zusammenhänge zwischen einzelnen Unterrichtsmerkmalen und Kriteriumsvariablen, z. B. Leistung von Lernenden, betrachteten

(Praetorius, 2014). International haben insbesondere John Carroll (1963) und Benjamin Bloom (1976) das Konstrukt der Unterrichtsqualität geformt. Beide entwickelten erste theoretische Modelle schulischen Lernens. Nach Carroll (1963) ist der Lernerfolg von dem Verhältnis der benötigten Lernzeit von Lernenden und der aufgewendeten Lernzeit (Unterrichtsqualität) abhängig (Gruehn, 2000). Nach Bloom (1976) wird Unterrichtsqualität durch vier Merkmale definiert: Hinweise zur Strukturierung der Lernstoffabfolge, lobende bzw. tadelnde Bekräftigung, aktive Beteiligung der Lernenden am Unterricht und Rückmeldung zu erbrachten Leistungen sowie Anregungen zur Korrektur (Schatz, 2021). In beiden Modellen spielt die Unterrichtsqualität eine wichtige Rolle und wird als zentrale Einflussvariable für schulischen Erfolg dargestellt, wenn auch bei Bloom (1976) die Unterrichtsqualität nur als indirekte Einflussvariable fungiert. Kritikpunkte an diesen ersten Modellen sind aber die noch recht schwammigen Definitionen von Unterrichtsqualität, nur die Berücksichtigung des indirekten Effekts des Lehrkrafthandelns auf den Lernerfolg und das Fehlen des Einbezugs von außerschulischen Einflussvariablen (Schatz, 2021). Als Weiterentwicklungen der Modelle können Walbergs (1966) *Produktivitätsmodell schulischen Lernens* und das deutlich später entwickelte *QAIT-Modell* von Slavin (1994, 2009) genannt werden. Beide Modelle berücksichtigen zusätzlich zu den Unterrichtsqualitätsmerkmalen auch Merkmale der Lernenden und Umweltmerkmale, wie zum Beispiel Familie und Klassenzugehörigkeit. Der Bericht von Coleman und Kolleg\*innen (1966), welcher Chancengleichheit und Bildungsgerechtigkeit in den USA thematisierte, kann als Startschuss für die empirische Forschung zum Thema Unterrichtsqualität gesehen werden. Auf Grundlage dieses Berichts startete die Debatte um Einflussfaktoren auf den schulischen Lernerfolg (Schatz, 2021).

Danach folgten erste systematische Übersichten von Unterrichtsqualitätsmerkmalen und von Unterrichtsmethoden, welche sich empirisch als wirksam für das Erreichen von Lernzielen erwiesen haben (Brophy & Good, 1986; Rosenshine & Fürst, 1971; Wallen & Travers, 1963; Wang, Haertel & Walberg, 1993). Unterrichtsqualität wird in diesen Zusammenhängen auch häufig über die Effektivität des Unterrichts definiert (z. B. Muijs et al., 2014; Seidel & Shavelson, 2007). Dementsprechend kann Unterricht dann als qualitativ beschrieben werden, wenn er zu erwünschten Entwicklungen seitens der Lernenden führt (Praetorius, Herrmann et al., 2020). Insbesondere international wird der Begriff der Effektivität für qualitativ vollen Unterricht immer noch verwendet (z. B. Boysen, 2020; Carpenter, Witherby & Tauber, 2020).

National haben sich die Merkmalskataloge zum guten Unterricht von Helmke (2010, 2022) und Meyer (2004, 2017) etabliert (siehe Tabelle 1). Helmke (2010) beschrieb zunächst zehn Merkmale guten Unterrichts (siehe Tabelle 1). Hierbei ist allerdings zu beachten, dass einige Unterrichtsqualitätsmerkmale, wie zum Beispiel Klassenführung, kognitive Aktivierung und Motivierung bereits gut empirisch untersucht sind, andere aber kaum empirisch fundiert sind (Lenski, Richter & Pant, 2015). In der revidierten Version nennt Helmke (2022) fünf fächerübergreifende Qualitätsbereiche, welche sich den älteren 10 Merkmalen guten Unterrichts zuordnen lassen. Meyer (2004, 2017) stellte 10 Merkmale guten Unterrichts unter anderem in Anlehnung an Werke von Bloom (1968, 1976) sowie Kounin (2006) und an bestehende empirische Befunde zusammen. Vergleichbar haben sich international die 12-fach- und stufenübergreifenden Merkmale von Brophy (2000) etabliert. Hierbei entwickelte Brophy (2000) den Merkmalskatalog guten Unterrichts anhand von bestehenden empirischen Befunden.

Tabelle 1

*Unterrichtsqualitätsmerkmale nach Helmke (2010, 2022), Meyer (2004, 2017) und Brophy (2000)*

Helmke (2010)	Helmke (2022)	Meyer (2004, 2017)	Brophy (2000)
1. Klassenmanagement	1. Klassenführung	2. Hoher Anteil echter Lernzeit 10. Vorbereitete Umgebung	2. Lerngelegenheiten (Opportunity to learn)
2. Lernförderliches Unterrichtsklima		3. Lernförderliches Klima	1. Unterstützendes Unterrichtsklima (supportive classroom climate)
3. Motivierung	3. Förderung der Lernbereitschaft	5. Sinnstiftendes Kommunizieren	8. Unterstützung der Lerntätigkeit (Scaffolding students' task engagement)
7. Schülerorientierung, Unterstützung			
4. Klarheit und Strukturierung	2. Erleichterung und Förderung der Informationsverarbeitung	1. Klare Strukturierung des Unterrichts 4. Inhaltliche Klarheit	5. Inhaltliche Kohärenz (Coherent content)

5. Wirkungs- und Kompetenzorientierung			
6. Umgang mit Heterogenität	5. Passung	7. Individuelles Fördern	10. Kooperatives Lernen (Co-operative learning)
8. Aktivierung	4. Kompetenzorientierung	5. Sinnstiftendes Kommunizieren	6. Durchdachte Gespräche (Thoughtful discourse)
			7. Praxis- und Anwendungsaktivitäten (Practice and application activities)
		9. Transparente Leistungserwartungen	4. Herstellung einer Lern- und Aufgabenorientierung (Establishing learning orientations)
			12. Leistungserwartungen (Achievement expectations)
9. Angemessene Methodenvariation		6. Methodenvielfalt	
10. Konsolidierung, Lernerfolgs-sicherung		8. Intelligentes Üben	9. Lehren von Lernstrategien (Strategy teaching)
			3. Orientierung am Lehrplan (Curricular alignment)
			11. Kriteriumsorientierte Beurteilungen (Goal-oriented assessment)

Diese populären Merkmalskataloge guten Unterrichts stimmen in vielen Punkten überein. So sind sich alle Autor\*innen einig, dass viel Zeit für das Lernen geschaffen und die

Lerntätigkeiten unterstützt werden sollen. Zusätzlich soll ein lernförderliches Klassenklima geschaffen werden. Darüber hinaus sollte der Unterricht klar strukturiert sein und kognitiv anregend gestaltet werden. Auch sollte die Lehrkraft verschiedene Methoden einsetzen und unter anderem auch im Hinblick auf die Heterogenität der Leistungen, Lernende individuell fördern.

Ein Kritikpunkt an Merkmalskatalogen guten Unterrichts ist, dass diese eine Verknüpfung von Lehren und Lernprozessen vermissen lassen (international: Snook, Clark, Harker, O'Neill & O'Neill, 2010; national: Decristan et al., 2020, Gruschka, 2007). Eine theoretische Konzeption, um die theoretische Lücke zu füllen, ist die Unterscheidung von *Oberfläche* und *Tiefe*. Demnach können Unterrichtsqualitätsmerkmale sowohl der Oberflächen- als auch der Tiefenstruktur von Unterricht zugeordnet werden. Hierbei geht es um die Unterscheidung zwischen greifbaren und dahinterliegenden Merkmalen und der Annahme, dass insbesondere dahinterliegende Merkmale zum Lernerfolg seitens der Lernenden führen (Decristan et al., 2020). Oser und Patry (1990) unterschieden erstmals zwischen einer Sichtstruktur und Basisstruktur (Tiefenstruktur) von Unterricht. Sie nahmen an, dass hinter den sichtbaren Handlungen verborgene Zusammenhänge von Lernschritten stehen. Oberflächenmerkmale, auf der Sichtstruktur, gelten als gut beobachtbar, verhaltensnah und somit für Außenstehende leicht erschließbar (Pauli & Reusser, 2006). Unter anderem zählen hierzu strukturelle Rahmenbedingungen, Aufgabenmaterialien, Sozialform, Sprechanteile und Medien (Kunter & Trautwein, 2013; Pauli & Reusser, 2006; Reusser, 2008). Konträr ist die Annahme bei Tiefenmerkmalen, dass diese nur indirekt beobachtbar sind. Hasselhorn und Gold (2017) verorten die Tiefenmerkmale auf der Mikroebene des unterrichtlichen Handelns. Der Tiefenstruktur werden die Interaktion zwischen der Lehrkraft und den Lernenden und mentale Verarbeitungsprozesse wie die Auseinandersetzung mit dem Lernstoff zugeordnet (Krumbacher, 2016; Kunter & Trautwein, 2013).

In den letzten Jahrzehnten wurden in der empirischen Unterrichtsforschung insbesondere generische Unterrichtsqualitätsdimensionen in den Blick genommen. Dabei war es das Ziel, allgemeingültige und Unterrichtsfach unabhängige Aspekte der Unterrichtsqualität zu identifizieren (Praetorius, Herrmann et al., 2020). Unterrichtsfach unabhängig haben national Klieme und Kolleg\*innen (2009) drei grundlegende Unterrichtsqualitätsdimensionen definiert, welche verschiedene Unterrichtsqualitätsmerkmale zusammenfassen: kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima. Aus wissenschaftstheoretischer Sicht wurde das Modell im Rahmen der Videostudie (1995) der

Trends in International Mathematics and Science Study (TIMSS) erarbeitet, um ein möglichst sparsames Modell zu entwickeln (Klieme, 2002; Praetorius, Rogh & Kleickmann, 2020). Hierfür wurden mehrere Skalen zur Erfassung von Unterrichtsqualitätsmerkmalen seitens Lernenden auf die Perspektive von externen Urteilenden übertragen. Im Anschluss wurden die Unterrichtsqualitätsmerkmale zu drei Basisdimensionen mit Hilfe von exploratorischer Faktorenanalyse zusammengefasst. Diese Basisdimensionen sollten über Fächer, Jahrgangsstufen und Schularten gelten (Praetorius, Rogh & Kleickmann, 2020). International vergleichbar entwickelten Pianta und Hamre (2009) das Classroom Assessment Scoring System und verweisen auf drei fächerübergreifende Dimensionen: instructional support, classroom organization und emotional support. Hierbei sind die Dimensionen theoretisch abgeleitet und empirisch unterstützt. Zunächst bezogen sich die Unterrichtsqualitätsdimensionen auf die Beschulung im Kindergartenalter. Allerdings gibt es inzwischen für das Classroom Assessment Scoring System Überarbeitungen auch für ältere Lernende. Demnach sollen diese Dimensionen auch Jahrgangsstufen und Unterrichtsfach übergreifend gültig sein (Pianta & Hamre, 2009; Pianta, La Paro & Hamre, 2008). Im angloamerikanischen Sprachraum ist auch das Danielson Framework for Teaching (FFT) weit verbreitet. Es beinhaltet vier Unterrichtsqualitätsdimensionen, welche ein breites Spektrum an Unterrichtsqualitätsmerkmalen abdecken: Planning and Preparation, Classroom Environment, Instruction und Professional Responsibilities (Danielson 2011). Hierbei sind die Dimensionen Planning and Preparation und Instruction der kognitiven Aktivierung (bzw. instructional support), Classroom Environment der Klassenführung (bzw. classroom organization) und einige Merkmale der Professional Responsibilities sowie einige der Classroom Environment dem unterstützenden Unterrichtsklima (emotional support) zuzuordnen.

Das theoretische Rahmenmodell der drei Basisdimensionen zur Unterrichtsqualität ist inzwischen gut etabliert und empirisch validiert. Die drei Unterrichtsqualitätsdimensionen bestehen aus Formulierungen generischer Unterrichtsqualitätsmerkmalen. Es ist aber zu beachten, dass sich eine Vielzahl an Studien zu der Thematik Unterrichtsqualität auf den naturwissenschaftlichen und mathematischen Unterricht fokussiert hat (Kunter & Ewald, 2016; Praetorius & Gräsel, 2021). Aktuelle Untersuchungen für den Mathematikunterricht zeigten, dass die drei generischen Unterrichtsqualitätsdimensionen (kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima) die Qualität des Mathematikunterrichts nicht umfassend genug beschreiben und um mathematikdidaktische Merkmale ergänzt werden müssen (z. B. Praetorius & Charalambous, 2018). Somit

erweiterten Praetorius und Charalambous (2018) das Modell der drei Basisdimensionen und schlugen auf Basis einer Synthese aus 12 international verbreiteten Frameworks zur Unterrichtsqualität sieben Unterrichtsqualitätsdimensionen vor. Dabei unterscheiden sie zwischen: 1. Auswahl und Thematisierung von Inhalten, 2. Kognitive Aktivierung, 3. Unterstützung des Übens, 4. formatives Assessment, 5. Unterstützung des Lernens aller Schülerinnen und Schüler, 6. sozio-emotionale Unterstützung und 7. Klassenführung. Im Vergleich zum Modell der drei Basisdimensionen von Klieme et al. (2009) sind hier Unterrichtsqualitätsmerkmale anders zusammengefasst und dadurch entstehen neue Schwerpunkte. So beschreiben Praetorius und Charalambous (2018) zum Beispiel den Umgang mit Heterogenität als eigenständiges Unterrichtsqualitätsmerkmal der Unterrichtsqualitätsdimension Unterstützung des Lernens aller Schülerinnen und Schüler, welcher jeweils in den drei Basisdimensionen indirekt mit einbezogen ist (Rakoczy, 2021).

Nach aktuellem Stand werden auch in anderen Fachdidaktiken weiterhin Merkmale für die Beschreibung guten Unterrichts gesucht und die Vollständigkeit durch die drei bisherigen Unterrichtsqualitätsmerkmale überprüft (z. B. für das Fach Sport: Herrmann & Gerlach, 2020; für das Fach Deutsch: Wiprächtiger-Geppert, Stahns & Riegler, 2021; für das Fach Latein: Lindl & Krämer, 2022). Ziel der allgemeinen Unterrichtsforschung ist es unter der Berücksichtigung der einzelnen Fachdidaktiken gemeinsame Unterrichtsqualitätsmerkmale zu finden, welche in fächerübergreifende Unterrichtsqualitätsdimensionen abgebildet werden können (Praetorius, Herrmann et al., 2020). Jentsch und Kolleg\*innen (2021) empfehlen für die Erfassung der Unterrichtsqualität im Mathematikunterricht, dass zusätzlich zu den drei generischen Basisdimensionen noch die fachdidaktische Strukturierung berücksichtigt werden sollte. Hierunter zählen Unterrichtsqualitätsmerkmale wie die Wissenssicherung und Umgang mit Fehlern.

Generell sollte der Begriff der Unterrichtsqualität nicht mit der Qualität der Lehrkraft gleichgestellt werden. Zwischen diesen beiden Konstrukten herrscht zwar eine systematische Beziehung, allerdings sind die beiden Konstrukte nicht austauschbar (Fauth et al., 2020b). Fauth und Kolleg\*innen (2020b) ließen im Rahmen einer Studie eine Lehrkraft den gleichen standardisierten Unterricht in 30 verschiedenen Klassen derselben Klassenstufe unterrichten. Trotz der vergleichbaren Ausgangssituation variierten die Wahrnehmungen von Lernenden und von externen Urteilenden bezüglich der Unterrichtsqualität erheblich zwischen den Klassen. Merkmale der Klasse, wie zum Beispiel die Motivation der Lernenden in einer Klasse, waren wichtige Prädiktoren für die Unterschiede in den Wahrnehmungen der

Unterrichtsqualität (Fauth et al., 2020b). Aufgrund dieser Resultate sollte die Qualität des Unterrichts nicht als alleiniger Indikator für die Qualität von Lehrkräften gesehen werden. Allerdings spielt die Kompetenz der Lehrkraft eine Rolle für die Stabilität der Unterrichtsqualität. So ist bei Lehrkräften mit einer hohen Kompetenz weniger Variabilität in der Qualität des unterrichtlichen Angebots zu finden (Voss, Zachrich, Fauth & Wittwer, 2022). Zusammengefasst lässt sich im Moment festhalten, dass die drei generischen Basisdimensionen durchaus noch als relevant angesehen werden. Unter der Berücksichtigung, dass Lernen in einem bestimmten Unterrichtsfach stattfindet, stellt sich die Frage, ob auch die fachspezifische Unterrichtsqualität für den Lernerfolg von Lernenden bedeutsam ist. Die drei Basisdimensionen stellen somit im Moment zunächst ein Gerüst dar, welches um fachspezifische Unterrichtsmerkmale ergänzt wird (Jentsch et al., 2021; Praetorius & Charalambous, 2018). Dies wird auch als hybride Erfassung von Unterrichtsqualität beschrieben.

## **1.2 Empirische Relevanz der Unterrichtsqualität**

Lange dominierte in der Unterrichtsforschung die Prozess-Produkt-Forschung, welche unabhängig von theoretischer Rahmung Zusammenhänge zwischen Unterrichtsmerkmalen (Prozess) und leistungsbezogenen, emotionalen und motivationalen Merkmalen von Lernenden (Produkt) in den Blick nahm. Ziel war es so Merkmale guten Unterrichts bestimmen zu können (Praetorius, 2014). Hierzu zählen auch die Meta-Analysen von Fraser, Walberg, Welch und Hattie (1987) und Scheerens und Bosker (1997), in denen sie positive Zusammenhänge zwischen Unterrichtsqualitätsmerkmalen und dem Lernerfolg aufwiesen. Fraser et al. (1987) betonten unter anderem besonders die Relevanz von Verstärkung, Hinweisen, Feedback, personalisierten Instruktionen und kooperativem Lernen für die Leistungen von Lernenden. Scheerens und Bosker (1997) nannten Verstärkung, Feedback, Differenzierung, kooperatives Lernen, aktive Lernzeit und strukturiertes Unterrichten als bedeutsam für die Lernendenvariablen. Zusammengefasst zeigten sich einige Unterrichtsqualitätsmerkmale (z. B. Verstärkung, Feedback und kooperatives Lernen) in beiden Meta-Analysen als effektiv.

Die Meta-Analyse von Seidel und Shavelson (2007) nahm neben dem Zusammenhang zwischen Unterrichtsqualitätsmerkmalen und kognitiven Lernergebnissen von Lernenden, auch Zusammenhänge zu Lernprozessen und motivational-affektiven Variablen von

Lernenden in den Blick. Sie gingen dem Ansatz nach, dass verschiedene Unterrichtsqualitätsmerkmale 1. auf das kognitive Wachstum, 2. auf die Motivationsentwicklung und 3. auf die Lernprozesse von Lernenden kurzfristige oder auch langfristige Auswirkungen haben können (Snow, Corno & Jackson, 1996). Hierfür fassten Seidel und Shavelson (2007) verschiedene Unterrichtsqualitätsmerkmale zu Komponenten zusammen. Den stärksten Effekt auf die kognitiven Lernergebnisse hatte die Komponente domänenspezifische Lernaktivitäten. Diese waren, neben den Komponenten soziale Erfahrungen, Zeit zum Lernen und Regulierung und Überwachung, relevant für die Ausprägung von motivational-affektiven Ergebnissen seitens der Lernenden. Unter der Komponente Regulierung und Überwachung werden unter anderem Feedback und Unterstützung seitens der Lehrkräfte gefasst. Die stärksten Zusammenhänge zu den Lernprozessen zeigten die Komponenten domänenspezifische Lernaktivitäten, Zeit zum Lernen und soziale Erfahrungen (Seidel & Shavelson, 2007).

Erweitert wurde der Prozess-Produkt-Forschungsansatz durch das Modell der drei Basisdimensionen von Unterrichtsqualität von Klieme und Kollege\*innen (2009), welche ihre Unterrichtsqualitätsdimensionen auf existierende Forschungstheorien (Konstruktivismus) und Theorien bezogen (ebd.). Die drei vorgeschlagenen Unterrichtsqualitätsdimensionen (kognitive Aktivierung, Klassenführung und konstruktive Unterstützung) wiesen sowohl in nationalen als auch in internationalen Studien einen Zusammenhang zu Leistungen und zu motivationalen sowie emotionalen Merkmalen von Lernenden auf (z. B. für Leistung: Allen et al., 2013; Dubberke, Kunter, McElvany, Brunner & Baumert, 2008; für motivationale Merkmale: Ohle-Peters et al., 2021; für emotionale Merkmale: Lazarides & Buchholz, 2019). Die Relevanz der Unterrichtsqualität für das Lernen wurde empirisch auch unter anderem von Hattie (2009) Befunden unterstrichen. In seiner quantitativen Synthese von über 800 Metaanalysen zum Zusammenhang von Unterrichtsqualität und Leistungen von Lernenden deckt Hattie auf, dass ca. 30.0 Prozent des kognitiven Lernerfolgs von Lernenden auf Unterrichtsmerkmale und Lehrkraftverhalten zurückzuführen sind. Bezogen auf die Unterrichtsqualitätsmerkmale kann zusammengefasst werden, dass zum Beispiel das Einholen von Feedback, effektive Klassenführung, eine positive Lernenden-Lehrkraft-Beziehung, Kreativitätsförderung einen mittleren bis starken Effekt auf den Lernerfolg haben (Hattie, 2009). Dies sind alles Unterrichtsqualitätsmerkmale, welche sich den drei Basisdimensionen von Klieme und Kolleg\*innen (2009) zuordnen lassen (siehe Tabelle 2).

Tabelle 2

*Zuordnung ausgewählter Hattie-Merkmale zu den Basisdimensionen guten Unterrichts nach Lotz und Lipowsky (2016)*

Kognitive Aktivierung	Klassenführung	Unterstützendes Unterrichtsklima
<ul style="list-style-type: none"> <li>• Feedback (0.73)</li> <li>• Meta-kognitive Strategien (0.69)</li> <li>• Kreativitätsförderung (0.65)</li> <li>• Problemlösen (0.61)</li> <li>• Lerntechniken (0.59)</li> </ul>	<ul style="list-style-type: none"> <li>• Klassenführung (0.52)</li> <li>• Aktive Lernzeit (0.38)</li> <li>• Reduzierung von Unterrichtsstörungen (0.34)</li> </ul>	<ul style="list-style-type: none"> <li>• Lernenden-Lehrkraft-Beziehung (0.72)</li> <li>• Nicht-Etikettieren von Lernenden (0.61)</li> </ul>

Auch in internationalen Vergleichsstudien, wie der internationale Grundschul-Lese-Untersuchung der Jahrgangsstufe 4 (IGLU), wurden national die drei Basisdimensionen eingesetzt, unter anderem mit dem Ziel den Zusammenhang der drei Dimensionen mit der Leseleistung und Lesemotivation zu ermitteln (2016: Hußmann, Wendt, Bos & Rieser, 2020; 2021: McElvany et al., 2023). So konnte ein Zusammenhang zwischen einer effizient eingeschätzten Klassenführung und hohen Leseleistungen sowie einer hohen Lesemotivation gefunden werden. Lernende, die die kognitive Aktivierung als gering einschätzten, wiesen auch geringere Leseleistungen auf. Lernende, die den Unterricht als sehr unterstützend wahrnahmen, zeigten höhere Leseleistungen und eine höhere Lesemotivation (Stahns, Rieser & Lankes, 2020). Auch im nationalen Bildungsmonitoring auf Basis der Bildungsstandards der KMK des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB-Bildungstrend), welches die mathematischen und naturwissenschaftlichen Kompetenzen am Ende der Sekundarstufe I untersucht, werden Unterrichtsqualitätsmerkmale, welche den drei Basisdimensionen und somit der Tiefenstruktur von Unterricht zugeordnet werden, erhoben. Zusätzlich werden auch Unterrichtsqualitätsmerkmale eingesetzt, die zur Sichtstruktur von Unterricht zählen. Im Vergleich zu den Unterrichtsqualitätsmerkmalen der Sichtstruktur, wiesen die Merkmale der Basisdimensionen höhere Zusammenhänge mit den mathematischen Kompetenzen und motivational-emotionalen Merkmalen (mathematikbezogenes Selbstkonzept, mathematisches Interesse und Mathematikangst) der Lernenden auf (Henschel, Rjosk, Holtmann & Stanat, 2019). Ebenfalls konnten die Zusammenhänge zwischen den drei Basisdimensionen und der Leistungen von Lernenden sowie der Motivation (intrinsische

Motivation) in der TIMSS-Studie 2019 zu den mathematischen und naturwissenschaftlichen Kompetenzen von Grundschulkindern in Deutschland bestätigt werden. Hier zeigte sich insbesondere die Verbindung zwischen der konstruktiven Unterstützung und den Leistungen sowie intrinsischen Motivation von Lernenden (Stang, Lepper, Steffensky & McElvany, 2020). In der PISA-Erhebung 2022 wurden Lernende der Sekundarstufe I zu verschiedenen Unterrichtsqualitätsmerkmalen (Kognitive Aktivierung: Mathematisches Argumentieren, Kognitive Aktivierung: Ermutigung zum mathematischen Denken, Disziplin im Klassenzimmer, Unterstützung durch die Lehrkraft) befragt, welche sich den Unterrichtsqualitätsdimensionen zuordnen lassen. Die Hälfte der Lernenden nahm den Unterricht als durchschnittlich unterstützend und moderat aktivierend wahr. Es zeigte sich ein Unterschied zwischen den Wahrnehmungen des Unterrichtsqualitätsmerkmals Kognitive Aktivierung: Mathematisches Argumentieren und Kognitive Aktivierung: Ermutigung zum mathematischen Denken, indem mehr Möglichkeiten zum mathematischen Argumentieren als zum eigenen mathematischen Denken wahrgenommen wurden. Die Angaben zu den Unterrichtsqualitätsmerkmalen Disziplin im Klassenzimmer und Unterstützung durch die Lehrkraft fielen im Vergleich zum OECD-Durchschnitt leicht unterdurchschnittlich aus. Die fünfzehnjährigen Lernenden, welche den Mathematikunterricht als unterstützend und aktivierend wahrnahmen, erreichten auch eine höhere mathematische Kompetenz. Ebenfalls gingen positive Wahrnehmungen der Unterrichtsqualität mit positiven Ausprägungen der Freude und des Interesses am Mathematikunterricht einher (Schiepe-Tiska et al., 2023).

Die Frage nach der Unterrichtsqualität in verschiedenen Unterrichtsfächern und die Relevanz für den Lernerfolg ist stets noch aktuell. So wird beispielsweise im Projekt Fachspezifische Lehrkraftkompetenzen – Prädiktive Validierung von Professionswissenstests für Lehrkräfte in sechs Unterrichtsfächern (Deutsch, Englisch, Evangelische Religion, Latein, Mathematik und Physik) (FLAKO) unter anderem der Frage nachgegangen, wie generische und fachspezifische Unterrichtsqualitätsmerkmale und das domänenspezifische Professionswissen von Lehrkräften mit dem Lernerfolg von Lernenden zusammenhängen (Krauss et al., 2017; Eine Verlängerung des Projekts wurde 2021 aus aktuellem Anlass vom Bundesministerium für Bildung und Forschung bewilligt).

### 1.3 Unterrichtsqualitätsdimensionen

Welche Unterrichtsqualitätsmerkmale guten Unterricht ausmachen, ist noch nicht final geklärt. Allerdings kann in Anlehnung an die theoretischen Herleitungen (z. B. Merkmalskataloge von Helmke [2010, 2021] und Meyer [2004, 2017] sowie Dimensionen, Subdimensionen und fachspezifische Ergänzungen in der Version des erweiterten Syntheseframeworks nach Praetorius, Herrmann und Kolleg\*innen [2020]) zusammenfassend festgehalten werden, dass Unterricht kognitiv aktivierend, gut strukturiert und unterstützend sein sollte, um als effektiv zu gelten. Unterstützt wird diese Annahme von empirischen Befunden. So unterstreicht zum Beispiel Hattie (2009) die Bedeutung von Unterrichtsmerkmalen für den Lernerfolg, welche sich der Unterrichtsqualitätsdimensionen von Klieme et al. (2009) zuordnen lassen. Auch in Hatties (2023) neusten Analysen zur Effektivität von unter anderem Unterrichtsqualitätsmerkmalen für den Lernerfolg erwiesen sich die Klassenführung ( $d = 0.43$ ), Zusammenhalt der Klasse ( $d = 0.66$ ), worunter er auch die soziale Unterstützung fasst, und die Beziehung zwischen Lernenden und Lehrkräften ( $d = 0.62$ ) als relevant.

Inwieweit die drei Unterrichtsqualitätsdimensionen zusammenhängen, wurde in verschiedenen Studien untersucht. Demnach konnten statistisch relevante Zusammenhänge, welche teilweise sehr hoch ausfielen, zwischen der kognitiven Aktivierung und dem unterstützenden Unterrichtsklima gefunden werden (Ciftci, 2015,  $r = .48$ ; Fauth et al., 2014b,  $r = .83$ ; Wagner, Göllner, Helmke, Trautwein & Lüdtke, 2013,  $r = .72$ ). Ein unterstützendes Lehrkraftverhalten steht demzufolge im Zusammenhang mit einem kognitiv anregenden Verhalten (Talić et al., 2022). Auch zeigten sich mittlere bis hohe Zusammenhänge zwischen der Klassenführung und des unterstützenden Unterrichtsklimas (Baier et al., 2019,  $r = -.61$ ; Wagner et al., 2013,  $r = .49$ ). Zwischen der kognitiven Aktivierung und Klassenführung zeigten sich eher niedrige bis mittlere Zusammenhänge (Baier et al., 2019,  $r = -.32$ ; Fauth et al., 2014b,  $r = .28$ ; Wagner et al., 2013,  $r = .59$ ). Diese Zusammenhänge zeigten sich sowohl für die Primar- als auch für die Sekundarstufe (Primarstufe: Fauth et al., 2014b; Sekundarstufe: Talić et al., 2022). Dass sich die Zusammenhänge zwischen den Unterrichtsqualitätsdimensionen auch Unterrichtsfach unabhängig zeigten, belegten Talić und Kolleg\*innen (2022).

### 1.3.1 Kognitive Aktivierung

Für die Gestaltung eines qualitätsvollen Unterrichts wird die kognitive Aktivierung als das relevanteste Merkmal beschrieben (Hasselhorn & Gold, 2017). Eine theoretische Annahme ist, dass kognitive Aktivierung in Form von unterrichtlichem Angebot, durch die Nutzung von Lernenden, zu einer tieferen Informationsverarbeitung und somit zu höheren Leistungen auf Seiten der Lernenden führt (Klieme & Rakoczy, 2008). Trotz einer großen Diversität an Operationalisierungen in Untersuchungen zum Unterrichtsmerkmal, stimmen auf abstrakter Ebene die Ideen, was kognitive Aktivierung abbilden soll, überein (Hugener, 2008; Praetorius, Klieme, Herbert & Pinger, 2018; Praetorius & Gräsel, 2021). Generell zielt sie darauf ab, kognitive Prozesse sowie eine tiefere Informationsverarbeitung mittels anregender Problemstellungen zu aktivieren (Bransford, Brown & Cocking, 2000; Lipowsky et al., 2009). Notwendig sind dabei das Aufzeigen von Gemeinsamkeiten und Unterschieden von Problemstellungen sowie das Auffordern zur Erläuterung und Verteidigung eigener Ideen. Das Stellen von herausfordernden Aufgaben ist ein besonders wichtiges Merkmal, da dies zu einer intensiven Auseinandersetzung mit dem Lernstoff und nachhaltigem Lernen führt (Klieme & Rakoczy, 2008; Mayer, 2004). Lernende nach ihrem Leistungsniveau zu differenzieren und das Anspruchsniveau anzupassen dient als Voraussetzung für eine kognitiv anregende Lernsituation (Leuders & Holzäpfel, 2011). In der praktischen Umsetzung kann dies bedeuten, dass unterschiedlich schwierige Aufgabenstellungen oder unterschiedlich lange Bearbeitungszeiten eingesetzt werden (Hertel, 2014; Lipowsky, Kastens, Lotz & Faust, 2011).

Die Bedeutsamkeit eines kognitiv anregenden Unterrichts für den Lernerfolg von Lernenden, konnte durch den positiven Zusammenhang zwischen Facetten der kognitiven Aktivierung und der Leistung von Lernenden in empirischen Studien gezeigt werden (z. B. Chinn, Anderson & Waggoner, 2001; McDonald Connor et al., 2009). Auch Hamre und Pianta (2005) stellten einen Zusammenhang zwischen der Nutzung von herausfordernden Aufgaben und höheren Leistungen von Lernenden fest. Den Unterricht kognitiv anregend zu gestalten, hat darüber hinaus auch positive Auswirkungen auf motivationale Merkmale von Lernenden (Ekmekci & Serrano, 2022). Empirisch konnte die Bedeutsamkeit der kognitiven Aktivierung zum Beispiel für Schullust und Interesse gezeigt werden (Fauth, Decristan, Rieser, Klieme & Büttner, 2014a; Lazarides & Raufelder, 2021; Schiefele, 2009). Ohle-Peters und Kolleg\*innen (2021) wiesen eine prädiktive Kraft des Unterrichtsqualitätsmerkmals Differenzierung, welches zu der Unterrichtsqualitätsdimension kognitive Aktivierung zugeordnet wird, für die intrinsische Lesemotivation von Lernenden nach. Somit zeigten Lernende eine positive

Ausprägung in der Lesemotivation, wenn Lehrkräfte ihnen zum Beispiel je nach ihrem Können unterschiedlich schwere Aufgaben stellten.

### 1.3.2 Klassenführung

Als weitere Voraussetzung für ein optimales Lernen in der Klasse gilt eine effiziente Klassenführung. Denn nur wenn eine gewisse Ruhe und Ordnung im Klassenraum stattfindet, können andere Unterrichtsqualitätsdimensionen (kognitive Aktivierung und unterstützendes Unterrichtsklima) ihre positive Wirkung zeigen (Lipowsky et al., 2009). Unterrichtsplanung ist für eine optimale Vorbereitung von Lehrkräften auf den Unterricht wichtig. So lassen sich zum Beispiel Ziele des Unterrichts und der Einsatz von Materialien im Voraus planen. Allerdings entfaltet sich der Akt des Lehrens zur Erreichung der beabsichtigten Ziele spontan als Reihe von Ereignissen. Einige Ereignisse sind nicht planbar und treten durch die Interaktion mit den Lernenden unerwartet auf. Diese Ereignisse zu verstehen, wie sie das Lernen beeinflussen und wie sie sich im Laufe des Unterrichts entwickeln können, sind entscheidende Faktoren für erfolgreiches Klassenmanagement (Wolff, Jarodzka & Boshuizen, 2021). Detaillierter beschrieben bedeutet das für die Lehrkraft, den Unterricht so zu steuern, dass effektiv mit Störungen umgegangen und möglichst viel Zeit für die Lerninhalte verwendet wird (Brophy, 1979; Emmer & Stough, 2001; Kuger, 2016). Eine längere Lernzeit kann dadurch hervorgerufen werden, dass die Lehrkraft eine ruhige Arbeitsatmosphäre schafft (Kounin, 1970). Hierfür benötigt die Lehrkraft ein breites Spektrum an Strategien für eine gelungene Klassenführung, die auf einen effektiven und konsequenten Umgang mit Störungen abzielen (Brophy, 1979). Es sollten früh Regeln und Routinen in Klassen eingeführt werden, die sich zum Beispiel auf Übergänge zwischen Unterrichtsphasen beziehen. Klassenregeln sind allgemeine Erwartungen und Normen in Bezug auf Verhalten von Lernenden und Unterrichtsabläufe. Daran anschließend sind die Klarheit und Einhaltung der Klassenregeln für eine störungsarme Lernatmosphäre von besonderer Bedeutung (Emmer & Stough, 2001; Rakoczy, 2021). Lehrkräfte gelten dann als gut strukturiert, wenn sie ihre Erwartungen, Regeln und Unterrichtsinhalte verständlich und klar formulieren können (Klieme et al., 2009). Im Rahmen der Meta-Analyse von Kyriakides und Kollegen (2013) wurde unter anderem gezeigt, dass insbesondere Unterrichtsqualitätsmerkmale der Klassenführung prädiktiv für den Lernerfolg von Lernenden war. So hingen Merkmale wie Zeitmanagement und Strukturierung mit höheren Leistungen seitens der Lernenden zusammen. Campbell, Kyriakides, Muijs und Robinson (2004) wiesen einen Zusammenhang zwischen der beobachteten Klassenführung

und insbesondere der tatsächlichen aktiven Lernzeit mit den Leistungen und Einstellungen zum Lernen von Lernenden auf. In Bezug auf die Bedeutung der Klassenführung für motivationale Merkmale, zeigten sich in Studien positive Zusammenhänge zwischen Lehrkräften, die als gut strukturiert von ihren Lernenden in den Klassen wahrgenommen wurden und Unterrichtsinhalte klar kommunizieren konnten, und der Lernfreude von Lernenden (Lazarides, Dietrich & Taskinen, 2019; Maulana, Opendakker & Bosker, 2016).

### **1.3.3 Unterstützendes Unterrichtsklima**

Das unterstützende Unterrichtsklima wird auch als motivational-emotionale Flankierung der kognitiven Prozesse bezeichnet (Hasselhorn & Gold, 2017). Die Berücksichtigung von Lernendenbedürfnissen und -interessen ist ein zentrales Merkmal der dritten Unterrichtsqualitätsdimension. Hierzu zählen auch die Förderung des Zugehörigkeitsgefühls von Lernenden und die Unterstützung eines angenehmen Unterrichtsklimas (Hamre & Pianta, 2010). Außerdem können auch eine positive Feedbackkultur, gegenseitiger Respekt und eine positive Beziehung zur Lehrkraft unter der Dimension gefasst werden (Patrick, Kaplan & Ryan, 2011; Pianta & Hamre, 2009; Reeve & Jang, 2006). Insbesondere weisen Unterrichtsqualitätsmerkmale der konstruktiven Unterstützung eine Relevanz für motivationale Merkmale von Lernenden auf. So konnte in verschiedenen Studien gezeigt werden, dass Lernende, die sich stärker unterstützt fühlten und eine positive Beziehung zur Lehrkraft wahrgenommen haben, auch positivere Ausprägungen in motivationalen und emotionalen Variablen, zum Beispiel intrinsische Motivation, Schullust und Ängstlichkeit (hier negativer Zusammenhang) aufwiesen (Lazarides & Buchholz, 2019; Roorda, Koomen, Spilt & Oort, 2011; Scherer, Nilsen & Jansen, 2017). Rosenfeld und Kollegen (2000) stellten fest, dass Lernende mit einer höher wahrgenommenen Unterstützung, höhere Schulzufriedenheit, Engagement, Selbstwirksamkeit und bessere Noten zeigten. Allerdings waren diese Zusammenhänge nur zu beobachten, wenn Lernende sich auch von ihren Eltern und Freunden unterstützt fühlten (Rosenfeld, Richman & Bowen, 2000).

Ein weiteres spezifisches Unterrichtsqualitätsmerkmal der konstruktiven Unterstützung ist die Motivierungsqualität. Diese zielt darauf ab, die Lernmotivation von Lernenden anzuregen und aufrecht zu halten (Kunter & Trautwein, 2013). Um zu erklären, wie die intrinsische Motivation von Lernenden gefördert werden kann, bietet die Selbstbestimmungstheorie einen wichtigen theoretischen Rahmen (Furtak & Kunter, 2012). In Bezug auf die Selbstbestimmungstheorie nach Deci und Ryan (2000) kann darin

unterschieden werden, ob Personen Handlungen intrinsisch oder extrinsisch motiviert initiieren. Von intrinsisch motivierten Handlungen bei Lernenden kann dann gesprochen werden, wenn Lernprozesse oder -aktivitäten selbstbestimmt sind, zum Beispiel basierend auf dem Interesse an den Lerninhalten (Deci & Ryan, 2000). Gegensätzlich dazu werden extrinsisch motivierte Handlungen aufgrund von anderen Gründen als der inhärenten Zufriedenheit durchgeführt (Ryan & Deci, 2020). Die intrinsische Motivation ist besonders bedeutsam für das Lernen in der Schule, da diese mit höheren Leistungen und größerem Engagement, welches wiederum zu höheren Leistungen führt, zusammenhängt (Froiland & Worrell, 2016; Taylor et al., 2014). Die intrinsische Motivation von Lernenden kann dann gefördert werden, wenn die Lehrkraft die Grundbedürfnisse nach Autonomieerleben, Kompetenzerleben und sozialer Eingebundenheit erfüllt (Gnambs & Hanfstingl, 2016; Roeser, Eccles & Sameroff, 2000; Ryan & Deci, 2000, 2020). Autonomieerleben spiegelt die Tendenz von Individuen wider, ihr eigenes Verhalten selbst zu organisieren und in Harmonie mit ihrem integrierten Selbst zu handeln. Somit sollte Lernenden zum Beispiel die Wahl zwischen interessanten Übungen und Lernthemen gelassen und ihnen Entscheidungsoptionen für Unterrichtsaufgaben gegeben werden. Ein weiteres Grundbedürfnis beschreibt das Kompetenzerleben. Dieses bezieht sich auf die Fähigkeit, die Umwelt beeinflussen und Ziele effektiv erreichen zu können. Zum Kompetenzerleben können positives Feedback und ein optimales Maß an Aufgabenanforderungen beitragen. Unter der sozialen Eingebundenheit kann der Wunsch, sich mit anderen verbunden zu fühlen und das Wissen, dass sich andere um einen selbst kümmern, gefasst werden. Bezogen auf das Lernen, ist für die Förderung der sozialen Eingebundenheit bedeutsam, dass persönliche Beziehungen innerhalb der Klassengemeinschaft unterstützt werden, damit sich die Lernenden integriert fühlen (Deci & Ryan, 2000, 2008; Gucciardi, Weixian, Gibson, Ntoumanis & Ng, 2020). Empirische Studien stellten fest, dass die Befriedigung der drei Grundbedürfnisse von Lernenden bereits in der Grundschule für den Schulerfolg relevant ist (Froiland & Oros, 2014; Oga-Baldwin & Fryer, 2017, Oga-Baldwin & Fryer, 2020). Auch konnten Zusammenhänge zwischen dem Erleben von Autonomie, Kompetenzerleben und sozialer Eingebundenheit und motivationalen Variablen (wie z. B. intrinsische Motivation) gefunden werden (Niemic & Ryan, 2009).

Zusammengefasst kann eine hohe Motivierung dann erreicht werden, wenn das Erleben von Selbstbestimmung gefördert wird. Hierfür ist es bedeutsam die psychologischen Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit zu erfüllen (Ryan & Deci, 2020).

Die Merkmale der Unterrichtsqualitätsdimensionen kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima lassen sich der Tiefenstruktur von Unterricht zuordnen, welche als besonders relevant für den Lernerfolg gilt (Pauli & Reusser, 2003). Allerdings sind Merkmale der Tiefenstruktur nur indirekt beobachtbar und können nur über die Wahrnehmung von Teilnehmenden des Unterrichts oder von externen Urteilenden erfasst werden (Decristan et al., 2020).

## **2 Wahrnehmung**

Um die Wahrnehmung von Unterrichtsqualität näher beleuchten zu können, muss im ersten Schritt geklärt werden, wie nach psychologischen Ansätzen Individuen die zu beurteilenden Personen oder Situationen wahrnehmen und welche Faktoren einen Einfluss auf die Wahrnehmung haben können.

### **2.1 Begriffsbestimmung der Wahrnehmung**

Die Wahrnehmung ist zentral für die menschliche Informationsverarbeitung und kann als Prozess beschrieben werden, in dem distale Reize aus proximalen Reizen erschlossen werden. Diese Reize sind immer mit Bedeutungen und Funktionen verbunden, welche einfacher erinnert werden können als die eigentliche Struktur des Reizes. Ohne die Wahrnehmung wäre eine Interaktion mit der Umwelt nicht möglich (Kiesel & Koch, 2018). Abzugrenzen ist die Wahrnehmung von der Beobachtung. Wahrnehmung geschieht unbeabsichtigt und beiläufig. Im Vergleich zur Wahrnehmung erfolgen Beobachtungen zielgerichtet und theoriegeleitet (Kloss, 2013). Die Wahrnehmung ist eng verbunden mit der Aufmerksamkeit. Denn diese wählt aus, welche Informationen überhaupt erst wahrgenommen werden (Kiesel & Koch, 2018). Ebenfalls gibt es einen Zusammenhang zwischen der Wahrnehmung und der kognitiven Fähigkeit. Jedoch lassen sich verschiedene Argumentationen für die Wirkrichtung vom Zusammenhang der kognitiven Fähigkeiten und der Wahrnehmung finden. Demnach kann die Wahrnehmung einerseits als Komponente des Intelligenzkonstrukts gesehen werden (Schweizer, 2006) und andererseits können die kognitiven Fähigkeiten einen Einfluss auf die Wahrnehmung haben, indem sie als Voraussetzung für bessere Leistungen in der selektiven Wahrnehmung gesehen werden (Melnick, Harrison, Park, Bennetto & Tadin, 2013). Bezüglich der Wahrnehmung lassen sich weitere Unterscheidungen treffen. So bezieht sich

zum Beispiel die soziale Wahrnehmung auf soziale Gegebenheiten wie Mitmenschen, Gruppen sowie Situationen und wird durch sozial erlernte Motive, Erwartungen, Einstellungen und Emotionen vorgeformt (Martin & Wawrinowski, 2014). Nach Martin und Wawrinowski (2014) kann die soziale Wahrnehmung von Fehlerquellen wie zum Beispiel von Emotionen, Bedürfnissen, Abwehrmechanismen, Projektionen, Rationalisierungen, Verleugnungen, Erinnerungstäuschungen und dem Halo-Effekt betroffen sein. Wie Personen wahrgenommen werden, hängt auch von Einstellungen und Erwartungen ab. Verhält sich eine Person eher erwartungskonform, werden eher positive Einschätzungen über die Person und ihr Verhalten getroffen, hingegen bei nicht erwartungskonformen Verhalten können Einschätzungen eher negativ ausfallen (Micholka-Metsch & Metsch, 2015). Eng mit diesem Mechanismus verbunden ist die soziale Kategorisierung und das damit verbundene Bilden von Stereotypen. Die soziale Kategorisierung strukturiert und vereinfacht unsere Wahrnehmung. Dabei werden Personen mittels wahrgenommener Ähnlichkeiten oder Unterschiede sozialen Gruppen zugeordnet (Degner, 2022). Ein Resultat aus der sozialen Kategorisierung ist auch das Bilden von Stereotypen. Stereotype kommen primär gegenüber Personengruppen, welche gesellschaftlicher Ungleichheit ausgesetzt sind, vor. Sie werden auch als typische Eigenschaften einer sozialen Gruppe verstanden. Dadurch dass soziale Kategorisierung und damit assoziierte Stereotype unsere Wahrnehmung beeinflussen, haben diese auch automatisch Auswirkungen auf unser Urteil und auf unsere Handlungen (vgl. Degner, 2022; Eckert, 2012). Der Zusammenhang zwischen der Wahrnehmung von Urteilen bzw. Handlungen wird in dem nächsten Kapitel 2.2 näher beleuchtet. Einstellungen und Erwartungen gegenüber Personen sind verwoben mit unserer kulturellen Erziehung und Erfahrungen. Somit lenkt im nächsten Schritt auch die Kultur unsere Wahrnehmung (Chen & Starosta, 2005; Micholka-Metsch & Metsch, 2015). Wahrnehmung von Personen, Verhalten von Personen und Situationen ist zusätzlich auch von anderen Personenmerkmalen wie Emotionen sowie Motivationen (Kuzmanovic, Jefferson, Bente & Vogeley, 2013), Geschlecht (Hofer et al., 2006) und Intelligenz (Roebbers & Schneider, 2001) abhängig. Beispielsweise konnte festgestellt werden, dass Kinder in Abhängigkeit ihrer Intelligenz, soziale Situationen divergierend wahrnehmen und diese auch verschieden erinnern (Roebbers & Schneider, 2001). Um den Wahrnehmungsprozess und mögliche Einflussfaktoren genauer zu verstehen, werden im Folgenden Informationsverarbeitungsmodelle beschrieben. Hierbei wird insbesondere die Rolle der Wahrnehmung in den Blick genommen.

## 2.2 Informationsverarbeitungsmodelle

Theorien zu Informationsverarbeitung fokussieren darauf, wie Menschen Umweltereignisse wahrnehmen, Informationen kodieren und diese mit Wissen im Gedächtnis verbinden, neues Wissen speichern und es dann bei Bedarf abrufen (Schunk, 2011). Es gibt mehrere Modelle, die den Prozess der Informationsverarbeitungen versuchen abzubilden (z. B. Heeg [2011]: Modell des Weges des Menschen von der Wahrnehmung zur Handlung; Moos [1979]: Ein Modell zur Beziehung zwischen Umwelt- und Personenvariablen sowie Kriteriumsvariablen des Schülerverhaltens; Wickens & Carswell [2021]: Model of human information processing [Adapted from Wickens, 1992]). Was alle Modelle vereint, ist der Versuch, den Prozess von der Wahrnehmung bis mindestens zum Urteil darzustellen. Anfangs steht häufig ein Reiz (Merkmal, Situation, Aufgabe, Problem), welcher zunächst wahrgenommen wird, bevor er bewertet und interpretiert wird. Am Ende mündet dieser Prozess in einer Reaktion in Form eines Urteils oder einer Handlung (Blömeke, Gustafsson & Shavelson, 2015; Heeg, 2011; Moos, 1979). Zwei Modelle, welche in der Pädagogischen Psychologie herangezogen werden, um Wahrnehmungs- und Urteilsprozesse zu beschreiben, werden detaillierter im Anschluss beschrieben.

Ein Modell, welches alle Arten von Wahrnehmungsvorgängen abdeckt und sowohl zur Wahrnehmung von Gegenständen als auch Personen herangezogen werden kann, ist das Linsenmodell nach Brunswik (1943). Die Grundidee des Modells basiert darauf, dass für das Verstehen des Wahrnehmungsprozesses des Menschen auch seine evolutionäre Entwicklung berücksichtigt werden muss. Demnach erhält ein Individuum ein besseres Abbild seiner Umwelt, wenn zusätzlich zu dem vorliegenden Reiz, Informationen zu weiteren Reizen herangezogen werden. Im Rahmen des Linsenmodells kann zwischen drei verschiedenen Komponenten unterschieden werden: Distales Merkmal, Proximales Merkmal und Manifestes Urteil (siehe Abbildung 1).

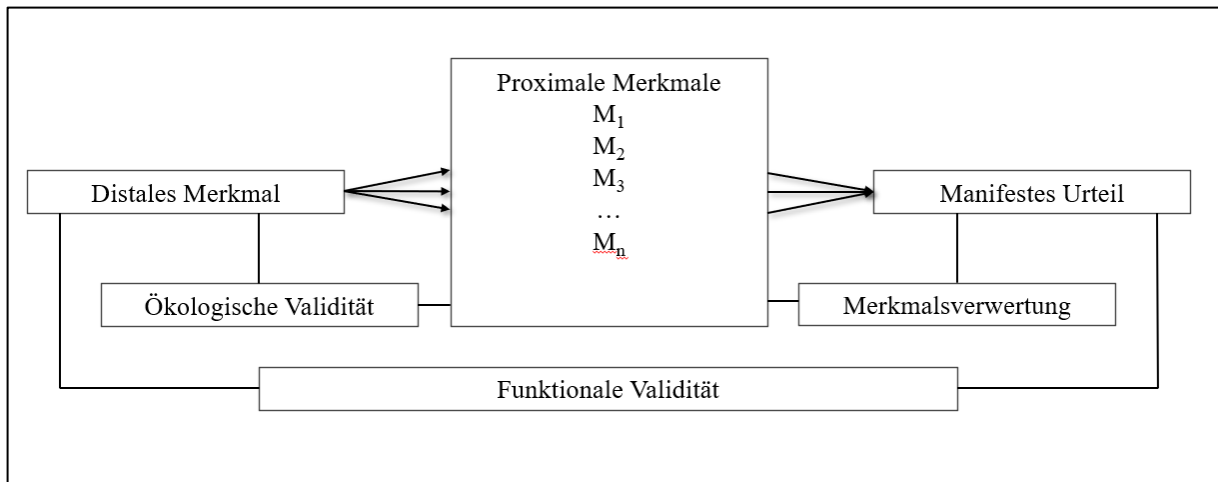


Abbildung 1. Linsenmodell nach Brunswik (1943)

Die distalen Merkmale sind nicht direkt zu beobachten oder zu beurteilen. Dementsprechend werden für die Wahrnehmung weitere proximale Merkmale, welche direkt zu beobachten sind, berücksichtigt. Schließlich kann aufgrund der Beobachtung der proximalen Merkmale ein manifestes Urteil getroffen werden (Henss, 1998). Dieser Vorgang ist auch die zentrale Aussage des Linsenmodells: die Beurteilung nicht beobachtbarer Merkmale anhand von beobachtbaren Merkmalen (Kürten, Greefrath & Hammann, 2020; Wiesner, 2020).

Ein Beispiel für das Linsenmodell wäre, dass die Intelligenz von Lernenden eingeschätzt werden soll (Distales Merkmal) und als Indikatoren für ordentliche Arbeitsweise, geordnetes Elternhaus, Erledigung der Hausaufgaben oder auch wacher Gesichtsausdruck (Proximale Merkmale) als Informationsquellen herangezogen werden. Das manifeste Urteil wäre das Lehrkrafturteil über die Intelligenz (Helmke, 2022). Bezogen auf die Unterrichtsqualität wäre ein Beispiel, dass die Nähe der Lernenden zur Lehrkraft (Distales Merkmal) von externen Urteilenden (Manifestes Urteil) eingeschätzt werden soll. Diese können Vermeidung des Blickkontakts, Häufigkeit der Kommunikation, körperliche Nähe oder auch Stimmung während der Kommunikation (Proximale Merkmale) als Indikatoren für ihr Urteil berücksichtigen.

Darüber hinaus sind noch drei korrelative Beziehungen zwischen den zwei Merkmalsarten und dem manifesten Urteil zu beachten. Die ökologische Validität steht für die Beziehung zwischen dem nicht beobachtbaren distalen Merkmal und dem beobachtbaren proximalen Merkmal. Die Merkmalsverwertung steht für die Beziehung zwischen dem proximalen Merkmal und dem manifesten Urteil. Die funktionale Validität beschreibt die Verbindung zwischen dem distalen Merkmal und dem manifesten Urteil (Klopp, 2012).

Das Linsenmodell ist sehr allgemein gehalten und ist auf alle Wahrnehmungsprozesse (z. B. Personen oder Situationen) übertragbar. So wird das Modell, neben der Wahrnehmung von Unterricht, auch häufig für die Beschreibung der diagnostischen Kompetenz und Urteilsgenauigkeit von Lehrkräften (z. B. Südkamp & Praetorius, 2017; Wiesner, 2020) oder zur Persönlichkeitswahrnehmung (z. B. Asendorpf, 2018) herangezogen.

Ein spezifischeres Modell ist das Modell von Greifeneder und Kollegen (2018) zu den Sequenzen der Informationsverarbeitung. Hierbei beziehen sie sich auf den Wahrnehmungsprozess bzw. Informationsverarbeitung bezüglich sozialer Situationen (siehe Abbildung 2). Die Interpretation einer sozialen Situation, das Verstehen des eigenen sozialen Umfeldes sowie die Reaktion auf diese können nur mittels miteinander verbundenen Teilaufgaben erfolgen. Diese Teilaufgaben können als Schritte bzw. Sequenzen angesehen werden, die für die Verknüpfung zwischen dem beobachtbaren Reiz und dem offensichtlichen Verhalten der Person verantwortlich sind. Im Rahmen der ersten Sequenz müssen die Reizereignisse wahrgenommen werden. Im nächsten Schritt wird die Wahrnehmung kodiert. Hierfür wird dem Reiz eine Bedeutung zugeschrieben. Grundlage für die Bedeutungszuschreibung ist das im Gedächtnis gespeicherte Vorwissen, welches für die Kodierung abgerufen wird. Schließlich wird die kodierte Wahrnehmung des Reizes im Gedächtnis gespeichert und beeinflusst möglicherweise die Wahrnehmung und Beurteilung nachfolgender Ereignisse. Für die weitere Verarbeitung, hierunter zählen die Urteilsbildung, Schlussfolgerungen und Entscheidungen, bilden nun der neu kodierte Reiz und das im Gedächtnis gespeicherte alte Wissen die Grundlage (Greifeneder et al., 2018).

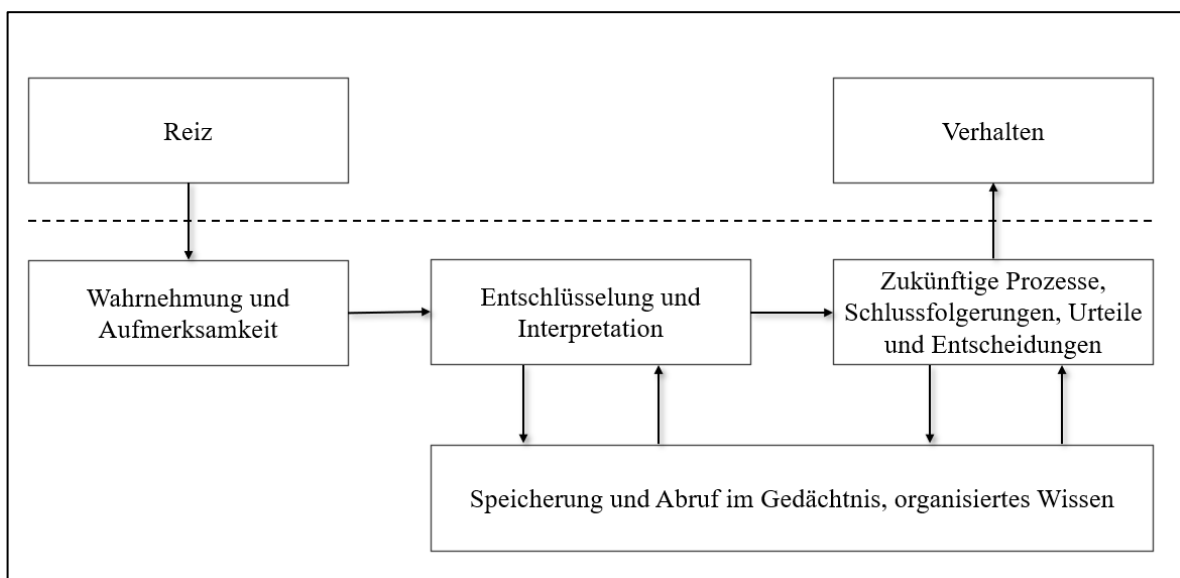


Abbildung 2. Sequenzen der Informationsverarbeitung nach Greifeneder et al. (2018)

Schon die Sequenz der Wahrnehmung und Aufmerksamkeit weist eine hohe Komplexität auf. In sozialen Situationen werden auf der einen Seite externale Reize wahrgenommen. Bezogen auf den Unterricht wären hier Beispiele: verbales und nonverbales Verhalten von Mitlernenden und der Lehrkraft, Gerüche, Nebengeräusche oder auch der Geschmack des Pausenbrots. Gleichzeitig können internale Reize registriert werden, wie beispielsweise Emotionen (glücklich, traurig, etc.), Schmerzen, Zittern oder auch Muskelverspannungen. Die Kapazität der Informationsverarbeitung ist allerdings begrenzt und in sozialen Situationen können nicht alle Reize verarbeitet werden, welche das sensorische System erreichen (Greifeneder et al., 2018). Mittels der Aufmerksamkeit verfügen Individuen über die Fähigkeit sich auf einige Reize in einer Situation zu konzentrieren und andere Reize von der weiteren Informationsverarbeitung auszuschließen. Diese Fähigkeit ist relevant für den Umgang der Einschränkung der Informationsverarbeitung (Broadbent, 1958). Empirisch konnten die Zusammenhänge zwischen der Wahrnehmung, Aufmerksamkeit und dem Gedächtnis bestätigt werden (Porto, Wouters & van Wieringen, 2023; Ophir, Hesselmann & Lamy, 2020). Beispielsweise zeigten Studien, dass in Bezug auf das Sprachverständnis trotz Lärm, Individuen mit höherer Arbeitsgedächtniskapazität Vorteile hatten die Sprache besser wahrzunehmen und verstehen zu können als Individuen mit weniger Arbeitsgedächtniskapazität (Holmer, Heimann & Rudner, 2016; McCreery, Walker, Spratford, Lewis & Brennan, 2019).

Darüber hinaus werden Reize aber auch aufgrund von anderen individuellen Merkmalen selektiert. Generell erregen Reize Aufmerksamkeit, welche sich von anderen abheben und hervorstechen. In einer bestimmten Situation kann ein Reiz vergleichend zu anderen Reizen sehr auffällig und unverwechselbar wirken. Weitere Gründe sind, dass ein Reiz aufgrund des Vorwissens und der Erwartungen von Individuen hervorsteht oder im Zusammenhang mit den aktuellen Zielen oder Bedürfnissen der Individuen, die Verarbeitung lenkt (Greifeneder et al., 2018; Fiske & Taylor, 2017). Somit wird die erste Sequenz der Informationsverarbeitung *Wahrnehmung und Aufmerksamkeit* durch individuelle Merkmale bedingt. In Bezug auf das Lernen in der Schule bedeutet dies, dass die Wahrnehmungen von dessen, was Lernende sehen und hören, in das Arbeitsgedächtnis eingeht. Was Lernende wahrnehmen, sollte demnach recht genau sein. Sind die Wahrnehmungen der Lernenden ungenau, so sind es letztlich auch die im Langzeitgedächtnis gespeicherten Langzeitinformationen (Eggen & Kauchak, 2014). Um sicherzustellen, dass Lernende die im Unterricht präsentierten Informationen richtig wahrnehmen, können Lehrkräfte an

Vorkenntnissen anknüpfen und Lernende aktiv in den Lernprozess mit einbeziehen (Fourie & Schlebusch, 2022). Reize werden aber nicht nur aufgrund von Vorwissen und Erwartungen unterschiedlich wahrgenommen, gespeichert und bewertet, sondern auch Emotionen haben einen Einfluss auf den Informationsverarbeitungsprozess, zum Beispiel aufgrund vom Zustand des Vertrauens (Kleiman, Sher, Elster & Mayo, 2015). So zeigten Posten und Gino (2021), dass im Zustand des Vertrauens erhöht Ähnlichkeiten von Elementen wahrgenommen und auf vorhandene Kategorien und Schemata zurückgegriffen werden und dadurch die Wahrnehmung und Erinnerung des neuen Reizes beeinträchtigt wird. In ihren Studien wurde der Zustand des Vertrauens dadurch verursacht, dass die Proband\*innen sich an eine Situation erinnern sollten, in der sie Personen vertraut bzw. misstraut haben. Im Anschluss wurden die Proband\*innen mittels Fragen zu ihrer Wahrnehmung, hinsichtlich eines vorab gesehenen Videoclips zu der gesehenen Person und gesehenen Situation, befragt. Im Zustand des Vertrauens wurde die Gedächtnisleistung verringert und mehr Fehlinformationen wurden wiedergegeben als im Zustand des Misstrauens. Die Antworten der Proband\*innen wiesen darauf hin, dass beim Beantworten eher auf im Gedächtnis gespeichertes Vorwissen zurückgegriffen wurde.

Zusammenfassend lässt sich festhalten, dass der Wahrnehmungsprozess sehr komplex ist. Bedingt wird die Komplexität insbesondere durch individuelle Merkmale, wie kognitive Fähigkeiten, Vorerfahrungen und Emotionen, welche die Wahrnehmung, Kategorisierung sowie Beurteilung neuer Reize (z. B. Personen und Situationen) beeinflussen. Neue Reize können nie direkt wahrgenommen werden und es werden immer mentale Repräsentationen der Umwelt aufgebaut (Jungermann, Pfister & Fischer, 2005). Nur über eine Bedeutungszuschreibung kann der neu wahrgenommene Reiz encodiert und interpretiert werden. Hierzu muss der neue Reiz mit den im Gedächtnis gespeicherten Informationen in Verbindung gebracht werden, welche in komplexen Strukturen (Kategorien, Schemata und Skripten) organisiert sind. Liegt ein größerer Speicher an Kategorien, Schemata und Skripten für die Verarbeitung des eingegangenen Reizes vor, so können akkuratere Beurteilungen über das Beobachtete getroffen werden und Beurteilungsfehler verringert werden (Greifeneder et al., 2018; Lau & Plessner, 2016).

Im Folgenden wird weiterhin die Begrifflichkeit *Wahrnehmung* verwendet. Dabei ist was am Ende in empirischen Studien gemessen werden kann ein Endprodukt, welches bei Greifeneder und Kollegen (2018) als Verhalten bezeichnet wird. Hinter der Begrifflichkeit der Wahrnehmung ist somit der ganze Informationsverarbeitungsprozess gemeint, welcher neben

der Wahrnehmung, auch die Entschlüsselung und Interpretation sowie die Schlussfolgerung und Urteilsbildung beinhaltet. So wird auch in internationalen Studien häufig der Begriff *perception* von zum Beispiel Lernenden (*students' perceptions*) verwendet, auch wenn die Wahrnehmung an sich nicht unbedingt das Konstrukt ist, welches in empirischen Studien gemessen werden kann. In Abgrenzung zu den Wörtern *rating* oder *judgment* lässt sich feststellen, dass in Suchmaschinen für englischsprachige Literatur doppelt so viele Treffer für den Begriff *perception* hinsichtlich der Lernenden von Unterricht gefunden wurde (Eigenrecherche, Stand 2023). Auch nationale Arbeitsgruppen, welche sich auf Aussagen über den Unterricht beziehen, verwenden häufig die Begrifflichkeit der *Wahrnehmung* bzw. *perception* (z. B. Fischer, Praetorius & Klieme, 2019; Wisniewski, Röhl & Fauth, 2022; Willems, 2022).

### **3 Wahrnehmung von Unterrichtsqualität**

In Bezug auf die Wahrnehmung von Unterricht, ist die Wahrnehmung ein kognitiver Prozess, der zu erwartungsbezogenen Urteilen führen kann (Dreesmann, 1979). Einfluss haben nach Dreesmann (1979) auf die Wahrnehmung von Unterricht Personenmerkmale und Merkmale der Interaktionssituationen. Zum letzteren zählen zum Beispiel die Rollenverteilung, das Machtverhältnis und die Zielorientierung. Wie schon in der Einleitung erwähnt, spielt auch in dem überarbeiteten Angebots-Nutzungs-Modell von Helmke (2022) die Wahrnehmung des unterrichtlichen Angebots eine zentrale Rolle für den Lernprozess. Kurz zusammengefasst: Auf der einen Seite steht das Angebot der Lehrkraft, welches über differenzierte Qualitätsmerkmale beschrieben werden kann und auf der anderen Seite sind die Lernenden mit ihren individuellen Lernvoraussetzungen zu berücksichtigen. Um der Komplexität von Lehr-Lern-Prozessen gerecht zu werden, müssen vermittelnde und interpretative Prozesse auf Lernendenseite berücksichtigt werden. Die Wirksamkeit des Lernens wird von zwei Mediationsprozessen auf Lernendenseite begleitet. Zunächst werden die Prozesse von den wahrgenommenen Erwartungen der Lehrkraft und dem Unterrichtsangebot sowie den darauffolgenden Interpretationen mediiert. Der zweite parallel verlaufende Vorgang beinhaltet die Auswirkungen der unterrichtlichen Maßnahmen auf die motivationalen, emotionalen und volitionalen Prozesse auf Seite der Lernenden (Helmke, 2022).

Die Komplexität von Unterrichtsqualität kommt unter anderem durch die Interaktion im Klassenraum zustande (Helmke, 2022). Im Unterricht spielen sich viele dynamische Prozesse ab. In kurzen Abständen wechseln Zuhörer\*innen und Sprecher\*innen sowie

Anweisungen, Erklärungen und Übungen (von Aufschnaiter & Welzel, 2001). Trotz der dynamischen Prozesse wird erwartet, dass Schüler\*innen Lernerfolge erzielen.

### 3.1 Methoden zur Erfassung von Unterrichtsqualität

Um Wirkmechanismen des unterrichtlichen Angebots zu untersuchen und die Unterrichtsqualität zu analysieren, werden entsprechende Methoden benötigt. Diese Methoden können klassifiziert werden in zum Beispiel *frei* (Berichte) vs. *gebunden* (Multiple-Choice-Fragebögen) oder auch *schriftlich* (Fragebögen) vs. *mündlich* (Interviews). Eine weitere Klassifizierung von Messungen unterscheidet zwischen einem geringen und hohen Ermessensspielraum; anders auch *niedrig-inferent* vs. *hoch-inferenten* Messungen (Helmke, 2006; Marquardt, 2011). Darüber hinaus lassen sich zwei Methoden unterscheiden, ob Unterrichtsqualitätsmerkmale direkt im Klassenraum wahrgenommen werden oder anhand von videografierten Unterrichtssequenzen (Herrle, Rauin & Engartner, 2016; Mashburn, Meyer, Allen & Pinata, 2014). Nähere Informationen folgen in Kapitel 3.2.3. Außerdem spielen noch der zugrunde gelegte Inferenzgrad, Beobachtungszeitraum und Itembezug eine wichtige Rolle bei der Erfassung von Unterrichtsqualität. Im Folgenden wird detailliert auf die einzelnen Aspekte von Messungen eingegangen.

#### 3.1.1 Inferenzgrad

Einschätzungen zur Unterrichtsqualität können entweder niedrig-inferent vs. hoch-inferent erfolgen (Praetorius, 2014). In diesem Zusammenhang meint die Inferenz den Grad der Schlussfolgerung, der bei den Urteilen notwendig ist. Je nach Fragestellung oder auch interessierendes Merkmal wird die Entscheidung getroffen, welche Art von Einschätzungen verwendet wird. Im Hinblick auf niedrig-inferente Einschätzungen wird der Grad der Interpretation möglichst geringgehalten und das Ziel ist eher die Beschreibung der Unterrichtsgestaltung. Bei niedrig-inferenten Einschätzungen wird sich fast ausschließlich an direkt beobachtbarem Verhalten orientiert und es werden eher Unterrichtsmerkmale der Sichtstruktur (im Sinne von Pauli & Reusser, 2006) des Unterrichts berücksichtigt. Für die Datengewinnung werden beispielsweise die Häufigkeit und Dauer von leicht beobachtbaren Unterrichtsereignissen erhoben (Lotz, Gabriel & Lipowsky, 2013). Konkret geht es bei niedrig-inferenten Einschätzungen um die Beschreibung unterrichtlicher Abläufe, einzelner Interaktionen und um die genaue Charakterisierung von spezifischen Prozessen, welche für

Lernende, Lehrkräfte und externe Urteilende direkt wahrnehmbar sind. Sollen Einschätzungen zur Tiefenstruktur von Unterricht erfolgen, sind hoch-inferente Einschätzungen notwendig. Bei diesen Einschätzungen kann nur bedingt auf direkt beobachtbares Verhalten zurückgegriffen werden und es sind Schlussfolgerungen und Interpretationen der urteilenden Person erforderlich (Lotz et al., 2013; Praetorius, 2014). Die Inhalte der hoch-inferenten Einschätzungen betreffen häufig die Bewertung der Qualität von Unterricht oder auch von einzelnen Unterrichtsabschnitten sowie einzelne Unterrichtsereignisse (Hugener, Rakoczy, Pauli & Reusser, 2006; Lotz et al., 2013). Um Daten zur Unterrichtsqualität zu erhalten, werden Schätzverfahren zur Ausprägung der Qualität des Unterrichtsmerkmals eingesetzt (Lotz et al., 2013). Aufgrund des Interpretationsgrades werden hoch-inferente Urteile im Vergleich zu niedrig-inferenten Einschätzungen eher mit Fehlervarianz erfasst und sind eher von Beurteilungsfehlern betroffen (Praetorius, 2014). Vorteile sind aber, dass komplexe, in Verbindung stehende Merkmale integriert und im gleichen Schritt beurteilt werden können. Mittels hoch-inferenten Einschätzungen wird auf die menschliche Urteilsfähigkeit zurückgegriffen, eine Vielfalt verschiedener Aspekte in einer Einschätzung zu analysieren und zu integrieren. Zugleich werden mit dieser Maßnahme Urteile standardisiert (Rakoczy, 2008). Schlussfolgernd muss für die Erfassung von Unterrichtsqualität eher auf hoch-inferente Einschätzungen zurückgegriffen werden, welche sich im Vergleich zu niedrig-inferenten Ratings der Sichtstruktur als prädiktiver für den Lernerfolg erwiesen (Klieme et al. 2009).

### **3.1.2 Beurteilungszeitraum**

Außerdem kann bei der Erfassung zwischen längerfristigen Aussagen über mehrere Unterrichtsstunden und Aussagen bezüglich einzelner Unterrichtsstunden unterschieden werden. Da sich die Qualität des Unterrichts in Routinen und stundenübergreifende Praktiken ausdrückt, welche nicht in einzelnen Unterrichtsstunden beobachtet werden können, sollten nach Helmke & Schrader (2008) qualitative Unterrichtsmerkmale durch längerfristige Urteile erfasst werden. Nicht nur, dass längerfristige Befragungen über mehrere Unterrichtsstunden die Fähigkeit voraussetzen, dass das Lehrkraftverhalten und die Unterrichtsqualitätsmerkmale abstrahiert werden können, ist auch zu beachten, dass Befragungen zum Unterricht oder zum Lehrkraftverhalten über einen längeren Beurteilungszeitraum häufiger von Verzerrungen betroffen sein können. Ein Argument für die Erfassung der Unterrichtsqualitätsmerkmale hinsichtlich einzelner Unterrichtsstunden ist die Instabilität der Merkmale. Aufgrund der Tatsache, dass die Unterrichtsqualitätsmerkmale, insbesondere das Merkmal kognitive

Aktivierung, nicht unbedingt über Unterrichtsstunden hinweg stabil sind, können keine übereinstimmenden oder generalisierbaren Urteile über das unterrichtliche Angebot oder das Lehrkraftverhalten getroffen werden (Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2014; Talić et al., 2022). Die Wahl des Beurteilungszeitraums ist abhängig von den Interessen und Schwerpunkten des Forschungsvorhabens und sollte sich auch an der gewählten Datenquelle orientieren (Praetorius, 2014). Insbesondere bei Lernenden jüngeren Alters sollte die Wahl des Beurteilungszeitraums genau bedacht werden.

Lernende speichern wahrgenommenen Unterricht in Form von Skripten (Schemata für Ereignisabläufe) (Goswami, 2001). In jedem Skript sind die wichtigen Personen und Requisiten dargestellt. Die Qualität, Länge und Komplexität dieser Schemata steigen mit dem Alter. Begünstigt wird die zunehmende Gedächtniskapazität durch die Entwicklung und Aneignung von Gedächtnisstrategien im Kindesalter (Kloss, 2013). Demzufolge steigert sich mit dem Alter auch die Qualität, Länge und Komplexität der gespeicherten Wahrnehmungen über den Unterricht und die Befragung von Grundschulkindern zum Unterricht sollte sich eher auf kürzere Ergebnisabläufe beziehen.

Marquardt (2011) kritisiert, wenn Kinder bei Befragungen ihr Urteil an einem größeren Zeitraum ausrichten sollen. Denn häufig bleibt der zugrunde gelegte Zeitraum für die Beurteilung einzelner Fragen unklar (Marquardt, 2011). Neben der Schwierigkeit der Abstraktion von längerfristigen Ereignissen und der Erinnerungsfähigkeit von jüngeren Kindern bezüglich konkreter Unterrichtsstunden, ist ein weiteres Argument für die Befragung von Lernenden zu einzelnen Unterrichtsstunden, die höhere Übereinstimmung der Lernendeneinschätzungen im Vergleich zu längerfristigen Befragungen über mehrere Unterrichtsstunden hinweg (Helmke et al., 2009). Bei stundenübergreifenden Urteilen ist auch zu beachten, dass kommunikative Prozesse der Lernenden untereinander zu einer Verfestigung der Wahrnehmung führen und anschließend die Urteile der Lernenden beeinflussen können. Demzufolge sollten Urteile von Lernenden, insbesondere von jüngeren Kindern, direkt im Anschluss an den zu beurteilenden Unterricht erfolgen und sich auf einzelne konkrete Unterrichtsstunden beziehen.

Ebenfalls sind auch Angaben zur wahrgenommenen Unterrichtsqualität methodisch zu unterscheiden, welche während des Unterrichts erfolgen (z. B. bei Unterrichtsbesuchen von [angehenden] Lehrkräften), retrospektiv erfolgen (z. B. Befragung der Lernenden oder Lehrkräfte zum gerade erfolgten Unterricht) und zeitlich losgelöste Angaben (z. B. von externen Urteilenden via Videographie). Bei retrospektiven Angaben ist zu berücksichtigen,

dass diese besonders Verzerrungen, wie Erinnerungseffekten, unterliegen können (Pieper, 2023). Aufgrund der Struktur und Kategorien, wie Informationen im Gedächtnis gespeichert werden, können bei der Einschätzung von Personen und Situationen Erinnerungseffekte auftreten. Informationen, welche nicht in unsere Kategorien, Schemata und Skripte passen, werden schlechter erinnert und eher in stereotypkonsistente und schon bekannte Informationen umgewandelt (Blair, Ma, & Lenton, 2001; Kleen, 2021). Ein größerer Speicher an Kategorien, Schemata und Skripten sorgt für genauere Wahrnehmungen und Einschätzungen (Greifeneder et al., 2018; Lau & Plessner, 2016). Demzufolge können Erinnerungseffekte in erster Linie bei jüngeren Lernenden auftreten, da diese über weniger Kategorien, Schemata und Skripte verfügen.

### 3.1.3 Itembezug

Bei der Itemformulierung für die Erfassung von Unterrichtsqualität ist zu beachten, dass sich Items auf verschiedene Inhalte beziehen können. Beispielsweise für den Bereich der Klassenführung gibt es Itemformulierungen, welche sich auf das Lehrkraftverhalten, Lernendenverhalten oder eine Kombination aus beidem beziehen (siehe Tabelle 3).

Tabelle 3

*Beispielitems mit unterschiedlichem Itembezug aus Fauth, Göllner, Lenske, Praetorius & Wagner (2018).*

Itembezug	Beispielitem	Quelle
Lehrkraftverhalten	Dem Lehrer gelingt es, Störungen schnell zu beenden.	(Projekt: KODEK, 2009-2012; Thiel, Ophardt & Piwowar, 2013)
Lernendenverhalten	Student behavior in this class is a problem.	(Tripod, Wallace, Kelcey & Ruzek, 2016)
Kombination Lehrkraft- und Lernendenverhalten	In Mathematik wird im Unterricht viel Zeit vertrödelt.	(Projekt: PISA, 2003; Baumert et al., 2009)

Aufgrund dessen haben Fauth und Kolleg\*innen (2020a) die Referenten-Perspektiven-Matrix entwickelt. Demnach sollte bei der Interpretation von Aussagen über Unterrichtsqualität

berücksichtigt werden, wer das Item beantworten soll (Lernende, Lehrkräfte oder externe Urteilende) und worauf sich das Item beziehen soll (auf das Lehrkraftverhalten, Lernendenverhalten oder eine Kombination aus beidem). Demnach ergibt sich eine Neun-Felder-Matrix: Drei Perspektiven (Lernende, Lehrkräfte oder externe Urteilende) und drei unterschiedliche Itembezüge (Lehrkraftverhalten, Lernendenverhalten oder eine Kombination aus beidem). Darüber hinaus ist die Antwort auf ein bestimmtes Item von der Qualität und Quantität der Verfügbarkeit der Informationen von den wahrnehmenden Personen abhängig sowie von dem Selbstbezug, den die Itemformulierung beinhaltet (Fauth, Göllner, Lenske, Praetorius & Wagner, 2020a). Bezieht sich das Item zum Beispiel auf das Lehrkraftverhalten, welches nicht direkt für Lernende und externe Urteilende beobachtbar ist, so müssen sich diese auf indirekte Verhaltensverweisen verlassen (z. B. „In Mathematik merkt unser Lehrer/unsere Lehrerin sofort, wenn Schüler/Schülerinnen beginnen, etwas anderes zu treiben“, Baumert et al., 2009, S. 211). Die Bewertung des Items würde aus Sicht der Lehrkraft wahrscheinlich anders ausfallen. Denn die Lehrkraft hat natürlich einen direkteren Zugriff auf ihre eigenen Gedanken und Wahrnehmungen (Fauth et al., 2020a). Dieser Aspekt ist insbesondere zu beachten, wenn die Übereinstimmungen verschiedener Perspektiven betrachtet werden.

Hinsichtlich Items der Klassenführung zeigten Studien, dass die Perspektiven zur Erfassung von Unterrichtsqualität am höchsten übereinstimmten, wenn sich die Items auf das Verhalten der Lernenden bezogen (Kunter & Baumert, 2006; Wagner et al., 2016; Wettstein, Ramseier, Scherzinger & Gasser, 2016). Aufgrund dessen empfehlen Fauth und Kolleg\*innen (2018) das Item so zu formulieren, dass es sich auf beobachtbares Verhalten der Lernenden bezieht. Für die Befragung von Lernenden ist neben des Itembezugs des beobachtenden Verhaltens auch die Formulierung *wir* oder *ich* in Items zu beachten. Für beide Varianten gibt es Vor- und Nachteile. Ein Vorteil von Angaben der Wahrnehmungen der ganzen Klasse (*wir*-Version) ist, dass diese die Validität der Urteile der Lernenden erhöhen können (Neuber, 2021). Nachteile der Angaben zu Wahrnehmungen der ganzen Klasse (*wir*-Version) sind, dass die Wahrnehmungen von den Mitlernenden aggregiert werden müssen. Dies setzt einen kognitiv aufwendigen Prozess bei den Lernenden voraus, welcher insbesondere für Lernende der Grundschule schwierig sein kann (Lenske, 2016). Außerdem müssen allen Lernenden auch die Informationen über den Urteilsgegenstand aller Mitlernenden der Klasse vorliegen. Auch wenn Lernende sich über ihre Wahrnehmungen bezüglich des Unterrichts austauschen, sind nicht unbedingt die Meinungen und Wahrnehmungen aller Mitlernenden der Klasse bekannt

(Piskol, 2008). Nach diesen Aspekten empfiehlt es sich eher Lernende nur nach eigenen Wahrnehmungen zu befragen, insbesondere im Grundschulalter.

Ein weiterer Vorteil für die Angaben ausschließlich zur eigenen Wahrnehmung (ich-Version) ist, dass Ereignishäufigkeiten ermittelt werden können. So kann beispielweise bei der Frage wie häufig ein Kind im Unterricht drangegenommen wurde, addiert und ermittelt werden wie viele Lernende mindestens einmal die Stunde von der Lehrkraft drangegenommen wurden. Allerdings muss beachtet werden, dass die Anonymität hierdurch nicht verletzt wird und durch die Formulierung keine Rückschlüsse auf einzelne Kinder möglich sind (Piskol, 2008).

Empirisch konnte gezeigt werden, dass sich Urteile durchaus unterscheiden, je nachdem, ob sie sich auf Aussagen über einzelne Lernende oder auf die Klasse beziehen. Insbesondere im Bereich kognitive Aktivierung waren signifikante Unterschiede zwischen Wahrnehmungen von Lernenden bezüglich eigener Wahrnehmungen der kognitiven Aktivierung (ich-Version) und der Wahrnehmung der Klasse in Bezug auf die kognitive Aktivierung (wir-Version) zu erkennen (Piskol, 2008).

Zusammengefasst gibt es zur Erfassung von Unterrichtsqualität verschiedene methodische Ansätze. Allerdings sind diese auch von der Perspektive (von Lernenden, Lehrkräften und externen Urteilenden) abhängig, welche zur Erfassung der Unterrichtsqualität herangezogen werden. Im Folgenden werden die Perspektiven zur Erfassung von Unterrichtsqualität und ihre möglichen Methoden näher beleuchtet.

### **3.2 Perspektiven zur Erfassung von Unterrichtsqualität**

Um an Informationen über die Qualität des unterrichtlichen Angebots zu gelangen, werden häufig die Perspektiven von Lernenden, Lehrkräften und externen Urteilenden eingeholt (siehe Abbildung 3). Generell ist nur eine geringe Übereinstimmung der Wahrnehmungen von Unterrichtsqualität von Lernenden, Lehrkräften und externen Urteilenden zu finden (vgl. Clausen, 2002; Fauth et al., 2014b; Kunter & Baumert, 2006). Dies kann, wie in Kapitel 3.1 beschrieben, an der Methode zur Erfassung der Unterrichtsqualität liegen, insbesondere an dem Itembezug. Der höchste Zusammenhang ist zwischen den drei Perspektiven im Bereich der Klassenführung festzustellen (Clausen, 2002; Fauth et al., 2014b; Kunter & Baumert, 2006). Gezeigt wurde die hohe Übereinstimmung im Bereich Klassenführung auch von Kunter und Kolleg\*innen (2013), indem sie sowohl Lehrkräfte und Lernende zur Unterrichtsqualität

befragten. Im Vergleich zu den anderen Unterrichtsqualitätsdimensionen fanden sie bezüglich der Klassenführung den größten Zusammenhang zwischen Lehrkraft- und Lernendenwahrnehmungen. Auch die Wahrnehmungen von Lernenden und externen Urteilenden wiesen den größten Zusammenhang im Bereich Klassenführung ( $r = .85$ ) auf. Hingegen fielen die Übereinstimmungen in den Bereichen kognitive Aktivierung ( $r = .39$ ) und unterstützendes Unterrichtsklima ( $r = .55$ ) deutlich geringer aus (Fauth et al., 2020b). Hinsichtlich des unterstützenden Unterrichtsklimas weisen die Wahrnehmungen von Lehrkräften und von Lernenden einen hohen Zusammenhang auf (Clausen, 2002; Fauth et al., 2014b; Kunter & Baumert, 2006). In einer Untersuchung von Begrich, Fauth und Kunter (2020) konnte eine Korrelation von  $r = .48$  zwischen den Wahrnehmungen von Lernenden und von externen Urteilenden bezüglich der kognitiven Aktivierung und hinsichtlich des unterstützenden Unterrichtsklimas eine Korrelation von  $r = .63$  gefunden werden. Gegensätzlich dazu konnte in vorigen Studien gezeigt werden, dass die Wahrnehmungen externer Urteilenden im Bereich des unterstützenden Unterrichtsklimas nicht mit den Lernenden- und Lehrkraftwahrnehmungen übereinstimmten (Fauth et al., 2014b; De Jong & Westerhof, 2001). Kearney und Peters (2013) berichteten generell über keinen signifikanten Zusammenhang von Lehrkraft- und Lernendenwahrnehmungen hinsichtlich des unterstützenden Unterrichtsklimas. Allerdings zeigte sich in dem der Dimension untergeordneten Unterrichtsqualitätsmerkmal Zusammenhalt im Klassenzimmer eine signifikante Korrelation der beiden Perspektiven. Insgesamt lässt sich festhalten, dass in den Bereichen der kognitiven Aktivierung und des unterstützenden Unterrichtsklimas die Befundlage zur Übereinstimmung der Perspektiven uneindeutig ist. Hingegen im Bereich Klassenführung einige Studien einen großen Zusammenhang der Wahrnehmungen der unterschiedlichen Perspektiven berichteten.

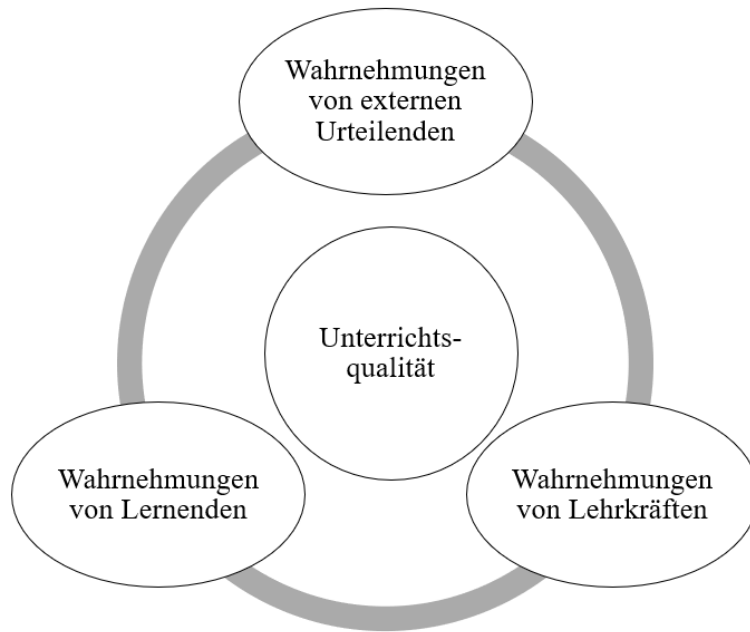


Abbildung 3. Perspektiven zur Erfassung von Unterrichtsqualität

### 3.2.1 Wahrnehmungen von Lernenden

Auch wenn sich die meisten Unterrichtsqualitätsmerkmale auf das kognitiv anregende, steuernde und unterstützende Verhalten von Lehrkräften beziehen, kann das Verhalten erst effektiv und fördernd wirken, wenn Lernende diese Merkmale wahrnehmen (Wubbels et al., 2015). Denn die Lernenden sind die Adressat\*innen von Unterricht und ihre Wahrnehmungen bezüglich des unterrichtlichen Angebots können als wichtiges Werkzeug für die Evaluation von Lehrkräften dienen. Dementsprechend haben sie auch eine praktische Bedeutung für die Unterrichtsreflexion (Helmke et al., 2009; Kultusministerkonferenz, 2004; Soebari & Aldridge, 2015). Geben Lernende Feedback an ihre Lehrkraft, kann dies wirksam für ein erfolgreiches und wertschätzendes Lehren sein. Mittels Wahrnehmungen von Lernenden erfahren Lehrkräfte, auf welchem Leistungsstand diese stehen, ob Verständnisprobleme vorliegen und an welcher Stelle Fehler gemacht werden. Die Diskrepanz zwischen Fremd- und Selbstwahrnehmung von Unterricht könnte so reduziert werden (Hattie, 2018; Zierer, Wisniewski, Schatz, Weckend & Helmke, 2019).

Die früheste Forschung zur Wahrnehmung von Lernenden bezüglich des Lernangebots geht auf die 1920er Jahre zurück und bezog sich noch eher auf die Urteile von Studierenden (z. B. Remmers & Brandenburg, 1927). Ziel war es schon damals, anhand der Wahrnehmung von Lernenden ein Maß zu erhalten, wie effektiv das Lernangebot oder eine Lehrkraft bei der

Förderung des Lernens ist (Carpenter et al., 2020). Im Bereich der Hochschulbildung sind Studierendenfeedbacks und Studierendenevaluationen schon weitverbreitet (z. B. Braga, Paccagnella, & Pellizzari, 2014; Brown & Atkins, 1990; Marsh, 2007). Für den Hochschulbereich konnten nicht nur Zusammenhänge zwischen den Wahrnehmungen von Lernenden über den Unterricht und des tatsächlichen Lernerfolgs gezeigt werden, sondern auch zwischen den Wahrnehmungen und der Urteilsgenauigkeit über ihren Wissenszuwachs. Demzufolge können Lernende ihren Wissenszuwachs akkurater einschätzen, wenn sie auch den Inhalt des Lernangebots als gut strukturiert, zum Beispiel in Form von Übergängen zwischen Inhalten, und positiver wahrnahmen (Titsworth & Kiewra, 2004). In der Sekundar- oder Primarstufe werden eher weniger Feedbacks von Lernenden seitens der Lehrkraft eingeholt (Fauth et al., 2014a). Dabei konnte die Relevanz der Unterrichtswahrnehmung von Lernenden für Leistungsvariablen sowie motivationale und emotionale Variablen in mehreren empirischen Studien auch für Kinder und Jugendliche gezeigt werden (z. B. Kunter & Baumert, 2006; Talić et al., 2022). Stang und McElvany (2020) bezogen sich in ihren Analysen auf die IGLU 2016-Teilstichprobe und betrachteten Wahrnehmungen von Lernenden der vierten Grundschulklasse hinsichtlich der Basisdimensionen von Unterrichtsqualität (Hußmann et al., 2017). Es zeigten sich gerichtete Zusammenhänge zwischen den Wahrnehmungen von Lernenden des unterstützenden Unterrichtsklimas sowie der kognitiven Aktivierung und der Lesemotivation. Demzufolge gaben Lernende an motivierter zu sein, wenn sie auch eine höhere Unterstützung seitens der Lehrkraft und einen kognitiv anregenden Unterricht wahrnahmen. Darüber hinaus konnte die statistische Relevanz von Wahrnehmungen von Lernenden der vierten Grundschulklasse bezüglich der Unterrichtsqualität für den Lernerfolg im Lesen gezeigt werden (z. B. McLean, Sparapani, Toste & Connor, 2016). Aufgrund der Bedeutung für den Lernerfolg und der Effizienz ist die Abfrage von Wahrnehmungen von Lernenden bezüglich der Unterrichtsqualität ein selbstverständlicher Bestandteil der großen Schulstudien geworden (Helmke, 2002). Beispiele dafür sind der IQB-Ländervergleich 2021 (Stanat et al., 2022), IGLU 2021 (McElvany et al., 2023), TIMSS 2019 (Stang et al., 2020) und PISA 2022 (Lewalter, Diedrich, Goldhammer, Köller & Reiss, 2023).

Befragungen von Lernenden zum Unterricht oder zur Lehrkraft erfolgen häufig mittels Fragebögen (Marquardt, 2011). Vermehrt beziehen sich die Befragungen zur Wahrnehmung von Lernenden zur Unterrichtsqualität sowohl national als auch international auf Lernende der

Sekundarstufe (z. B. Greene, Miller, Crowson, Duke & Akey, 2004; Kunter et al., 2008; Rakoczy et al., 2022; Schatz, 2021; Talić et al., 2022, Wild, 2015).

Aus entwicklungspsychologischer Perspektive entwickeln Kinder ab einem Alter von etwa neun Jahren eine Urteilsfähigkeit, die es ermöglicht, Situationen und Personen zutreffend einschätzen zu können (von UsLAR, 2007). Empirisch konnte bestätigt werden, dass auch schon Lernende im Grundschulalter die Fähigkeit besitzen valide Bewertungen von Unterricht abzugeben (vgl. Ditton, 2002; Gruehn, 2000). So können bereits Lernende im Grundschulalter zwischen Urteilen der Popularität der Lehrkraft und Urteilen der Unterrichtsqualität unterscheiden (Fauth, Decristan, Rieser, Klieme & Büttner, 2016).

Aus methodischer Sicht kann bei der Erfassung von Wahrnehmungen von Lernenden zwischen zwei Ebenen unterschieden werden. Um Aussagen über Einflüsse verschiedener Lernumgebungen auf Unterrichtseinschätzungen treffen zu können, wird die Klassenebene und damit die Varianzaufklärung zwischen den Klassen betrachtet (Lüdtke, Robitzsch, Trautwein & Kunter, 2009). Die Wahrnehmungen der Lernenden können allerdings innerhalb derselben Klassen divergieren (Wagner et al., 2016). Die Individualebene kann betrachtet werden, um Aufschlüsse über die Varianz eines Merkmals bzw. der Wahrnehmung innerhalb der Klasse zu erlangen (Lüdtke et al., 2009). Da Merkmale der Interaktionssituation mit dem Wahrnehmungs- und Urteilsprozess verbunden sind, hängen Urteile von Lernenden derselben Klasse, welche die gleiche Situation erleben, höher zusammen als Urteile von Lernenden unterschiedlicher Klassen (Dreesmann, 1979).

#### *Prädiktoren für divergierende Wahrnehmungen innerhalb der Gruppe*

Die Wirksamkeit von Unterricht hängt neben dem unterrichtlichen Angebot auch von der Wahrnehmung und Nutzung der einzelnen Lernenden ab, welche von den Vorerfahrungen und individuellen Merkmalen beeinflusst werden (Jansen, Decristan & Fauth, 2022; Vieluf et al., 2020). Demzufolge kann das gleiche unterrichtliche Angebot zu unterschiedlichen Wahrnehmungen und Urteilen auf Seiten der Lernenden führen. In Studien wurde vermehrt berichtet, dass Wahrnehmungen von Lernenden über den Unterricht nur gering innerhalb einer Klasse übereinstimmen (Lüdtke et al., 2009; Wagner et al., 2016). *Intraklassenkorrelationen* ( $ICC_1$ ), welche Informationen über die Zuverlässigkeit von Urteilen geben, lagen in Studien zwischen  $ICC_1 = .05$  (*unterstützendes Unterrichtsklima*) und  $ICC_1 = .32$  (*Klassenführung*). Oftmals wurden allerdings Wahrnehmungen und Urteile von Lernenden der Sekundarschule herangezogen sowie sich auf den naturwissenschaftlichen Unterricht bezogen (vgl. Lüdtke, Trautwein, Kunter & Baumert, 2006; Praetorius et al., 2017; Wagner et al., 2016). Untersuchungen,

welche sich auf Lernende im Grundschulalter bezogen, wiesen allerdings vergleichbare Werte auf und die größte Übereinstimmung von Wahrnehmungen von Lernenden konnte im Bereich Klassenführung festgestellt werden (z. B. Stang & McElvany, 2020).

Dass es in Bezug auf die Unterrichtsqualitätsdimensionen zu unterschiedlicher Übereinstimmung zwischen den Wahrnehmungen von Lernenden kommt, kann an der Beobachtbarkeit der Unterrichtsqualitätsmerkmale liegen. Wie in Kapitel 3.1 beschrieben, sind einige Unterrichtsqualitätsmerkmale wie zum Beispiel Merkmale der Klassenführung (Störung im Unterricht) relativ direkt erfassbar und können dementsprechend auch zu höheren Übereinstimmungen von Wahrnehmungen von Lernenden führen. Für Einschätzungen in den Dimensionen unterstützendes Unterrichtsklima und kognitive Aktivierung ist häufig ein größerer Interpretationsaufwand notwendig (Clausen, Reusser & Klieme, 2003). Somit müssen Lernende für diese Einschätzungen auf vorhandenes Wissen zurückgreifen, um Urteile auch über nicht direkt beobachtbare Unterrichtsqualitätsmerkmale fällen zu können. Daraus lässt sich auch schlussfolgern, dass bei der Einschätzung von Unterrichtsqualitätsmerkmalen Lernende auf ihre individuellen Ressourcen zurückgreifen müssen, um Urteile fällen zu können. So wird zum Beispiel der Unterricht von Lernenden positiver wahrgenommen und beurteilt, wenn diese ein höheres Vorwissen (häufig gemessen durch die letzte Fachnote) aufweisen (Babad, 1996; Levy, Den Brok, Wubbels & Brekelmans, 2003). Ein hoher Zusammenhang konnte zwischen der wahrgenommenen Unterstützung und der Nähe zur Lehrkraft nachgewiesen werden (Fauth et al., 2016; Madill, Gest & Rodkin, 2014). Insbesondere bei den Befragungen von Lernenden in Grundschulen ist auch die Nähe zur Lehrkraft zu berücksichtigen. Eine besonders intensive Begleitung durch die Lehrkraft spielt in jüngeren Jahren eine bedeutendere Rolle als in der Sekundarstufe I. In einer Untersuchung von Jerome, Hamre und Pianta (2009) wurde gezeigt, dass die eingeschätzte Nähe zur Lehrkraft stetig bis zur sechsten Schulklasse abnimmt. Auch motivationale Merkmale wie das Interesse können die Wahrnehmung und Urteile lenken (Jansen et al., 2022; Willems, 2022). Lernende mit positiv ausgeprägtem situationalem Interesse schätzten die Unterrichtsqualität signifikant positiver ein, als Lernende mit geringerem Interesse (Willems, 2022). In der Empirie zeigten sich divergierende Urteile von Lernenden in Abhängigkeit von demografischen Merkmalen wie Geschlecht (Gentry, Gable & Rizza, 2002; Stang & McElvany, 2020), Migrationshintergrund (Levy et al., 2003) und sozialer Hintergrund (Matthews, 2008). So gibt es unter anderem Befunde zu der Tendenz, dass Schülerinnen die Qualität sprachlicher Unterrichtsfächer positiver beurteilen und im Gegensatz dazu Schüler

die Qualität naturwissenschaftlicher Unterrichtsfächer positiver einschätzen (Ditton, 2002; Wagner, 2008).

### *Forschungsdesiderat*

Betrachtet man die Resultate zu Übereinstimmungen zwischen den Wahrnehmungen von Lernenden ist festzuhalten, dass diese nur gering auch innerhalb einer Klasse übereinstimmen (z. B. Gruehn, 2000; Stang & McElvany, 2020; Wagner et al., 2016). Theoretische Ansätze und empirische Befunde verweisen auf mögliche individuelle Einflussfaktoren, welche Unterschiede in den Wahrnehmungen von Lernenden verursachen können. Vermehrt beziehen sich Studien auf Befragungen von Lernenden der Sekundarstufe. Für den Primarbereich liegen bisher weniger empirische Studien zur Wahrnehmung bezüglich der Unterrichtsqualität vor. Der Primarbereich zeichnet sich durch andere Strukturen aus, wie zum Beispiel im Gegensatz zu dem gegliederten System der Sekundarstufe die gemeinsame Beschulung aller Kinder und entwicklungsbedingt durch andere pädagogische Schwerpunkte, wie die intensivere Beziehung zur Lehrkraft. Dadurch können Ergebnisse der Sekundarstufe nicht direkt übertragen und die Wahrnehmungen von Lernenden in Grundschulen über ihren Unterricht müssen näher beleuchtet werden.

Um ebenfalls Urteile von Beteiligten des Unterrichts zu erhalten, kann auch auf Wahrnehmungen von Lehrkräften zurückgegriffen werden.

### **3.2.2 Wahrnehmungen von Lehrkräften**

Im Vergleich zu Lernenden wird bei Lehrkräften vorausgesetzt, dass sie über das methodisch-didaktische Wissen verfügen, um das Unterrichtsgeschehen wahrnehmen und einschätzen zu können (Praetorius, 2014; Marquardt, 2011). Ein weiterer Vorteil der Befragungen von Lehrkräften ist, dass sie sich ebenfalls wie Lernende auf einen großen Beobachtungszeitraum für ihre Wahrnehmungen über den Unterricht beziehen können (Wubbels, Brekelmans & Hooymayers, 1992). Aufgrund ihrer beruflichen Tätigkeit gelten Lehrkräfte als Expert\*innen für die Planung und Durchführung unterrichtlichen Angebots (Kunter & Baumert, 2006). Zu der beruflichen Professionalität von Lehrkräften zählen die angemessene Wahrnehmung des eigenen Unterrichts und die fortlaufende Überprüfung und Verbesserung der eigenen Handlungskompetenzen (Gruber, 2004; Praetorius, 2014; Schwindt, 2008). Genau diesbezüglich werden Lehrkraftwahrnehmungen zur Selbstevaluation des eigenen unterrichtlichen Angebots sowie für die Aus- und Weiterbildung häufig verwendet. Hierbei wird insbesondere mit dem Einsatz von Videos gearbeitet. In der Lehrkraftausbildung wird

häufig der Begriff *Professionelle Wahrnehmung* für die Wahrnehmung der Lehrkräfte vom Unterricht verwendet (Gold, Pfirmann, & Holodynski, 2021; Xu, Widjaja & Ferguson, 2019). Unter der Professionellen Wahrnehmung von Unterricht wird die Kompetenz verstanden, lernförderliche und lernhinderliche Faktoren von Unterrichtsereignissen zu erkennen und analysieren zu können. Darauf aufbauend können im nächsten Schritt präziser lernwirksame Unterrichtshandlungen ausgewählt werden (Hammerness, Darling-Hammond & Shulman, 2002). In der Lehrkraftausbildung hat somit die Professionelle Wahrnehmung eine wichtige Bedeutung, da die Fähigkeit der Lehrkräfte das unterrichtliche Angebot wahrzunehmen und in der Qualität einschätzen zu können, sich zum Beispiel auf danach folgendes Handeln auswirken kann. Daran schließt die Frage an, wie gut Lehrkräfte ihren eigenen Unterricht eigentlich wahrnehmen können. Eng verbunden mit der Professionellen Wahrnehmung von Unterricht ist das Konstrukt der Urteilsgenauigkeit von Lehrkräften, welches eines von drei Komponenten der diagnostischen Kompetenz ist (Kramer, Förtsch, Boone, Seidel & Neuhaus, 2021). Unter dieser wird die Fähigkeit verstanden, Merkmale, in den meisten Fällen Eigenschaften von Lernenden, akkurat einschätzen zu können. Die Genauigkeit wird mittels der Übereinstimmung von Einschätzungen von Lehrkräften mit den beispielsweise objektiven Einschätzungen zu Merkmalen bestimmt (Artelt & Gräsel, 2009; Ready & Wright, 2011). Die Urteilsgenauigkeit kann als Voraussetzung für das Lehren gesehen werden, um das unterrichtliche Angebot an die Bedürfnisse von Lernenden anzupassen (Gabriele, Joram & Park, 2016; Moser Opitz & Nührenböcker, 2015; Vogt & Rogalla, 2009). Hierfür müssen Lehrkräfte ihren eigenen Unterricht angemessen wahrnehmen und einschätzen können, um eine Passung des unterrichtlichen Angebots mit den Bedürfnissen der Lernenden gewährleisten zu können. Für eine gute Passung müssen Lehrkräfte zusätzlich wahrnehmen können, wie das Lernangebot bei den Lernenden ankommt und wie unterschiedlich Lernende auf das Angebot reagieren können. Unterricht sollte an die Lernvoraussetzungen der Lernenden angepasst sein (Praetorius & Südkamp, 2019). Somit ist die Urteilsgenauigkeit bedeutsam für die Differenzierung im Unterricht und für die Unterrichtsqualität (Hill & Chin, 2018).

Im Hinblick auf die Vorhersagekraft von Lehrkraftwahrnehmungen für den Lernerfolg von Lernenden, konnten leichte Zusammenhänge zwischen den Lehrkraftwahrnehmungen über die Klassenführung und Leistungen von Lernenden gefunden werden (Fauth et al., 2014b). Darüber hinaus konnte auch gezeigt werden, dass die Lehrkraftwahrnehmungen bezüglich des unterstützenden Unterrichtsklimas, besonders Wahrnehmungen über den

Zusammenhalt innerhalb der Klasse, prädiktiv für die Zufriedenheit von Lernenden sind (Kearney & Peters, 2013). Prewett und Kollegen (2019) zeigten, dass Lehrkraftwahrnehmungen über die emotionale Unterstützung seitens der Lehrkraft und bezüglich der Nähe zur Lehrkraft mit dem Interesse und mit der Selbstwirksamkeit im Bereich Mathematik von Lernenden zusammenhängen.

#### *Prädiktoren für divergierende Wahrnehmungen innerhalb der Gruppe*

In der empirischen Bildungsforschung wird eher weniger ausschließlich auf Wahrnehmungen von Lehrkräften über das unterrichtliche Angebot zurückgegriffen, um Fragen über Charakteristiken über effektiven Unterricht und deren Auswirkungen auf den Lernerfolg zu beantworten (Marquardt, 2011; Watson, Miller, Davis & Carter, 2010). Ein Grund dafür kann unter anderem sein, dass Wahrnehmungen von Lehrkräften selbstwertdienlichen Verzerrungen unterliegen können (Wubbelset al., 1992). Generell ist das Lehren sowie auch das Lernen eng mit emotionalen Erfahrungen verbunden (Hargreaves, 1998). Auch die persönlichen Identitäten werden durch Emotionen geformt, insbesondere in Berufsfeldern, bei denen von Individuen starkes Engagement erwartet wird, wie es bei Lehrkräften auch zutrifft (Nias, 1986; Zembylas, 2003). Oftmals werden die Identitäten zur Erklärung für und Rechtfertigung von Herangehensweisen an den Unterricht herangezogen (Hargreaves, 1998). Aber nicht nur das Unterrichten an sich ist demzufolge von Emotionen abhängig, sondern auch die Wahrnehmungen von Lehrkräften. Denn die Fähigkeit kritisch zu denken und Entscheidungen zu treffen wird angetrieben von den eigenen Emotionen und Motivationen (Zembylas, 2003). In der Praxis konnte gezeigt werden, dass Lehrkraftwahrnehmungen häufig von Emotionen wie Angst begleitet werden, welche mit Erwartungen und Befürchtungen verbunden sind (Bromme & Rheinberg, 2006; Pohlmann, 2009). Hierzu zählt auch unter anderem die Selbstwirksamkeitserwartung, welche den subjektiven Glauben in eigene Fähigkeiten meint, um den Schulalltag erfolgreich bestreiten zu können (Florin, 2020). Demnach werden Lehrkraftwahrnehmungen über den eigenen Unterricht immer auch von ihren Motivationen, Emotionen und Erwartungen gesteuert und können demzufolge unterschiedlich ausfallen. Lehrkräfte sind selbst die Akteure des eigenen Unterrichts und können nicht losgelöst von ihrer Rolle ihre eigene Handlung beurteilen (Praetorius, 2014). Die Tatsache, dass Lehrkräfte selbst in die Unterrichtssituation eingebunden sind, erschwert es zusätzlich die Komplexität von Unterricht ganzheitlich wahrzunehmen und einordnen zu können (Praetorius, 2014; Schwindt, 2008). Auch Blömeke, Gustafsson und Shavelson (2015) beschäftigten sich mit Unterrichtswahrnehmung von Lehrkräften. Anhand ihres Modells zur

Transformation von Kompetenz in Performanz wird deutlich, dass die Wahrnehmung von Lehrkräften bezüglich Unterrichtsmerkmalen abhängig von Motivationen und Erwartungen ist. Diese sind wiederum eng verbunden mit Einstellungen, mentalen Modellen und Schemata, welche notwendig sind, um Reize einordnen zu können (Stamouli, Schmid & Gruber, 2010). Somit kann sich die Wahrnehmung mit zunehmender Erfahrung der Lehrkraft mit ähnlichen Situationen entwickeln und verändern (Hecker, Falkenstern & Lemmrich, 2020). Empirisch konnte neben der Bedeutung der Erwartungen für die Unterrichtsqualität, auch der Zusammenhang zwischen Motivationen von Lehrkräften und der Qualität des unterrichtlichen Angebots festgestellt werden (z. B. Bleck, 2019). Generell konnte gezeigt werden, dass die Erfahrung bedeutsam ist, wie Unterrichtsqualität von Lehrkräften wahrgenommen wird und dass sich berufserfahrene Lehrkräfte und Anfänger\*innen in der Wahrnehmung von Unterrichtssituationen unterscheiden (Chi, 2006; Cortina, Miller, McKenzie & Epstein, 2015). Zusammengefasst haben sowohl motivational-emotionale als auch kognitive Merkmale (z. B. Vorwissen) von Lehrkräften einen Einfluss auf die Informationsverarbeitung von Lehrkräften (Endsley, 2006; Hecker et al., 2020).

Würden sich aber Lehrkräfte nur auf ihre eigene Wahrnehmung von Unterrichtsqualität verlassen, können aufgrund von Verzerrungen in der Wahrnehmung und nicht passenden Feedbacks des eigenen Handelns, eine Lernumgebung erschaffen oder aufrechterhalten werden, welche nicht den Bedürfnissen der Lernenden entspricht. Reflektierende Praxis beinhaltet daher Vergleiche der Selbstwahrnehmung von Lehrkräften und externen Daten, wie zum Beispiel Wahrnehmungen von Lernenden (Wisniewski et al., 2022). Um die Selbst- und Fremdwahrnehmung im Bereich Unterrichtsqualität vergleichen zu können, werden oftmals standardisierte Feedback-Fragebögen eingesetzt. Studien zeigten, dass die Selbstwahrnehmung der Lehrkräfte mit der Fremdwahrnehmung, gemessen durch Lernende oder externe Urteilende, nur gering bis moderat zusammenhängen (Fauth et al., 2014b; Maulana, Opdenakker, Brok & Bosker, 2012; Wisniewski et al., 2022). Übereinstimmungen der Selbst- und Fremdwahrnehmung fallen niedriger aus, wenn Einschätzungen einen hohen Inferenzgrad bzw. Interpretationsaufwand voraussetzen, also bei hoch-inferenten Einschätzungen (Clausen, 2002). Items zur Unterrichtsqualität sind häufig nicht einfach zu beobachten und resultieren aus der Interaktion im Unterricht (Wisniewski, Zierer, Dresel & Daumiller, 2020).

Letztlich wird beispielsweise von McCaffrey und Kolleg\*innen (2001) empfohlen, Lehrkräfte in Blick auf die Häufigkeit des Einsatzes von Unterrichtspraktiken zu befragen,

jedoch nicht unbedingt bezüglich der Unterrichtsqualität. Doch selbst in Bezug auf den eigenen Anteil der gesamten Sprechzeit, konnte gezeigt werden, dass Lehrkräfte diesen massiv unterschätzen und ihre Einschätzungen nicht mit der Fremdwahrnehmung übereinstimmen. Diese Ergebnisse zeigten sich sowohl für die Sekundarstufe als auch für die Primarstufe (für die Sekundarstufe: Helmke et al., 2008a; für die Primarstufe: Helmke et al., 2008b).

#### *Forschungsdesiderat*

Lehrkräfte sind selbst in die Unterrichtssituation eingebunden, dies erschwert es die Komplexität von Unterricht ganzheitlich wahrzunehmen. Außerdem können sie nur bedingt die Qualität des Unterrichts oder ihr Verhalten losgelöst von ihrer Rolle wahrnehmen. Fasst man die Befundlage zusammen ist festzustellen, dass die Wahrnehmungen von Lehrkräften insbesondere aufgrund von motivationalen sowie emotionalen Merkmalen und Erwartungen beeinflusst sein können. Nur wenige Studien zur Wahrnehmung von Unterrichtsqualität beruhen auf Daten der Lehrkräfte, da unter anderem von vorhinein mit selbstwertdienlichen Verzerrungen gerechnet wird. Bezüglich des Abgleichs der Selbst- bzw. Fremdwahrnehmung zeigten bisherige Studien nur einen geringen Zusammenhang. Für eine erfolgreiche Unterrichtsplanung ist es daher wichtig, dass die Wahrnehmung von Lehrkräften, ihre Selbstwahrnehmung, mit der Wahrnehmung anderer, zum Beispiel der Lernenden, übereinstimmt. Für eine erfolgreiche Passung des unterrichtlichen Angebots mit den Bedürfnissen von Lernenden, ist es nicht nur bedeutsam, dass die Perspektiven übereinstimmen, sondern dass Lehrkräfte einschätzen können, wie Lernende ihren Unterricht wahrnehmen. In diesem Forschungsfeld liegen bisher noch nicht viele Untersuchungen vor, die die Kompetenz von Lehrkräften, die wahrgenommene Unterrichtsqualität auf Lernendenseite einzuschätzen, in den Blick nehmen.

Wichtig ist aber, dass Lehrkräfte trainieren können, die Unterrichtsqualität wahrnehmen zu können. So zeigten Gold und Kollegen (2013) in einer experimentellen Studie, dass angehende Lehrkräfte durch den Einsatz eines Trainingsseminars ihre Fähigkeit verbesserten, Klassenführung in Unterrichtsvideos zu erkennen und zu interpretieren. An dieser Stelle ist natürlich zu erwähnen, dass es sich nicht um den eigenen Unterricht handelte und die angehenden Lehrkräfte die Perspektive von externen Urteilenden einnahmen, um die Qualität der Klassenführung einzuschätzen. Der Einsatz von Unterrichtsvideos und der Perspektivenwechsel von angehenden Lehrkräften und schon praktizierenden Lehrkräften zu externen Urteilenden ist in der Aus- und Weiterbildung inzwischen ein gängiges Verfahren

und hat stark zugenommen (z. B. Gaudin & Chaliès, 2015), aber auch in der empirischen Bildungsforschung wird häufiger auf diese Methode zurückgegriffen.

### 3.2.3 Wahrnehmungen von externen Urteilenden

Laut Helmke (2022) hat keine andere Perspektive oder Methode das Potenzial der Unterrichtsbeobachtung von externen Urteilenden. Bei der Erfassung von Unterrichtsqualität gilt die Perspektive von externen Urteilenden auch als Königsweg. Gründe dafür sind die hohe Validität und die starke Prädiktion für die Leistungen von Lernenden (Fauth et al., 2014b). Der emotionale Abstand zum Unterrichtsgeschehen sowie ein höheres Ausmaß an Vergleichsmöglichkeiten sind weitere Vorteile (Rakoczy, 2008). Bei Unterrichtswahrnehmungen lassen sich zwei Typen der Beobachtungen unterscheiden: die Beobachtung, welche direkt im Klassenraum stattfindet und welche anhand von videografierten Unterrichtssequenzen erfolgt. Direkte Beobachtungen ermöglichen den Beobachter\*innen Gespräche und Interaktionen aufzufassen, welche auf Videoaufzeichnungen verborgen bleiben. Allerdings sind im Gegensatz zu Beobachtungen mittels Videoaufzeichnungen Grenzen bezüglich der Anzahl der Externen gesetzt, die sich physisch im Klassenraum aufhalten können. Weitere Vorteile videografiertter Unterrichtsszenen sind, dass diese differenziert beurteilt, die Bewertungen auch zeitversetzt stattfinden und wiederholt angesehen werden können (Herrle et al., 2016; Mashburn et al., 2014). Eine der ersten groß angelegten Video-Studien ist die 1997 veröffentlichte TIMSS (Third International Mathematics and Science) Videostudie, welche ein wichtiger Impulsgeber für weitere Forschungsprogramme war (Seidel & Thiel, 2017; Stigler et al., 1996). Seitdem hat sich die videobasierte Lehr-Lernforschung sowohl national als auch international etabliert, unter anderem, da Videoaufzeichnungen qualitative sowie quantitative Analysen ermöglichen und wiederholte Beobachtungen und Bewertungen zulassen (Helmke, 2006; Roth, 2009; Seidel & Thiel, 2017). In der Praxis wird die Videografie häufig als Instrument der Professionalisierung von Lehrkräften zur Unterrichtsentwicklung oder zur Wahrnehmung und Beurteilung professionellen Handelns verwendet (Clausen, Weingarten & Wegner, 2013; Gold & Windscheid, 2022). In der Unterrichtsforschung gelten Videostudien als recht zeitaufwendig und teuer (Gargani & Strong, 2014). Unter anderem auch deswegen, da neben der Schulung von Externen auch viel Zeit für das Schauen der Unterrichtsvideos eingeplant werden muss. Neueste Untersuchungen konnten zeigen, dass auch Wahrnehmungen von externen Urteilenden reliabel und prädiktiv für die Leistungen sein können, wenn auch nur

kurze Videos für die Wahrnehmung der Unterrichtsqualität gezeigt werden (Begrich et al., 2020).

Wie in Kapitel 3.1 beschrieben, sind Unterrichtsqualitätsmerkmale der Tiefenstruktur nicht direkt beobachtbar und somit setzen Beurteilungen externer Urteilenden bezüglich Unterrichtsqualitätsmerkmale einen Inferenzschluss voraus, wobei von direkt beobachtbaren Merkmalen auf überdauernde Merkmale des Unterrichts geschlossen werden muss (Begrich et al., 2017; Klieme, 2006). Für die Bildung eines Gesamturteils wird sich häufig an verhaltensnahen definierten Indikatoren orientiert, die in Manualen beschrieben sind (Rakoczy & Pauli, 2006). Um das Ausmaß an Interpretationen zu kontrollieren und zu übereinstimmenden Wahrnehmungen in nicht direkt beobachtbaren Merkmalen zu gelangen, werden in der Unterrichtsforschung Trainings eingesetzt, bei denen die externen Urteilenden subjektive Interpretationen diskutieren (Kane et al., 2012; Lotz et al., 2013).

#### *Prädiktoren für divergierende Wahrnehmungen innerhalb der Gruppe*

Doch obwohl in Studien Trainings und Kodiermanuale eingesetzt werden, gehen teilweise hohe Varianzanteile (bis zu 41.0 %) von Wahrnehmungen auf Beurteilungsfehler zurück und beziehen sich somit nicht auf das einzuschätzende Merkmal (Pietsch & Tosana, 2008; Praetorius, 2014). Dieser Befund wird durch eine Metanalyse von Hoyt und Kerns (1999) unterstützt, in der gezeigt wurde, dass 37.0 Prozent der Varianz in Wahrnehmungen von externen Urteilenden auf individuelle Beurteilungsfehler zurückzuführen sind. Aufgrund der Komplexität der kognitiven Prozesse beim Wahrnehmen von Personen und Situationen können subjektiv verzerrte Fehleinschätzungen bzw. Beurteilungsfehler auftreten (Marquardt, 2011). Beurteilungsfehler bei Unterrichtsbeobachtungen von Externen können auf der einen Seite aufgrund von individuellen Merkmalen und auf der anderen Seite aufgrund von methodischen Methoden auftreten. Unterschiede in der Wahrnehmung von Externen können zum Beispiel aufgrund von der Erfahrung von externen Urteilenden zustande kommen (Gabriel-Busse, Groß-Mlynek, Feldhoff & Haring, 2020; Wolff, Jarodzka & Boshuizen, 2017). In Bezug auf die Erfahrung konnte festgestellt werden, dass Externe mit einem unterschiedlichen Grad an Erfahrung und Expertise sich insbesondere in der Wahrnehmung der Motivierung im Unterricht unterschieden. Externe mit weniger Erfahrung schätzten die Motivierungsqualität unvoreingenommener ein (z. B. Praetorius, Lenske & Helmke, 2012). Wolff und Kolleginnen (2017) konnten einen Zusammenhang zwischen der Erfahrung von Externen und der unterschiedlichen Fokussierung von Unterrichtsqualitätsmerkmalen und der Tiefe der Schlussfolgerungen bei der Wahrnehmung von Unterrichtsvideos nachweisen.

Außerdem können aktuelle Zustände der Externen, beispielsweise die Stimmung, Auswirkungen auf die Wahrnehmung haben, indem diese Zustände insbesondere die Selektion von eingehenden Reizen bzw. Informationen beeinflussen (Martin & Wawrinowski, 2014). Auch kann zum Beispiel die eigene Stimmung als Informationsquelle für Urteile dienen. So können Reize positiv wahrgenommen werden, da Personen generell positiv gestimmt sind (Stroebe, 2014).

Dass Wahrnehmungen von externen Urteilenden hinsichtlich Unterrichtsqualität auch aufgrund methodischer Merkmale divergieren können, wurde schon in einigen wenigen Studien gezeigt (z. B. zu Itemformulierungen: Clausen et al., 2003). Ein methodisches Merkmal, welches die Wahrnehmung von Externen lenken kann, ist die Darbietungsordnung der Unterrichtsszenen. Dieser Beurteilungsfehler wird auch als Reihenfolgeeffekt betitelt. Er tritt bei Gedächtnisprozessen auf und ist auf die Arbeitsweise des Kurzzeitgedächtnisses zurückzuführen (Ebbinghaus 1885; Kooken et al., 2017). Hierbei hat die sequenzielle Position, in der Informationen präsentiert werden, einen Einfluss auf die Wahrnehmung und Beurteilung von Informationen (Cushman & Mele, 2008). Eine Erklärung für den Reihenfolgeeffekt ist, dass die Informationsverarbeitung mit dem vorigen Reiz auf die Verfügbarkeit bestimmter Gedächtnisinhalte wirkt (Sudman, Bradburn & Schwarz, 1996). In der empirischen Bildungsforschung gibt es bisher nicht viele Studien, die untersucht haben, welche Auswirkung die Reihenfolge der zum Beispiel zu beurteilenden Unterrichtsvideos auf die Wahrnehmung von externen Urteilenden haben kann. Eine Untersuchung hierzu zeigte, dass die Varianz, welche auf die Verzerrungen der Wahrnehmung der Externen bezogen werden konnte, verringert wurde, wenn die externen Urteilenden die Videos in einer zufälligen Reihenfolge präsentiert bekamen (Mashburn et al., 2014). Eine Randomisierung von Unterrichtsszenen ist natürlich nur mittels des Beobachtungsverfahrens via Videos möglich. In Live-Beobachtungen lässt sich der Reihenfolgeeffekt weniger gut kontrollieren (Mashburn et al., 2014). Zusätzlich können individuelle Merkmale wie beispielsweise Geschlecht, Alter, Studiengang (Praetorius, 2014), die Konzentrationsfähigkeit (Hogarth & Einhorn, 1992), aktuelle Stimmung (Murphy & De Shon, 2000) sowie Erfahrungen im schulischen Bereich (Strong, Gargani & Hacifazlioglu, 2011) das Auftreten von Beurteilungsfehlern beeinflussen.

### *Forschungsfazit*

Analysen von Unterricht anhand von externen Beobachtungen liefern wichtige Informationen zur Qualität und damit auch zur Verbesserung von Unterricht. Dies hat zur Folge, dass der Bedarf nach Analysen über Faktoren, die die Beurteilungen von Externen beeinflussen, und

somit auch nach Verbesserungen der bisherigen Verfahren der Unterrichtsbeobachtung aufkommt (Mashburn et al., 2014). Die fehlende Stabilität von externen Wahrnehmungen ist ein großer Nachteil. Die Wahrnehmungen sind also nicht nur von tatsächlichen Unterschieden in den Unterrichtsvideos abhängig, sondern auch von den Unterrichtsqualität unabhängigen Faktoren, was die Validität der Ergebnisse einschränkt und zu Fehlinterpretationen von Zusammenhängen mit anderen Konstrukten, wie zum Beispiel Leistungen von Lernenden führen kann. Allerdings wurden bisher noch nicht ausreichend Gründe, sowohl individuelle als auch methodische, für die Instabilität der Wahrnehmung von externen Urteilenden bezüglich Unterrichtsqualität beleuchtet.

## **4 Forschungsanliegen und Forschungsfragen**

Um an Informationen über die Unterrichtsqualität in Klassen zu gelangen, müssen Wahrnehmungen von Lernenden, Lehrkräften oder externen Urteilenden eingeholt werden. Unterrichtsqualität ist nicht direkt beobachtbar und Merkmale der Unterrichtsqualität sind auf der Tiefenstruktur von Unterricht zu verordnen, welche eher durch hoch-inferente Verfahren erfasst werden können. Es konnte aber gezeigt werden, dass sich jeweils die Wahrnehmungen von Lernenden, Lehrkräften und externen Urteilenden hinsichtlich der Begutachtung der Unterrichtsqualität ein und desselben Unterrichts unterscheiden. Für jede Perspektive ist noch nicht hinreichend geklärt, wie diese Unterschiede in den Wahrnehmungen entstehen. Insbesondere für den Bereich der Primarstufe fehlen Studien, welche die Wahrnehmung von Lernenden hinsichtlich Unterrichtsqualität untersuchen. Im Rahmen der vorliegenden kumulativen Dissertation werden die divergierenden Wahrnehmungen von Unterrichtsqualität innerhalb einer Gruppe von Urteilenden fokussiert und individuelle sowie methodische Prädiktoren für die Wahrnehmung von Lernenden, Lehrkräften und externen Urteilenden erforscht. Darüber hinaus wird in den weiterführenden Analysen die Übereinstimmung zwischen der Wahrnehmung von Lernenden und Lehrkräften in den Blick genommen. Die Grundlage für die Ziele und Annahmen der Dissertation bieten hierfür die bisherigen Theorien und empirische Forschung zur Thematik Wahrnehmung von Unterrichtsqualität.

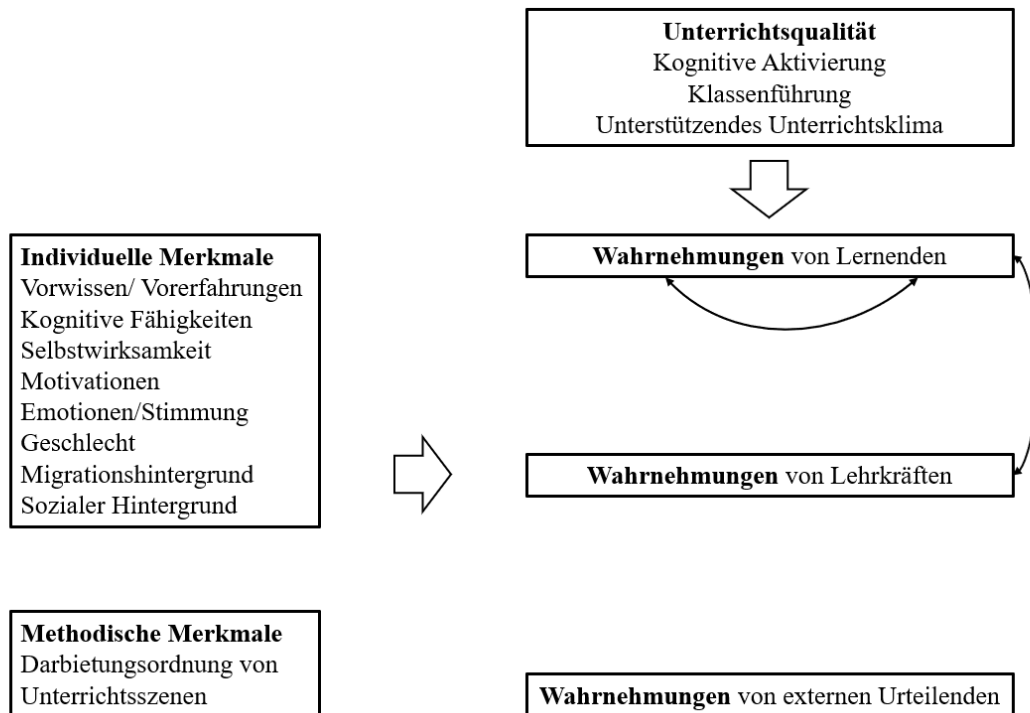


Abbildung 4. Untersuchungsmodell der vorliegenden kumulativen Dissertation

In der Abbildung 4 werden die Forschungsdesiderate und deren Verbindungen zusammengefasst. Bei der Erstellung der Darstellung wurde sich an dem Angebots-Nutzungs-Modell von Helmke (2022) orientiert. In unterschiedlichen Studiendesigns sollen die Wahrnehmungen von Lernenden, Lehrkräften und externen Urteilenden bezüglich der Unterrichtsqualitätsdimensionen kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima in Grundschulklassen beleuchtet werden. Um mögliche Unterschiede in den Wahrnehmungen erklären zu können, werden individuelle Merkmale wie auch methodische Merkmale in den Blick genommen. Hierzu gehören kognitive Merkmale wie das Vorwissen, Erfahrungen und kognitive Fähigkeiten. Aber auch motivationale und emotionale Merkmale wie die Motivationen, Selbstwirksamkeit, Emotionen und Stimmung werden hinsichtlich ihrer Vorhersagekraft für die Wahrnehmung von Unterrichtsqualität untersucht. Zusätzlich werden Zusammenhänge zwischen den demografischen Merkmalen wie beispielsweise das Geschlecht, Migrationshintergrund und der soziale Hintergrund und der Wahrnehmung von Unterrichtsqualität näher beleuchtet. Ergänzend wird das methodische Merkmal Darbietungsordnung von Unterrichtsszenen hinsichtlich seiner prädiktiven Kraft für die Wahrnehmung von externen Urteilenden überprüft und in Verbindung mit individuellen Merkmalen gebracht.

Das Forschungsanliegen lässt sich insgesamt zu zwei übergeordneten Themen zuordnen: Übereinstimmungen von Wahrnehmungen in Grundschulklassen und individuelle und methodische Prädiktoren für die Wahrnehmungen verschiedener Perspektiven. Zu allen Themenbereichen wurden Forschungsfragen formuliert. Generell verfolgt die Dissertation das Ziel, die Wahrnehmung von Unterrichtsqualität näher zu beleuchten und folgende Forschungsfragen sollen im Rahmen der Dissertation beantwortet werden:

*Übereinstimmungen von Wahrnehmungen*

- 1) Wie hoch übereinstimmen die Wahrnehmungen von Lernenden einer Grundschulklasse bezüglich des gleichen Unterrichts?
- 2) Wie genau schätzen Lehrkräfte ein, wie Lernende in ihren Grundschulklassen die Unterrichtsqualität wahrnehmen?

*Individuelle und methodische Merkmale als Prädiktoren für die Wahrnehmungen verschiedener Perspektiven*

- 3) Welche individuellen Merkmale von Lernenden im Grundschulalter bedingen ihre Wahrnehmung von Unterrichtsqualität?
- 4) Welche individuellen Merkmale von Lehrkräften sind prädiktiv für ihre Wahrnehmung von Unterrichtsqualität?
- 5) Welche individuellen Merkmale von Externen bedingen ihre Wahrnehmung von Unterrichtsqualität?
- 6) Welche methodischen Merkmale haben einen Effekt auf die Wahrnehmung von Externen bezüglich Unterrichtsqualität?

Um die Forschungsfragen beantworten zu können, wurden insgesamt vier Beiträge erstellt. Nachfolgend werden die vier Beiträge der Dissertation kurz angerissen. Auch werden Resultate, welche nicht direkt zur Beantwortung der übergeordneten Forschungsfragen notwendig sind, in Kürze dargestellt. Damit ergibt sich ein Überblick über die gesamten empirischen Ergebnisse der Beiträge sowie deren Nutzen bzw. Implikationen für Forschung und Praxis.

## **5 Zusammenfassungen der Einzelbeiträge**

Für die Beantwortung der übergeordneten Fragestellungen dieser Dissertation wurden unterschiedliche Datenquellen herangezogen, welche in den Zusammenfassungen ausführlicher beschrieben werden. Der erste Beitrag fokussierte die Perspektive der Lernenden und untersuchte divergierende Wahrnehmungen von Grundschüler\*innen und Prädiktoren für die Unterschiede in der Wahrnehmung. Im zweiten Beitrag wurde der Frage nachgegangen, inwiefern Motivationen (intrinsische Motivation und Motivation sich an innovativen Programmen zu beteiligen) und Selbstwirksamkeitserwartungen die Wahrnehmung der Unterrichtsqualität von Lehrkräften vorhersagt. Im dritten Beitrag wurde die Perspektive der externen Urteilenden in den Blick genommen und untersucht, ob Wahrnehmungen von Externen aufgrund der Reihenfolge der Unterrichtsszenen verzerrt werden und welche Rolle die Stimmung und die Vorerfahrungen dabei spielen. Der vierte Beitrag konzentrierte sich auf die Frage, ob Lehrkräfte wahrnehmen und akkurat einschätzen können, wie Lernende die Motivierungsqualität ihres eigenen Unterrichts wahrnehmen. Die Daten für den ersten Beitrag sowie die Videosequenzen aus dem dritten Beitrag stammten aus dem Projekt „Entwicklung und Überprüfung von Kompetenzmodellen zur integrativen Verarbeitung von Texten und Bildern“ (01.10.2008 – 30.09.2014; Förderung: Deutsche Forschungsgemeinschaft [DFG]) (vgl. McElvany et al., 2012; Ohle et al., 2017). Grundlage für die Analysen der Beiträge II und IV waren Daten des Projekts „Evaluation von Konzepten und Maßnahmen der fächerübergreifenden Leseförderung im Primarbereich“ (Projektlaufzeit: 01.08.2015 – 31.12.2018; Förderung: Bundesministerium für Bildung und Forschung BMBWF) (McElvany et al., 2018). Die statistischen Analysen wurden mittels der Statistikprogramme SPSS und Mplus durchgeführt.

Die nachfolgenden Zusammenfassungen verfolgen den gleichen Aufbau und enthalten Informationen zu folgenden Aspekten: 1) Theoretischer Hintergrund, 2) Forschungsanliegen, 3) Methode, 5) Ergebnisse und 6) Schlussfolgerungen und Implikationen.

## 5.1 Beitrag I. Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität

*Theoretischer Hintergrund.* Die Wahrnehmung von Lernenden hinsichtlich Unterrichtsqualität ist prädiktiv für ihre Lernergebnisse (z. B. Kunter & Baumert, 2006), variiert jedoch oft erheblich innerhalb des Klassenzimmers (Wagner et al., 2016). Dieses international anerkannte Problem wirft die Frage auf, welche Faktoren diese interindividuellen Unterschiede verursachen. Zur Variabilität der Wahrnehmung von Lernenden in der Grundschule und zur Frage, welche Faktoren die Varianz beeinflussen, ist bisher wenig geforscht worden.

*Forschungsanliegen.* Die vorliegende Studie untersuchte daher die Wahrnehmung von Unterrichtsqualität durch Lernende der Grundschule und analysierte demografische (Geschlecht, Migrationshintergrund, sozialer Hintergrund), sowie kognitive (kognitive Fähigkeit), motivationale (Schulfreude) und emotionale (Nähe zur Lehrkraft) Merkmale von Lernenden als Prädiktoren für Wahrnehmungsvarianzen.

*Methode.* Die Stichprobe bestand aus  $N = 647$  Viertklässler\*innen (50.8 % weiblich) aus 33 Klassen, die direkt nach einer Unterrichtseinheit zu ihrer Wahrnehmung des Unterrichts befragt wurden. Die Lernenden bewerteten drei Kriterien der Unterrichtsqualität mit jeweils fünf Items: Herausfordernde Aufgaben, Störungen im Klassenraum und Motivierung durch die Lehrkraft – und deckten damit Aspekte aus allen Bereichen der Unterrichtsqualität ab. Für die Untersuchung möglicher Prädiktoren wurden Mehrebenenregressionsmodelle unter Berücksichtigung der geclusterten Datenstruktur gerechnet.

*Ergebnisse.* Ergebnisse zeigten, dass die Wahrnehmungen von Lernenden in Grundschulklassen nur gering übereinstimmten ( $.14 \leq ICC \leq .26$ ). Die Wahrnehmungen der herausfordernden Aufgaben wurden durch das Vorwissen, die kognitiven Fähigkeiten, die motivationale Orientierung und den sozialen Hintergrund vorhergesagt. Der soziale Hintergrund war auch prädiktiv für die Wahrnehmungen von Störungen im Klassenraum. Die Wahrnehmung der Motivierung durch die Lehrkraft wurde durch die kognitiven Fähigkeiten und der emotionalen Bindung vorhergesagt.

*Schlussfolgerungen und Implikationen.* Aus psychologischen Theorien zum Wahrnehmungsprozess ist bekannt, dass kognitive Merkmale (z. B. das Vorwissen) und Wahrnehmung sich gegenseitig bedingen. Die Resultate des Beitrags deuteten auf die Relevanz von motivationalen, emotionalen und demografischen Merkmalen für die Wahrnehmung von Lernenden bezüglich Unterrichtsqualität in Grundschulklassen.

## **5.2 Beitrag II. Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht**

*Theoretischer Hintergrund.* In verschiedenen Kontexten konnte die empirische Bedeutung motivationaler Lehrkraftmerkmale (z. B. Selbstwirksamkeitserwartung) für das Lehrkraftverhalten und die wahrgenommenen Unterrichtsqualitätsmerkmale belegt werden (u.a. Holzberger, Philipp & Kunter, 2013).

*Forschungsanliegen.* Ziel dieses Beitrags ist die Frage zu beantworten, inwiefern lesebezogene motivationale Lehrkraftmerkmale, sowohl die wahrgenommene Unterrichtsqualität als auch die Motivation für die Beteiligung an innovativen Programmen, gemessen durch die Quantität und Intensität der Arbeit an Projekten wie der Bund-Länder-Initiative „Bildung durch Sprache und Schrift (BiSS)“, vorhersagen und diese wiederum die wahrgenommene Qualität des Leseunterrichts.

*Methode.* Dafür wurden  $N = 40$  Lehrkräfte (94.9 % weiblich) aus sieben Bundesländern im Mai 2016 zu ihren motivationalen Merkmalen (Selbstwirksamkeitserwartung bezüglich Diagnostik im Leseunterricht und Motivation bezüglich Diagnostik im Leseunterricht), qualitativer (Intensität der Zusammenarbeit) und quantitativer (Zeitaufwand) Beteiligung an innovativen Programmen und zur wahrgenommenen Qualität des Leseunterrichts (Differenzierung im Leseunterricht und Motivierung von Lernenden) befragt. Für die Überprüfung der Fragestellungen wurden Strukturgleichungsmodelle und Mediationsmodelle spezifiziert.

*Ergebnisse.* Die Werte der Variablen der Motivation und Selbstwirksamkeitserwartung der befragten BiSS-Lehrkräfte wiesen hohe Ausprägungen aus. In Regressionsmodellen zeigte sich unter Einbezug aller Variablen, dass die Selbstwirksamkeitserwartung bezüglich

Diagnostik im Leseunterricht prädiktiv für die wahrgenommene Motivierungsqualität von Lehrkräften und die Motivation bezüglich Diagnostik im Leseunterricht prädiktiv für die Differenzierung im Leseunterricht war. Dieser Zusammenhang zwischen individuellen Merkmalen von Lehrkräften und der wahrgenommenen Unterrichtsqualität wurde in Mediationsmodellen von der Beteiligung an innovativen Programmen mediiert.

*Schlussfolgerungen und Implikationen.* Es wurden nur Lehrkräfte befragt, die sich aktiv im BiSS-Projekt beteiligten. Die Auswahl der Stichprobe vergleichend zur Gesamtpopulation der Grundschullehrkräfte kann als selektiv angesehen werden. Hinsichtlich dieser Annahme, dass eher motivierte Lehrkräfte teilnahmen, können auch die gerichteten Zusammenhänge der motivationalen Merkmale von Lehrkräften überschätzt worden sein.

### **5.3 Beitrag III. Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt**

*Theoretischer Hintergrund.* Aufgrund der hohen Validität und Prädiktion für die Lernergebnisse von Lernenden (Fauth et al., 2014a), werden häufig externe Urteilende zur Unterrichtsqualität befragt. Ein Nachteil ist allerdings die fehlende Stabilität von externen Wahrnehmungen. Studien wiesen auf hohe Varianzanteile hin, welche auf Beurteilungsverzerrungen, wie den Reihenfolgeeffekt zurückzuführen sind (Pietsch & Tosana, 2008; Praetorius, 2014).

*Forschungsanliegen.* Im Rahmen des Beitrags soll die Frage geklärt werden, ob in Abhängigkeit von der sequenziellen Ordnung Unterrichtsvideos mit einer mittleren Qualität (a) nach einer vorigen Bewertung eines Unterrichtsvideos mit einer niedrigeren Qualität positiver oder (b) nach einer vorigen Bewertung eines Unterrichtsvideos mit einer höheren Qualität negativer bewertet werden. Zusätzlich wird auch der Einfluss von individuellen Merkmalen der externen Urteilenden (Stimmung [Müdigkeit] und die Vorerfahrung in der Schule) für die wahrgenommene Unterrichtsqualität und für das Auftreten des Reihenfolgeeffekts untersucht.

*Methode.* An der Studie nahmen  $N = 69$  Studierende ( $n = 34$  im Lehramt und  $n = 35$  in anderen Studiengängen) teil ( $M_{Alter} = 24.92$ ,  $SD = 4.40$ ;  $M_{Semester} = 7.83$ ;  $SD = 4.66$ ). Die Studierenden wurden drei Versuchsgruppen zugeordnet, welche zwei 10-minütige Videos in

unterschiedlicher Reihenfolge der Qualitätsniveaus präsentiert bekamen: Gruppe A (niedrig/mittel), B (mittel/hoch) und C (hoch/mittel). Direkt nach jedem Video gaben die Studierenden ihre Einschätzungen für die Qualität der kognitiven Aktivierung, Klassenführung und Motivierung.

*Ergebnisse.* In Bezug auf den Reihenfolgeeffekt konnte festgestellt werden, dass die kognitive Aktivierung und Motivierung negativer nach einer vorigen Beurteilung eines Videos hohen Qualitätsniveaus wahrgenommen wurden, als wenn vorher kein anderes Video eingeschätzt wurde (signifikante Unterschiede zwischen den Bedingungen B und C). Eine höhere Erfahrungszeit in der Schule und eine geringe Müdigkeit hingen mit einer negativeren Wahrnehmung der Motivierungsqualität zusammen. Zwischen der Müdigkeit und der experimentellen Bedingung A (niedrig/mittel) konnte ein Interaktionseffekt gefunden werden. Das individuelle Merkmal Müdigkeit verstärkte den Reihenfolgeeffekt und müdere Studierende, welche zuvor ein Video niedriger Qualität bewerteten, orientierten sich stärker an diesem und überschätzten die Motivierung des Videos mittlerer Qualität stärker als Studierende, welche ihre Müdigkeit geringer wahrnahmen.

*Schlussfolgerungen und Implikationen.* In zukünftigen Videostudien zu Unterrichtsqualität kann ein Auftreten eines Reihenfolgeeffekts vermindert werden, indem die gezeigten Unterrichtsszenen mit unterschiedlichen Qualitätsniveaus randomisiert werden. Auch die Stimmung, zum Beispiel die Müdigkeit, und Vorerfahrungen sollten von externen Urteilenden immer berücksichtigt und zusätzlich erfasst werden.

#### **5.4 Beitrag IV (weiterführende Analysen). Students' perceived motivational support in elementary school – How accurate are teachers' judgments?**

*Theoretischer Hintergrund.* Um den Unterricht bestmöglich an die Bedürfnisse von Lernenden anzupassen und möglichst für alle Lernenden motivierend gestalten zu können, müssen Lehrkräfte die diagnostische Kompetenz besitzen die wahrgenommene Motivierung seitens der Lernenden einzuschätzen. Empirische Studien zur diagnostischen Kompetenz von Lehrkräften haben bisher vor allem die Genauigkeit der Leistungsbeurteilung durch Lehrkräfte im Unterricht fokussiert und nur selten Bereiche wie zum Beispiel die Unterrichtswahrnehmung (Südkamp, Kaiser & Möller, 2012). Bisher zeigte sich, dass für

latente und leistungsfernere Konstrukte Lehrkräfte Schwierigkeiten aufweisen, diese akkurat wahrnehmen und einschätzen zu können.

*Forschungsanliegen.* Ziel des Beitrags war es, die Beurteilungsgenauigkeit von Lehrkräften hinsichtlich der wahrgenommenen Motivationsunterstützung der Lernenden ihrer Grundschulklassen und deren Auswirkungen auf die intrinsische Lesemotivation sowie auf die Entwicklung der intrinsischen Lesemotivation der Lernenden im Laufe eines Schuljahrs zu untersuchen.

*Methode.* Um die Genauigkeit der Lehrkräfteeinschätzungen zu untersuchen, bewerteten die Lehrkräfte, wie ihre Schüler\*innen auf Klassenebene ( $N = 50$  Lehrkräfte) die Unterstützung von Autonomieerleben, Kompetenzerleben und soziale Eingebundenheit wahrnahmen und wie 10 Schüler\*innen auf individueller Ebene ( $N = 49$  Lehrkräfte) diese wahrnahmen. Gleichzeitig bewerteten die Lernenden die tatsächlich wahrgenommene Motivationsunterstützung und ihre intrinsische Lesemotivation. Für die Überprüfung der Urteilsgenauigkeit wurden die drei Urteilskomponenten nach Helmke und Schrader (1987) sowie das globale Abweichungsmaß berechnet.

*Ergebnisse.* Die Resultate zeigten geringe Übereinstimmungen zwischen den Lehrkräfteeinschätzungen der wahrgenommenen Motivationsunterstützung und den tatsächlichen Wahrnehmungen von Lernenden. Insbesondere im Bereich der wahrgenommenen Unterstützung des Autonomieerlebens gab es Ungenauigkeiten im Urteil. Wie theoretisch angenommen, erwies sich die Urteilsgenauigkeit bezüglich der wahrgenommenen Motivierungsqualität der Lernenden als prädiktiv für die (Entwicklung der) Lesemotivation.

*Schlussfolgerungen und Implikationen.* Die Ergebnisse des Beitrags deuten auf Schwierigkeiten von Lehrkräften hin, die wahrgenommene Motivationsunterstützung ihrer Lernenden zu beurteilen. Aufgrund der Bedeutung und Auswirkungen der Genauigkeit des Lehrkrafturteils, wie zum Beispiel auf die (Entwicklung der) Lesemotivation, sollten Lehrkräfte und Lehramtsstudierende für dieses Thema sensibilisiert und darin geschult werden.

## Literaturverzeichnis I

- Allen, J. P., Gregory, A., Mikami, A., Lun, J., Hamre, B. & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—secondary. *School Psychology Review*, 42(1), 76–98. <https://doi.org/10.1080/02796015.2013.12087492>
- Artelt, C. & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(3), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>
- Asendorpf, J. B. (2018). *Persönlichkeit: was uns ausmacht und warum*. Berlin: Springer. <https://doi.org/10.1007/978-3-662-56106-5>
- Aufschnaiter, S. v. & Welzel, M. (2001). *Nutzung von Videodaten zur Untersuchung von Lehr-Lern-Prozessen*. Münster: Waxmann.
- Babad, E. (1996). How high is „high inference“? Within classroom differences in student’s perceptions of classroom interaction. *Journal of Classroom Interaction*, 31, 1–9.
- Baier, F., Decker, A.-T., Voss, T., Kleickmann, T., Klusmann, U. & Kunter, M. (2019). What makes a good teacher? The relative importance of mathematics teachers’ cognitive ability, personality, knowledge, beliefs, and motivation for instructional quality. *British Journal of Educational Psychology*, 89, 767–786. <https://doi.org/10.1111/bjep.12256>
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M. & Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente (Materialien aus der Bildungsforschung Nr. 83)*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Begrich, L., Fauth, B. & Kunter, M. (2020). Who sees the most? Differences in students’ and educational research experts’ first impressions of classroom instruction. *Social Psychology of Education*, 23, 673–699. <http://doi.org/10.1007/s11218-020-09554-2>
- Begrich, L., Fauth, B., Kunter, M. & Klieme, E. (2017). Wie informativ ist der erste Eindruck? Das Thin-Slices-Verfahren zur videobasierten Erfassung des Unterrichts. *Zeitschrift für Erziehungswissenschaften*, 20(1), 23–47.

- Bieg, M., Goetz, T., Sticca, F., Brunner, E., Becker, E., Morger, V. & Hubbard, K. (2017). Teaching methods and their impact on students' emotions in mathematics: an experience-sampling approach. *ZDM Mathematics Education*, 49, 411–422. <https://doi.org/10.1007/s11858-017-0840-1>
- Biesta, G. J. J. & Stengel, B. S. (2016). Thinking Philosophically About Teaching. In D. Gitomer & C. Bell, (Hrsg.), *Handbook of Research on Teaching* (pp. 7–67). Washington, DC: American Educational Research Association. [https://doi.org/10.3102/978-0-935302-48-6\\_1](https://doi.org/10.3102/978-0-935302-48-6_1)
- Blair, I. V., Ma, J. E. & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, 81, 828–841. <https://doi.org/10.1037/0022-3514.81.5.828>
- Bleck, V. (2019). *Lehrerenthusiasmus: Entwicklung, Determinanten, Wirkungen*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-23102-6>
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (2015). Beyond dichotomies: Viewing competence as a continuum. *Zeitschrift für Psychologie*, 223, 3–13.
- Bloom, B. S. (1968). Mastery learning. In *Evaluation comment* (Vol. 1, No. 2). Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation of Instructional Programs.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Boysen, G. A. (2020). The multidimensional nature of teaching and student evaluations: Commentary on students' judgements of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition*, 9(2), 152–156. <https://doi.org/10.1016/j.jarmac.2020.02.001>
- Braga, M., Paccagnella, M. & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71–88.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school: Expanded ed.* Washington, DC: National Academy Press.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.

- Bromme, R. & Rheinberg, F. (2006). Lehrende in Schulen. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (5.überarb. Aufl., S. 296–334). Weinheim & Basel: Beltz.
- Brophy, J. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71, 733–750.
- Brophy, J. (2000). *Teaching (Educational Practices Series—1)*. Geneva, Switzerland: International Academy of Education and International Bureau of Education, UNESCO.
- Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Third handbook of research on teaching* (pp.328–375). New York: Macmillan.
- Brown, G. & Atkins, M. (1990). *Effective teaching in higher education*. London, UK: Routledge.
- Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, 50, 255–272.
- Campbell, R. J., Kyriakides, L., Muijs, R. D. & Robinson, W. (2004). *Assessing teacher effectiveness: developing a differentiated model*. London, New York: RoutledgeFalmer.
- Carpenter, S. K., Witherby, A. E. & Tauber, S. K. (2020). On students' (mis)judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition*, 9(2), 137–151. <https://doi.org/10.1016/j.jarmac.2019.12.009>
- Carroll, J. B. (1963). A model of school learning. *Teacher College Record*, 64, 723–733.
- Chen, G.-M. & Starosta, W. J. (2005). *Foundations of Intercultural Communication*. University Press of America, Lanhan.
- Cheng, Y. C. & Cheung, W. M. (2004). Four types of school environment: Multilevel self-management and educational quality. *Educational Research and Evaluation*, 10(1), 71–100.
- Chi, M. T. H. (2006). Laboratory Methods for Assessing Experts' and Novices' Knowledge. In K. A. Ericsson (Hrsg.), *The Cambridge Handbook of Expertise and Expert Performance* (S. 167–183). Cambridge, MA: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.010>

- Chinn, C. A., Anderson, R. C. & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, 36(4), 378–411.
- Ciftci, S. K. (2015). Effects of Secondary School Students' Perceptions of Mathematics Education Quality on Mathematics Anxiety and Achievement. *Educational Science Theory & Practice*, 15(6), 1487–1501.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.
- Clausen, M., Weingarten, J. & Wegner, H. (2013). Unterrichtsqualität an einer besonderen Schule: Videobasierte Evaluation eines Oberstufen-Internats für leistungsstarke und hoch motivierte Schülerinnen und Schüler. *Gruppendynamik und Organisationsberatung*, 44(3), 301–321.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M. & Weinfeld, F. D. (1966). *Equality of educational opportunity*. Washington, DC.: National Center for Educational Statistics.
- Cortina, K. S., Miller, K. F., McKenzie, R. & Epstein, A. (2015). Where Low and High Inference Data Converge: Validation of CLASS Assessment of Mathematics Instruction Using Mobile Eye Tracking with Expert and Novice Teachers. *International Journal of Science and Mathematics Education*, 13(2), 389–403. <https://doi.org/10.1007/s10763-014-9610-5>
- Creemers, B. P. M. & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. Routledge: London.
- Cruickshank, D. (1985). Profile of an effective teacher. *Educational Horizons*, 90–92.
- Cushman, F. & Mele, A. (2008). Intentional action: two-and-a-half folk concepts? In J. Knobe & S. Nichols (Hrsg.), *Experimental philosophy* (pp. 171–188). Oxford: Oxford University Press.

- Danielson, C. (2011). *The Framework for Teaching Evaluation Instrument*. NJ: Danielson Group.
- Deci, E. L. & Ryan, R. M. (2000). The “what” and “why” of goalpursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Deci, E. L. & Ryan, R. M. (2008). Self-determination theory. A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, 49(3), 182–185.
- Decristan, J., Hess, M., Holzberger, D. & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale. Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift für Pädagogik*, 66, 102–116.
- Degner, J. (2022). *Vorurteile: haben immer nur die anderen*. Berlin, Heidelberg: Springer.
- De Jong, R. & Westerhof, K. J. (2001). The quality of student ratings of teacher behavior. *Learning Environments Research*, 4, 51–85.
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48(2), 262–286.
- Dreesmann, H. R. (1979). *Das Unterrichtsklima als situative Bedingung für kognitive Prozesse und das Leistungsverhalten von Schülern*. Heidelberg, Dissertation.
- Dubberke, T., Kunter, M., McElvany, N., Brunner, M. & Baumert, J. (2008). Lerntheoretische Überzeugungen von Mathematiklehrkräften: Einflüsse auf die Unterrichtsgestaltung und den Lernerfolg von Schülerinnen und Schülern. *Zeitschrift für Pädagogische Psychologie*, 22(3–4), 193–206.
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Eckert, C. (2012). *Beeinflusst Stereotype Threat die Leseleistung von Jungen?* Unveröffentlichte Dissertation, Johannes Gutenberg-Universität Mainz. Retrieved from <https://openscience.ub.uni-mainz.de/handle/20.500.12030/4417>

- Eggen, P. & Kauchak, D. (2014). *Educational psychology: Windows on classrooms* (9th ed.). New Jersey: Pearson Education Inc.
- Ekmekci, A. & Serrano, D. M. (2022). The impact of teacher quality on student motivation, Achievement, and Persistence in Science and Mathematics. *Education Sciences*, 12 (10), 649–669. <https://doi.org/10.3390/educsci12100649>
- Emmer, E. T. & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, 36, 103–112.
- Endsley, M. R. (2006). Expertise and Situation Awareness. In K.A. Ericsson (Hrsg.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 633–653). Cambridge, MA: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.036>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014a). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014b). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2016). Urteile zur Unterrichtsqualität von Grundschüler\_innen: Mehr als „Ich mag meine Lehrerin“? In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler & R. Schmidt (Hrsg.), *Facetten grundschulpädagogischer und grundschuldidaktischer Forschung*. Wiesbaden: Springer.
- Fauth, B., Göllner, R., Lenske, G., Praetorius, A.-K. & Wagner, W. (2018). *Warum (und inwiefern) ist Unterrichtsqualität eine Frage der Perspektive? Auf dem Weg zu einem besseren Verständnis perspektivenspezifischer Unterrichtsbeurteilungen (Symposiumsbeitrag)*. Basel, GEBF. Verfügbar unter <https://unterrichtsforschung.dipf.de/de/arbeitsergebnisse/tagungsbeitraege/beitraege-als-pdf/geb-2018-fauth-gollner-lenske-praetorius-wagner.pdf>

- Fauth, B., Göllner, R., Lenske, L., Praetorius, A. & Wagner, W. (2020a). Who sees what? Theoretical considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik*, *66*, 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>
- Fend, H. (1981). *Theorie der Schule*. München: Urban & Schwarzenberg.
- Fischer, J., Praetorius, A.-K. & Klieme, E. (2019). The impact of linguistic similarity on cross-cultural comparability of students' perceptions of teaching quality. Educational Assessment, Evaluation and Accountability. *Studies in Educational Evaluation*, *65*, 100861. <https://doi:10.1007/s1092-019-09295-7>
- Fiske, S. T. & Taylor, S. E. (2017). *Social cognition: From brains to culture (3rd ed.)*. New York: McGraw-Hill.
- Florin, M. (2020). *Die Bedeutung der Selbstwirksamkeit von Lehrpersonen für ihre professionelle Handlungskompetenz und für die Unterrichtsqualität*. Universität Zürich, Dissertation. Verfügbar unter [https://www.zora.uzh.ch/id/eprint/188375/1/Florin\\_Margaretha\\_Dissertation.pdf](https://www.zora.uzh.ch/id/eprint/188375/1/Florin_Margaretha_Dissertation.pdf)
- Fourie, M. & Schlebusch, G. (2022). Information processing ability and its implications for teaching and learning. *International e-Journal of Educational Studies*, *6*(12), 110–123. <https://doi.org/10.31458/iejes.1130846>
- Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, *11*, 145–252.
- Froiland, J. M. & Oros, E. (2014). Intrinsic motivation, perceived competence and classroom engagement as longitudinal predictors of adolescent reading achievement. *Educational Psychology*, *34*, 119–132.
- Froiland, J. M. & Worrell, F. C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. *Psychology in the Schools*, *53*(3), 321–336.

- Furtak, E. & Kunter, M. (2012). Effects of Autonomy Supportive Teaching on Student Learning and Motivation. *Journal of Experimental Education*, 80(3), 284–316.
- Gabriel-Busse, K., Groß-Mlynek, L., Feldhoff, T. & Haring, M. (2020). Eine Unterrichtssequenz – unterschiedliche Einschätzungen. Analyse videografiertener Unterrichtssequenzen als Bestandteil einer evidenzbasierten Lehrer/innenausbildung. In I. Gogolin, B. Hannover & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung* (ZfE-Edition, Bd. 4, S. 291–314). Wiesbaden: Springer VS.
- Gabriele, A. J., Joram, E. & Park, K. H. (2016). Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes? *Learning and Instruction*, 45, 49–60.
- Gargani, J. & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education*, 65(5), 389–401. <https://doi.org/10.1177/0022487114542519>
- Gaudin, C. & Chaliès, S. (2015). Video viewing in teacher education and professional development: a literature review. *Educational Research Review*, 16, 41–67. <https://doi.org/10.1016/j.edurev.2015.06.001>
- Gentry, M., Gable, R. K. & Rizza, M. G. (2002). Students' perceptions of classroom activities: Are there grade-level and gender differences? *Journal of Educational Psychology*, 94(3), 539–544.
- Gnambs, T. & Hanfstingl, B. (2016). The decline of academic motivation during adolescence: An accelerated longitudinal cohort analysis on the effect of psychological need satisfaction. *Educational Psychology*, 36(9), 1691–1705.
- Gold, B., Förster, S. & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, 27(3), 141–155.
- Gold, B., Pfirrmann, C. & Holodynski, M. (2021). Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *Journal of Teacher Education*, 72(4), 31–447.

- Gold, B. & Windscheid, J. (2022). 360°-Videos in der Lehrer\*innenbildung – Die Rolle der Videoart und des Beobachtungsschwerpunktes für das Präsenzerleben und die kognitive Belastung. In J. Windscheid. & B. Gold (Hrsg.), *360°-Videos in der empirischen Sozialforschung. Ein interdisziplinärer Überblick zum Einsatz von 360°-Videos in Forschung und Lehre* (S. 165–191). Springer VS.
- Goswami, U. (2001). *So denken Kinder. Einführung in die Psychologie der kognitiven Entwicklung*. Bern: Huber.
- Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L. & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation, *Contemporary Educational Psychology*, 29(4), 462–482. <https://doi.org/10.1016/j.cedpsych.2004.01.006>
- Greifeneder, R., Bless, H. & Fiedler, K. (2018). *Social cognition: How individuals construct social reality*. New York: Routledge.
- Gruber, H. (2004). Kompetenzen von Lehrerinnen und Lehrern - Ein Blick aus der Expertiseforschung (Forschungsbericht Nr. 13). In A. Hartinger & M. Fölling-Albers (Hrsg.), *Lehrerkompetenzen für den Sachunterricht* (S.21–33). Bad Heilbrunn: Klinkhardt.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Gruschka, A. (2007). Was ist guter Unterricht? Über neue Allgemein-Modellierungen aus dem Geiste der empirischen Unterrichtsforschung. *Pädagogische Korrespondenz*, 36, 10–43.
- Gucciardi, D. F., Weixian, J. C., Gibson, W., Ntoumanis, N. & Ng, L. (2020). Motivational climate in the classroom: Factorial and convergent validity evidence of the Need-Supportive Behaviors Scale with health science students. *European Journal of Psychological Assessment*, 36(2), 324–335. <https://doi.org/10.1027/1015-5759/a000524>
- Hammerness, K., Darling-Hammond, L. & Shulman, L. S. (2002). Toward expert thinking: How curriculum case writing prompts the development of theory-based professional knowledge in student teacher education. *Teacher Education*, 13, 221–245.

- Hamre, B. K. & Pianta, R. C. (2005). Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure? *Child Development*, 76, 949–967. <http://dx.doi.org/10.1111/j.1467-8624.2005.00889>
- Hamre, B. K. & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization, measurement, & improvement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). New York, NY: Routledge.
- Hargreaves, A. (1998). The emotional practice of teaching. *Teaching and Higher Education*, 14, 835–854. [https://doi.org/10.1016/S0742-051X\(98\)00025-0](https://doi.org/10.1016/S0742-051X(98)00025-0)
- Hasselhorn, M. & Gold, A. (2017). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* (4. Aufl.). Stuttgart: Kohlhammer.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Hattie, J. (2018). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von „Visible Learning“* (4. Aufl.). Baltmannsweiler: Schneider.
- Hattie, J. (2023). *Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement*. Taylor & Francis.
- Hecker, S.-L., Falkenstern, S. & Lemmrich, S. (2020). Zwischen DaZ-Kompetenz und Performanz. Zur Relevanz situationsspezifischer Fähigkeiten und zum Stand ihrer Ausbildung bei Lehrkräften aller Fächer in der Domäne DaZ. *Zeitschrift zur Konzeption, Gestaltung und Diskussion (HLZ)*, 3(1), 565–584.
- Heeg, F. J. (2011). Geleitete Reflexion als Coaching-Methodik für erfolgreiche betriebliche Innovationsprozesse. In R. Giesen (Hrsg.), *Band 2 der Schriftenreihe TRANSFER des DVNLP*.
- Helmke, A. (2002). Kommentar: Unterrichtsqualität und Unterrichtsklima: *Perspektiven und Sackgassen*. *Unterrichtswissenschaft*, 30, 261–277.
- Helmke, A. (2006). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern* (4. Aufl.). Seelze: Kallmeyersche Verlagsbuchhandlung.

- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung. Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Seelze: Klett-Kallmeyer.
- Helmke, A., Helmke, T., Schrader, F. W., Wagner, W., Nold, G. & Schröder, K. (2008a). Alltagspraxis des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 371–381). Weinheim: Beltz.
- Helmke, A., Helmke, T., Heyne, N., Hosenfeld, A., Hosenfeld, I., Schrader, F. W. & Wagner, W. (2008b). Zeitnutzung im Grundschulunterricht: Ergebnisse der Unterrichtsstudie „VERA–Gute Unterrichtspraxis“. *Zeitschrift für Grundschulforschung*, 1, 23–36.
- Helmke, A., Piskol, K., Pikowsky, B. & Wagner, W. (2009). Schüler als Experten von Unterricht. Unterrichtsqualität aus Schülerperspektive. *Lernende Schule*, 46-47, 98–105.
- Helmke, A. & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98.
- Helmke, A. & Schrader, F.-W. (2008). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. *SEMINAR – Lehrerbildung und Schule*, 3, 17–47.
- Henschel, S., Rjosk, C., Holtmann, M. & Stanat, P. (2019). Merkmale der Unterrichtsqualität im Fach Mathematik. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 355–383). Münster: Waxmann.
- Henss, R. (1998). *Gesicht und Persönlichkeitseindruck*. Göttingen: Hogrefe.
- Herrle, M., Rauin, U. & Engartner, T. (2016). Videos als Ressourcen zur Generierung von Wissen über Unterrichtsrealität(en). In U. Rauin & M. Herrle (Hrsg.), *Videoanalysen in der Unterrichtsforschung – Methodische Vorgehensweisen und Anwendungsbeispiele* (S. 8–28). Weinheim: Beltz Juventa.

- Herrmann, C. & Gerlach, E. (2020). Unterrichtsqualität im Fach Sport – Ein Überblicksbeitrag zum Forschungsstand in Theorie und Empirie. *Unterrichtswissenschaft*, 48(3), 361–384. <https://doi.org/10.1007/s42010-020-00080-w>
- Hertel, S. (2014). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. In B. Kopp & B. Arend (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft [21. Jahrestagung der Kommission Grundschulforschung und Pädagogik der Primarstufe 2012 in Nürnberg]* (Jahrbuch Grundschulforschung, Bd. 17, S. 19–34). Wiesbaden: Springer.
- Hill, H. C. & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal*, 55(5), 1076–1112. <https://doi.org/10.3102/0002831218769614>
- Hofer, A., Siedentopf, C. M., Ischebeck, A., Rettenbacher, M. A., Verius, M., Felber, S. & Fleischhacker, W. W. (2006). Gender differences in regional cerebral activity during the perception of emotion: a functional MRI study. *NeuroImage*, 32, 854–862.
- Hogarth, R. M. & Einhorn, H. J. (1992). Order effect in belief updating: The belief – adjustment model. *Cognitive Psychology*, 24, 1–55.
- Holmer, E., Heimann, M. & Rudner, M. (2016). Imitation, sign language skill and the developmental ease of language understanding (d-elu) model. *Frontiers in Psychology*, 7, 107. <https://doi.org/10.3389/fpsyg.2016.00107>
- Holzberger, D., Philipp, A. & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality. A longitudinal analysis. *Journal of Educational Psychology*, 105(3), 774–786.
- Hoyt, W. T. & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: a metaanalysis. *Psychological Methods*, 4, 403–424.
- Huber, C. (2021). Lehrkraftfeedback im Unterricht – wie Förderbedarf, Feedbackvalenz und soziale Integration in Grundschulklassen zusammenhängen. *Empirische Sonderpädagogik*, 13(4), S. 289-311. <http://dx.doi.org/10.25656/01:24107>
- Hugener, I. (2008). Inszenierungsmuster im Unterricht und Lernqualität. Münster: Waxmann.

- Hugener, I., Rakoczy, K., Pauli, C. & Reusser, K. (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In S. Rahm, I. Mammes & M. Schratz (Hrsg.), *Schulpädagogische Forschung, Unterrichtsforschung, Perspektiven innovativer Ansätze* (S. 41–53). Innsbruck: Studien Verlag.
- Huitt, W. (1995). A systems framework of the teaching/learning process. *Educational Psychology Interactive*. Valdosta State University. Retrieved from <http://www.edpsycinteractive.org/materials/tchlrnmd.html>
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T. C. & Valtin, R. (Hrsg.) (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Hußmann, A., Wendt, H., Bos, W. & Rieser, S. (Hrsg.) (2020). *IGLU 2016 – Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Münster: Waxmann.
- Jansen, N., Decristan, J. & Fauth, B. (2022). Individuelle Nutzung unterrichtlicher Angebote: Zur Bedeutung von Lernvoraussetzungen und Unterrichtseteiligung. *Unterrichtswissenschaft*, 50(2), 157–183. <https://doi.org/10.1007/s42010-021-00141-8>
- Jentsch, A., Schlesinger, L., Heinrichs, H., Kaiser, G., König, J. & Blömeke, S. (2021). Erfassung der fachspezifischen Qualität von Mathematikunterricht: Faktorenstruktur und Zusammenhänge zur professionellen Kompetenz von Mathematik Lehrpersonen. *Journal für Mathematik-Didaktik*, 42, 97–121. <https://doi.org/10.1007/s13138-020-00168-x>
- Jerome, E. M., Hamre, B. K. & Pianta, R. C. (2009). Teacher-child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development*, 18(4), 915–945. <https://doi.org/10.1111/sode.2009.18.issue-410.1111/j.1467-9507.2008.00508.x>
- Jungermann, H., Pfister, H.-R. & Fischer, K. (2005). *Die Psychologie der Entscheidung. Eine Einführung*. München: Elsevier Spektrum Akademischer Verlag.
- Kane, T., Staiger, D., McCaffrey, D., Cantrell, S., Archer, J., Buhayar, S., et al. (2012). *Gathering feedback for teaching: Combining high-quality observations with student*

*surveys and achievement gains. Technical report.* Seattle: Bill & Melinda Gates Foundation, Measures of Effective Teaching Project.

Kearney, W. S. & Peters, S. (2013). A comparison of teacher and student perceptions of elementary classroom climate. *National Forum of Educational Administration and Supervision*, 31(1), 20–37.

Kiesel, A. & Koch, I. (2018). Wahrnehmung und Aufmerksamkeit. In A. Kiesel & H. Spada (Hrsg.), *Lehrbuch Allgemeine Psychologie* (S. 35–114). Bern: Hogrefe.

Kleiman, T., Sher, N., Elster, A. & Mayo, R. (2015). Accessibility is a matter of trust: Dispositional and contextual distrust blocks accessibility effects. *Cognition*, 142, 333–344. <https://doi.org/10.1016/j.cognition.2015.06.001>

Klieme, E. (2002). Was ist guter Unterricht? Ergebnisse der TIMSS-Videostudie im Fach Mathematik. In W. Bergsdorf (Hrsg.), *Herausforderungen der Bildungsgesellschaft* (S. 89–113). Weimar: Rhino Verlag.

Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765–773.

Klieme, E. (2020). Guter Unterricht - auch und besonders unter Einschränkungen der Pandemie? *Die Deutsche Schule. Beiheft*, 16, 117–135. <https://doi.org/10.31244/9783830992318>

Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts Pythagoras. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 128–146). Münster: Waxmann.

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 (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.

- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Kleen, H. (2021). *Einstellungen und Urteile von angehenden Lehrkräften gegenüber türkischstämmigen Schülern. Die Relevanz gesellschaftlicher Integration*. Wiesbaden: Springer.
- Klopp, E. (2012). *Das Linsenmodell von Brunswik*. Verfügbar unter <https://www.eric-klopp.de/texte/das-linsenmodell-der-wahrnehmung-nach-egon-brunswik.php>
- Kloss, J. (2013). Grundschüler als Experten für Unterricht – Empirische Überprüfung der Validität von Unterrichtsbeurteilungen durch Schüler der dritten und vierten Jahrgangsstufe. In E. Witruk, *Beiträge zur Pädagogischen und Rehabilitationspsychologie* (Volume 6). Frankfurt am Main: Peter Lang.
- Kooken, J., Welsh, M. E., McCoach, D. B., Miller, F. G., Chafouleas, S. M., Riley-Tillmann, T. C. & Fabio, G. (2017). Test order in teacher-rated behavior assessments: is counterbalancing necessary? *Psychological Assessment*, 29(1), 98–109.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Kounin, J. S. (2006). *Techniken der Klassenführung* (Original der deutschen Ausgabe, 1976). Münster: Waxmann.
- Kramer, M., Förtsch, C., Boone, W. J., Seidel, T. & Neuhaus, B. J. (2021). Investigating pre-service biology teachers' diagnostic competences: relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. *Education Sciences*, 11(89). <https://doi.org/10.3390/educsci11030089>
- Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Göhring, A., Hofmann, B., Kirchhoff, P. & Mulder, R. H. (Hrsg.) (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Münster: Waxmann.
- Krumbacher, C. (2016). *Die Relevanz lernprozessorientierter Sequenzierung im physikbezogenen Sachunterricht – eine Videostudie zur Berücksichtigung von*

*Tiefenstrukturen beim Experimentieren.* Duisburg-Essen: Universitätsbibliothek Duisburg-Essen.

Kürten, R., Greefrath, G. & Hammann, M. (Hrsg.) (2020). *Komplexitätsreduktion in Lehr-Lern-Laboren. Innovative Lehrformate in der Lehrerbildung zum Umgang mit Heterogenität und Inklusion.* Münster: Waxmann. <https://doi.org/10.31244/9783830989905>

Kuger, S. (2016). Curriculum and Learning Time in International School Achievement Studies. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing Contexts of Learning* (S. 395–422). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-45357-6\\_16](https://doi.org/10.1007/978-3-319-45357-6_16)

Kultusministerkonferenz, (2004). *Standards für die Lehrerbildung: Bildungswissenschaften.* Bonn: KMK.

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.

Kunter, M. & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (Dortmunder Symposium der Empirischen Bildungsforschung, Bd. 1, S. 9–31). Münster: Waxmann.

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. & Trautwein, U. (2013). *Psychologie des Unterrichts.* Paderborn: Ferdinand Schöningh.

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

- Kuzmanovic, B., Jefferson, A., Bente, G. & Vogeley, K. (2013). Affective and motivational influences in person perception. *Frontiers in Human Neuroscience*, 7, Article 266. <https://doi.org/10.3389/fnhum.2013.00266>
- Kyriakides, L., Christoforou, C. & Charalambous, C.Y. (2013). What matters for student learning outcomes: a meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152.
- Kyriakides, L., Creemers, B. P., Panayiotou, A. & Charalambous, E. (2020). *Quality and Equity in Education*. Routledge: London.
- Lau, A. & Plessner, H. (2016). *Sozialpsychologie und Sport – Ein Lehrbuch in 12 Lektionen*. Aachen: Meyer & Meyer.
- Lazarides, R. & Buchholz, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction*, 61, 45–59.
- Lazarides, R., Dietrich, J. & Taskinen, P. H. (2019). Stability and change in students' motivational profiles in mathematics classrooms: The role of perceived teaching. *Teaching and Teacher Education*, 79, 164–175.
- Lazarides, R. & Raufelder, D. (2021). Control-Value Theory in the Context of Teaching: Does Teaching Quality Moderate Relations Between Academic Self-Concept and Achievement Emotions? *British Journal of Educational Psychology*, 91(1), 127–147. <https://doi.org/10.1111/bjep.12352>
- Lenske, G. (2016). *Schülerfeedback in der Grundschule. Untersuchungen zur Validität*. Münster: Waxmann.
- Lenski, A. E., Richter, D. & Pant, H. A. (2015). Kompetenzorientierung im Unterricht aus der Perspektive von Lehrkräften und Schülerinnen und Schülern. *Zeitschrift für Pädagogik*, 61, 712–737.
- Leuders, T. & Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht. *Unterrichtswissenschaft*, 39(3), 213–230.

- Levy, J., Den Brok, P., Wubbels, T. & Brekelmans, M. (2003). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research*, 6(1), 1–31.
- Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O. & Reiss, K. (Hrsg.) (2023). *PISA 2022. Analyse der Bildungsergebnisse in Deutschland*. Münster: Waxmann
- Lindl, A. & Krämer, A. (2022). Was macht guten Lateinunterricht aus? Eine Delphi-Studie zu Unterrichtsqualität im Fach Latein. *Die Alten Sprachen im Unterricht*, 69(1), 9–15.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (2. Aufl., S. 69–105). Heidelberg: Springer.
- Lipowsky, F., Kastens, C., Lotz, M. & Faust, G. (2011). Aufgabenbezogene Differenzierung und Entwicklung des verbalen Selbstkonzepts im Anfangsunterricht. *Zeitschrift für Pädagogik*, 57(6), 868–884.
- Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K. & Pauli, C. (2009). Quality of geometry instruction and its short-term impact on students? Understanding of Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537.
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, 59(3), 357–380.
- Lotz, M. & Lipowsky, F. (2016): Die Hattie-Studie und ihre Bedeutung für den Unterricht. Ein Blick auf ausgewählte Aspekte der Lehrer-Schüler-Interaktion. In G. Mehlhorn, F. Schulz & K. Schöppe (Hrsg.), *Begabungen entwickeln & Kreativität fördern* (S. 97–136). München: kopaed.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings in multilevel modelling. *Contemporary Educational Psychology*, 34, 120–131.
- Lüdtke, O., Trautwein, U., Kunter, M. & Baumert, J. (2006). Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20, 85–96.

- Madill, R. A., Gest, S. D. & Rodkin, P. C. (2014). Students' perceptions of relatedness in the classroom: The roles of emotionally supportive teacher-child interactions, children's aggressive disruptive behaviors, and peer social preference. *School Psychology Review*, 43(1), 86–105.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry, & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education* (pp. 319–383). Dordrecht: Netherlands: Springer.
- Martin, E. & Wawrinowski, U. (2014). *Beobachtungslehre. Theorie und Praxis reflektierter Beobachtung und Beurteilung* (6. aktual. erw. Aufl). Weinheim: Beltz Juventa
- Marquardt, C. (2011). *Videoanalysen in der Unterrichtsforschung: Momentaufnahme oder typischer Schulalltag? Eine Betrachtung der DESI-Videostudie*. Hamburg: Diplomica.
- Mashburn, A. J., Meyer, J. P., Allen, J. P. & Pinata, R. C. (2014). The Effect of Observation Length and Presentation Order on the Reliability and Validity of an Observational Measure of Teaching Quality. *Educational and Psychological Measurement*, 74(3), 400–422.
- Matthews, C. E. (2008). *Socio-economic status, student perceptions and college readiness*. Washington State University, Dissertation. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.427.3797&rep=rep1&type=pdf>
- Maulana, R., Opdenakker, M.-C. & Bosker, R. (2016). Teachers' instructional behaviors as important predictors of academic motivation: Changes and links across the school year. *Learning and Individual Differences*, 50, 147–156. <https://doi.org/10.1016/j.lindif.2016.07.019>
- Maulana, R., Opdenakker, M., Brok, P. & Bosker, R. J. (2012). Teacher–student interpersonal relationships in Indonesian lower secondary education: Teacher and student perceptions. *Learning Environments Research*, 15, 251–271. <https://doi.org/10.1007/s10984-012-9113-7>
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1), 14–19.

- McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D. & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education*, 32(5), 493–517. <https://doi.org/10.2307/749803>
- McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D. & Brennan, M. (2019). Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss. *Frontiers in Neuroscience*, 13, 1093. <https://doi.org/10.3389/fnins.2019.01093>
- McDonald Connor, C., Piasta, S. B., Fishman, B. J., Glasney, S., Schatschneider, C., Crowe, E. et al. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development*, 80(1), 77–100.
- McElvany, N., Lorenz, R., Frey A., Goldhammer, F., Schilcher, A. & Stubbe, T. C. (Hrsg.) (2023). *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre*. Münster: Waxmann.
- McElvany, N., Ohle-Peters, A., Iglar, J., Schlitter, T., Teerling, A., Asseburg, R. & Köller, O. (2018). Evaluation der Leseförderung an Grundschulen im Rahmen von "Bildung durch Sprache und Schrift" (BiSS). In F. Schwabe, N. McElvany, W. Bos & H. G. Holtappels (Hrsg.), *Jahrbuch der Schulentwicklung. Bd. 20: Schule und Unterricht in gesellschaftlicher Heterogenität* (1. Auflage, S. 258–280). Weinheim: Beltz Juventa.
- McElvany, N., Schroeder, S., Baumert, J., Schnotz, W., Horz, H. & Ullrich, M. (2012). Cognitively demanding learning materials with texts and instructional pictures: Teachers' diagnostic skills, pedagogical beliefs and motivation. *European Journal of Psychology of Education*, 27(3), 403–420.
- McIlrath, D. & Huitt, W. (1995). The teaching-learning process: A discussion of models. *Educational Psychology Interactive*. Valdosta State University. Retrieved from <http://www.edpsycinteractive.org/papers/modeltch.html>
- McLean, L., Sparapani, N., Toste, J. R. & Connor, C. M. (2016). Classroom quality as a predictor of first graders' time in non-instructional activities and literacy achievement. *Journal of School Psychology*, 56, 45–58. <https://doi.org/10.1016/j.jsp.2016.03.004>

- Melnick, M. D., Harrison, B. R., Park, S., Bennetto, L. & Tadin, D. (2013). A strong interactive link between sensory discriminations and intelligence. *Current Biology*, 23, 1013–1017.
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelson Verlag GmbH.
- Meyer, H. (2017). *Was ist guter Unterricht?* (12. Aufl.). Berlin: Cornelson Verlag GmbH.
- Micholka-Metsch, J. & Metsch, M.-C. (2015). *Strategien für die deutsch-chinesische Geschäftsbeziehung. Erfolgreich verhandeln und Konflikte lösen.* Wiesbaden: Springer.
- Moos, R. H. (1979). *Evaluating educational environments.* San Francisco: Jossey-Bass.
- Moser Opitz, E. & Nührenbörger, M. (2015). Diagnostik und Leistungsbeurteilung. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 491–512). Berlin: Springer.
- Muijs, D., Kyriakides, L., Werf, G. van der, Creemers, B., Timperley, H. & Earl, L. (2014). State of the art—Teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256.
- Murphy, K. R. & De Shon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 53, 873–900.
- Neuber, K. (2021). *Schülerrückmeldungen zum Unterricht von Lehramtsstudierenden. Eine mehrperspektivische Analyse der Nutzung von Schülerrückmeldungen im Praxissemester.* Münster: Waxmann.
- Nias, J. (1986). *Teacher socialization: The individual in the system.* Victoria, Canada: Deakin University Press.
- Niemiec, C. P. & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom. *Theory and Research in Education*, 7(2), 133–144.
- Ohle, A., McElvany, N., Schnotz, W., Wagner, I., Horz, H., Ullrich, M. et al. (2017). Development and Evaluation of a Competency Model for Teaching Integrative Processing of Texts and Pictures (BiTe). In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence Assessment in Education – Research, Models and Instruments* (pp. 167–180). Dordrecht: Springer.

- Ohle-Peters, A., Iglar, J., Schlitter, T., Teerling, A., Köller, O. & McElvany, N. (2021). Unterrichtsqualität und intrinsische Lesemotivation im Kontext der Bund-Länder-Initiative „Bildung durch Sprache und Schrift“ (BiSS). *Zeitschrift für Erziehungswissenschaft*, 24, 861–882. <https://doi.org/10.1007/s11618-021-01022-7>
- Oga-Baldwin, W. L. Q. & Fryer, L. K. (2017). Exploring motivational profiles in public elementary school English classes. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in language education*. Tokyo: JALT.
- Oga-Baldwin, W. L. Q. & Fryer, L. K. (2020). Girls show better quality motivation to learn languages than boys: latent profiles and their gender differences. *Heliyon*, 6(5), e0405. <https://doi.org/10.1016/j.heliyon.2020.e04054>
- Ophir, E. A., Hesselmann, G. & Lamy, D. (2020). The attentional blink unveils the interplay between conscious perception, spatial attention and working memory encoding. *Conscious. Cognit*, 85, 103008. <https://doi.org/10.1016/j.concog.2020.103008>
- Oser, F. & Patry, J.-L. (1990). *Choreographien unterrichtlichen Lernens: Basismodelle des Unterrichts*. Berichte zur Erziehungswissenschaft (Nr. 89). Freiburg (CH): Pädagogisches Institut der Universität Freiburg.
- Patrick, H., Kaplan, A. & Ryan, A. M. (2011). Positive classroom motivational environments: Convergence between mastery goal structure and classroom social climate. *Journal of Educational Psychology*, 103(2), 367–382.
- Pauli, C. & Reusser, K. (2003). Unterrichtsskripts im schweizerischen und im deutschen Mathematikunterricht. *Unterrichtswissenschaft*, 31, 238–272.
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52, 774–798.
- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109–119. <https://doi.org/10.3102/0013189X09332374>
- Pianta, R. C., La Paro, K. & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore: Paul H. Brookes.

- Pieper, V. (2023). *(Lehrkraft-)Feedback im Klassenraum – Eine qualitative Videostudie zu Interaktionen von Lehrkräften und Schüler\*innen im Mathematikunterricht vierter Grundschulklassen in Niedersachsen*. Veröffentlichte Dissertation. Verfügbar unter <https://voado.uni-vechta.de/bitstream/handle/21.11106/469/Vanessa%20Pieper%20Dissertation%202023.pdf?sequence=2&isAllowed=y>
- Pietsch, M. & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Piskol, K. (2008). *Unterrichtsqualität aus der Schülerperspektive: Ein Verfahren zur Unterrichtsentwicklung*. Veröffentlichte Diplomarbeit, Universität Mannheim. Verfügbar unter [https://madoc.bib.uni-mannheim.de/2664/1/Piskol\\_2008.pdf](https://madoc.bib.uni-mannheim.de/2664/1/Piskol_2008.pdf)
- Pohlmann, S. (2009). *Der Übergang am Ende der Grundschulzeit, Zur Formation der Übergangsempfehlung aus der Sicht der Lehrkräfte*. Münster: Waxmann.
- Porto, L., Wouters, J. & Wieringen, A. van (2023). Speech perception in noise, working memory, and attention in children: A scoping review. *Hearing Research*, 439, 108883. <http://dx.doi.org/10.1016/j.heares.2023.108883>
- Posten, A.-C. & Gino, F. (2021). How trust and distrust shape perception and memory. *Journal of Personality and Social Psychology*, 121(1), 43–58. <https://doi.org/10.1037/pspa0000269>
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Praetorius, A.-K. & Charalambous, C. Y. (2018). Classroom observation frameworks for studying teaching quality: looking back and looking forward. *ZDM Mathematics Education*, 50(3), 535–553.
- Praetorius, A.-K. & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? *Unterrichtswissenschaft*, 49, 167–188. <https://doi.org/10.1007/s42010-021-00119-6>
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifität. *Unterrichtswissenschaft*, 48(3), 409–446.

- Praetorius, A.-K., Lenske, G. & Helmke, A. (2012). Observer ratings of instructional quality: do they fulfill what they promise? *Learning and Instruction*, 22, 387–400.
- Praetorius, A.-K., Lauermann, F., Klassen, R. M., Drexler, K., Nitsche, S., Janke, S., Dickhäuser, O. & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student reported teaching quality. *Teaching and Teacher Education*, 65, 241–254.
- Praetorius, A.-K. & Kleickmann, T. (2022). Nutzung von Lerngelegenheiten im Unterricht: konzeptuelle und methodische Zugänge und Herausforderungen - Einführung in den Thementeil. *Unterrichtswissenschaft*, 50, 149–155. <https://doi.org/10.1007/s42010-022-00145-y>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of the three basic dimensions. *ZDM: The International Journal on Mathematics Education*, 50, 407–426.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Praetorius, A.-K., Rogh, W. & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48(3), 303–318.
- Praetorius, A.-K. & Südkamp, A. (2019). Diagnostische Kompetenz von Lehrpersonen hinsichtlich motivationaler Merkmale von Schülerinnen und Schülern. In H. Gaspard, U. Trautwein, M. Hasselhorn, *Diagnostik und Förderung von Motivation und Volition* (S. 35–49). Göttingen: Hogrefe.
- Prewett, S. L., Bergin, D. A. & Huang, F. L. (2019). Student and teacher perceptions on student-teacher relationship quality: A middle school perspective. *School Psychology International*, 40(1), 66–87.
- Proctor, C. P. (1984). Teacher expectations: A model for school improvement. *The Elementary School Journal*, 84(4), 469–481. Retrieved from <http://www.jstor.org/stable/1001371>

- Rakoczy, K. (2021). Unterrichtsqualität aus Sicht der Forschung. In K. Seifried, S. Drewes & M. Hasselhorn (Hrsg.), *Handbuch Schulpsychologie: Psychologie für die Schule* (S. 298–307). Stuttgart: Kohlhammer.
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht – Unterricht aus der Perspektive von Lernenden und Beobachtern*. Münster: Waxmann.
- Rakoczy, K., Frick, U., Weiß-Wittstadt, S., Tallon, M. & Wagner, E. (2022). Einmal begeistert, immer begeistert? Eine Experience-Sampling Studie zur wahrgenommenen Unterrichtsqualität und Motivation von Schülerinnen und Schülern im Kunstunterricht. *Unterrichtswissenschaft*, 50, 211–236. <https://doi.org/10.1007/s42010-022-00147-w>
- Rakoczy, K. & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch -deutschen Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis". Teil 3. Videoanalysen* (S. 206–233). Frankfurt: GFPF/DIPF.
- Ready, D. D. & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360. <https://doi.org/10.3102/0002831210374874>
- Reeve, J. & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, 98(1), 209–218.
- Remmers, H. H. & Brandenburg, G. C. (1927). Experimental data on the Purdue ratings scale for instructors. *Educational Administration and Supervision*, 13, 519–527.
- Reusser, K. (2008). Empirisch fundierte Didaktik – didaktisch fundierte Unterrichtsforschung. In M. A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik* (9. Sonderheft der Zeitschrift für Erziehungswissenschaft, S. 219–237). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.),

- Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 9–32). Münster: Waxmann.
- Roebbers, C. M. & Schneider, W. (2001). Individual Differences in Children's Eyewitness Recall: The Influence of Intelligence and Shyness. *Applied Developmental Science*, 5, 9–20.
- 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. *Elementary School Journal*, 100, 443–471.
- 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. *Review of Educational Research*, 81(4), 493–529.
- Rosenfeld, L. B., Richman, J. M. & Bowen, G. L. (2000). Social support networks and school outcomes: The centrality of the teacher. *Child & Adolescent Social Work Journal*, 17(3), 205–226.
- Rosenshine, B. & Fürst, N. (1971). Research on teacher performance criteria. In B. O. Smith (Hrsg.), *Research in teacher education* (S. 37–72). Englewood Cliffs, NJ: Prentice-Hall.
- Roth, K. (2009). Using video studies to compare and understand science teaching: results from the TIMSS 1999 video study of 8th grade science teaching. In T. Janik, & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 7–19). Münster: Waxmann.
- Ryan, R. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Schatz, C. (2021). *Unterrichtsqualität aus Sicht der Lernenden. Eine empirische Untersuchung zur Wahrnehmung der Entwicklung von Unterrichtsqualität aus der Perspektive der Lernenden an den Projektschulen „Schulen zum Leben“ in Mecklenburg-Vorpommern*. Stuttgart: Schneider Verlag.

- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.
- Scherer, R., Nilsen, T. & Jansen, M. (2017). 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>
- Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197–222). New York/London: Routledge.
- Schiepe-Tiska, A., Heinle, A., Heine, J.-H., Todtenhöfer, P., Reinhold, F., Krauss, S., Holzberger D., Lewalter, D. & Reiss, K. (2023). Wie sieht der aktuelle Mathematikunterricht in Deutschland aus? Befunde aus PISA 2022 und PISA-Ceco. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller & K. Reiss (Hrsg.), *PISA 2022. Analyse der Bildungsergebnisse in Deutschland* (S. 199–230). Münster: Waxmann.
- Schunk, D. H. (2011). *Learning theories: An educational perspective* (6th ed.). London: Pearson.
- Schwindt, K. (2008). *Lehrpersonen betrachten Unterricht: Kriterien für die kompetente Unterrichtswahrnehmung*. Münster: Waxmann.
- Schweizer, Karl (Hrsg.) (2006). *Leistung und Leistungsdiagnostik*. Heidelberg: Springer.
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie: Integration von Struktur und Prozessparadigma. *Zeitschrift für Pädagogik*, 60, 850–866.
- 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.
- Seidel, T. & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaften, Sonderheft*, 32, 1–21.
- Slavin, R.E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21(2), 141–157. [http://doi/10.1016/0883-0355\(94\)90029-9](http://doi/10.1016/0883-0355(94)90029-9)

- Slavin, R. E. (2009). *Educational psychology: Theory and practice* (9th ed.). Upper Saddle River, NJ: Pearson Education.
- Snook, I., Clark, J., Harker, R., O'Neill, A.-M. & O'Neill, J. (2010). Critic and conscience of society: A reply to John Hattie. *Journal of Educational Studies*, 45, 93–98.
- Snow, R. E., Corno, L. & Jackson, D. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243–310). New York: MacMillan.
- Soebari, T. S. & Aldridge, J. M. (2015). Using student perceptions of the learning environment to evaluate the effectiveness of a teacher professional development programme. *Learning Environments Research*, 18, 163–178.
- Stahns, R., Rieser S. & Lankes, E.-M. (2020). Unterrichtsführung, Sozialklima und kognitive Aktivierung im Deutschunterricht in vierten Klassen. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe & R. Valtin, *Internationale Grundschul-Lese-Untersuchung 2016 (IGLU 2016) (Version 1) [Datensatz]* (S. 251–275). Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen. [http://doi.org/10.5159/IQB\\_IGLU\\_2016\\_v1](http://doi.org/10.5159/IQB_IGLU_2016_v1)
- Stamouli, E., Schmid, C. & Gruber, H. (2010). Expertiseerwerb: Jagt die Lehrerinnen- und Lehrerfortbildung einem Phantom hinterher? In F.H. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 107–124). Münster: Waxmann.
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S. & Henschel, S. (Hrsg.). (2022). *IQB-Bildungstrend 2021 Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Münster: Waxmann.
- Stang, J., Lepper, C., Steffensky, M. & McElvany, N. (2020). Einblicke in die Gestaltung des Mathematik- und naturwissenschaftsbezogenen Sachunterrichts an Grundschulen in Deutschland. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt. (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 187–208). Münster: Waxmann.

- Stang, J. & McElvany, N. (2020). Unterschiede in der Wahrnehmung der Qualität des Deutschunterrichts zwischen Grundschülerinnen und Grundschulern. *Zeitschrift für Pädagogische Psychologie*, 35, 261–273.
- Stroebe, W. (2014). Strategien zur Einstellungs- und Verhaltensänderung. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 231–268). Berlin: Springer.
- Strong, M., Gargani, J. & Hacifazlioglu, Ö. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62, 367–382.
- Sudman, S., Bradburn, N. M. & Schwarz, N. (1996). *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Südkamp, A., Kaiser J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, 743–762.
- Südkamp, A. & Praetorius, A.-K. (Hrsg.) (2017). *Diagnostische Kompetenz von Lehrkräften – Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.
- Talić, I., Scherer, R., Marsh, H. W., Greiff, S., Möller, J. & Niepel, C. (2022). Uncovering everyday dynamics in students' perceptions of instructional quality with experience sampling. *Learning and Instruction*, 81, 1–14. <https://doi.org/10.1016/j.learninstruc.2022.101594>
- Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S. & Koestner, R. (2014). A self-determination theory approach to predicting school achievement overtime: The unique role of intrinsic motivation. *Contemporary Educational Psychology*, 39, 342–358.
- Terhart, E. (1994). Unterricht. In D. Lenzen (Hrsg.), *Erziehungswissenschaft. Ein Grundkurs* (S. 133–158). Reinbek: Rowohlt.
- Thiel, F., Ophardt, D. & Piwowar, V. (2013). *Kompetenzen des Klassenmanagements (KODEK): Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement; Abschlussbericht des Projekts*. [Freie Universität Berlin]. <https://doi.org/10.2314/GBV:771646968>. Verfügbar unter <https://www.ewi-psy.fu->

berlin.de/erziehungswissenschaft/arbeitsbereiche/schulentwicklungsforschung/downloads/Abschlussbericht-KODEK.pdf

Titsworth, B. S. & Kiewra, K. A. (2004). Spoken organizational lecture cues and student note-taking as facilitators of student learning. *Contemporary Educational Psychology*, 29, 447–461.

Uslar, D. v. (2007). *Existenz und Lebensgeschichte. Der Mensch und seine Welt im Spiegel der Psychologie. Grundzüge einer Psychologischen Anthropologie*. Würzburg: Königshausen und Neumann.

Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts – Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift für Pädagogik, Beiheft*, 66, 63–80.

Vogt, F. & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, 25, 1051–1060.

Voss, T., Zachrich, L., Fauth, B. & Wittwer, J. (2022). The same yet different? Teaching quality differs across a teacher's classes, but teachers with higher knowledge make teaching quality more similar. *Learning and Instruction*, 80, 101614. <https://doi.org/10.1016/j.learninstruc.2022.101614>

Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht – am Beispiel der Studie DESI (Deutsch Englisch Schülerleistungen International) der Kultusministerkonferenz*. Universität Koblenz-Landau, Dissertation. Verfügbar unter [https://kola.opus.hbz-nrw.de/frontdoor/deliver/index/docId/194/file/Diss\\_\(Publikation\).pdf](https://kola.opus.hbz-nrw.de/frontdoor/deliver/index/docId/194/file/Diss_(Publikation).pdf)

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(5), 705–721.

- Walberg, H. J. (1966). *Classroom Climate Questionnaire*. Cambridge: Harvard University.
- 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, 1834–1868.
- Wallen, N. E. & Travers, R. M. W. (1963). Analyses and investigation of teaching methods. In N. L. Gage (Hrsg.), *Handbook of research on teaching* (S. 448–505). Chicago, IL: Rand McNally.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294. <https://doi.org/10.2307/1170546>
- Watson, S., Miller, T., Davis, L. & Carter, P. (2010). Teachers' perceptions of the effective teacher. *Research in the schools*, 17(2), 11–22.
- Wettstein, A., Ramseier, E., Scherzinger, M. & Gasser, L. (2016). Unterrichtsstörungen aus Lehrer- und Schülersicht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48, 171–183.
- Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: Harper-Collins.
- Wickens, C. D. & Carswel, C. M. (2021). Information Processing. In G. Salvendy & W. Karwowski, *Handbook of human factors* (pp. 114–158). New Jersey: John Wiley & Sons.
- Wiesner, C. (2020). Diagnostische Urteile durch Evidenzen im Unterricht. *Erziehung & Unterricht*, 7–8(170), 591–599.
- Wild, A. (2015). Relationships between high school chemistry students' perceptions of a constructivist learning environment and their STEM career expectations. *International Journal of Science Education*, 37(14), 2284–2305, <https://doi.org/10.1080/09500693.2015.1076951>
- Willems, A. S. (2022). Individuelle Schüler\*innenprofile des situationalen und dispositionalen Interesses und ihre Bedeutung für die Wahrnehmung der Unterrichtsqualität im Fach Mathematik. *Zeitschrift für Erziehungswissenschaft*, 25(2), 377–404.

- Wiprächtiger-Geppert, M., Stahns, R. & Riegler, S. (2021). Fachspezifität von Unterrichtsqualität in der Deutschdidaktik. *Unterrichtswissenschaft*, 49(2), 203–209. <http://doi.org/10.1007/s42010-021-00109-8>
- Wisniewski, B., Röhl, S. & Fauth, B. (2022). The perception problem: A comparison of teachers' self-perceptions and students' perceptions of instructional quality. *Learning Environments Research*, 25, 775–802. <https://doi.org/10.1007/s10984-021-09397-4>
- Wisniewski, B., Zierer, K., Dresel, M. & Daumiller, M. (2020). Obtaining students' perceptions of instructional quality – Two-level structure and measurement invariance. *Learning and Instruction*, 66, 101303. <https://doi.org/10.1016/j.learninstr uc.2020.101303>
- Wolff, C.E., Jarodzka, H. & Boshuizen, H. P. A. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, 66, 295–308.
- Wolff, C. 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>
- Wubbels, T., Brekelmans, M., Den Brok, P., Wijsman, L., Mainhard, T. & Tartwijk, J. van (2015). Teacher-student relationships and classroom management. In E. T. Emmer, & E. J. Sabornie (Eds.), *Handbook of classroom management* (2 ed., pp. 363–386). Routledge/Taylor & Francis.
- Wubbels, T., Brekelmans, M. & Hooyman, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching and Teacher Education*, 8, 47–58.
- Xu, L., Widjaja, W. & Ferguson, J. (2019). Seeing through the eyes of the teacher? Investigating primary school teachers' professional noticing through a video-based research methodology. *International Journal of Research and Method in Education*, 42(5), 470–484. <https://doi.org/10.1080/1743727X.2018.1499016>
- Zembylas, M. (2003). Emotions and teacher identity: A poststructural perspective. *Teachers and Teaching: Theory and Practice*, 9, 213–238.

Zierer, K., Wisniewski, B., Schatz, C., Weckend, D. & Helmke, A. (2019). Wie kann Feedback der Lernenden die Unterrichtsqualität verbessern? *Journal für LehrerInnenbildung*, 19(1), 26–40. <https://doi.org/10.25656/01:18108>

## 6 Beiträge der kumulativen Dissertation

### 6.1 Beitrag I. Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität

Jennifer Iglér<sup>1</sup>, Annika Ohle-Peters<sup>1</sup> und Nele McElvany<sup>1</sup>

<sup>1</sup> Institut für Schulentwicklungsforschung der Technischen Universität Dortmund

#### *Zusammenfassung*

Schülereinschätzungen von Unterrichtsqualität haben sich als prädiktiv für ihre Lernergebnisse erwiesen, jedoch deuten empirische Studien auf große Unterschiede in der Einschätzung desselben Unterrichts innerhalb von Klassen hin. Merkmale von Unterrichtsqualität können in drei Dimensionen unterteilt werden: kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima. Für die Identifizierung von Prädiktoren für divergierende Schülereinschätzungen des Unterrichts in derselben Klasse können Theorien zur Informationsverarbeitung herangezogen werden. Vor diesem Hintergrund untersuchte die vorliegende Studie die Varianz der Schülereinschätzungen und analysierte in Anlehnung an psychologische Befunde zur Informationsverarbeitung die Prädiktion von Einschätzungsdifferenzen durch kognitive Voraussetzungen (Vorwissen, kognitive Fähigkeiten), motivationale Orientierung (Schullust) und emotionale Bindungen (Nähe zur Lehrkraft) sowie durch demographische Merkmale (Geschlecht, Migrationshintergrund, sozialer Hintergrund). Die Stichprobe bestand aus 647 Schülerinnen und Schülern (50.8 % weiblich) aus 33 vierten Klassen, die direkt im Anschluss an eine Unterrichtseinheit zu ihrer Einschätzung der Unterrichtsqualität befragt wurden, welche anhand von drei Merkmalen mittels jeweils fünf Items erfasst wurde: herausfordernde Aufgaben, Störungen im Unterricht und Motivierung durch die Lehrkraft. Für die Überprüfung der Fragestellungen wurden Intraklassenkorrelationen und Mehrebenenregressionsmodelle unter Berücksichtigung der Clusterung der Daten eingesetzt. Die höchste Übereinstimmung innerhalb der Klassen zeigte sich für das Merkmal Störungen im Unterricht ( $ICC_1 = .26$ ), etwas geringer fielen die Übereinstimmungen bei den Merkmalen Motivierung durch die Lehrkraft ( $ICC_1 = .19$ ) und herausfordernde Aufgaben ( $ICC_1 = .14$ ) aus. Bei gleichzeitiger Berücksichtigung aller

Schülermerkmale in einem Modell waren für die abweichenden Einschätzungen im Bereich herausfordernde Aufgaben die Merkmale Vorwissen, kognitive Fähigkeiten, motivationale Orientierung und sozialer Hintergrund prädiktiv. Schülereinschätzungen von Störungen im Unterricht wurden vom sozialen Hintergrund vorhergesagt und Einschätzungen im Bereich der Motivierung von den kognitiven Fähigkeiten und der emotionalen Bindung. Das Geschlecht und der Migrationshintergrund wiesen keine bedeutsame Prädiktionskraft auf. Diese Befunde sind sowohl für die Unterrichtsforschung als auch für die pädagogische Praxis von großer Relevanz.

Schlagwörter: Schülereinschätzungen, Unterrichtsqualität, Wahrnehmung der Lernumgebung, Heterogenität, Grundschule

*Through the eyes of a primary school child – Individual predictors of students' ratings of teaching quality*

*Abstract*

Students' ratings of teaching quality, which are predictive for their learning outcomes, often vary substantially within classes. Criteria for teaching quality can be categorized by three dimensions: instructional support, classroom organization, and emotional support. Identifying predictors of the diverging ratings of teaching in the same class, theories of information processing can be considered. In this context, the present study examined the variance of students' ratings of teaching quality and analyzed cognitive (prior knowledge, cognitive ability), motivational (school pleasure) and emotional (relationship with teacher), as well as demographic (gender, migrant background, social background) student characteristics as predictors for perceptual variances. The sample consisted of 647 fourth graders (50.8 % female) from 33 classes, who were asked about their perceptions of teaching quality directly after an instructional unit. Students rated three criteria of teaching quality with five items each: cognitive demand of tasks, classroom-disruptions, and motivational support by teacher – thus covering aspects from all dimensions of teaching quality. For the examination of the questions and hypotheses, intraclass correlations and multilevel regression models were used considering the clustering of the data. Results from multilevel regression models showed that there were differences in the ratings of teaching quality, and that individual student characteristics predicted part of the variance. The highest intraclass correlation was found for the criterion classroom-disruptions ( $ICC_1 = .26$ ), while the correlation for the criteria

motivation by the teacher ( $ICC_1 = .19$ ) and challenging tasks ( $ICC_1 = .14$ ) were slightly lower. Considering simultaneously all students' characteristics in one model, prior knowledge, cognitive abilities, motivational orientation and social background were predictive for students' ratings of cognitive demand. Students' ratings of classroom disturbances were predicted by social background and students' ratings of motivational support by cognitive abilities and emotional attachment. Gender and migrant background had no significant predictive power for divergent students' ratings of teaching quality. These results are relevant both for instructional research and practice.

Keywords: students' ratings, teaching quality, perception of learning environment, heterogeneity, primary school

## **1 Einleitung**

Die Befragung von Schülerinnen und Schülern ist ein etabliertes Verfahren in der empirischen Bildungsforschung zur Erfassung der Unterrichtsqualität neben der Einschätzung von Lehrkräften oder externen Beobachterinnen und Beobachtern. Die prädiktive Bedeutung der Unterrichtseinschätzung durch Schülerinnen und Schüler für den Lernerfolg konnte bereits in mehreren Studien empirisch belegt werden (z. B. Kunter & Baumert, 2006). Viele dieser Untersuchungen zu Schülereinschätzungen über ihren Unterricht beziehen sich jedoch auf die Sekundarstufe. Für den Primarbereich, der sich durch andere Strukturen, wie beispielsweise die gemeinsame Beschulung aller Kinder im Gegensatz zu dem gegliederten System der Sekundarstufe und entwicklungsbedingt durch andere pädagogische Schwerpunkte wie unter anderem die intensivere Beziehung zur Lehrkraft auszeichnet, liegen bisher nur wenige empirische Untersuchungen zu Unterrichtseinschätzungen vor. Dabei konnte in den wenigen Studien, die sich auf die Primarstufe beziehen, gezeigt werden, dass auch in dieser Altersklasse die Schülereinschätzungen der Unterrichtsqualität prädiktiv für deren Lernerfolg sind (z. B. Fauth, Decristan, Rieser, Klieme & Büttner, 2014).

Obwohl alle Lernenden in einer Klasse dasselbe Lernangebot von ein und derselben Lehrperson erhalten, haben Untersuchungen eine große Varianz in den Schülereinschätzungen der Unterrichtsqualität innerhalb von Klassen aufgezeigt. Aus psychologischer Perspektive können divergierende Schülereinschätzungen aufgrund von individuellen Unterschieden in der Wahrnehmung und in der Interpretation von Schülerinnen und Schülern auftreten. Der Wahrnehmung und Interpretation werden zentrale Rollen für die Informationsverarbeitung und die Entscheidungsfindung zugesprochen, welche von weiteren kognitiven sowie

motivationalen und emotionalen Prozessen beeinflusst werden (Bless, Fiedler & Strack, 2004). Insbesondere wurden noch nicht hinreichend gesicherte Faktoren identifiziert, die die individuelle Einschätzung der Lernenden in Grundschulen vorhersagen und somit die Varianz der Schülerangaben zur Unterrichtsqualität innerhalb von Grundschulklassen bedingen. Daher werden in der vorliegenden Studie die Übereinstimmung von Schülereinschätzungen in der Grundschule bezüglich der Unterrichtsqualität untersucht und Prädiktoren für den Wahrnehmungsprozess und die daraus resultierende Einschätzung in den Blick genommen.

## **2 Theoretischer Hintergrund**

### **2.1 Basisdimensionen von Unterrichtsqualität**

Bei der Beschreibung qualitativ vollen Unterrichts wird zwischen der Oberflächen- und Tiefenstruktur unterschieden, wobei sich letztere als besonders relevant für den Lernerfolg von Schülerinnen und Schülern erwies (Pauli & Reusser, 2003). Unterrichtsqualität wurde und wird in theoretischen Modellen und empirischen Studien über verschiedene Merkmale operationalisiert. Im deutschsprachigen Raum haben sich drei Basisdimensionen von Unterrichtsqualität etabliert: *kognitive Aktivierung*, *Klassenführung* und *unterstützendes Unterrichtsklima* (vgl. Klieme & Rakoczy, 2008; Kunter et al., 2013; vergleiche international: *instructional support*, *classroom organization* und *emotional support*; Pianta & Hamre, 2009).

Die kognitive Aktivierung wird in diesem Zusammenhang als wichtigste Dimension für erfolgreiches domänenspezifisches Lernen beschrieben und zielt darauf ab, kognitive Prozesse sowie eine tiefere Informationsverarbeitung mittels anregender Problemstellungen zu aktivieren (Bransford, Brown & Cocking, 2000; Lipowsky et al., 2009). In diesem Kontext ist das Stellen von *herausfordernden Aufgaben* ein besonders wichtiges Merkmal, da dies zu einer intensiven Auseinandersetzung mit dem Lernstoff und nachhaltigem Lernen führt (Klieme & Rakoczy, 2008; Mayer, 2004).

Die Basisdimension Klassenführung nimmt eine besondere Rolle ein, da sie die Voraussetzung für ein effektives Lernen im Klassenraum ist. Der Unterricht soll möglichst so gesteuert werden, dass wenige Störungen auftreten und die Unterrichtszeit effektiv genutzt werden kann (Emmer & Stough, 2001). Dazu gehört ein breites Spektrum an Strategien für eine gelungene Klassenführung, die unter anderem auf einen *effektiven Umgang mit Störungen* zielt (Brophy, 1979). Demzufolge kann eine längere Lernzeit hervorgerufen werden, indem die

Lehrkraft eine ruhige Arbeitsatmosphäre schafft und Störungen im Unterricht verhindert (Kounin, 1970).

Die dritte Dimension, unterstützendes Unterrichtsklima, bezieht sich auf das Erkennen von Schülerbedürfnissen und der Beachtung von Schülerinteressen. Generell sollte allen am Schulleben beteiligten Personen, sowie auch ihren Äußerungen, eine wertschätzende Haltung entgegengebracht werden (Hamre & Pianta, 2010). Ferner ist es die Aufgabe von Lehrkräften das Erfahren von Autonomie, Kompetenz und sozialer Eingebundenheit seitens der Schülerinnen und Schüler zu fördern, welches zu einer höheren Wahrnehmung von Selbstbestimmung und damit einhergehend zu einer höheren Motivation führt (Ryan & Deci, 2000). Ein wichtiges Merkmal für das Erleben von Selbstbestimmung ist die *Motivierung durch die Lehrkraft*.

Unterrichtsqualitätsmerkmale werden in der Empirie über verschiedene Zugänge untersucht: Lehrkräfte, externe Beobachterinnen und Beobachter sowie Schülerinnen und Schüler. Generell ist nur eine geringe Übereinstimmung der Einschätzungen von Unterrichtsqualität von Schülerinnen und Schülern zu den Perspektiven der Lehrkräfte und externen Beobachterinnen und Beobachtern zu finden (vgl. Clausen, 2002; Fauth et al., 2014; Kunter & Baumert, 2006). Der höchste Zusammenhang ist zwischen den drei Perspektiven im Bereich der Klassenführung festzustellen. Hinsichtlich des unterstützenden Unterrichtsklimas weisen die Urteile von Lehrkräften und von Schülerinnen und Schülern einen Zusammenhang auf. Im Bereich der kognitiven Aktivierung ist die Befundlage nicht eindeutig (Clausen, 2002; Fauth et al., 2014; Kunter & Baumert, 2006). Lehrkräften wird genügend pädagogischdidaktische Expertise für das Urteilen über den Unterricht zugesprochen, jedoch können diese selbstwertdienlich verzerrt sein (Clausen, 2002). Vorteile der Perspektive externer Beobachterinnen und Beobachtern sind sowohl die hohe Objektivität als auch der geschulte methodisch-didaktische Blick. Allerdings haben externe Beobachterinnen und Beobachter nur einen kontextunabhängigen Zugang zu der Unterrichtssituation. Schülerinnen und Schüler verfügen zwar nicht über das methodisch-didaktische Wissen, weisen jedoch eine breite Basis an Erfahrungen über viele Unterrichtsstunden auf (Praetorius, 2014).

## **2.2 Schülerperspektive der Einschätzung von Unterrichtsqualität**

### **2.2.1 Schülereinschätzung als Maß für Unterrichtsqualität**

Die Betrachtung der Schülerperspektive ist aus dem Grund bedeutsam, da Schülerinnen und Schüler die Adressatinnen und Adressaten des Unterrichts sind (Helmke, Piskol, Pikowsky &

Wagner, 2009). Daher gelten Schülerurteile über den Unterricht als wichtiges Werkzeug für die Evaluation von Lehrkräften und haben somit auch eine praktische Relevanz für die Unterrichtsreflexion (Kultusministerkonferenz, 2004; Soebari & Aldridge, 2015). Aufgrund des geringen Erhebungsaufwandes, der niedrigen Kosten und der ökonomischen Durchführung, hat sich die Schülerbefragung außerdem als besonders effizient für empirische Untersuchungen erwiesen (Clausen, 2002; Helmke, 2015).

Ab einem Alter von etwa neun Jahren entwickeln Kinder eine Urteilsfähigkeit, die es ermöglicht, Situationen und Personen zutreffend einschätzen zu können (von Uslar, 2007). Empirische Untersuchungen haben gezeigt, dass Grundschülerinnen und Grundschüler die Fähigkeit für valide Bewertungen von Unterricht aufweisen (vgl. Ditton, 2002); so lassen sich zum Beispiel Schülerurteile in der Grundschule bezüglich der Unterrichtsqualität durchaus von Urteilen der Popularität der Lehrkraft trennen und stehen im Zusammenhang mit der Leistungsentwicklung von Schülerinnen und Schülern (vgl. Fauth, Decristan, Rieser, Klieme & Büttner, 2016; Fauth et al., 2014). Bei Untersuchungen über Schülereinschätzungen des aktuellen Unterrichts ist zu beachten, dass Kinder den Unterricht in Form von Schemata für Ereignisabläufe (Skripten) wahrnehmen, wobei die Qualität, Länge und Komplexität dieser mit dem Alter steigt (Kloss, 2013).

Bei der Erfassung von Schülerurteilen kann zwischen längerfristigen Aussagen über mehrere Unterrichtsstunden und Aussagen bezogen auf einzelne Unterrichtsstunden unterschieden werden. Nach Helmke & Schrader (2008) sollten qualitative Merkmale des Unterrichts durch längerfristige Beurteilungen erfasst werden, da sich die Qualität des Unterrichts in Routinen und stundenübergreifenden Praktiken ausdrückt und diese nicht in einer einzelnen Unterrichtsstunde beobachtet werden können. Allerdings setzt die längerfristige Befragung über mehrere Unterrichtsstunden die Fähigkeit von Schülerinnen und Schülern voraus, das Lehrkraftverhalten sowie Merkmale des Unterrichts zu abstrahieren (Kloss, 2013). Demzufolge kann besonders bei jüngeren Schülerinnen und Schülern die Befragung zum Unterricht oder zum Lehrkraftverhalten über einen längeren Beurteilungszeitraum zu Verzerrungen führen. Ein weiterer Nachteil längerfristiger Schülerbefragungen über mehrere Unterrichtsstunden ist, dass kommunikative Prozesse der Schülerinnen und Schüler untereinander zu einer Verfestigung der Wahrnehmungen führen und somit die Aussagen der Kinder beeinflussen. Zudem kann eine Variabilität der zu bewertenden Qualitätsmerkmale über die Unterrichtsstunden hinweg dazu führen, dass keine generalisierbaren oder übereinstimmenden Aussagen über den Unterricht und das

Lehrkraftverhalten von den Schülerinnen und Schülern getroffen werden können (vgl. für den Bereich kognitive Aktivierung: Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2014).

Ein weiteres Argument für die Erfassung von Schülereinschätzungen zu einzelnen Unterrichtsstunden in der Grundschule ist neben der besseren Erinnerungsfähigkeit von jüngeren Kindern bezüglich konkreter Unterrichtsstunden und der Schwierigkeit der Abstraktion von längerfristigen Ereignissen die höhere Übereinstimmung zwischen den Schülereinschätzungen im Vergleich zu stundenübergreifenden Urteilen (Helmke et al., 2009). Um potentielle Verzerrungseffekte zu reduzieren, sollten demzufolge Schülereinschätzungen in der Grundschule möglichst direkt im Anschluss an den zu beurteilenden Unterricht erfolgen und sich auf einzelne konkrete Unterrichtsstunden beziehen.

### **2.2.2 Divergierende Schülereinschätzungen**

Schülerurteile über den Unterricht setzen sich aus zwei Komponenten zusammen: dem Anteil der individuellen Wahrnehmung des Schülers oder der Schülerin und dem Anteil der geteilten Wahrnehmungen der Klasse (Lüdtke, Robitzsch, Trautwein & Kunter, 2009; Wagner et al., 2016). In früheren Studien wurden oftmals entweder Analysen auf Individualebene ohne Berücksichtigung der geteilten und nicht-geteilten Komponente durchgeführt oder der Fokus von Studien lag auf den geteilten (aggregierten) Schülerwahrnehmungen, um generelle Aussagen über den Unterricht treffen zu können (Göllner, Wagner, Eccles & Trautwein, 2018). Die Betrachtung des Anteils der individuellen Wahrnehmung kann von besonderem Interesse in Bezug auf die individuellen Lernbedürfnisse von Schülerinnen und Schülern und die Anpassung des Unterrichts sein.

Um Aussagen über die Zuverlässigkeit von aggregierten Schülereinschätzungen zu treffen, werden Intraklassenkorrelationen ( $ICC_1$ ) betrachtet. Diese lagen in einigen Studien zur Unterrichtsqualität zwischen  $ICC_{1(\text{unterstützendes Unterrichtsklima})} = .05$  und  $ICC_{1(\text{Klassenführung})} = .32$  (vgl. Lüdtke, Trautwein, Kunter & Baumert, 2006; Praetorius et al., 2017; Wagner et al., 2016). Demzufolge sind 5.0 Prozent bis 32.0 Prozent der Schülerangaben zur Unterrichtsqualität auf die Zugehörigkeit zu einer Schule bzw. zu einer Klasse zurückzuführen. Unterschiede zwischen den  $ICC_1$  der einzelnen Unterrichtsqualitätsdimensionen können aufgrund der Beobachtbarkeit der Kriterien (zum Beispiel Störungen im Unterricht vs. subjektivere Wahrnehmungen zur Schüler-Lehrer-Beziehung) auftreten. Bisher wurden häufig die Intraklassenkorrelationen in Bezug auf den naturwissenschaftlichen Unterricht und in der Sekundarstufe betrachtet (ebd.).

Zu divergierenden Schülereinschätzungen können unter anderem methodische Aspekte der Untersuchung, wie die Itemformulierung, führen. Bei der Itemformulierung sind besonders im Grundschulalter die Beobachtbarkeit der einzuschätzenden Merkmale, die Eingrenzung der Interpretationsspanne sowie die Wahrnehmungsperspektive (ich vs. wir) zu beachten (Piskol, 2008). Neben methodischen Aspekten der Untersuchung können Unterschiede in der Wahrnehmung zwischen Einschätzungen konzeptionell durch zwei Erklärungsstränge begründet werden: Zum einen können Unterschiede zwischen den Schülerurteilen durch verschiedene Wahrnehmungen, zum anderen aber auch aufgrund einer tatsächlich unterschiedlichen Behandlung durch die Lehrkraft zustande kommen. Besonders im Zuge der Differenzierung und des individualisierten Unterrichts passen Lehrkräfte ihr Lernangebot den Lernvoraussetzungen und -fortschritten der Schülerinnen und Schüler an. Dadurch sammeln Schülerinnen und Schüler individuelle Erfahrungen in Bezug auf den Unterricht und auf das Lehrkraftverhalten, was wiederum den Grad der Übereinstimmung der Schülereinschätzungen beeinflussen kann. Dies betrifft insbesondere nicht direkt beobachtbare Unterrichtsmerkmale, wie die Motivierung von Lehrkräften (Gärtner, 2010; Wenger, Lüdtke & Brunner, 2018). Aus pädagogisch-psychologischer Perspektive ist besonders die Betrachtung von individuellen Wahrnehmungsunterschieden als Grund für differierende Unterrichtsbeurteilungen relevant. Generell erfolgt die Einschätzung von Personen oder Situationen über einen Informationsverarbeitungsprozess, welcher drei Schritte umfasst (Bless et al., 2004; Wyer & Srull, 1989): Wahrnehmung des Reizes, Kategorisierung des Reizes und Urteilsbildung. In dem Prozess können interindividuelle Unterschiede aufgrund kognitiver, motivationaler und emotionaler Merkmale auftreten, die am Ende zu unterschiedlichen Einschätzungsergebnissen führen können (Bless et al., 2004). Da die Unterrichtseinschätzung von Schülerinnen und Schülern auch auf deren Wahrnehmung des unterrichtlichen Geschehens basiert, kann angenommen werden, dass die Mechanismen des allgemein formulierten Informationsverarbeitungsprozesses auch auf die Wahrnehmung und die später daraus resultierende Einschätzung von Unterricht übertragbar sind.

### **2.2.3 Prädiktoren für divergierende Schülereinschätzungen**

Zur Identifikation von Prädiktoren der divergierenden Einschätzungen von Unterricht bei gleichem Lernangebot in einer Klasse kann, wie bereits erläutert, auf Theorien zur Informationsverarbeitung zurückgegriffen werden, bei denen die individuellen Merkmale als

Einflussfaktoren auf die Einschätzung von Personen und Situationen in drei Bereiche eingeteilt werden können.

### **Kognitive Voraussetzungen (Bereich I)**

Auch wenn der Stimulus einer Situation konstant ist, kann eine hohe Varianz in dem *Vorwissen* der Personen, mit welchem sie dem Reizereignis begegnen, auftreten. Dementsprechend können Personen aufgrund von unterschiedlichem Vorwissen die gleichen Reize unterschiedlich kognitiv verarbeiten und verschieden auf diese reagieren (Bless et al., 2004; Walther, Fiedler & Nickel, 2003). Unumstritten ist, dass auch die *kognitiven Fähigkeiten* mit der Wahrnehmung und Einschätzung zusammenhängen. Die kognitiven Fähigkeiten können beispielsweise auf das Gedächtnis von sozialen Situationen wirken (Roebbers & Schneider, 2001). Bisherige Studien in der Unterrichtsforschung zeigten die Bedeutung des Vorwissens von Schülerinnen und Schülern für die Beurteilung des erlebten Unterrichts auf Individualebene. Der Unterricht wurde von Schülerinnen und Schülern positiver beurteilt, wenn diese ein höheres Vorwissen (häufig gemessen durch die letzte Fachnote) aufwiesen (Ditton, 2002; Levy, den Brok, Wubbels & Brekelmans, 2003; Seidel, 2006). So nahmen zum Beispiel Grundschülerinnen und Grundschüler mit einem geringeren Vorwissen den kognitiven Anspruch höher wahr (Fauth et al., 2016). Dies konnte auch für die kognitiven Fähigkeiten gezeigt werden. Schülerinnen und Schüler mit höheren Leistungszuwächsen beurteilten Unterrichtsmerkmale in der Sekundarstufe positiver als Schülerinnen und Schüler mit niedrigeren Leistungszuwächsen (Wagner, 2008).

### **Motivationale Orientierung und emotionale Bindung (Bereich II)**

Für die Einschätzung von Situationen und Personen kann ebenfalls die *motivationale Orientierung* der urteilenden Person eine Rolle spielen. So konnte in Experimenten gezeigt werden, dass innere Motivation mit der Wahrnehmung von Situationen zusammenhängt (Balcetis & Dunning, 2006; Bruner & Goodman, 1947). Die *emotionale Bindung* wirkt auf die Verarbeitung kognitiver Inhalte (Bower, 1981). Beispielsweise kann das Sympathieempfinden des Gegenübers die Personenwahrnehmung und -einschätzung lenken und sich dementsprechend auch auf die Kommunikation auswirken (Herkner, 1991). Nach Erkenntnissen der Unterrichtsforschung konnten Zusammenhänge zwischen der motivationalen Orientierung und den Beurteilungen von Unterricht, speziell bezüglich der Dimensionen *Klassenführung*, sowie *unterstützendes Unterrichtsklima*, gefunden werden (z. B. Ditton, 2002; Klieme & Rakoczy, 2003; Kloss, 2013; Seidel, 2006). Zum Beispiel schätzten

Lernende mit einer höheren Ausprägung bezüglich der motivationalen Orientierung (z. B. Selbstkonzept oder Interesse für das Fach Deutsch) den Unterricht in der Sekundarstufe systematisch als unterstützender ein (Seidel, Jurik, Häusler & Stubben, 2016; Schenke, Ruzek, Lam, Karabenick & Eccles, 2018). Ein weiterer Zusammenhang konnte zwischen den Einschätzungen von Unterrichtsqualität und der Nähe der Schülerinnen und Schüler zur Lehrkraft gefunden werden, welche ein positives Urteil über den Unterricht sowohl in der Grundschule als auch in der Sekundarstufe hervorrufen kann (Kloss, 2013; Kunter & Baumert, 2006). Unter dem Begriff *Nähe zur Lehrkraft* sind die Unterstützung seitens der Lehrkräfte sowie die positive Schüler-Lehrer-Beziehung zu verstehen. Speziell zwischen der Nähe zur Lehrkraft und der wahrgenommenen Unterstützung konnte ein hoher Zusammenhang nachgewiesen werden (Fauth et al., 2016; Madill, Gest & Rodkin, 2014).

### **Demographische Merkmale (Bereich III)**

Neben den zuvor genannten Faktoren lenkt auch das *Geschlecht* die Informationsverarbeitung (Cahill, 2006). Es konnten Geschlechterdifferenzen in kognitiven Prozessen, z. B. im Wahrnehmen und Erinnern von Situationen, festgestellt werden (Barkley & Gabriel, 2007). So wurde ein Vorteil für Frauen im Bereich des verbalen episodischen Gedächtnisses und in der Verarbeitungsgeschwindigkeit gefunden (Lepach, Reimers, Pauls, Petermann & Daseking, 2015). Zusätzlich werden Wahrnehmungs- und Denkprozesse von den Erfahrungen aufgrund der eigenen Kultur und des *Migrationshintergrunds* gelenkt (Nisbett, Peng, Choi & Norenzayan, 2001). Insbesondere werden Differenzen in der Wahrnehmung von informativen Teilen einer Szene auf unterschiedliche kulturelle Erfahrungen zurückgeführt (Nisbett & Miyamoto, 2005; Segall, Campbell & Herskovit, 1966). Die Wahrnehmung sozialer Situationen ist zudem von sozial erlernten Motiven (z. B. Erwartungen, Einstellungen oder Emotionen) abhängig (Martin & Wawrinowski, 1993), welche in der Kindheit von Eltern und der Lernumgebung vermittelt werden. Aufgrund von existierenden unterschiedlichen *sozialen Hintergründen* werden schon Kinder mit Unterschieden in sozialen Interaktionen und Erfahrungen konfrontiert, welche zu Differenzen in der Informationsverarbeitung führen können (Ahnert, 2013; Bruner & Goodman, 1947).

Auch in der Unterrichtsforschung konnte die prädiktive Kraft des *Geschlechts* für die Schülereinschätzung von Unterrichtsqualität gezeigt werden (Gentry, Gable & Rizza, 2002; Wagner, 2008). Bezüglich des Deutsch- sowie Englischunterrichts gaben Schülerinnen positivere Urteile über den Unterricht in der Sekundarstufe an als Schüler (Wagner, 2008). Im Gegensatz dazu konnte für das Fach Mathematik die Tendenz ermittelt werden, dass

männliche Lernende positivere Einschätzungen über die Qualität des Unterrichts in der Sekundarstufe berichteten (Ditton, 2002). Hinsichtlich der Qualität der Klassenführung schätzten Schülerinnen Lehrkräfte in der Sekundarstufe dominanter und kooperativer ein als männliche Lernende sowohl in der Grundschule als auch in der Sekundarstufe (Goh & Fraser, 1995; Wubbels & Levy, 1993). Ebenfalls wies der *Migrationshintergrund* eine Bedeutung für die Einschätzungen von interpersonellem Verhalten und Charakteristiken von Lehrkräften auf. Die Strenge einer Lehrkraft in Sekundarstufen wurde unterschiedlich von Schülerinnen und Schülern je nach Herkunft eingeschätzt (Den Brok, Levy, Wubbels & Rodriguez, 2003; Levy et al., 2003). Unabhängig von ihrer Aufenthaltsdauer nahmen Schülerinnen und Schüler, die zu Hause eine andere Sprache sprachen, die Lehrkräfte weniger streng und unterstützender wahr als Schülerinnen und Schüler, die zu Hause die Landessprache sprachen (Den Brok et al., 2003).

Darüber hinaus zeigten empirische Untersuchungen Zusammenhänge zwischen dem *sozialen Hintergrund* und dem wahrgenommenen Lehrkraftverhalten in der Sekundarstufe. Demnach nahmen Schülerinnen und Schüler mit einem höheren sozioökonomischen Status die Erwartungen und den Anspruch der Lehrkräfte sowie die Wertschätzungen, die ihnen entgegengebracht wurden, positiver wahr als Lernende mit einem niedrigeren sozioökonomischen Status. Hingegen schätzten letztere aber die Feedbackkultur von Lehrkräften besser ein (Matthews, 2008).

Fasst man die Befundlage zusammen, ist festzustellen, dass es teilweise nur niedrige Übereinstimmung zwischen den Schülereinschätzungen von Unterricht innerhalb von Klassen gibt (Clausen, 2002; Gruehn, 2000; Lüdtke et al., 2006). Theoretische Modelle wie auch empirische Befunde geben Hinweise auf mögliche individuelle Merkmale, die zu divergierenden Einschätzungen beitragen.

### **3 Forschungsfragen und Hypothesen**

In Anlehnung an das Angebots-Nutzungs-Modell müssen Schülerinnen und Schüler, bevor sie Lernangebote nutzen können, den Unterricht zunächst wahrnehmen und interpretieren (Helmke, 2015). Dieses Modell setzt die Merkmale von Unterrichtsqualität mit Eigenschaften von Lehrpersonen und den Wirkungsweisen und Zielkriterien des Unterrichts in Verbindung. Jedoch werden in diesem Modell noch keine Verbindungen zwischen individuellen Schülermerkmalen und der Wahrnehmung sowie Interpretation der unterrichtlichen Angebote berücksichtigt. In der Annahme, dass die oben theoretisch sowie auch empirisch herausgearbeiteten Faktoren auch für

die Schülereinschätzungen des Unterrichts in der Grundschule relevant sind und bestimmte Merkmale eine besonders wichtige Bedeutung für einzelne Dimensionen der Unterrichtsqualität aufweisen, werden folgende Forschungsfragen und Hypothesen formuliert:

F1 Wie hoch ist die Übereinstimmung der Einschätzungen der Schülerinnen und Schüler bezüglich der Unterrichtsqualitätsmerkmale herausfordernde Aufgaben, Störungen im Unterricht und Motivierung durch die Lehrkraft in der Grundschule innerhalb von Klassen und lassen sich Unterschiede in den Übereinstimmungen der Schülereinschätzungen zwischen den Unterrichtsqualitätsmerkmalen finden? Im Hinblick auf Untersuchungen zur Übereinstimmung der Schülereinschätzungen innerhalb von Sekundarklassen, wird eine ähnlich geringe Übereinstimmung der Aussagen von Grundschülerinnen und Grundschülern angenommen. Bezüglich der unterschiedlich hohen Übereinstimmung in den einzelnen Unterrichtsqualitätsdimensionen wird sich an empirischen Befunden orientiert.

H1.1 Es wird nur eine geringe mittlere Übereinstimmung der Schülereinschätzungen innerhalb der Grundschulklassen erwartet.

H1.2 Die größte mittlere Übereinstimmung der Schülereinschätzungen wird bezüglich des Merkmals Störungen im Unterricht angenommen.

H1.3 Die niedrigste mittlere Übereinstimmung der Schülereinschätzungen wird bezüglich des Merkmals Motivierung durch die Lehrkraft erwartet.

F2 Stehen die Unterschiede in den Einschätzungen von den Merkmalen herausfordernde Aufgaben, Störungen im Unterricht und Motivierung durch die Lehrkraft innerhalb von Schulklassen mit individuellen Schülermerkmalen (Vorwissen, kognitive Fähigkeiten, motivationale Orientierung, emotionale Bindung, Geschlecht, Migrationshintergrund und sozialer Hintergrund) in systematischem Zusammenhang? Für die einzelnen Unterrichtsqualitätsmerkmale wird untersucht, ob ein Zusammenhang zu individuellen Schülermerkmalen besteht, welche zuvor aus der Theorie und Empirie hergeleitet wurden. Dabei können für die drei Unterrichtsqualitätsmerkmale verschiedene Prädiktoren auf Schülerseite angenommen werden:

H2.1 Die Einschätzung des domänenspezifischen Unterrichtmerkmals herausfordernde Aufgaben wird insbesondere durch die Schülermerkmale Vorwissen und kognitive Fähigkeiten mitbedingt.

H2.2 Die Einschätzung des Unterrichtmerkmals Störungen im Unterricht wird insbesondere durch die Schülermerkmale motivationale Orientierung, emotionale Bindung, Geschlecht, Migrationshintergrund und sozialer Hintergrund mitbedingt.

H2.3 Die Einschätzung des Unterrichtsmerkmals Motivierung durch die Lehrkraft wird insbesondere durch die Schülermerkmale motivationale Orientierung und emotionale Bindung mitbedingt.

## **4 Methodisches Vorgehen**

### **4.1 Design und Stichprobe**

Die Datengrundlage stammt aus dem DFG geförderten Projekt „Entwicklung und Überprüfung von Kompetenzmodellen zur integrativen Verarbeitung von Texten und Bildern“ (vgl. McElvany et al., 2012; Ohle et al., 2017; Schnotz et al., 2010). Im April 2012 (Rheinland-Pfalz) und 2013 (Nordrhein-Westfalen) wurde in jeder teilnehmenden vierten Klasse eine Unterrichtseinheit von drei Schulstunden mit dem Schwerpunkt Lesen von Texten mit instruktionalen Bildern zur Thematik „Südamerika“ durchgeführt. Hierfür wurden den Lehrkräften Arbeitsmaterialien zur Verfügung gestellt, die aus Texten mit instruktionalen Abbildungen bestanden und die die Möglichkeit boten, das Lesen dieses Materials zu üben. Zum konkreten Einsatz der Materialien und der methodischen Planung wurden keine Vorgaben gemacht. Direkt im Anschluss an die Unterrichtseinheit wurden die Schülerinnen und Schüler mittels eines Fragebogens nach ihrer Einschätzung des gerade erlebten Unterrichts befragt. Variablen zur kognitiven Fähigkeit, Vorwissen (hier: Leseverständnis), motivationalen Orientierung und emotionalen Bindung sowie demographische Angaben der Kinder wurden zu einem früheren Messzeitpunkt im November des jeweiligen Schuljahres erfasst. Insgesamt umfasste die Stichprobe  $N = 647$  Schülerinnen und Schüler der vierten Klassenstufe aus  $N = 33$  Klassen ( $M_{Alter} = 9.42$ ,  $SD_{Alter} = 0.58$ ; 50.8 % weiblich). 33.5 Prozent wiesen einen Migrationshintergrund auf.

### **4.2 Instrumente**

Für die Erfassung der Unterrichtsqualitätseinschätzung wurden drei Skalen, die den drei Basisdimensionen der Unterrichtsqualität zugeordnet werden können, in Anlehnung an Instrumente aus der COACTIV Studie (Baumert et al., 2009) entwickelt und eingesetzt. Die Items wurden in der ersten Projektphase für die Sekundarstufe hinsichtlich eines Unterrichts zu Bild-Text-Integration umformuliert, pilotiert und eingesetzt, bevor sie in einer weiteren Projektphase in der Grundschule zum Einsatz kamen (Anhang A). Die Einschätzungen wurden auf einer vierstufigen Skala ( $1 = \text{trifft nicht zu} - 4 = \text{trifft zu}$ ) gegeben. Das Unterrichtsmerkmal

herausfordernde Aufgaben wurde über fünf Items abgefragt ( $M = 2.06$ ,  $SD = 0.78$ ;  $\alpha = .84$ ,  $\omega = .85$ ; Beispielitem „Wir Schüler mussten ganz schön denken, bis wir den Text und das Bild verstanden haben.“). Für die Erfassung der Störungen im Unterricht wurden fünf Items eingesetzt ( $M = 1.75$ ,  $SD = 0.71$ ;  $\alpha = .84$ ,  $\omega = .85$ ; Beispielitem „Im Unterricht zu Südamerika wurde der Unterricht oft gestört.“). Die Skala zur Motivierung durch die Lehrkraft wurde über fünf Items operationalisiert ( $M = 3.05$ ,  $SD = 0.73$ ;  $\alpha = .85$ ,  $\omega = .86$ ; Beispielitem „Im Unterricht gerade hat unser/ e Lehrer/ in das Lesen der Texte mit Bildern sehr spannend gestaltet.“).

Um das Vorwissen im Bereich Leseverständnis zu erfassen, wurde eine im Rahmen des Projekts adaptierte Version des Lesetests aus der Lernausgangslagenuntersuchung (LAU; Behörde für Schule und Berufsbildung, 2011) eingesetzt. Hierbei wurde das Textverständnis der Schülerinnen und Schüler mittels vier kurzer Lesetexte und insgesamt 23 Verständnisfragen getestet. Die Reliabilität des Tests lag bei  $\alpha = .72$  und  $\omega = .72$ . Durchschnittlich beantworteten die Schülerinnen und Schüler  $M = 10.72$  ( $SD = 3.96$ ) Fragen richtig. Die kognitiven Fähigkeiten wurden mit dem Subtest zur Figural-räumlichen Fähigkeit des revidierten Kognitiven Fähigkeitstests für die 4.–12. Klasse erfasst (KFT 4 – 12 + R; vgl. Heller & Perleth, 2000). In dieser Untersuchung bearbeiteten Schülerinnen und Schüler in der Testversion A ( $\alpha = .75$  und  $\omega = .75$ ) bzw. B ( $\alpha = .76$  und  $\omega = .76$ ) 20 Aufgaben zu Wortklassifikationen. Im Mittel erreichten die Schülerinnen und Schüler einen T-Wert von  $M = 43.61$  ( $SD = 7.77$ ). Das Merkmal motivationale Orientierung wurde mit einer Skala Schullust abgebildet (drei Items;  $M = 3.73$ ,  $SD = 0.90$ ;  $\alpha = .65$ ;  $\omega = .66$ ; Beispielitem „Es gibt in der Schule viele Dinge, die mir wirklich Spaß machen.“; adaptiert aus Quellenberg, 2009). Die Skala Nähe zur Lehrkraft, als Teilbereich der emotionalen Bindung, wurde über fünf Items erhoben ( $M = 3.40$ ,  $SD = 0.47$ ;  $\alpha = .67$ ;  $\omega = .68$ ; Beispielitem „Insgesamt komme ich mit meinen Lehrern/ Lehrerinnen gut aus.“; adaptiert aus Quellenberg, 2009). Die Items beider Skalen wurden über eine 4-stufige Likert-Skala erfasst (1 = *trifft nicht zu*–4 = *trifft zu*). Die Geschlechtszugehörigkeit wurde über die Frage „Bist du ein Mädchen oder Junge?“ ermittelt (0 = Mädchen, 1 = Junge). Die Operationalisierung des Migrationshintergrunds erfolgte über die in der Familie gesprochene Sprache. 66.5 Prozent der Schülerinnen und Schüler sprachen immer Deutsch zu Hause, 19.4 Prozent sprachen fast *immer Deutsch*, 12.5 Prozent sprachen *manchmal Deutsch* und 1.6 Prozent sprachen *nie Deutsch* zu Hause. Ein Migrationshintergrund wurde definiert, wenn die Schülerin oder der Schüler auch oder nur eine andere Sprache als Deutsch zu Hause sprach. Für die Differenzierung wurde eine

Dummy-Variable gebildet (0 = *nur Deutsch*; 1 = *Deutsch und eine andere Sprache oder nur eine andere Sprache*; adaptiert aus Bos et al., 2010). Als Indikator für das Merkmal sozialer Hintergrund wurde die Anzahl der zu Hause verfügbaren Bücher erfragt (Beispielitem „Wie viele Bücher hast du zu Hause?“; adaptiert aus Bos et al., 2010). Für die Beantwortung der Frage konnte zwischen sieben bildlich unterstützten Antwortalternativen 1 = keine bis 7 = mehr als 500 gewählt werden (*keine Bücher* = 0.7 %, *1–10 Bücher* = 6.6 %, *11–50 Bücher* = 16.7 %, *51–100 Bücher* = 21.0 %, *101–250 Bücher* = 25.1 %, *251–500 Bücher* = 15.8 %, *mehr als 500 Bücher* = 14.2 %). Die Interkorrelationen zwischen den Skalen der Schülerwahrnehmung und den individuellen Merkmalen lagen im niedrigen bis mittleren Bereich (Tabelle 1).

Tabelle 1

*Interkorrelation zwischen den Schülereinschätzungen der Unterrichtsqualität und den Schülermerkmalen auf Within-Level*

Variablen	Interkorrelationen									
	2	3	4	5	6	7	8	9	10	
1 Herausfordernde Aufgaben	.32**	-.06	-.30**	-.29**	-.18**	-.11 <sup>+</sup>	-.07	.07	-.28**	
2 Störungen im Unterricht		-.34**	-.11 <sup>+</sup>	-.13*	-.13*	-.15*	-.14*	.11 <sup>+</sup>	-.17**	
3 Motivierung durch die Lehrkraft			.00	-.05	.22**	.27**	.08	-.09*	.03	
4 Vorwissen				.49**	.07	.17**	.04	-.21**	.33**	
5 Kognitive Fähigkeiten					.07	.10 <sup>+</sup>	.09*	-.15**	.32**	
6 Motivationale Orientierung						.55**	.23**	-.06	.13*	
7 Emotionale Bindung							.18**	-.15**	.15*	
8 Geschlecht								-.05	.01	
9 Migrationshintergrund									-.16**	
10 Sozialer Hintergrund									-	

*Anmerkungen.* CFI = .96, RMSEA = .03, SRMR<sub>Within</sub> = .04, SRMR<sub>Between</sub> = .14,  $\chi^2 = 615.39$ ,  $df = 427$ , AIC = 46245.46, BIC = 46813.44; <sup>+</sup>  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$

### 4.3 Statistische Methoden

Die Analysen zur Skalenreliabilität sowie den Deskriptiva wurden mit Hilfe des Programms SPSS 24 durchgeführt. Für die Überprüfung der Fragestellungen und Hypothesen wurden Intraklassenkorrelationen und Mehrebenenregressionsmodelle mit Mplus Version 7.11 (Muthén & Muthén, 1998–2015) gerechnet. Die Intraklassenkorrelation wurde berechnet, um die Varianzaufklärung der Konstrukte auf Klassenebene zu prüfen (H1.1). Verglichen wurden die Intraklassenkorrelationen anhand der Berechnung des Unterschieds einer Korrelation von einem festen Wert (Eid, Gollwitzer & Schmitt, 2011) (H1.2– H1.3). Für die Überprüfung der Zusammenhänge von Schülermerkmalen für die Einschätzungen von Unterrichtsqualität wurden Mehrebenenregressionsmodelle spezifiziert, in welchen die Schülereinschätzungen der Unterrichtsmerkmale auf Within-Level die abhängigen Variablen und die individuellen Schülermerkmale auf Within-Level die unabhängigen Variablen bildeten. Zusätzlich wurden auf Between-Level die Schülereinschätzungen der Unterrichtsmerkmale modelliert (H2.1– H2.3). In den Analysen wurden Interkorrelationen zwischen den unabhängigen Variablen sowie zwischen den abhängigen Variablen auf beiden Ebenen zugelassen. In den Modellen wurden die intervallskalierten Variablen (herausfordernde Aufgaben, Störungen im Unterricht, Motivierung durch die Lehrkraft, motivationale Orientierung und emotionale Bindung) latent modelliert. Die Faktorladungen der einzelnen Skalen wurden für die Bestimmung der Korrelationen und in den Mehrebenenanalysen auf die Werte im Messmodell fixiert. Die Clusterung der Daten wurde über den Befehl *cluster = id\_class* berücksichtigt. Für fehlende Werte wurde die Funktion der multiplen Imputation in SPSS 24 gewählt. Für die Analysen in Mplus wurden unter Berücksichtigung des Befehls *type = imputation* die aus der Imputation entstandenen fünf Datensätze berücksichtigt. Die gerichteten Zusammenhänge der sieben Schülermerkmale zu den drei eingeschätzten Unterrichtsqualitätsmerkmalen wurden zunächst für jedes Schülermerkmal einzeln untersucht. Anschließend wurden alle Schülermerkmale im Gesamtmodell als Prädiktoren auf die eingeschätzten Unterrichtsmerkmale simultan aufgenommen, wobei Interkorrelationen zwischen den Schülermerkmalen zugelassen wurden.

## 5 Ergebnisse

### 5.1 Übereinstimmung der Schülerwahrnehmungen

Bezüglich des Unterrichtsmerkmals Störungen im Unterricht konnte die höchste Übereinstimmung gefunden werden ( $ICC_1 = .26$ ). Die Übereinstimmung der Schülereinschätzung hinsichtlich des Merkmals Motivierung der Lehrkraft wies einen Wert von  $ICC_1 = .19$  auf. Die niedrigste Übereinstimmung zeigte sich in den Einschätzungen der herausfordernden Aufgaben ( $ICC_1 = .14$ ). Durch Vergleiche der ICCs konnten signifikante Unterschiede zwischen den Übereinstimmungen in den Bereichen herausfordernde Aufgaben und Störungen im Unterricht ( $z = -3.18; p = .001$ ) sowie Störungen im Unterricht und Motivierung durch die Lehrkraft ( $z = 1.87; p = .031$ ) festgestellt werden. Ein auf 10 %-Niveau signifikanter Unterschied zeigte sich zwischen den Intraklassenkorrelationen in den Bereichen herausfordernde Aufgaben und Motivierung durch die Lehrkraft ( $z = -1.31; p = .095$ ). Die Resultate stützen die aus der Theorie abgeleiteten Hypothesen der eher geringen mittleren Übereinstimmung der Schülereinschätzungen (H 1.1) und der größten mittleren Übereinstimmung bezüglich der Störungen im Unterricht (H1.2). Die Hypothese 1.3, dass die niedrigste Übereinstimmung bezüglich der Motivierung zu finden ist, kann nicht empirisch unterstützt werden. Auf Within-Level sind bei dem Merkmal herausfordernde Aufgaben 12.9 % Varianz, bei Störungen im Unterricht 12.4 % Varianz und bei Motivierung durch die Lehrkraft 16.8 % Varianz aufzuklären.

### 5.2 Prädiktoren für die Schülereinschätzungen von Unterrichtsqualität

Für die Untersuchung der zweiten Fragestellung, welche individuellen Schülermerkmale die Unterschiede in der Einschätzung von Unterrichtsqualität vorhersagen, wurde die Prädiktion der einzelnen Schülermerkmale auf die eingeschätzten Unterrichtsqualitätsmerkmale untersucht (Tabelle 2).

Tabelle 2

*Individuelle Schülermerkmale als Prädiktoren für die Einschätzung von Unterrichtsqualität ( $\beta$ -Koeffizienten und Standardfehlern)*

Prädiktoren	Einschätzungen von Merkmalen der Unterrichtsqualität					
	Herausfordernde Aufgaben		Störungen im Unterricht		Motivierung durch die Lehrkraft	
	$\beta$	SE	$\beta$	SE	$\beta$	SE
Vorwissen	-.29**	0.05	-.10 <sup>+</sup>	0.06	.00	0.05
$R^2$	.09	0.03	.01	0.01	.00	0.00
<i>Anmerkungen.</i> CFI = .98, RMSEA = .02, SRMR <sub>Within</sub> = .03, SRMR <sub>Between</sub> = .14, $\chi^2 = 282.71$ , $df = 210$ , AIC = 22364.24, BIC = 22632.58; <sup>+</sup> $p < .10$ , ** $p < .01$						
Kognitive Fähigkeiten	-.29**	0.05	-.13*	0.05	-.05	0.04
$R^2$	.08	0.03	.02	0.01	.00	0.00
<i>Anmerkungen.</i> CFI = .98, RMSEA = .03, SRMR <sub>Within</sub> = .04, SRMR <sub>Between</sub> = .14, $\chi^2 = 303.51$ , $df = 210$ , AIC = 22361.10, BIC = 22629.44; * $p < .05$ , ** $p < .01$						
Motivationale Orientierung	-.19**	0.05	-.13*	0.06	.21**	0.06
$R^2$	.04	0.02	.02	0.02	.05	0.02
<i>Anmerkungen.</i> CFI = .98, RMSEA = .02, SRMR <sub>Within</sub> = .04, SRMR <sub>Between</sub> = .13, $\chi^2 = 315.23$ , $df = 242$ , AIC = 28131.95, BIC = 28431.59; * $p < .05$ , ** $p < .01$						
Emotionale Bindung	-.10	0.06	-.14*	0.06	.27**	0.05
$R^2$	.01	0.01	.02	0.02	.07	0.03
<i>Anmerkungen.</i> CFI = .98, RMSEA = .02, SRMR <sub>Within</sub> = .04, SRMR <sub>Between</sub> = .14, $\chi^2 = 369.89$ , $df = 279$ , AIC = 28954.00, BIC = 29271.54; * $p < .05$ , ** $p < .01$						
Geschlecht	-.07	0.06	-.14*	0.07	.07	0.06
$R^2$	.00	0.01	.02	0.02	.01	0.01
<i>Anmerkungen.</i> CFI = .97, RMSEA = .03, SRMR <sub>Within</sub> = .04, SRMR <sub>Between</sub> = .13, $\chi^2 = 295.56$ , $df = 210$ , AIC = 19309.68, BIC = 19569.25; * $p < .05$						
Migrationshintergrund	.05	0.06	.11 <sup>+</sup>	0.06	-.09*	0.04
$R^2$	.00	0.01	.01	0.01	.01	0.01
<i>Anmerkungen.</i> CFI = .97, RMSEA = .03, SRMR <sub>Within</sub> = .04, SRMR <sub>Between</sub> = .13, $\chi^2 = 296.11$ , $df = 210$ , AIC = 19406.58, BIC = 19666.47; <sup>+</sup> $p < .10$ , * $p < .05$						
Sozialer Hintergrund	-.27**	0.05	-.16**	0.06	.02	0.05
$R^2$	.07	0.03	.03	0.02	.00	0.00

---

*Anmerkungen.* CFI = .98, RMSEA = .02, SRMR<sub>Within</sub> = .03, SRMR<sub>Between</sub> = .14,  $\chi^2 = 286.08$ ,  $df = 210$ , AIC = 22367.15, BIC = 22635.50; \*\*  $p < .01$

---

In den Mehrebenenregressionsmodellen für jedes zu untersuchende Schülermerkmal zeigten die Schülermerkmale Vorwissen, kognitive Fähigkeiten, motivationale Orientierung und sozialer Hintergrund für das eingeschätzte Unterrichtsmerkmal herausfordernde Aufgaben eine prädiktive Kraft. Schülerinnen und Schüler mit einem höheren Vorwissen, einer höheren kognitiven Fähigkeit, einer höheren motivationalen Orientierung sowie einem höheren sozialen Familienhintergrund schätzten im Vergleich zu ihren Mitschülerinnen und Mitschülern in der gleichen Klasse die Aufgaben als weniger herausfordernd ein. Dieses Resultat unterstützt empirisch nur teilweise die Hypothese 2.1, da neben Vorwissen und kognitiven Fähigkeiten auch motivationale Orientierung und sozialer Familienhintergrund eine Prädiktionskraft aufwiesen (vgl. auch deren Relevanz im folgenden Gesamtmodell). Die Einschätzung zu den Störungen im Unterricht wurde durch die Schülermerkmale Vorwissen, kognitive Fähigkeiten, motivationale Orientierung, emotionale Bindung, Geschlecht, Migrationshintergrund sowie sozialer Hintergrund vorhergesagt. Demzufolge nahmen Schülerinnen und Schüler mit niedrigerem Vorwissen, niedrigeren kognitiven Fähigkeiten, niedrigerer motivationaler Orientierung und emotionaler Bindung, einem Migrationshintergrund, einem niedrigeren sozialen Hintergrund und männlichen Geschlechts mehr Störungen im Unterricht wahr als andere Schülerinnen und Schüler in ihrer Klasse. Die Ergebnisse unterstützen nur teilweise die Annahmen der Hypothese 2.2, da sich zwar alle angenommenen Merkmale als signifikante Prädiktoren erwiesen, zusätzlich aber auch das Vorwissen und die kognitiven Fähigkeiten relevant waren. Für die eingeschätzte Motivierung der Lehrkraft erwiesen sich die motivationale Orientierung und emotionale Bindung sowie der Migrationshintergrund als prädiktiv. Schülerinnen und Schüler mit einer höheren motivationalen Orientierung und emotionalen Bindung sowie ohne Migrationshintergrund schätzten die Motivierung durch die Lehrkraft höher ein als andere Schülerinnen und Schüler in ihrer Klasse. Die Hypothese 2.3 kann als teilweise unterstützt angesehen werden, da zusätzlich das Merkmal Migrationshintergrund prädiktiv war.

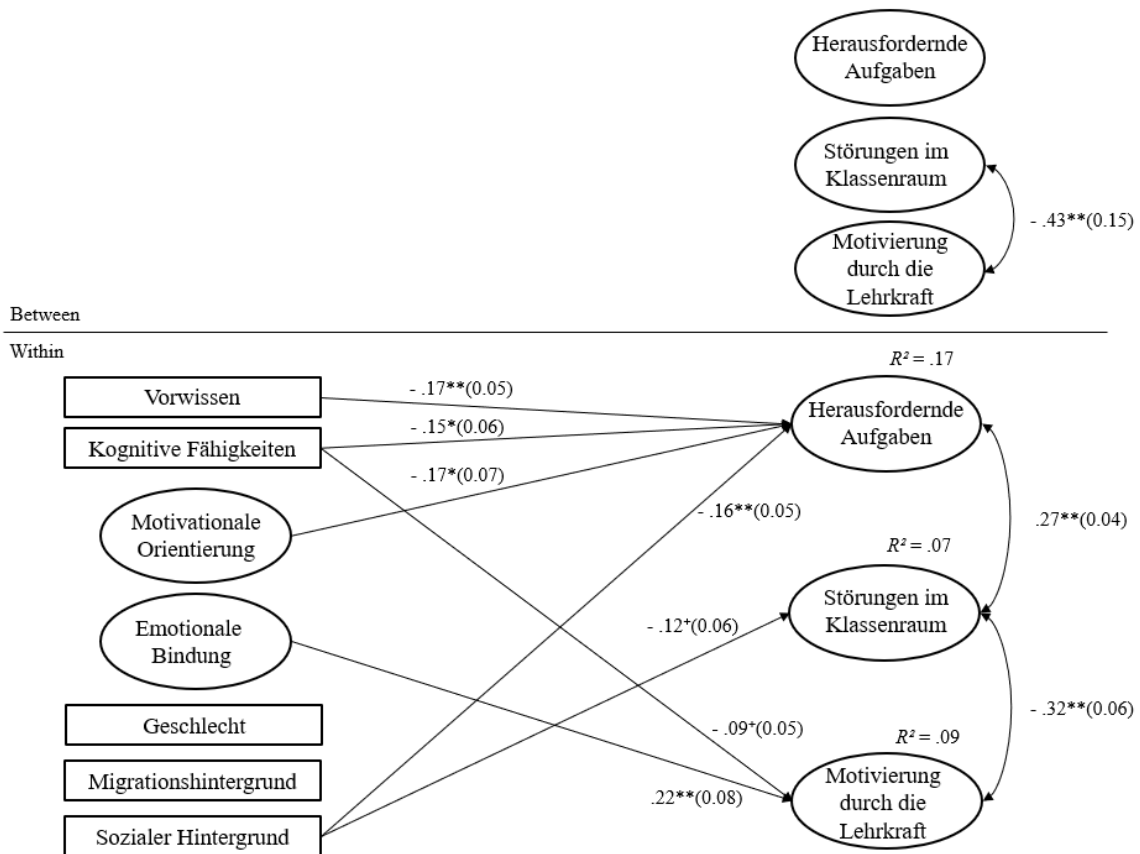


Abbildung 1. Prädiktoren für die Schülereinschätzungen von Unterricht (Angaben zu standardisierten  $\beta$ -Koeffizienten und Standardfehlern; dargestellt ohne Interkorrelationen zwischen den Prädiktoren).

Anmerkungen. CFI = .96, RMSEA = .03, SRMR<sub>Within</sub> = .04, SRMR<sub>Between</sub> = .14,  $\chi^2 = 615.35$ ,  $df = 427$ , AIC = 46245.50; BIC = 46813.49; +  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$

Im Gesamtmodell zeigten sich bei gleichzeitiger Berücksichtigung aller Schülermerkmale gerichtete Zusammenhänge zwischen dem Vorwissen, den kognitiven Fähigkeiten, der motivationalen Orientierung und dem sozialen Hintergrund und der Wahrnehmung herausfordernder Aufgaben innerhalb von Klassen (Abbildung 1). Hinsichtlich der Einschätzung von Störungen im Unterricht erwies sich nur der soziale Hintergrund als prädiktiv. Bezüglich der Motivierung durch die Lehrkraft zeigten sich im Gesamtmodell die kognitiven Fähigkeiten und die emotionale Bindung als signifikante Prädiktoren.

## 6 Diskussion

Die vorliegende Studie beschäftigte sich mit den Fragen, ob Unterschiede in Schülereinschätzungen von Unterrichtsqualität innerhalb von Grundschulklassen vorliegen und welche individuellen Schülermerkmale prädiktiv für abweichende Einschätzungen sind. In dieser Untersuchung konnte auch für Grundschul Kinder eine hohe klasseninterne Varianz für ihre Einschätzungen für alle drei Unterrichtsqualitätsmerkmale gezeigt werden. Bei der Interpretation der Übereinstimmungen ist zu beachten, dass sich die Items hinsichtlich ihres Inferenzgrades unterscheiden. Bei niedrig-inferenten Items werden beobachtbare Verhaltensaspekte thematisiert und der Gebrauch dieser Items führt zu einer relativ hohen Übereinstimmung innerhalb von Klassen. Hierunter fallen auch die in der Studie verwendeten Items zum Merkmal Störungen im Unterricht. Im Gegensatz dazu sind für die Beantwortung von hoch-inferenten Items kognitive Schlussfolgerungen und Interpretationen notwendig (Clausen et al., 2003; Dreesmann, 1982). Der unterschiedliche Abstraktionsgrad der Items der vorliegenden Studie kann die differierenden Übereinstimmungen der Schülereinschätzungen innerhalb der Klasse in den drei Unterrichtsqualitätsmerkmalen und das Ergebnis, dass die größte Übereinstimmung der Schülereinschätzungen in dem Bereich der Klassenführung zu finden war, unterstützt haben.

Ergebnisse der Mehrebenenregressionsmodelle liefern Hinweise auf die Bedeutung der individuellen Schülermerkmale für die Unterschiede in den Einschätzungen von Unterrichtsqualität. Insbesondere erwiesen sich die Schülermerkmale Vorwissen, kognitive Fähigkeiten, motivationale Orientierung und emotionale Bindung sowie soziale Herkunft als prädiktiv für die Schülereinschätzung hinsichtlich der Unterrichtsqualität. Wie herausfordernd und schwierig Aufgaben im Unterricht von Schülerinnen und Schülern eingeschätzt wurden, hing mit dem Vorwissen in dem Unterrichtsfach, den generellen kognitiven Fähigkeiten, dem sozialen Hintergrund und der motivationalen Orientierung zusammen. Die Bedeutung einer hohen motivationalen Orientierung für eine positivere Schülereinschätzung von Unterricht konnte auch in anderen Studien gezeigt werden (z. B. Seidel et al., 2016). Die Störungen im Klassenraum wurden besonders hoch von Schülerinnen und Schülern mit einem niedrigen sozialen Hintergrund eingeschätzt. Die Motivierung durch die Lehrkraft schätzten Schülerinnen und Schüler mit einer höheren kognitiven Fähigkeit und mit einer positiveren emotionalen Bindung zur Lehrkraft höher ein. Das Ergebnis, dass eine höhere emotionale Bindung der Schülerinnen und Schüler, gemessen durch die Nähe zur Lehrkraft, zu einer

höheren eingeschätzten Motivierung durch die Lehrkraft in einer Grundschulklasse führte, bestärkt die Relevanz einer positiven Lehrkraft-Schüler-Interaktion. Die Bedeutung einer positiven Beziehung zwischen der Lehrkraft und den Schülerinnen und Schülern für die Motivierung konnte bereits mehrfach in Studien bestätigt werden (Looser, 2011). Es ist aber zu berücksichtigen, dass die Schülerangaben bezüglich der Nähe zur Lehrkraft und der Motivierungsqualität vom vorigen Beziehungsverhältnis zwischen der Lehrkraft und den Schülerinnen und Schülern abhängig sein können.

Verglichen mit den bisherigen theoretischen und empirischen Befunden ist es auffällig, dass sich die Schülermerkmale Geschlecht und Migrationshintergrund als weniger bedeutsam für die Schülereinschätzung von Grundschulunterricht erwiesen haben. Eine Erklärung hierfür kann die Relevanz der Identitätsbildung in der Jugend (13–18 Jahre) sein (Erikson, 1968). Die intensive Auseinandersetzung mit der Frage nach der eigenen Identität (Geschlecht, Migration, etc.) im Jugendalter kann somit die Einschätzungen der Schülerinnen und Schüler vom Unterricht sowie auch die Wahrnehmung der Lehrkräfte in der Sekundarstufe möglicherweise stärker lenken als in der Grundschule. Weitergehend ist hervorzuheben, dass insbesondere die wahrgenommene Herausforderung von Aufgaben durch verschiedene, kognitive, motivationale und soziale Schülermerkmale multiple vorhergesagt wurde. Dieses Merkmal war zugleich dasjenige, mit der geringsten Übereinstimmung der Einschätzungen innerhalb der Klassen, was wiederum konsistent ist mit der Heterogenität der Schülerschaften innerhalb von Grundschulklassen. Zudem kann die multiple Vorhersage von individuellen Hintergrundmerkmalen im Bereich herausfordernde Aufgaben zusätzlich durch die Perspektive der Items erklärt werden, da sich die Items explizit auf die Schülerinnen und Schüler bezogen und somit hinsichtlich der Wahrnehmungsperspektive am wenigsten Interpretationsfähigkeit im Vergleich zu den anderen Items voraussetzten.

## **6.1 Limitationen der Studie**

Im Schülerfragebogen zur wahrgenommenen Unterrichtsqualität wurde bei der Formulierung der Items zwischen einem Fokus auf die Aufgabengestaltung („Wir Schüler mussten ganz schön denken, bis wir den Text und das Bild verstanden haben.“), Lehrkraft-Fokus („Im Unterricht gerade hat unser/ e Lehrer/ in das Lesen der Texte mit Bildern sehr spannend gestaltet.“) und Schüler-Fokus („Im Unterricht zu Südamerika wurde der Unterricht oft gestört.“) gewechselt. Wird die „wir“-Perspektive in der Itemformulierung berücksichtigt, so kann die Einschätzung der Mitschülerinnen und Mitschüler nicht bekannt sein oder auch im

Widerspruch zur eigenen Meinung des Schülers oder der Schülerin stehen (Piskol, 2008). Durch den Gebrauch von „ich“/„mir“ anstelle von „wir“/„uns“ in Itemformulierungen werden auch die individuellen Wahrnehmungen berücksichtigt (Stapleton, Yang & Hancock, 2016; Wenger et al., 2018). Um eine mögliche Verzerrung der Ergebnisse aufgrund der unterschiedlichen Formulierung der Items zu vermeiden, sollte die Formulierung in nachfolgenden Studien zu individuellen Wahrnehmungsunterschieden konstant gehalten werden sowie auf die „ich“-Perspektive bei Itemformulierungen zurückgegriffen werden. Dadurch kann einerseits ein besserer Bezug zur individuellen Unterrichtseinschätzung und deren Einflussfaktoren genommen werden und andererseits wird für Grundschülerinnen und Grundschüler weniger kognitive Verarbeitungskapazität benötigt (Piskol, 2008). Die Items zur Variable Nähe zur Lehrkraft bezogen sich nicht direkt auf die Lehrkraft, die hinsichtlich der Unterrichtsqualität eingeschätzt wurde, sondern generell auf die Intensität der Lehrkraft-Schüler-Interaktionen des Kindes zu allen Lehrkräften an ihrer Schule.

## **6.2 Implikationen für Forschung und Praxis**

Diese Untersuchung leistet einen Beitrag zur Theoriebildung, indem heuristische Modelle von Lehr-Lernprozessen um den Aspekt des Wahrnehmungs-/Urteilsprozesses sowie der Prädiktoren von Urteilsprozessen erweitert werden. Schülerinnen und Schüler müssen zunächst das Angebot der Lehrkräfte wahrnehmen, bevor sie es nutzen können. Dieser Prozess wird von individuellen Schülermerkmalen bedingt. Psychologische Theorien über den Wahrnehmungs- oder Urteilsprozess fokussieren häufig das Wechselspiel von individuellen kognitiven Merkmalen (z. B. Vorwissen) mit der Urteilsbildung (vgl. Bless et al., 2004). Allerdings konnte in dieser Studie besonders auch die Bedeutsamkeit der motivationalen und emotionalen Merkmale für das Schülerurteil über den Unterricht aufgezeigt werden. Demzufolge sollten diese individuellen Faktoren auch in weiteren Untersuchungen berücksichtigt werden.

Für Forschung im Bereich der Unterrichtswahrnehmung von Lernenden ist relevant, dass in der vorliegenden Studie eine hohe klasseninterne Variabilität gefunden wurde, jedoch nur ein geringer Anteil der Gesamtvarianz erklärt werden konnte. Dementsprechend blieb ein großer Anteil der Varianz auf Within-Level unerklärt (kognitive Aktivierung: 83.0 %, Klassenführung: 93.0 % und Motivierung: 91.0 %). In zukünftigen Studien sollten zum einen die Heterogenität der Schülerurteile über den Unterricht bei der Interpretation berücksichtigt, sowie zum anderen auch weitere Prädiktoren seitens der Schülerinnen und Schüler, der

Lehrkräfte sowie der Schule erfasst und untersucht werden. In vielen Untersuchungen zur Unterrichtsqualität werden die Schülereinschätzungen in oftmals aggregierter Form auf Klassen- oder Schulebene, z. B. durch Bildung des Klassenmittelwertes, betrachtet. Nach den berichteten Ergebnissen ist die Methode der Aggregation allerdings kritisch zu hinterfragen, da dabei die Heterogenität der Schülerurteile in den Klassen nicht berücksichtigt wird. Dadurch, dass die Unterrichtsqualität allein durch Schülereinschätzungen erfasst wurde, kann nicht die Frage beantwortet werden, inwiefern die divergierenden Einschätzungen auf unterschiedlichen individuellen Wahrnehmungen oder einem tatsächlich unterschiedlichen Verhalten der Lehrkraft beruhen. Neben der gezielten, pädagogisch oder fachdidaktisch angezeigten Differenzierung von Lehrkräften können auch persönliche Vorurteile und Überzeugungen, die sich in ihren Handlungen ausdrücken, zu differierendem Verhalten der Lehrkraft und somit zu den Unterschieden in den Schülerwahrnehmungen führen (Schweig, 2016). Hinsichtlich des Informationsverarbeitungsprozesses würde dies bedeuten, dass Unterschiede unabhängig von der Verarbeitung von Schülerinnen und Schülern schon auf der Reizebene vorlägen. Für die wissenschaftliche Untersuchung der Unterrichtsqualität hat die grundsätzliche Frage, ob beobachtete Unterschiede in Schüleraussagen zum Unterricht auf Wahrnehmungsunterschiede oder auf differenzielles Lehrkraftverhalten zurückgehen, wichtige Implikationen. Diese betreffen, wie in der vorliegenden Studie, die Interpretation bzw. Interpretierbarkeit von Schülerurteilen sowie in Abhängigkeit von den Forschungsdesideraten auch die zu wählenden Methoden. Um der Frage nachzugehen, ob differenzielles Verhalten von Lehrkräften zu den divergierenden Schülereinschätzungen führt, könnte in nachfolgenden Studien beispielsweise zusätzlich das Lehrkraftverhalten von externen Beobachterinnen und Beobachtern mittels Videographie bewertet werden.

In Bezug auf die Praxis unterstreichen die Ergebnisse der heterogenen Unterrichtswahrnehmung, die zumindest in Teilen systematisch auf konkrete Schülermerkmale zurückgeführt werden konnte, die Bedeutung der Berücksichtigung der Heterogenität von Schülereigenschaften für die Planung eines erfolgreichen Unterrichts. Die Resultate unterstützen die Relevanz der Anpassung der Aufgabenschwierigkeit und die Vermeidung des Gefühls der Überforderung unter Berücksichtigung des Vorwissens und der kognitiven Fähigkeiten der Schülerinnen und Schüler. Neben einem geringen Vorwissen und geringen kognitiven Fähigkeiten kann auch eine niedrige Motivation zu einer Tendenz der wahrgenommenen Überforderung führen. Dementsprechend sollten motivationale

Orientierungen von Schülerinnen und Schülern von der Lehrkraft beobachtet und gegebenenfalls gefördert werden.

Die vorliegende Studie trägt dazu bei, die Heterogenität bei der Wahrnehmung des Unterrichts in Grundschulklassen näher zu beschreiben und Prädiktoren hierfür zu identifizieren.

### **Danksagung**

Wir möchten uns bei Matthias Trendtel für die Unterstützung im Bereich der statistischen Methodenwahl während der Vorbereitung des Manuskripts bedanken.

### **Literatur**

Ahnert, L. (2013). Entwicklungs- und Sozialisationsrisiken bei jungen Kindern. In L. Fried & S. Roux (Hrsg.), *Handbuch der Pädagogik der Frühen Kindheit* (S. 75–85). Weinheim: Beltz.

Balcetis, E. & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91* (4), 612–625.

Barkley, C. L. & Gabriel, K. I. (2007). Sex differences in cue perception in a visual scene: Investigation of cue type. *Behavioral Neuroscience*, *121*, 291–300.

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M. & Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente (Materialien aus der Bildungsforschung Nr. 83)*. Berlin: Max-Planck-Institut für Bildungsforschung.

Behörde für Schule und Berufsbildung (2011). *LAU – Aspekte der Lernausgangslage und der Lernentwicklung Klassenstufen 5, 7 und 9* (Bd. 8, HANSE – Hamburger Schriften zur Qualität im Bildungswesen). Münster: Waxmann.

Bless, H., Fiedler, K. & Strack, F. (2004). *Social Cognition: How individuals construct social reality*. Philadelphia: Psychology Press. Bos, W., Strietholt, R., Goy, M., Stubbe T. C., Tarelli, I. & Hornberg, S. (Hrsg.) (2010). *IGLU 2006 – Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

- Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129–148.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school: Expanded ed.* Washington, DC: National Academy Press.
- Brok, P. den, Levy, J., Wubbels, T. & Rodriguez, M. (2003). Cultural influences on students' perceptions of videotaped lessons. *International Journal of Intercultural Relations*, 27, 355–374.
- Brophy, J. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71, 733–750.
- Bruner, J. S. & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, 42, 33–44.
- Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, 7, 477–484.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen: Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31, 122–141.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Dreesmann, H. (1982). *Unterrichtsklima: Wie Schüler den Unterricht wahrnehmen.* Weinheim: Beltz.
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48 (2), 262–286.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden Lehrbuch.* Weinheim: Beltz.
- Emmer, E. T. & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, 36, 103–112.
- Erikson, E. H. (1968). *Identität und Lebenszyklus.* Frankfurt: Suhrkamp.

- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2016). Urteile zur Unterrichtsqualität von Grundschüler\_innen: Mehr als „Ich mag meine Lehrerin“? In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler & R. Schmidt (Hrsg.), *Facetten grundschulpädagogischer und grundschuldidaktischer Forschung*. Wiesbaden: Springer.
- 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.
- Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen. *Zeitschrift für Pädagogische Psychologie, 24* (2), 111–122.
- Gentry, M., Gable, R. K. & Rizza, M. G. (2002). Students' perceptions of classroom activities: Are there gradelevel and gender differences? *Journal of Educational Psychology, 94* (3), 539–544.
- 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.
- Goh, S. C. & Fraser, B. J. (1995). *Learning environment and student outcomes in primary mathematics classrooms in Singapore*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Hamre, B. K. & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization, measurement, & improvement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). New York, NY: Routledge.
- Heller, K. A. & Perleth, C. (2000). *KFT 4 – 12 + R – Kognitiver Fähigkeits-Test für 4. bis 12. Klassen, Revision*. Göttingen: Beltz.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (6. überarb. Aufl.). Seelze: Klett-Kallmeyer.

- Helmke, A., Piskol, K., Pikowsky, B. & Wagner, W. (2009). Schüler als Experten von Unterricht. Unterrichtsqualität aus Schülerperspektive. *Lernende Schule*, 46, 98–105.
- Helmke, A. & Schrader, F.-W. (2008). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. *SEMINAR – Lehrerbildung und Schule*, 3, 17–47.
- Herkner, W. (1991). *Lehrbuch Sozialpsychologie*. Bern: Hans Huber.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222–237.
- Klieme, E. & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, E., J. Neubrand, M. Prenzel, U. Schiefele, W. Schneider & K.-J. Tillmann (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 334–359). Opladen: Leske + Budrich.
- Kloss, J. (2013). Grundschüler als Experten für Unterricht – Empirische Überprüfung der Validität von Unterrichtsbeurteilungen durch Schüler der dritten und vierten Jahrgangsstufe. In E. Witruk, *Beiträge zur Pädagogischen und Rehabilitationspsychologie* (Volume 6). Frankfurt am Main: Peter Lang.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Bonn: KMK.
- 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.
- Kunter, M., Baumert, J., Voss, T., Klusmann, U., Richter, D. & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 10, 805–820.
- Lepach, A. C., Reimers, W., Pauls, F., Petermann, F. & Daseking, M. (2015). Geschlechtseffekte bei Intelligenz und Gedächtnisleistungen. *Zeitschrift für Neuropsychologie*, 26 (1), 5–16.

- Levy, J., den Brok, P., Wubbels, T. & Brekelmans, M. (2003). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research*, 6(1), 1–31.
- Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K. & Pauli, C. (2009). Quality of geometry instruction and its shortterm impact on students? Understanding of Pythagorean Theorem. *Learning and Instruction*, 19 (6), 527–537.
- Looser, D. (2011). Soziale Beziehungen und Leistungsmotivation. Die Bedeutung von Bezugspersonen für die längerfristige Aufrechterhaltung der Lern- und Leistungsmotivation. Opladen: Budrich UniPress.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings in multilevel modelling. *Contemporary Educational Psychology*, 34, 120–131.
- Lüdtke, O., Trautwein, U., Kunter, M. & Baumert, J. (2006). Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20, 85–96.
- Madill, R. A., Gest, S. D. & Rodkin, P. C. (2014). Students' perceptions of relatedness in the classroom: The roles of emotionally supportive teacher-child interactions, children's aggressive disruptive behaviors, and peer social preference. *School Psychology Review*, 43 (1), 86–105.
- Martin, E. & Wawrinowski, U. (1993). *Beobachtungslehre. Theorie und Praxis reflektierter Beobachtung und Beurteilung* (2. überarb. Aufl.). Weinheim: Juventa.
- Matthews, C. E. (2008). *Socio-economic status, student perceptions and college readiness* (Doctoral dissertation, Washington State University). Zugriff am 28.07.2017. Verfügbar unter <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.427.3797&rep=rep1&type=pdf>.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59 (1), 14–19.
- McElvany, N., Schroeder, S., Baumert, J., Schnotz, W., Horz, H. & Ullrich, M. (2012). Cognitively demanding learning materials with texts and instructional pictures: Teachers'

- diagnostic skills, pedagogical beliefs and motivation. *European Journal of Psychology of Education*, 27 (3), 403–420.
- Muthén, L. K. & Muthén, B. O. (1998–2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nisbett, R. E. & Miyamoto, Y. (2005). The influence of culture on holistic versus analytic perception. *Trends in Cognitive Science*, 10, 467–473.
- Nisbett, R. E., Peng, K., Choi, I. & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, 108, 291–310.
- Ohle, A., McElvany, N., Schnotz, W., Wagner, I., Horz, H., Ullrich, M. et al. (2017). Development and Evaluation of a Competency Model for Teaching Integrative Processing of Texts and Pictures (BiTe). In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence Assessment in Education – Research, Models and Instruments* (pp. 167–180). Dordrecht: Springer.
- Pauli, C. & Reusser, K. (2003). Unterrichtsskripts im schweizerischen und im deutschen Mathematikunterricht. *Unterrichtswissenschaft*, 31, 238–272.
- Pianta, R. C. & Hamre, B. K. (2009). Measurement and improvement of teacher-child interactions: Implications for policy and accountability frameworks of standardized observation. In G. Sykes, B. Schneider, D. N. Plank & T. G. Ford (Eds.), *Handbook of education policy research* (pp. 652–660). New York: Routledge.
- Piskol, K. (2008). *Unterrichtsqualität aus der Schülerperspektive: Ein Verfahren zur Unterrichtsentwicklung*. Unveröffentlichte Diplomarbeit, Universität Mannheim.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Praetorius, A.-K., Lauermaun, F., Klassen, R. M., Drexler, K., Nitsche, S., Janke, S., Dickhäuser, O. & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education*, 65, 241–254.

- Quellenberg, H. (2009). Studie zur Entwicklung von Ganztagschulen (StEG) ausgewählte Hintergrundvariablen, Skalen und Indices der ersten Erhebungswelle, In Zusammenarbeit mit dem StEG-Konsortium und den Mitarbeiter\_innen des StEG-Teams. *Materialien zur Bildungsforschung: Vol. 24*. Frankfurt, Main: Gesellschaft zur Förderung Pädagogischer Forschung / Deutsches Institut für Internationale Pädagogische Forschung.
- Roebers, C. M. & Schneider, W. (2001). Individual differences in children's eyewitness recall: The influence of intelligence and shyness. *Applied Developmental Science, 5*, 9–20.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.
- 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.
- Schnotz, W., Horz, H., McElvany, N., Schroeder, S., Ullrich, M., Baumert, J., Hachfeld, A. & Richter, T. (2010). Das BITE-Projekt: Integrative Verarbeitung von Texten und Bildern in der Sekundarstufe I. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik* (S. 143–153). Weinheim: Beltz.
- Schweig, J. D. (2016). Moving beyond means: revealing features of the learning environment by investigating the consensus among student ratings. *Learning Environments Research, 19*, 441–462.
- Segall, M. H., Campbell, D.T. & Herskovits, M. J. (1966). *The influence of culture on visual perception*. New York: Bobbs-Merrill Co.
- Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research, 9* (3), 253–271.
- Seidel, T., Jurik, V., Häusler, J. & Stubben, S. (2016). Mikro-Umwelten im Klassenverband: Wie sich kognitive und motivational-affektive Schülervoraussetzungen auf die Wahrnehmung und das Verhalten im Fachunterricht auswirken. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 65–88). Münster: Waxmann.

- Seidel, T. & Shavelson, R. (2007). Teaching effectiveness research in the past decade. *Review of Educational Research*, 77, 454–499.
- Soebari, T. S. & Aldridge, J. M. (2015). Using student perceptions of the learning environment to evaluate the effectiveness of a teacher professional development programme. *Learning Environments Research*, 18, 163–178.
- 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.
- Uslar, D. von (2007). *Existenz und Lebensgeschichte. Der Mensch und seine Welt im Spiegel der Psychologie. Grundzüge einer Psychologischen Anthropologie*. Würzburg: Königshausen und Neumann.
- Walther, E., Fiedler, K. & Nickel, S (2003). The influence of prior knowledge on constructive biases. *Swiss Journal of Psychology*, 62 (4), 219–231.
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht – am Beispiel der Studie DESI (Deutsch Englisch Schülerleistungen International) der Kultusministerkonferenz*. Dissertation am Fachbereich 8, Psychologie, Universität Koblenz-Landau.
- 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 (5), 705–721.
- Wenger, M., Lüdtke, O. & Brunner, M. (2018). Übereinstimmung, Variabilität und Reliabilität von Schülerurteilen zur Unterrichtsqualität auf Schulebene. *Zeitschrift für Erziehungswissenschaft*, 21, 929–950.
- Wubbels, T. & Levy, J. (1993). *Do you know what you look like? Interpersonal relationships in education*. London: Falmer Press.
- Wyer, R. S. & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Erlbaum.

## Anhang

### Skalen zur Erfassung der Unterrichtswahrnehmung

#### *Herausfordernde Aufgaben*

In dieser Frage geht es darum, wie schwer es für euch war, die Bilder und Texte zu verstehen.

- a. Wir Schüler mussten ganz schön denken, bis wir den Text und das Bild verstanden haben.
- b. Wir Schüler mussten gut aufpassen, damit wir den Text und das Bild verstehen.
- c. Für uns Schüler war es manchmal ganz schön kompliziert, den Text und das Bild zu verstehen.
- d. Wir Schüler mussten uns ziemlich anstrengen, um den Text und das Bild zu verstehen.
- e. Wir Schüler mussten gut überlegen, damit wir den Text und das Bild verstehen.

#### *Störungen im Unterricht*

In dieser Frage geht es darum, wie der Unterricht zu Südamerika ablief.

Im Unterricht zu Südamerika

- a. wurde der Unterricht oft gestört.
- b. wurde fortwährend laut gequatscht.
- c. wurde andauernd Blödsinn gemacht.
- d. war meistens viel Chaos in der Klasse.
- e. konnte man sich nur schwer konzentrieren, weil so viele Störungen passierten.

#### *Motivierung durch die Lehrkraft*

In dieser Frage geht es darum, wie eure Lehrerin/ euer Lehrer euer Interesse wecken konnte.

Im Unterricht gerade ...

- a. hat unser/ e Lehrer/ in das Lesen der Texte mit Bildern sehr spannend gestaltet.
- b. konnte unser/ e Lehrer/ in uns Schülerinnen und Schüler manchmal richtig zum Lesen der Texte mit den Bildern begeistern.
- c. hat unser/ e Lehrer/ in interessante Texte mit Bildern zum Lesen ausgewählt.
- d. hat unser/ e Lehrer/ in dafür gesorgt, dass wir Spaß dabei hatten, die Texte mit den Bildern zu verstehen.
- e. waren wir Schülerinnen und Schüler beim Lesen der Texte mit den Bildern fast immer mit Begeisterung dabei.

## **6.2 Beitrag II. Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht**

Jennifer Iglér<sup>1</sup>, Annika Ohle<sup>1</sup>, Theresa Schlitter<sup>1</sup>, Annika Teerling<sup>2</sup>, Regine Asseburg<sup>2</sup>, Nele McElvany<sup>1</sup> & Olaf Köller<sup>2</sup>

1 Institut für Schulentwicklungsforschung der Technischen Universität Dortmund  
2 Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Universität Kiel

### *Zusammenfassung*

In der vorliegenden Studie wird die Mediation der Vorhersagekraft von motivationalen Orientierungen von Lehrkräften für Merkmale qualitätvollen Leseunterrichts durch die Beteiligung an innovativen Programmen untersucht. Als Merkmale qualitätvollen Unterrichts haben sich in zahlreichen empirischen Studien unter anderem die Differenzierung von Lernangeboten im Unterricht sowie die Motivierung von Schülerinnen und Schülern erwiesen. Inwiefern die Beteiligung von Lehrkräften an innovativen Programmen, die auf eine Verbesserung von Diagnose und Förderung im Leseunterricht abzielen, prädiktiv für Merkmale qualitätvollen Leseunterrichts im Unterricht ist und welche Vorhersagekraft motivationale Lehrkraftmerkmale für eben diese Beteiligung haben, ist bislang ungeklärt.

Im Rahmen des Projekts „Evaluation von Konzepten und Maßnahmen der fachübergreifenden Leseförderung im Primarbereich“ (BiSS-EvalLesen) wurden  $N = 40$  Lehrkräfte aus sieben Bundesländern zu ihren motivationalen Orientierungen, der Beteiligung an Maßnahmen der Bund-Länder-Initiative „Bildung durch Sprache und Schrift (BiSS)“ sowie zu Qualitätsmerkmalen ihres Leseunterrichts befragt. Ergebnisse von manifesten Strukturgleichungsmodellen zeigten hohe lineare Zusammenhänge der motivationalen Orientierungen von Lehrkräften und der Beteiligung an innovativen Programmen mit den Merkmalen qualitätvollen Leseunterrichts. Für die Motivierung von Schülerinnen und Schülern zeigte die Beteiligung an innovativen Programmen eine hohe Prädiktionskraft, die sich außerdem als Mediator für den Zusammenhang zwischen Lehrkraftmotivation und Motivierung von Schülerinnen und Schülern erwies. Dies deutet auf die Effektivität der BiSS-

Maßnahmen hin, wobei die angenommene Wirkrichtung in einer Längsschnittstudie geprüft werden muss.

Schlagwörter: Leseunterricht, Unterrichtsqualität, Schulprogramme, Innovationen, Motivationale Lehrkraftmerkmale.

*Abstract*

The present study investigates whether teachers' motivational orientations influence teaching quality in reading lessons, which is mediated by the participation in innovative programs. Numerous empirical studies have shown differentiated instruction and pupils' motivation as characteristics of good instructional quality. Up to now the relation of teachers' participation in innovative programs – aiming for improving diagnostics and promotion of students' reading competences – and instructional quality is not clarified yet. Furthermore, there is little empirical evidence on the prediction of teachers' motivational orientations on participation in innovative programs.

In the context of the project "Evaluation of concepts and measures of interdisciplinary reading support in primary education" (BiSS-EvalLesen),  $N = 40$  teachers from seven federal states filled in questionnaires about their motivational orientation, participation in measures of the initiative "Bildung durch Sprache und Schrift (BiSS)" as well as quality characteristics of their reading lessons. Results of manifest structural equation models (SEM) showed high linear relationships between teachers' motivational orientations and instructional quality (differentiated instruction and pupils' motivation) and between aspects of participation in innovative programs and instructional quality measures. The motivation of pupils is predicted by the aspects of participation in innovative programs, which also mediates the effect between the motivation of the teachers and the motivation of the pupils. This indicates the effectiveness of the BiSS-measures, whereby the assumed direction of action must be examined in a longitudinal study.

keywords: reading lessons, instructional quality, school programs, innovations, motivational orientations

## 1 Einleitung

Die Wirkungskette zwischen der Qualität von Unterricht und Lernprozessergebnissen im Bereich Lesen wird einerseits in theoretischen Modellen zu Unterrichtsprozessen angenommen und konnte darüber hinaus in zahlreichen empirischen Studien nachgewiesen werden (z. B. Helmke, 2012; Lipowsky, 2006; für den Bereich Lesen: Taylor, Pearson, Peterson & Rodriguez, 2003). Jedoch verdeutlichen Studien, dass Lehrkräfte unterschiedlich erfolgreich sind, einen an die individuellen Leseleistungen angepassten und zugleich qualitativollen Leseunterricht mit dem Ziel der individuellen Verbesserung im Bereich Lesen anzubieten (z. B. Bos, Tarelli, Bremerich-Vos & Schwippert, 2012). Lerngelegenheiten optimal an die Schülerinnen und Schüler anzupassen, ist ein zentraler Bestandteil professionellen Lehrkrafthandelns.

Um die Passung von Leseförderung für alle Schülerinnen und Schüler zu optimieren, wurden verschiedene innovative Förderprogramme entwickelt. Hierzu zählen unter anderem das Programm *ProLesen – Auf dem Weg zur Leseschule* (Bayerisches Staatsministerium für Unterricht und Kultus & Staatsinstitut für Schulqualität und Bildungsforschung, 2010) und die Bund-Länder-Initiative „Bildung durch Sprache und Schrift“ BiSS (Becker-Mrotzek, Hasselhorn, Roth & Stanat, 2016). Ziel des BiSS-Programms ist die wissenschaftliche Überprüfung der Wirksamkeit, Effizienz und Erprobung eingesetzter Maßnahmen und Methoden zur Förderung und Diagnose von sprachlichen Fertigkeiten im Elementar-, Primar- und Sekundarbereich sowie die Weiterentwicklung dieser Maßnahmen und die Weiterqualifikation des pädagogischen Personals. Für die Umsetzung des Programms schlossen sich je drei bis zehn Bildungseinrichtungen, Kindertageseinrichtungen und Schulen sowie auch beispielsweise Bibliotheken und Universitäten zu einem Verbund zusammen; im Jahr 2017 waren dies über 100 aktive Verbünde. Ein besonderer Aspekt des Programms ist die formative und summative Evaluation ausgewählter Verbünde mit dem Ziel, Erkenntnisse unter anderem über die Diagnose- und Förderkonzepte zu erhalten (Becker-Mrotzek et al., 2016; Schneider et al., 2012). Die vorliegende Studie ist Teil eines Evaluationsprojektes, welches die Wirksamkeit von Maßnahmen in den Modulen „Diagnose und Förderung der Leseflüssigkeit und ihrer Voraussetzungen“ (P3) und „Diagnose und Förderung des Leseverständnisses“ (P4) in Grundschulen in sieben Verbänden untersucht.

## 2 Theoretischer Hintergrund

### 2.1 Merkmale qualitätvollen Leseunterrichts

Allgemein lassen sich Qualitätsmerkmale von Unterricht in drei Dimensionen einordnen, die sich in ähnlicher Operationalisierung im nationalen wie auch im internationalen Kontext etabliert haben. Hierzu gehören als domänenspezifisches Merkmal die *kognitive Aktivierung* und unterrichtsfachübergreifend die *Klassenführung, Regelklarheit und Struktur* sowie das *unterstützende Unterrichtsklima* (vgl. Klieme & Rakoczy, 2008; Pianta & Hamre, 2009). Für erfolgreiches Lernen gilt die kognitive Aktivierung als bedeutsamstes Merkmal. Ein Ziel der kognitiven Aktivierung ist es, eine tiefere Informationsverarbeitung und hochwertige kognitive Prozesse mittels anregenden Aufgaben zu aktivieren (Bransford, Brown & Cocking, 2000; Lipowsky et al., 2009). Dabei dienen die korrekte Einschätzung des Leistungsniveaus von Schülerinnen und Schülern und die entsprechende Anpassung des Anspruchsniveaus als notwendige Voraussetzungen für eine kognitiv anregende Lernsituation (Leuders & Holzäpfel, 2011). Hierfür ist das Differenzieren nach dem individuellen Lernstand und das Anknüpfen an die individuellen Voraussetzungen von zentraler Bedeutung (Bransford, Brown & Cocking, 2000; Lipowsky et al., 2009; Pietsch, 2010). Allgemein wird unter *Differenzierung* die angemessene Reaktion der Lehrkraft auf die Bedürfnisse der Schülerinnen und Schüler mit unterschiedlichen Hintergründen, Fähigkeiten und Interessen verstanden. Lehrkräfte können hierbei auf unterschiedlichen Wegen einen differenzierten, personalisierten und reaktionsfähigen Unterricht gestalten (Tomlinson & Allan, 2000). Ziel der Differenzierung ist, allen Schülerinnen und Schülern die bestmöglichen Lernchancen zu bieten, welche durch die Anpassung des Lernangebots an einzelne Schülergruppen realisiert werden können (Bohl, Batzel & Richey, 2011; Coubergs, Struyven, Vanthournout & Engels, 2017). Häufig werden leistungsbezogene Differenzierungsmaßnahmen in Form von unterschiedlich schwierigen Aufgabenstellungen oder unterschiedlich langen Bearbeitungszeiten eingesetzt (Hertel, 2014; Lipowsky, Kastens, Lotz & Faust, 2011). Studien zum Leseunterricht zeigten, dass sich Facetten eines kognitiv anregenden Unterrichts positiv auf die Leseleistungen von Schülerinnen und Schülern auswirken (Chinn, Anderson & Waggoner, 2001; McDonald Connor et al., 2009), wobei die Differenzierung eine entscheidende Rolle für den Lernprozess im Bereich Lesen einnimmt (Lankes, 2004; Taylor, 2007). Für eine erfolgreiche Differenzierung ist es notwendig, das Unterrichten ausschließlich im Klassenverband zu lösen und stattdessen das Arbeiten in Kleingruppen in den Unterricht

zu integrieren. Ebenfalls sind Phasen selbständiger Lernarbeit sowie die Unterstützung einzelner Schülerinnen und Schüler sinnvolle methodische Maßnahmen, um differenziertes Lernen zu fördern (Tarelli, Lankes, Drossel & Gegenfurtner, 2012). Voraussetzung für eine angemessene Differenzierung sind diagnostische Kompetenzen zur Einschätzung von Leistungsniveaus und der Schwierigkeit des Lernmaterials seitens der Lehrkraft (Christie, Enz & Vukelich, 2007).

Das unterrichtsfachübergreifende Qualitätsmerkmal, unterstützendes Unterrichtsklima, ermöglicht ein erfolgreiches Lernen und ist Voraussetzung für den Lernfortschritt von Schülerinnen und Schülern (Klieme & Rakoczy, 2008; Klieme, Lipowsky, Rakoczy & Ratzka, 2006; Lenske, 2016). Zu dieser Dimension zählen eine positive Lehrer-Schüler-Beziehung sowie ein positiver Umgang mit Fehlern. Das unterstützende Unterrichtsklima hat sich vor allem als Prädiktor für die *Motivierung von Schülerinnen und Schülern* erwiesen (Klieme & Rakoczy, 2008). Durch das Erleben von Autonomie, Kompetenz und sozialer Eingebundenheit im Unterricht kann die Motivation von Schülerinnen und Schülern gesteigert werden (Ferdinand, 2014; Ryan & Deci, 2000). Um einen Unterricht möglichst motivierend zu gestalten und die Bereitschaft der Schülerinnen und Schüler zu fördern, können für das Alter interessante Übungen und Lerngegenstände sowie eigene Entscheidungsmöglichkeiten bezüglich Unterrichtsaufgaben gewählt werden (Lankes & Carstensen, 2007; Wellenreuther, 2008). Eine motivierende Unterrichtsgestaltung führt zu höherer Teilnahmemotivation von Schülerinnen und Schülern am Leseunterricht. Dabei zeigte sich ebenfalls, dass sich die Teilnahmemotivation durch ein stetiges Abrufen des Interesses von Schülerinnen und Schülern stabil entwickeln kann (Hidi & Renninger, 2006). Dementsprechend ist Unterricht interaktiv und bedarf einer Anpassung des Lernangebots an Schülergruppen. Studien zeigten, dass die Qualitätsmerkmale wie beispielsweise die Differenzierung nicht zufriedenstellend im Leseunterricht umgesetzt werden (z. B. Tarelli, Lankes, Drossel & Gegenfurtner, 2012). Für eine Optimierung und Anpassung des Unterrichts an die Lernkultur sind Veränderungen und Weiterentwicklungen der Lernangebote unabdingbar.

## **2.2 Beteiligung an innovativen Programmen**

Veränderungen von Unterricht und somit gegebenenfalls auch eine Verbesserung der Unterrichtsqualität können durch innovative Programme ausgelöst werden (Goldenbaum, 2012). Unter Innovationen werden Neuerungen verstanden (Emmrich, 2010), die im

schulischen Bereich zumeist das Ziel haben, Kompetenzen von Schülerinnen und Schülern zu spezifischen Zeitpunkten in der Schullaufbahn durch kompetenzorientierte Unterrichtsentwicklung zu fördern (Pant, Vock, Pöhlmann & Köller, 2008). Entsprechend ist es Ziel des BiSS-Programms, Kinder in den für den individuellen Bildungserfolg zentralen sprachlichen Kompetenzen vom Elementarbereich über die Grundschule bis zur Sekundarstufe zu fördern. Hierfür wird untersucht, welche Maßnahmen sich in der Praxis in den verschiedenen Etappen der institutionellen Betreuung bewähren (vgl. Becker-Mrotzek et al., 2016).

Eine erfolgreiche Implementation von innovativen Programmen wird zumeist in der Implementationsforschung auf Lehrkräfte und das Kollegium zurückgeführt (Goldenbaum, 2012). Jedoch bestätigten Lehrkräfte mehrfach, zusätzliche Aufgaben in Bezug auf das Unterrichten der Klasse, wie die Beteiligung an Innovationen, als fordernd und anstrengend wahrzunehmen (z. B. Chaplain, 2008). Die Klarheit des relativen Vorteils der Implementation gegenüber der bisherigen Praxis sowie die Kongruenz mit bestehenden Werten, Überzeugungen und subjektiven Theorien der einzelnen Lehrkraft können die Beteiligung und Übernahme der Innovationen in die Praxis begünstigen (Rogers, 2003). Dabei spielt auch die Kooperation von Lehrkräften für eine erfolgreiche Umsetzung von innovativen Programmen eine besondere Rolle (Axtell et al., 2000; Clegg, Unsworth, Epitropaki & Parker, 2002). Generell gilt eine intensive Zusammenarbeit als fester Bestandteil professionellen Handelns (Schwartz, Klusmann, Ihme & Möller, 2012). Diese Zusammenarbeit von Lehrkräften hat eine positive Wirkung auf die Qualität des Unterrichts sowie den Lernerfolg von Schülerinnen und Schülern (Holtappels, Pfeiffer, Röhrich & Voss, 2008; Jang, 2006; Terhart, 2002). Ebenfalls geht eine intensive Lehrkraftkooperation mit einer positiven Einschätzung der Zielerreichung in der individuellen Förderung einher (Holtappels & Heerdegen, 2005). In einigen nationalen Studien konnte gezeigt werden, dass Lehrkräfte im Mittel nur in einem geringen Maß kooperieren (Fussangel & Gräsel, 2012; Steinert et al., 2006). Für eine gelingende Zusammenarbeit sind die Verständigungen über ein gemeinsames Ziel sowie eine Verteilung von Teilaufgaben von Lehrkräften jedoch grundlegend notwendig. Eine intensivere Form der Zusammenarbeit weist eine gemeinsame Analyse und Reflektion über gemeinsame Unterrichtsbestandteile sowie die gemeinsame Erarbeitung neuer Inhalte auf (Gräsel, Fussangel & Pröbstel, 2006). Einen positiven Effekt auf eine intensive Kooperation können das Lehren in festen Teams und Doppelbesetzung haben (Holtappels & Heerdegen, 2005). In Bezug auf Studien zu Innovationen zeigte sich, dass Personen, die eine stärkere

Gruppenzugehörigkeit wahrnehmen, häufiger innovative Programme umsetzen (Axtell et al., 2000). Kooperationen zwischen Lehrkräften tragen darüber hinaus zur Implementationstiefe von Innovationen bei (Goldenbaum, 2012).

Für das Planen und Umsetzen von innovativen Programmen wird Lehrkräften ein hohes Engagement abverlangt. Mitarbeiterengagement beinhaltet eine positive Haltung bezüglich der Arbeit und ist mit einem gewissen zeitlichen Einsatz verbunden (Bakker, Schaufeli, Leiter & Taris, 2008; Schaufeli, Salanova, González-Romá & Bakker, 2002). Im Durchschnitt lag bei einer britischen Studie von Butt und Lance (2005) der *Zeitaufwand* von Lehrkräften für ihre komplexen Tätigkeiten, wozu neben dem Unterricht selbst zum Beispiel auch die Vor- und Nachbereitung des Unterrichts, die Durchführung von Tests und Diagnostik sowie die Elternarbeit zählten, bei ca. 50 Stunden pro Woche und war somit im Vergleich zu anderen Berufsgruppen hoch. Insgesamt wurden in sechs Prozent der Arbeitszeit von diesen Lehrkräften schulorganisatorische Aufgaben, wie auch die Arbeit an innovativen Programmen, bewältigt (Gunter et al., 2005). In Deutschland wenden Lehrkräfte ebenfalls primär Zeit für das Lehren von Schülerinnen und Schülern auf und werten die Unterstützung von Programmarbeit an der Schule als weniger zeitraubend (Bauer et al., 2007). In deutschsprachigen Untersuchungen gaben Lehrkräfte für ihre durchschnittliche Arbeitszeit in der Woche im Mittel einen Wert von 32–45,5 Stunden an (Philipp & Kunter, 2013; Schaarschmidt, Sieland, Fischer & Rahm, 2007). Generell forderten Lehrkräfte in Befragungen die Ausweitung von Zeitkontingenten für außerunterrichtliche Aufgaben sowie für die Vor- und Nachbereitung von Unterricht und im gleichen Zuge die Reduktion von Pflichtstunden (vgl. Schaarschmidt et al., 2007). Darüber hinaus zeigte sich in einer Studie von Philipp und Kunter (2013), dass im Mittel 69,6 Prozent der Lehrkräfte der Sekundarstufe I angaben, in Vollzeit zu arbeiten sowie etwa eine Stunde pro Woche für die Beteiligung an Programmen aufzuwenden. Beteiligten sich Lehrkräfte zeitlich weniger an Programmen, so fühlten sie sich weniger emotional erschöpft. Ein positives Schulklima konnte das zeitliche Engagement von Lehrkräften an außerunterrichtlichen Aktivitäten unterstützen (Somech & Ron, 2007). Folglich sind das zeitliche Engagement sowie die Beteiligung an innovativen Programmen zentral für eine gelungene Umsetzung von Innovationen und können dabei auch von Lehrkraftmerkmalen beeinflusst werden (Altrichter & Wiesinger, 2005).

### **2.3 Motivationale Lehrkraftmerkmale**

Nach dem Modell Professioneller Kompetenz von Lehrkräften zählt neben der *intrinsischen*

*Motivation* auch die *Selbstwirksamkeitserwartung* zu den motivationalen Orientierungen von Lehrkräften (Kunter et al., 2008). So finden motivationale Merkmale auch häufig Berücksichtigung bei der Erklärung der Intensität der Beteiligung an innovativen Programmen sowie bei der Erklärung von gelungener Umsetzung von Innovationen. Neben kollektiven Faktoren konnte in Studien zur Implementation gezeigt werden, dass individuelle Lehrkraftmerkmale, wie zum Beispiel *intrinsische Motivation* und *Selbstwirksamkeitserwartung*, eine wichtige Bedeutung für die Beteiligung an Innovationen und deren Umsetzung haben (Fullan, 2007; Gräsel, 2010). Im Hinblick auf die Akzeptanz von Innovationen nahmen motivationale Lehrkraftmerkmale einen substantiellen Einfluss (Emmrich, 2010; Jäger, 2004). Des Weiteren erwies sich die Motivation auch als bedeutsamer Faktor für einen qualitativollen Unterricht und den Lernerfolg (Roth, Assor, Kanat-Maymon & Kaplan, 2007). Allgemein wird unter der intrinsischen Motivation das Handeln aufgrund von Interesse und Spaß ohne externe Belohnung verstanden (Deci & Ryan, 2000). Bei intrinsisch motivierten Lehrkräften dient der Spaß an Aspekten der Berufstätigkeit selbst, zum Beispiel am Lehren oder an der Diagnostik, als antreibende Kraft für ihr Handeln in der Klasse (Kyriacou & Coulthard, 2000; Roness, 2011; Thomson, Turner & Nietfeld, 2012; Watt & Richardson, 2008). In weiteren Untersuchungen zeigte sich, dass Lehrkräfte, die motiviert sind und mit Freude ihren Beruf ausüben, ein höheres Engagement auch außerhalb des Unterrichts aufweisen (Long & Hoy, 2006).

Dabei kann die intrinsische Motivation von Lehrkräften nicht nur Auswirkungen auf die Unterrichtsqualität bzw. das Gelingen von innovativen Programmen haben, sondern auch die Motivation von Schülerinnen und Schülern beeinflussen (Bieg & Mittag, 2009; Fischer & Rustemeyer, 2007). Diese Annahme unterstützen Ergebnisse verschiedener Studien, die Zusammenhänge zwischen der intrinsischen Motivation von Lehrkräften und autonomieunterstützendem Lehrkrafthandeln zeigten, welches zu einer höheren Motivierung von Schülerinnen und Schülern führte (Müller, Hanfstingl & Andreitz, 2009; Roth et al., 2007).

In Bezug auf die sozial-kognitive Theorie von Bandura (1997) beschreibt die Selbstwirksamkeitserwartung die subjektive Gewissheit, neue oder schwierige Aufgaben aufgrund von den eigenen Kompetenzen zu bewältigen. Die Selbstwirksamkeitserwartung hat einen Einfluss auf das Denken, Fühlen, Handeln und auf motivationale Aspekte. Dementsprechend können diese Erwartungen auch als das Vertrauen in die eigene Kompetenz für das Bewältigen schwieriger Anforderungssituationen bezeichnet werden (Schwarzer &

Schmitz, 1999). Im Kontext der Innovationsforschung konnte die Bedeutung der Selbstwirksamkeitserwartung für die Umsetzung von Innovationen gezeigt werden (vgl. Edelstein, 2002). Eine hohe Selbstwirksamkeitserwartung von Lehrkräften unterstützt die Beteiligung an innovativen Programmen, einen gelingenden Umgang mit Innovationen und deren Implementation (Emmrich, 2010; Gräsel, 2010; Tierney & Farmer, 2004). Die Selbstwirksamkeitserwartung von Lehrkräften hat nicht nur eine prädiktive Bedeutung für das Engagement für innovative Lehrmaßnahmen (Lohman, 2006), sondern auch für ein qualitativvolles unterrichtliches Angebot und für Schülerleistungen (Caprara, Barbaranelli, Steca & Malone, 2006; Holzberger, Philipp & Kunter, 2013).

### **3 Forschungsanliegen**

Die hier vorgestellte Studie ist Teil eines Evaluationsprojektes zur Bund- Länder-Initiative „Bildung durch Sprache und Schrift“ (BiSS). Eine Besonderheit der Innovationsmaßnahmen im Rahmen von BiSS ist, dass sich Schulen freiwillig mit eigenen, individuellen Maßnahmen, unter anderem zu Diagnostik im Leseunterricht, um eine Teilnahme bewerben konnten. Hinzu bestand für die Schulen die Möglichkeit, angebotene Maßnahmen zu nutzen und gegebenenfalls auch zu übernehmen. Wie im Theorieteil hergeleitet wurde, sind das Nutzen und die Implementation von innovativen Programmen sowohl von der Zusammenarbeit und dem zeitlichen Engagement von Lehrkräften sowie auch von individuellen motivationalen Lehrkraftmerkmalen abhängig. Inwiefern die theoretisch abgeleiteten Zusammenhänge zwischen motivationalen Lehrkraftmerkmalen, Beteiligung an innovativen Programmen und Merkmalen qualitativvollen Leseunterrichts auch in dem Kontext des BiSS-Programms empirisch nachweisbar sind, ist bislang ungeklärt und wird in der vorliegenden Studie entlang der folgenden Forschungsfragen und entsprechenden Annahmen untersucht:

1.1 Lassen sich Merkmale qualitativvollen Leseunterrichts durch die Beteiligung an innovativen Programmen vorhersagen?

H1.1 Es wird angenommen, dass (a) die Häufigkeit der Kooperation im Rahmen des Programms und (b) der Zeitaufwand der Lehrkräfte für das BiSS-Programm positive Prädiktoren für die Differenzierung und Motivierung von Schülerinnen und Schülern im Leseunterricht sind.

1.2 Lässt sich die Beteiligung an innovativen Programmen durch motivationale Lehrkraftmerkmale vorhersagen?

H1.2 Es wird angenommen, dass (a) die Motivation bezüglich Diagnostik und (b) die Selbstwirksamkeitserwartung bezüglich Diagnostik positive Prädiktoren für die Häufigkeit der Kooperation und den Zeitaufwand für das BiSS-Programm sind.

1.3 Lassen sich Merkmale qualitätvollen Leseunterrichts durch motivationale Lehrkraftmerkmale vorhersagen?

H1.3 Es wird angenommen, dass (a) die Motivation bezüglich Diagnostik und (b) die Selbstwirksamkeitserwartung bezüglich Diagnostik positive Prädiktoren für die Differenzierung und Motivierung von Schülerinnen und Schülern im Leseunterricht sind.

2 Lassen sich Merkmale qualitätvollen Leseunterrichts durch motivationale Lehrkraftmerkmale vermittelt über die Beteiligung an innovativen Programmen vorhersagen?

H2 Es wird eine Mediation der Prädiktion motivationaler Lehrkraftmerkmale von Merkmalen qualitätvollen Leseunterrichts durch die Beteiligung an innovativen Programmen erwartet.

## 4 Methode

### 4.1 Stichprobe

Die Daten wurden im Rahmen des Projekts *Evaluation von Konzepten und Maßnahmen der fachübergreifenden Leseförderung im Primarbereich (BiSS-EvalLesen)* erhoben. Für die Befragung wurden digitale Schulleitungs- und Lehrkraftfragebögen an allen BiSS-Schulen in den sieben zu evaluierenden Verbänden eingesetzt und gezielt an Lehrkräfte, die sich aktiv am BiSS-Programm beteiligen, sowie Schulleitungen adressiert. Insgesamt nahmen an  $N = 33$  Grundschulen  $N = 71$  Lehrkräfte auf freiwilliger Basis an der Erhebung im Sommer 2016 teil. Hiervon beteiligten sich  $n = 44$  Grundschullehrkräfte von den  $n = 30$  Grundschulen an der Befragung zu ihrem Leseunterricht und zu motivationalen Aspekten bezüglich des Leseunterrichts sowie bezüglich des BiSS-Programms. Durch die unterschiedliche Anzahl der im Rahmen von BiSS aktiv beteiligten Lehrkräfte am BiSS-Programm selbst variiert die Anzahl der befragten Lehrkräfte pro Schule. Aufgrund von vollständig fehlenden Werten in zentralen Variablen zur Beantwortung der Forschungsfragen ( $n = 3$ ) und einer Abweichung von mehr als zwei Standardabweichungen in einer zentralen Variable ( $n = 1$ ), mussten  $n = 4$

Lehrkräfte von den Analysen ausgeschlossen werden. Im finalen Analysesample von  $N = 40$  Lehrkräften (5.0% 25–29 Jahre, 30.0% 30–39 Jahre, 32.5% 40–49 Jahre, 27.5% 50–59 Jahre, 5.0% über 60 Jahre; 94.9% weiblich) waren  $n = 6$  Lehrkräfte als Schulleitung an der jeweiligen Grundschule tätig. Die Befragung wurde in sieben Bundesländern durchgeführt, aufgrund der vier ausgeschlossenen Lehrkräfte entfiel eine Schule (vgl. Tabelle 1).

Tabelle 1: Anzahl der befragten Lehrkräfte nach Bundesländern (Analysesample)

Bundesland	Anzahl der Schulen	Anzahl der befragten BiSS-Lehrkräfte
Brandenburg	6	8
Berlin	5	5
Baden-Württemberg	1	1
Bayern	5	9
Hessen	3	3
Rheinland-Pfalz	5	9
Thüringen	4	5

Im Durchschnitt wiesen die Lehrkräfte eine Berufserfahrung (ohne Referendariat) von  $M = 18.26$  Jahren ( $SD = 11.73$ ) auf. 72.5 Prozent der BiSS- Lehrkräfte gaben an, während ihres Lehramtsstudiums das Fach Deutsch studiert zu haben und 92.5 Prozent hatten ein Studium für die Schulform Grundschule absolviert. Im Mittel waren die Lehrkräfte  $M = 23.91$  Monate am BiSS-Programm beteiligt ( $SD = 9.92$ ,  $Min = 8.00$ ,  $Max = 40.00$ ), wobei sich der Großteil der Lehrkräfte (76.5%) freiwillig für die Teilnahme an der Projektgruppe gemeldet hatte ( $n = 6$  Schulleitungen bei dieser Angabe ausgeschlossen). An mehr als der Hälfte der Schulen (79.2%,  $n = 24$ ) gab es vor der Teilnahme am BiSS-Programm ein Leseförderkonzept. Der Großteil der Lehrkräfte (87.2%) hatte bisher schon an einem vergleichbaren Programm zu BiSS teilgenommen.

## 4.2 Instrumente

Für die Erfassung von Merkmalen der Qualität des Leseunterrichts wurden zwei Fragebogenskalen – *Differenzierung im Leseunterricht* und *Motivierung von Schülerinnen und Schülern* – eingesetzt. Die Skalen wurden unter Berücksichtigung der Itemmittelwerte bei

einer Toleranz für 50.0 Prozent fehlender Itemwerte gebildet. Die Skala *Differenzierung im Leseunterricht* wurde mittels sechs Items erfasst (adaptiert nach Hußmann, Wendt, Bos & Rieser, 2018, i. Vorb.). Die Lehrkräfte gaben auf einer vierstufigen Likertskala von „Nie oder fast nie (1)“ bis „Jeden Tag oder fast jeden Tag (4)“ an, wie häufig sie im Leseunterricht differenzieren (Beispielitem: „Ich differenziere in meinem Leseunterricht in der untersuchten Klasse für unterschiedliche Schülergruppen nach Lesefähigkeit die Leseinhalte.“). Die Skala zur Differenzierung wies eine Reliabilität von Cronbachs  $\alpha = .83$  auf.

Die Entwicklung der *Motivierung von Schülerinnen und Schülern* seit der Teilnahme am BiSS-Programm wurde mittels der Skala *Wahrgenommene Entwicklung in Bezug auf die Schülerschaft* erfasst (adaptiert aus Ostermeier, 2004). Hierfür schätzten die Lehrkräfte über drei Items auf einer vierstufigen Likertskala von „Stimme nicht zu (1)“ bis „Stimme zu (4)“ die Entwicklung ein (Beispielitem: „Seit der Teilnahme am BiSS-Projekt nehme ich wahr, dass meine Schülerinnen und Schüler sich aktiver am Leseunterricht beteiligen.“). Bei dieser Skala lag die Reliabilität in der hier verwendeten Stichprobe bei Cronbachs  $\alpha = .88$ .

Die Beteiligung an innovativen Programmen wurde über die Angabe der *Häufigkeit der Kooperation* und des *Zeitaufwandes für BiSS* erfasst. Für die Skala *Häufigkeit der Kooperation* schätzten Lehrkräfte ein, wie häufig sie im Rahmen von BiSS mit anderen BiSS-Lehrkräften zu verschiedenen Aspekten kooperieren (adaptiert aus Ramm et al., 2006). Die vierstufige Antwortskala für insgesamt sieben Items reichte von „Nie oder fast nie (1)“ bis „Jeden Tag oder fast jeden Tag (4)“ (Beispielitem: „Wie oft kooperieren Sie mit anderen Lehrkräften im Kontext von BiSS bei der Einführung und Erprobung von Diagnostikmethoden?“). Hier lag die Reliabilität in der hier verwendeten Stichprobe bei Cronbachs  $\alpha = .78$ . Der *Zeitaufwand in BiSS* wurde über ein Item erfasst, bei dem die Lehrkräfte angaben, wie viele Stunden sie durchschnittlich im Monat für die Arbeit an BiSS verwenden (Eigenentwicklung).

Zur Erfassung der motivationalen Lehrkraftmerkmale wurden Skalen zur intrinsischen *Motivation bezüglich Diagnostik im Leseunterricht* und *Selbstwirksamkeitserwartung bezüglich Diagnostik im Leseunterricht* eingesetzt; jeweils mit vierstufigen Likertskalen von „Stimme nicht zu (1)“ bis „Stimme zu (4)“. Die *Motivation bezüglich Diagnostik im Leseunterricht* wurde über vier Items operationalisiert (Beispielitem: „Schülerinnen und Schüler hinsichtlich ihrer Lesekompetenz einzuschätzen macht mir Spaß.“; adaptiert nach McElvany et al., 2010). Die Skala erreichte in der hier verwendeten Stichprobe eine Reliabilität von Cronbachs  $\alpha = .81$ . Für die *Selbstwirksamkeitserwartung bezüglich*

*Diagnostik im Leseunterricht* wurden sechs Items eingesetzt (Beispielitem: „Ich traue mir die Diagnose von Leseflüssigkeit zu.“; Eigenentwicklung). Die Skala wies in der hier verwendeten Stichprobe eine Reliabilität von Cronbachs  $\alpha = .86$  auf.

### **4.3 Analysestrategien**

Deskriptiva wurden mit SPSS 24 und Strukturgleichungsmodelle zur Überprüfung der Prädiktion der Merkmale des Leseunterrichts (Forschungsfragen 1.1–1.3) mit Mplus 7.1 (Muthén & Muthén, 1998–2015) berechnet. Alle Konstrukte waren aufgrund der Stichprobengröße manifest spezifiziert, eine Interkorrelation der Prädiktoren, Mediatoren sowie der abhängigen Variablen untereinander wurde zugelassen. Abschließend wurde ein manifestes Mediationsmodell spezifiziert (Forschungsfrage 2). Hierfür wurden neben direkten Effekten der motivationalen Lehrkraftmerkmale und der Beteiligung an innovativen Programmen auf die Differenzierung im Leseunterricht und die Motivierung von Schülerinnen und Schülern, indirekte Effekte motivationaler Lehrkraftmerkmale über die Beteiligung an innovativen Programmen auf die Merkmale des Leseunterrichts überprüft. Da an einigen Schulen mehr als eine Lehrkraft befragt wurde, wurde eine Clusterung nach Schulen in den Modellen mittels des Befehls *cluster = ID\_schul* in Mplus berücksichtigt ( $.33 < ICC < .60$ ). Fehlende Werte wurden mittels des *Full Information Likelihood Algorithmus* (FIML) geschätzt. Aufgrund der Stichprobengröße von  $N = 40$  wurde auf eine MLR-Schätzung wegen ihrer Robustheit auch bei kleineren Stichproben zurückgegriffen (Urban & Mayerl, 2003).

## **5 Ergebnisse**

### **5.1 Deskriptive Ergebnisse**

Die deskriptiven Ergebnisse zeigten, dass BiSS-Lehrkräfte im Mittel einschätzten, im Leseunterricht ein- bis zweimal pro Woche Schülerinnen und Schüler nach der Lesefähigkeit zu differenzieren (vgl. Tabelle 2). Am häufigsten differenzierten sie in Bezug auf die zu bearbeitende Stoffmenge (Textlänge und Aufgabenstellungen zum Text) ( $M = 3.32$ ,  $SD = 0.78$ ,  $Min = 1.00$ ,  $Max = 4.00$ ) sowie die zur Verfügung stehende Bearbeitungszeit ( $M = 3.32$ ,  $SD = 0.78$ ,  $Min = 1.00$ ,  $Max = 4.00$ ). Im Bereich der Textgenres differenzierten die Lehrkräfte am wenigsten nach der Lesefähigkeit der Schülerinnen und Schüler ( $M = 2.05$ ,  $SD = 0.91$ ,  $Min = 1.00$ ,  $Max = 4.00$ ). Hinsichtlich der wahrgenommenen Entwicklung der Motivierung von Schülerinnen und Schülern stimmten die Lehrkräfte durchschnittlich eher

zu, dass sich die Kinder am Leseunterricht häufiger beteiligten ( $M = 2.90$ ,  $SD = 0.72$ ,  $Min = 1.00$ ,  $Max = 4.00$ ), mehr Freude ( $M = 2.76$ ,  $SD = 0.86$ ;  $Min = 1.00$ ,  $Max = 4.00$ ) und Interesse zeigten ( $M = 2.97$ ,  $SD = 0.78$ ,  $Min = 1.00$ ,  $Max = 4.00$ ). Im Mittel gaben die Lehrkräfte an, ein- bis zweimal im Monat mit Lehrkräften im Kontext von BiSS zu kooperieren. Für die Kooperation mit anderen Lehrkräften bezüglich Diskussionen von Schülerleistungen im Bereich Lesen ( $M = 2.15$ ,  $SD = 0.84$ ,  $Min = 1.00$ ,  $Max = 4.00$ ) und Einführung sowie Erprobung neuer Unterrichtsideen ( $M = 2.03$ ,  $SD = 0.72$ ,  $Min = 1.00$ ,  $Max = 4.00$ ) gaben die befragten Lehrkräfte die höchste Ausprägung an. Am wenigsten arbeiteten Lehrkräfte im Hinblick auf gegenseitige Unterrichtsbeobachtungen zusammen ( $M = 1.31$ ,  $SD = 0.57$ ,  $Min = 1.00$ ,  $Max = 3.00$ ). Im Mittel wendeten die Lehrkräfte  $M = 7.09$  Stunden im Monat für die Arbeit im BiSS-Programm auf. Ihre eigene Motivation bezüglich Diagnostik im Leseunterricht wurde von den Lehrkräften durchschnittlich hoch eingeschätzt. Materialien für die Leseförderung dahingehend zu kontrollieren, ob die Anforderungen dem Leistungsniveau der Klasse entsprechen, bereitete den Lehrkräften besonders viel Freude ( $M = 3.33$ ,  $SD = 0.62$ ,  $Min = 2.00$ ,  $Max = 4.00$ ). Am wenigsten motiviert waren Lehrkräfte, die Schwierigkeit von Materialien für die Leseförderung ( $M = 3.13$ ,  $SD = 0.73$ ,  $Min = 2.00$ ,  $Max = 4.00$ ) sowie die Lesekompetenz der Schülerinnen und Schüler einzuschätzen ( $M = 3.13$ ,  $SD = 0.61$ ,  $Min = 2.00$ ,  $Max = 4.00$ ). Ebenfalls fiel die Einschätzung für die Selbstwirksamkeitserwartung bezüglich Diagnostik im Leseunterricht im Durchschnitt eher hoch aus. Insbesondere trauten sich BiSS-Lehrkräfte die Diagnose der Leseflüssigkeit ( $M = 3.46$ ,  $SD = 0.60$ ,  $Min = 2.00$ ,  $Max = 4.00$ ) und des Leseverständnisses ( $M = 3.31$ ,  $SD = 0.70$ ,  $Min = 2.00$ ,  $Max = 4.00$ ) zu. Eher weniger trauten sich Lehrkräfte die Diagnose des Wortschatzes ( $M = 2.85$ ,  $SD = 0.67$ ,  $Min = 2.00$ ,  $Max = 4.00$ ) sowie der Lesestrategien ( $M = 2.92$ ,  $SD = 0.77$ ,  $Min = 1.00$ ,  $Max = 4.00$ ) zu.

Tabelle 2: Deskriptiva, Interkorrelation zwischen den Variablen

Variablen	Deskriptiva			Interkorrelationen					
	<i>M</i>	<i>SD</i>	Fehlen de Werte (%)	2	3	4	5	6	
1 Differenzierung im Leseunterricht <sup>1</sup>	2.85	0.64	5.0	.35*	.20	.46**	.47**	.30	
2 Motivierung von Schülerinnen und Schülern <sup>2</sup>	2.88	0.70	2.5		.67**	.28	.60**	.47**	
3 Häufigkeit der Kooperation <sup>1</sup>	1.79	0.47	5.0			.38**	.53**	.23	
4 Zeitaufwand für BiSS <sup>3</sup>	7.09	3.44	17.5				.38**	.20	
5 Motivation bezüglich Diagnostik <sup>2</sup>	3.21	0.54	2.5					.47**	
6 Selbstwirksamkeitserwartung bezüglich Diagnostik <sup>2</sup>	3.11	0.53	2.5					-	

Anmerkungen. <sup>1</sup> = „Nie oder fast nie (1)“ bis „Jeden Tag oder fast jeden Tag (4)“, <sup>2</sup> = „Stimme nicht zu (1)“ bis „Stimme zu (4)“, <sup>3</sup> = Stunden/Monat; \*  $p < .05$ , \*\*  $p < .01$

Die Interkorrelationen zwischen den Skalen sind in Tabelle 2 dargestellt. Jedes motivationale Lehrkraftmerkmal sowie ein Faktor der Beteiligung an innovativen Programmen wies mindestens zu einem Merkmal qualitätvollen Leseunterrichts eine mittelhohe Korrelation auf. Hinsichtlich der Motivation der Lehrkräfte bezüglich Diagnostik im Leseunterricht wurden signifikante mittlere bis hohe Zusammenhänge zu allen anderen Skalen nachgewiesen, während die Selbstwirksamkeitserwartung lediglich mit der Motivierung von Schülerinnen und Schülern und der Lehrkraftmotivation statistisch signifikant korrelierte. Die Maße zur Beteiligung an innovativen Programmen wiesen jeweils nur zu einem Merkmal qualitätvollen Leseunterrichts signifikante Zusammenhänge auf.

## 5.2 Prädiktoren für die Merkmale des Leseunterrichts sowie für die Beteiligung an innovativen Programmen

Das für die Beantwortung der Frage 1.1 spezifizierte manifeste Strukturgleichungsmodell I untersuchte, ob die Maße der Beteiligung an innovativen Programmen eine prädiktive Kraft für die Merkmale des Leseunterrichts aufweisen. Es zeigte signifikante Koeffizienten für die Häufigkeit der Kooperation bezüglich der Motivierung von Schülerinnen und Schülern sowie

für den Zeitaufwand für BiSS bezüglich der Differenzierung im Leseunterricht (vgl. Tabelle 3). Lehrkräfte, die häufiger kooperierten, schätzten die Entwicklung der Motivation von Schülerinnen und Schülern höher ein. Lehrkräfte mit einem höheren Zeitaufwand für das BiSS-Programm differenzierten häufiger im Leseunterricht. Die beiden unabhängigen Variablen korrelierten mittelhoch miteinander ( $r = .41, p < .01$ ). Für die abhängigen Variablen zeigte sich keine statistisch signifikante Korrelation. Die Resultate unterstützen empirisch die Hypothese 1.1.

In Bezug auf die Prüfung der Annahme, dass sich die Beteiligung an innovativen Programmen durch motivationale Lehrkraftmerkmale vorhersagen lässt (H 1.2), zeigten sich signifikant positiv gerichtete Zusammenhänge zwischen der Motivation bezüglich Diagnostik im Leseunterricht und den beiden Maßen der Beteiligung an innovativen Programmen. Für die Selbstwirksamkeitserwartung hingegen konnte kein Zusammenhang gefunden werden (vgl. Tabelle 3). Lehrkräfte mit einer höheren Motivation bezüglich Diagnostik wendeten mehr Zeit für BiSS auf und arbeiteten häufiger im Kollegium zusammen. Im Strukturgleichungsmodell II korrelierten die Motivation bezüglich Diagnostik und die Selbstwirksamkeitserwartung bezüglich Diagnostik in mittlerer Höhe miteinander ( $r = .47, p < .00$ ). Die abhängigen Variablen zeigten keine signifikante Korrelation untereinander. Dieses Ergebnis stützte nur teilweise die aus der Theorie abgeleiteten Annahmen der Hypothese 1.2.

Das manifeste Strukturgleichungsmodell III zeigte hinsichtlich der Annahme der Hypothese 1.3, dass Merkmale qualitätvollen Leseunterrichts durch motivationale Lehrkraftmerkmale vorhergesagt werden können, lediglich für die Motivation bezüglich Diagnostik signifikant gerichtete Zusammenhänge (vgl. Tabelle 3). Eine höhere Motivation bezüglich Diagnostik ging mit einer häufigeren Differenzierung im Leseunterricht sowie mit einer höheren wahrgenommenen Motivierung von Schülerinnen und Schülern einher. Auch in diesem Strukturgleichungsmodell III korrelierten die motivationalen Lehrkraftmerkmale ( $r = .47, p < .01$ ), und die abhängigen Variablen wiesen untereinander keine signifikante Korrelation auf. Aufgrund der fehlenden Prädiktionskraft der Selbstwirksamkeitserwartung bezüglich Diagnostik für die Merkmale des Leseunterrichts kann die Hypothese 1.3 nur eingeschränkt unterstützt werden.

Tabelle 3: Prädiktoren von Beteiligung an innovativen Programmen für die Merkmale des Leseunterrichts (Angaben zu standardisierten  $\beta$ -Koeffizienten und Standardfehlern)

Modell	Prädiktoren	Merkmale qualitätvollen Leseunterrichts				Beteiligung an innovativen Programmen				
		Differenzierung im Leseunterricht		Motivierung von Schülerinnen und Schülern		Häufigkeit der Kooperation		Zeitaufwand für BiSS		
		$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	
I	Häufigkeit der Kooperation	-.02	0.14	.65**	0.10	-	-	-	-	
	Zeitaufwand für BiSS	.48**	0.15	.06	0.13	-	-	-	-	
	R <sup>2</sup>	.22		.45**						
	<i>Anmerkungen.</i> $\chi^2 = 24.93$ , $df = 5$ , $p < .01$ ; $CFI = 1.00$ ; $RMSEA = .00$ ; $SRMR = .00$ ; $AIC = 365.70$ ; $BIC = 389.35$ ; * $p < .05$ , ** $p < .01$									
II	Motivation bezüglich Diagnostik	-	-	-	-	.53**	0.12	.35*	0.14	
	Selbstwirksamkeitserwartung bezüglich Diagnostik	-	-	-	-	-.00	0.14	.06	0.20	
	R <sup>2</sup>					.28*		.15		
	<i>Anmerkungen.</i> $\chi^2 = 15.15$ , $df = 5$ , $p < .05$ ; $CFI = 1.00$ ; $RMSEA = .00$ ; $SRMR = .00$ ; $AIC = 342.82$ ; $BIC = 366.46$ ; * $p < .05$ , ** $p < .01$									
III	Motivation bezüglich Diagnostik	.42**	0.11	.50**	0.15	-	-	-	-	
	Selbstwirksamkeitserwartung bezüglich Diagnostik	.09	0.17	.23	0.12	-	-	-	-	
	R <sup>2</sup>	.22		.40						
	<i>Anmerkungen.</i> $\chi^2 = 24.25$ , $df = 5$ , $p < .01$ ; $CFI = 1.00$ ; $RMSEA = .00$ ; $SRMR = .00$ ; $AIC = 264.62$ ; $BIC = 288.26$ ; * $p < .05$ , ** $p < .01$									

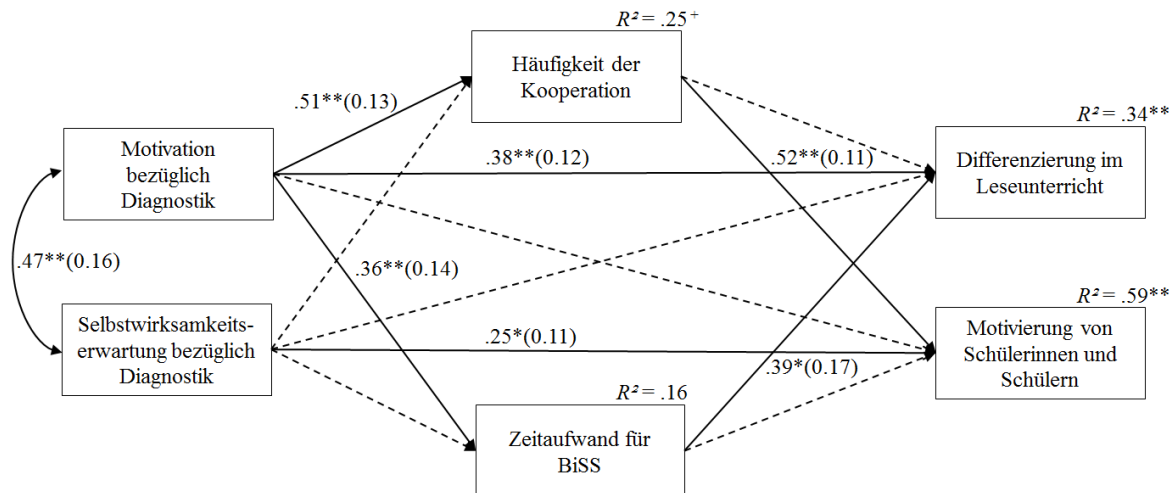
### 5.3 Die Bedeutung der motivationalen Lehrkraftmerkmale unter Berücksichtigung der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht

Das Mediationsmodell zeigte für die Beantwortung der zweiten Forschungsfrage, ob sich zentrale Aspekte der Qualität des Leseunterrichts durch motivationale Lehrkraftmerkmale über die Beteiligung an innovativen Programmen vorhersagen lassen, signifikante gerichtete Zusammenhänge. Die motivationalen Lehrkraftmerkmale wiesen eine Prädiktionskraft für die Beteiligung an innovativen Programmen und für die Merkmale des Leseunterrichts auf sowie die Beteiligung an innovativen Programmen für die Merkmale des Leseunterrichts (vgl.

Abbildung 1).

Abbildung 1: Mediationsanalyse zur Vorhersagekraft motivationaler Lehrkraftmerkmale über die Beteiligung an innovativen Programmen für die Qualität des Leseunterrichts

(Angaben zu standardisierten a-Koeffizienten und Standardfehlern).



Anmerkungen.  $\chi^2 = 57.54$ ,  $df = 14$ ,  $p < .01$ ;  $CFI = 1.00$ ;  $RMSEA = .00$ ;  $SRMR = .00$ ;  $AIC = 474.27$ ;  $BIC = 518.87$ ;  $^{+} p < .10$ ,  $^{*} p < .05$ ,  $^{**} p < .01$

Für die Annahme, dass sich indirekte Effekte für motivationale Lehrkraftmerkmale auf die Merkmale des Leseunterrichts über die Beteiligung an innovativen Programmen finden lassen, zeigte die Mediationsanalyse einen signifikanten indirekten Pfad zwischen der Motivation bezüglich Diagnostik im Leseunterricht und der wahrgenommenen Entwicklung der Motivierung von Schülerinnen und Schülern im Unterricht, der über die Häufigkeit der Kooperation vermittelt wurde (indirect =  $.26$ ,  $p = .01$ ) (vgl. Abbildung 1). Bezüglich der weiteren untersuchten indirekten Pfade zwischen den motivationalen Lehrkraftmerkmalen und den Merkmalen qualitätvollen Leseunterrichts über die Beteiligung an innovativen Programmen, die in der Abbildung 1 gestrichelt dargestellt sind, konnten keine signifikanten Pfade identifiziert werden. Die erklärte Varianz der Merkmale des Unterrichts in dem Gesamtmodell betrug  $R^2 = .34$  für die Differenzierung im Leseunterricht und  $R^2 = .59$  für die wahrgenommene Entwicklung der Motivierung von Schülerinnen und Schülern im Leseunterricht. Die Hypothese 2 kann aufgrund keiner vollständigen Mediation zwischen der Beteiligung an innovativen Maßnahmen und Aspekten der Qualität des Leseunterrichts sowie nur eines tendenziell signifikanten indirekten Pfades mit Einschränkungen als empirisch unterstützt angesehen werden.

## 6 Diskussion

Die vorliegende Studie beschäftigte sich mit der Vorhersagekraft von motivationalen Lehrkraftmerkmalen über die Beteiligung an innovativen Programmen für Merkmale qualitätvollen Leseunterrichts. Hypothesenkonforme Ergebnisse der Strukturgleichungsmodelle wiesen auf Zusammenhänge der Motivation von Lehrkräften sowohl mit einem differenzierten Lernangebot für Schülerinnen und Schüler als auch mit der Motivierung von Schülerinnen und Schülern hin. Ferner zeigte sich im Mediationsmodell, dass Lehrkräfte mit einer höheren Ausprägung ihrer Selbstwirksamkeitserwartung bezüglich Diagnostik im Leseunterricht die Steigerung der Motivierung von Schülerinnen und Schülern in ihrem Unterricht höher einschätzten. Die Motivation von Lehrkräften hatte ebenfalls eine große Relevanz für die Beteiligung an innovativen Programmen.

Unerwartet erwies sich die Selbstwirksamkeitserwartung in den Regressionsanalysen als nicht signifikant bedeutsam für die Beteiligung an innovativen Programmen sowie für die Merkmale qualitätvollen Leseunterrichts. Die Höhe der Koeffizienten erreichte jedoch teilweise einen Wert von  $\beta = .23$ ; bei einer Replikation der Studie mit einer größeren Stichprobe könnten diese Zusammenhänge ein Signifikanzniveau von 5.0 Prozent erreichen.

Die relativ hohen Werte der Motivation und Selbstwirksamkeitserwartung der befragten BiSS-Lehrkräfte können darauf hinweisen, dass sich primär Lehrkräfte mit hohen motivationalen Orientierungen in der Umsetzung von innovativen Programmen engagieren und in Folge an der Befragung teilnahmen. Darüber hinaus wurden Schulen untersucht, die sich freiwillig mit ihren Innovationsmaßnahmen an BiSS beteiligt haben. Trotz dieser möglichen Positivselektion der Stichprobe und der damit gegebenenfalls einhergehenden Varianzeinschränkung konnten viele der abgeleiteten Zusammenhangsmuster empirisch gefunden werden. In früheren empirischen Studien wurden ebenfalls motivationale Lehrkraftmerkmale als wichtige Prädiktoren für die Akzeptanz und Umsetzungen von innovativen Programmen (z. B. Emmrich, 2010) sowie für einen qualitätvollen Unterricht identifiziert (Roth et al., 2007) nachgewiesen. Vergleichbar hohe Werte wurden auch in der Studie von McElvany et al. (2010) für die Motivation von Lehrkräften gefunden ( $M = 3.22$ ,  $SD = 0.53$ ), deren Teilnahme an einer Videographiestudie ebenfalls freiwillig war.

Die Beteiligung an innovativen Programmen nahm in der Studie eine besondere Rolle als Mediator zwischen den motivationalen Lehrkraftmerkmalen und der Qualität des Leseunterrichts ein. Der indirekte Effekt aus der Mediationsanalyse unterstützt die Wichtigkeit der Motivation von Lehrkräften bezüglich Diagnostik für die Motivierung von

Schülerinnen und Schülern, die über die Häufigkeit der Kooperation mit anderen BiSS-Lehrkräften mediiert wird. An dieser Stelle lassen die Ergebnisse die Vermutung aufkommen, dass eine häufigere Differenzierung im Leseunterricht mit einem hohen Zeitaufwand für BiSS durch eine intensivere Auseinandersetzung mit Förder- und Diagnosemaßnahmen im Rahmen des BiSS-Programms begründet sind. Das Resultat, dass Lehrkräfte, die intensiver im Rahmen des BiSS-Programms zusammenarbeiteten, eine stärkere Entwicklung der Motivierung von Schülerinnen und Schülern wahrnahmen, bestätigt die wichtige Rolle der Kooperation für eine erfolgreiche Umsetzung von innovativen Programmen, welche zu einem verbesserten Unterricht führen können (Holtappels, 2013).

Die Angaben der Lehrkräfte bezüglich ihres Zeitaufwandes für das BiSS- Programm lagen mit einem Mittelwert von etwas unterhalb von zwei Stunden die Woche nur leicht über der in der Literatur ermittelte durchschnittlich aufgewendete Zeit für Programme von ca. einer Stunde pro Woche (Philipp & Kunter, 2013). Auch waren die etwa zwei Zeitstunden in der Woche für die Arbeit in BiSS im Bereich der von Gunter et al. (2005) beschriebenen sechs Prozent der Arbeitszeit (ca. drei Stunden pro Woche bei einer Vollzeitstelle), die für schulorganisatorische Angelegenheiten wie zum Beispiel Programmarbeit von Sekundarlehrkräften aufgebracht wird. Dementsprechend gaben Lehrkräfte für die Beteiligung am BiSS-Programm einen durchschnittlichen Zeitaufwand für das außerunterrichtliche Engagement für die Programmarbeit an. Jedoch konnte in dieser Befragung nicht ermittelt werden, ob sich die Lehrkräfte mehr Zeit und Freiraum für die Beteiligung am BiSS-Programm wünschen. Bei der Betrachtung der Ergebnisse ist zu berücksichtigen, dass 82.1 Prozent der befragten Lehrkräfte eine Vollzeitstelle haben.

Für die Häufigkeit der Kooperation gaben die Lehrkräfte an, lediglich ein- bis zweimal Mal pro Monat zusammen im Kontext von BiSS zu kooperieren. Die geringe Zusammenarbeit konnte jedoch auch in anderen Umfragen zur Kooperation von Lehrkräften an deutschen Schulen festgestellt werden (Bos et al., 2016). Da aber eine intensive Zusammenarbeit auch für das Umsetzen innovativer Maßnahmen und deren Implementation von großer Bedeutung ist, besteht der Bedarf auf die mangelnde Kooperation aufmerksam zu machen und diese zu fördern. Hierfür würden sich insbesondere die Fort- und Weiterbildungsangebote im Rahmen von BiSS anbieten. Auch in der SINUS-Untersuchung wurden nicht durchgängig hohe Kooperationsniveaus nachgewiesen und festgestellt, dass sich die Lehrerverkooperation erst im Laufe der Projektteilnahme entwickelte (Ostermeier, 2004). Des Weiteren gaben in der Befragung unter 15.0 Prozent an, gemeinsam schon in anderen Projekten oder Teams

zusammengearbeitet zu haben. Dies kann ein Grund für eine bisher eher geringe Kooperation im Team sein. Weitere Gründe für eine geringe Kooperation können eine unterschiedliche Zielvorstellung sowie Unterrichtsstile, die Einschränkung der individuellen Arbeitsweise von Lehrkräften und der Wunsch nach mehr Zeit für Kooperationsgespräche mit den Kolleginnen und Kollegen sein (Gräsel et al., 2006; Schaarschmidt et al., 2007).

Insgesamt deuten die positiv gerichteten Zusammenhänge zwischen der Beteiligung an innovativen Programmen und der Qualität des Leseunterrichts auf die Effektivität der BiSS-Maßnahmen hin, wobei die angenommene Wirkrichtung in einer Längsschnittstudie untersucht werden muss. Inwieweit die motivationalen Lehrkraftmerkmale nicht nur die Beteiligung an innovativen Maßnahmen, wie zum Beispiel BiSS-Maßnahmen, vorhersagt, sondern auch durch diese beeinflusst werden, könnte in Folgestudien untersucht werden.

Im Hinblick auf zukünftige Forschung ist zu bedenken, dass in der vorliegenden Untersuchung alle Variablen auf Selbstaussagen von Lehrkräften beruhen. Jedoch muss aufgrund der Gewährleistung der Anonymität nicht von einem großen Effekt der sozialen Erwünschtheit und einer hohen Antwortverzerrung ausgegangen werden (Franzen & Pointner, 2015). Für eine Folgestudie zur Überprüfung des Modells könnten diese und weitere Unterrichtsqualitätsmerkmale über Schüleraussagen oder videobasierte Unterrichtsbeobachtungen erfasst werden. Zudem muss die Stichprobengröße von  $N = 40$  als Einschränkung gesehen werden. So kann eine zu kleine Stichprobengröße in Strukturgleichungsmodellen zu Verzerrungen der Teststatistik und zu einer Verschlechterung der Modellschätzungen führen (Algina & Olejnik, 2000; Muthén & Muthén, 2002). Es wurden nur Lehrkräfte, die im BiSS-Programm aktiv sind, befragt, sodass die Auswahl der Stichprobe selektiv im Vergleich zur Gesamtpopulation der Grundschullehrkräfte war. Die Ergebnisse können aufgrund der Annahme, dass an der Erhebung eher motivierte Lehrkräfte der Schulen teilgenommen haben, verzerrt und besonders gerichtete Zusammenhänge hinsichtlich der motivationalen Orientierungen überschätzt werden. Die angenommene Wirkrichtung der Effekte muss in einer Längsschnittstudie geprüft werden.

Die Studie leistet insbesondere durch die Identifikation der Bedeutung der Beteiligung an BiSS-Maßnahmen für einen qualitätvollen Leseunterricht, wichtige erste Ergebnisse des Evaluationsprojekts BiSS-EvalLesen. Aus den Resultaten lassen sich konkrete Implikationen für die Praxis ableiten, welche im noch laufenden BiSS-Programm umgesetzt werden können.

## Literatur

- Algina, J. & Olejnik, S. (2000). Determining Sample Size for Accurate Estimation of the Squared Multiple Correlation Coefficient. *Multivariate Behavioral Research*, 35, 119–136.
- Altrichter, H. & Wiesinger, S. (2005). Implementation von Schulinnovationen – aktuelle Hoffnungen und Forschungswissen. *Journal für Schulentwicklung*, 9 (4), 28–36.
- Axtell, C. M., Holman, D. J., Unsworth, K. L., Wall, T. D., Waterson, P. E. & Harrington, E. (2000). Shopfloor innovation: Facilitating the suggestion and implementation of ideas. *Journal of Occupational and Organizational Psychology*, 73 (3), 265–285.
- Bakker, A. B., Schaufeli, W. B., Leiter, M. P. & Taris, T. W. (2008). Work engagement. An emerging concept in occupational health psychology. *Work & Stress*, 22 (3), 187–200.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bauer, J., Unterbrink, T., Hack, A., Pfeifer, R., Buhl-Griesshaber, V., Muller, U. et al. (2007). Working conditions, adverse events and mental health problems in a sample of 949 German teachers. *International Archives of Occupational and Environmental Health*, 80 (5), 442–449.
- Bayerisches Staatsministerium für Unterricht und Kultus & Staatsinstitut für Schulqualität und Bildungsforschung (Hrsg.). (2010). *ProLesen. Auf dem Weg zur Leseschule Leseförderung in den gesellschaftswissenschaftlichen Fächern. Aufsätze und Materialien aus dem KMK-Projekt „ProLesen“*. Donauwörth: Auer.
- Becker-Mrotzek, M., Hasselhorn, M., Roth, H.-J. & Stanat, P. (2016). Die Bund-Länder Initiative „Bildung durch Sprache und Schrift (BISS)“. *Psychologie in Erziehung und Unterricht*, 63 (1), 75–77.
- Bieg, S. & Mittag, W. (2009). Die Bedeutung von Unterrichtsmerkmalen und Unterrichtsemotionen für die selbstbestimmte Lernmotivation. Effects of classroom conditions and learning emotions on self-determined motivation of adolescents. *Empirische Pädagogik*, 23 (2), 117–142.
- Bohl, T., Batzel, A. & Richey, P. (2011). Öffnung – Differenzierung – Individualisierung – Adaptivität. Charakteristika, didaktische Implikationen und Forschungsbefunde

- verwandter Unterrichtskonzepten zum Umgang mit Heterogenität. *Schulpädagogik heute*, 2 (4), 1–23.
- Bos, W., Pietsch, M., List, M. K., Guill, K., Gröhlich, C., Scharenberg, K. et al. (2016). *KESS 4: Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) an der Universität Kiel.
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, D.C.: National Academy Press.
- Butt, G. & Lance, A. (2005). Secondary teacher workload and job satisfaction. *Educational management administration & leadership*, 33 (4), 401–422.
- Caprara, G. V., Barbaranelli, C., Steca, P. & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement. A study at the school level. *Journal of School Psychology*, 44 (6), 473–490.
- Chaplain, R. P. (2008). Stress and psychological distress among trainee secondary teachers in England. *Educational Psychology*, 28 (2), 195–209.
- Chinn, C. A., Anderson, R. C. & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, 36 (4), 378–411.
- Christie, J. F., Enz, B. & Vukelich, C. (2007). *Teaching language and literacy. Preschool through the elementary grades*. Boston, MA: Pearson.
- Clegg, C., Unsworth, K., Epitropaki, O. & Parker, G. (2002). Implicating trust in the innovation process. *Journal of Occupational & Organizational Psychology*, 75 (4), 409–422.
- Coubergs, C., Struyven, K., Vanthournout, G. & Engels, N. (2017). Measuring teachers' perceptions about differentiated instruction: The DI-Quest instrument and model. *Studies in Educational Evaluation*, 53, 42–54.

- Deci, E. L. & Ryan, R. M. (2000). The what and why of goal pursuits: human needs and the self-determination of behavior. *Psychology Inquiry*, 11, 227–268.
- Edelstein, W. (2002). Selbstwirksamkeit, Innovation und Schulreform. Zur Diagnose der Situation. In M. Jerusalem & D. Hopf (Hrsg.), *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (S. 13–27). Weinheim: Beltz.
- Emmrich, R. (2010). *Motivstrukturen von Lehrerinnen und Lehrern in Innovations- und Transferkontexten* (Berliner Beiträge zur Pädagogik, Bd. 8.). Frankfurt a. M.: Lang.
- Ferdinand, H. D. (2014). Entwicklung von Fachinteresse. Längsschnittstudie zu Interessenverläufen und Determinanten positiver Entwicklung in der Schule (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 8). Münster: Waxmann.
- Fischer, N. & Rustemeyer, R. (2007). Motivationsentwicklung und schülerperzipiertes Lehrkraftverhalten im Mathematikunterricht. *Zeitschrift für Pädagogische Psychologie*, 21 (2), 135–144.
- Franzen, A. & Pointner, S. (2015). Anonymisierungstechniken in sozialwissenschaftlichen Laborexperimenten. In M. Keuschnigg & T. Wolbring (Hrsg.), *Experimente in den Sozialwissenschaften* (Sonderband Soziale Welt, Bd. 22, S. 155–173). Baden-Baden: Nomos.
- Fullan, M. G. (2007). *The new meaning of educational change* (4. Aufl.). New York, NY: Teachers College Press.
- Fussangel, K. & Gräsel, C. (2012). Lehrerkooperation aus der Sicht der empirischen Schulforschung. In E. Baum, T.-S. Idel & H. Ullrich (Hrsg.), *Kollegialität und Kooperation in der Schule. Theoretische Konzepte und empirische Befunde* (S. 29–40). Wiesbaden: VS Verlag.
- Goldenbaum, A. (2012). Innovationsmanagement in Schulen. Eine empirische Untersuchung zur Implementation eines Sozialen Lernprogramms (Dissertation, Freie Universität Berlin, 2011). Wiesbaden: VS Verlag.
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13 (1), 7–20.

- Gräsel, C., Fussangel, K. & Pröbstel, C. (2006). Lehrkräfte zur Kooperation anregen – eine Aufgabe für Sisyphos? *Zeitschrift für Pädagogik*, 52 (2), 205–219.
- Gunter, H., Rayner, S., Thomas, H., Fielding, A., Butt, G. & Lance, A. (2005). Teachers, time and work. Findings from the Evaluation of the Transforming the School Workforce Pathfinder Project. *School Leadership & Management*, 25 (5), 441–454.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. überarbeitete Aufl., Schule weiterentwickeln – Unterricht verbessern. Orientierungsband). Seelze: Klett-Kallmeyer.
- Hertel, S. (2014). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. In B. Kopp & B. Arend (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft [21. Jahrestagung der Kommission Grundschulforschung und Pädagogik der Primarstufe 2012 in Nürnberg]* (Jahrbuch Grundschulforschung, Bd. 17, S. 19–34). Wiesbaden: Springer.
- Hidi, S. & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41 (2), 111–127.
- Holtappels, H. G. (2013). Schulentwicklung und Lehrerkooperation. In N. McElvany & H. G. Holtappels (Hrsg.), *Empirische Bildungsforschung. Theorien, Methoden, Befunde und Perspektiven* (Festschrift für Wilfried Bos, S. 35–62). Münster: Waxmann.
- Holtappels, H. G., Pfeiffer, H., Röhrich, T. & Voss, A. (2008). Einfluss von Prozessmerkmalen der Organisations- und Unterrichtsentwicklung auf die Lernleistungen. In H. G. Holtappels, K. Klemm & H.-G. Rolff (Hrsg.), *Schulentwicklung durch Gestaltungsautonomie. Ergebnisse der Begleitforschung zum Modellvorhaben Selbstständige Schule in Nordrhein-Westfalen* (S. 314–330). Münster: Waxmann.
- Holtappels, H. G. & Heerdegen, M. (2005). Schülerleistungen in unterschiedlichen Lernumwelten im Vergleich zweier Grundschulmodelle in Bremen. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther, (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 361–397). Münster: Waxmann.

- Holzberger, D., Philipp, A. & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality. A longitudinal analysis. *Journal of Educational Psychology*, 105 (3), 774–786.
- Hußmann, A., Wendt, H., Bos, W. & Rieser, S. (Hrsg.). (2018, i. Vorb.). *IGLU 2016. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Münster: Waxmann.
- Jäger, M. (2004). *Transfer in Schulentwicklungsprojekten*. Wiesbaden: Springer.
- Jang, S.-J. (2006). Research on the effects of team teaching upon two secondary school teachers. *Educational Research*, 48 (2), 177–194.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222–237.
- 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.
- Kyriacou, C. & Coulthard, M. (2000). Undergraduates' views of teaching as a career choice. *Journal of Education for Teaching*, 26 (2), 117–126.
- Lankes, E.-M. (2004). Leseunterricht in der Grundschule. Unterschiede zwischen Lehrkräften im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 7 (4), 551–568.
- Lankes, E.-M. & Carstensen, C. H. (2007). Der Leseunterricht aus der Sicht der Lehrkräfte. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 161–193). Münster: Waxmann.

- Lenske, G. (2016). *Schülerfeedback in der Grundschule. Untersuchung zur Validität* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 92). Münster: Waxmann.
- Leuders, T. & Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht. *Unterrichtswissenschaft*, 39 (3), 213–230.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenz und Kompetenzentwicklung von Lehrerinnen und Lehrern. Zeitschrift für Pädagogik*, 51, Beiheft (S. 47–70). Weinheim: Beltz.
- Lipowsky, F., Kastens, C., Lotz, M. & Faust, G. (2011). Aufgabenbezogene Differenzierung und Entwicklung des verbalen Selbstkonzepts im Anfangsunterricht. *Zeitschrift für Pädagogik*, 57 (6), 868–884.
- Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K. & Pauli, C. (2009). Quality of geometry instruction and its short-term impact on students? Understanding of Pythagorean Theorem. *Learning and Instruction*, 19 (6), 527–537.
- Lohman, M. C. (2006). Factors influencing teachers' engagement in informal learning activities. *Journal of Workplace Learning*, 18 (3), 141–156.
- Long, J. F. & Hoy, A. W. (2006). Interested instructors: A composite portrait of individual differences and effectiveness. *Teaching and Teacher Education*, 22 (3), 303–314.
- McDonald Connor, C., Piasta, S. B., Fishman, B. J., Glasney, S., Schatschneider, C., Crowe, E. et al. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development*, 80 (1), 77–100.
- McElvany, N., Schroeder, S., Richter, T., Hachfeld, A., Baumert, J., Schnotz, W., Horz, H. & Ullrich, M. (2010). Texte mit instruktionalen Bildern als Unterrichtsmaterial – Kompetenzen der Lehrkräfte. *Unterrichtswissenschaft*, 38 (2), 98–116.
- Müller, F. H., Hanfstingl, B. & Andreitz, I. (2009). Bedingungen und Auswirkungen selbstbestimmter Lehrermotivation. *Erziehung und Unterricht*, 159 (1/2), 142–152.
- Muthén, L. K. & Muthén, B. O. (1998–2015). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

- Muthén, L. K. & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620.
- Ostermeier, C. (2004). *Kooperative Qualitätsentwicklung in Schulnetzwerken. Eine empirische Studie am Beispiel des BLK-Programms „Steigerung der Effizienz des mathematisch-natur- wissenschaftlichen Unterrichts“* (SINUS). Münster: Waxmann.
- Pant, H., Vock, M., Pöhlmann, C. & Köller, O. (2008). Offenheit für Innovationen. Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, 54 (6), 827–845.
- Philipp, A. & Kunter, M. (2013). How do teachers spend their time? A study on teachers' strategies of selection, optimisation, and compensation over their career cycle. *Teaching and Teacher Education*, 35 (1), 1–12.
- 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.
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 13 (1), 121–148.
- Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D et al. (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Rogers, E. M. (2003). *Diffusion of innovations* (5. Aufl.). New York, NY: Free Press.
- Roness, D. (2011). Still motivated? The motivation for teaching during the second year in the profession. *Teaching and Teacher Education*, 27 (3), 628–638.
- Roth, G., Assor, A., Kanat-Maymon, Y. & Kaplan, H. (2007). Autonomous motivation for teaching. How self-determined teaching may lead to self-determined learning. *Journal of Educational Psychology*, 99 (4), 761–774.
- 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.
- Schaarschmidt, U., Sieland, B., Fischer, A. & Rahm, T. (2007). *Die Arbeitszeit der Lehrerinnen und Lehrer in Nordrhein-Westfalen. Ergebnisse und Vorschläge der*

*Projektgruppe QuaGiS zur Entwicklung eines zukunftsfähigen Arbeitszeitmodells.*  
Dortmund: VBE NRW.

Schaufeli, W. B., Salanova, M., González-Romá, V. & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3 (1), 71–92.

Schneider, W., Baumert, J., Becker-Mrotzek, M., Hasselhorn, M., Kammermeyer, G., Rauschenbach, T. et al. (2012). *Expertise. „Bildung durch Sprache und Schrift (BISS)“ (Bund- Länder-Initiative zur Sprachförderung, Sprachdiagnostik und Leseförderung).* Zugriff am 01.06.2017. Verfügbar unter <http://www.biss-sprachbildung.de/pdf/BiSS-Expertise.pdf>

Schwartz, K., Klusmann, U., Ihme, T. & Möller, J. (2012). Will kooperieren gelernt sein? Ein Quasi-Experiment im Lehramtsstudium. *Zeitschrift für Pädagogische Psychologie*, 26 (4), 263–273.

Schwarzer, R. & Schmitz, G. S. (1999). Kollektive Selbstwirksamkeitserwartung von Lehrern. Eine Längsschnittstudie in zehn Bundesländern. *Zeitschrift für Sozialpsychologie*, 30 (4), 262–274.

Somech, A. & Ron, I. (2007). Promoting organizational behavior in schools: The impact of individual and organizational characteristics. *Educational Administration Quarterly*, 43 (1), 38–66.

Steinert, B., Klieme, E., Maag Merki, K., Döbrich, P., Halbheer, U. & Kunz, A. (2006). Lehrerkooperation in der Schule: Konzeption, Erfassung, Ergebnisse. *Zeitschrift für Pädagogik*, 52 (2), 185–203.

Tarelli, I., Lankes, E.-M., Drossel, K. & Gegenfurtner, A. (2012). Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 137–169). Münster: Waxmann.

Taylor, B. M. (2007). *The what and the how of good classroom reading instruction in the elementary grades.* Minneapolis, MN: University of Minnesota Center for Reading Research.

- Taylor, B. M., Pearson, P. D., Peterson, D. S. & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal*, 104 (1), 3–28.
- Terhart, E. (2002). *Standards für die Lehrerbildung: Eine Expertise für die Kultusministerkonferenz*. Universität Münster: Zentrale Koordination Lehrerbildung.
- Thomson, M. M., Turner, J. E. & Nietfeld, J. L. (2012). A typological approach to investigate the teaching career decision. Motivations and beliefs about teaching of prospective teacher candidates. *Teaching and Teacher Education*, 28 (3), 324–335.
- Tierney, P. & Farmer, S. M. (2004). The Pygmalion process and employee creativity. *Journal of Management*, 30 (3), 413–432.
- Tomlinson, C. A. & Allan, S. D. (2000). *Leadership for differentiating schools & classrooms*.
- Alexandria, VA: Association for Supervision and Curriculum Development.
- Urban, D. & Mayerl, J. (2003). Wie viele Fälle werden gebraucht? Ein Monte-Carlo-Verfahren zur Bestimmung ausreichender Stichprobengrößen und Teststärken(power) bei Strukturgleichungsanalysen mit kategorialen Indikatorvariablen. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 53, 42–69.
- Watt, H. M. & Richardson, P. W. (2008). Motivations, perceptions, and aspirations concerning teaching as a career for different types of beginning teachers. *Learning and Instruction*, 18 (5), 408–428.
- Wellenreuther, M. (2008). *Lehren und Lernen – aber wie? Empirisch-experimentelle Forschungen zum Lehren und Lernen im Unterricht*. Baltmannsweiler: Schneider.

### **6.3 Beitrag III. Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt**

Jennifer Iglér<sup>1</sup>, Annika Ohle-Peters<sup>1</sup> und Nele McElvany<sup>1</sup>

<sup>1</sup> Institut für Schulentwicklungsforschung der Technischen Universität Dortmund

#### *Zusammenfassung*

Unterrichtsqualität wird häufig über Urteile von Externen erfasst. Mit Blick auf die Güte dieser Urteile stellt sich die Frage, ob diese aufgrund von Darbietungsmodalitäten sowie individuellen Faktoren systematisch verzerrt sein können. In der vorliegenden Studie wurden daher der Einfluss der Reihenfolge von Unterrichtsvideos mit unterschiedlichen Qualitätsniveaus sowie von individuellen Faktoren wie Vorerfahrung und mentaler Zustand (Müdigkeit) auf Unterschiede in den Urteilen und Interaktionen untersucht. Hierfür wurden  $N=69$  Studierende zu drei Versuchsgruppen zugeordnet, die zwei zehnmündige Videos in variierender Reihenfolge der Qualitätsniveaus präsentiert bekamen: Gruppe A (niedrig/mittel), B (mittel/hoch) und C (hoch/mittel). Die Analysen zeigten Unterschiede in den mittleren Ratings aufgrund der Reihenfolge sowie der Vorerfahrung und der Müdigkeit. Überdies wurden Interaktionen zwischen der experimentellen Bedingung und der Müdigkeit identifiziert. Die Ergebnisse wiesen auf Verzerrungen der Urteile von Externen hin, besonders im Bereich Motivierung. Konsequenzen für die Interpretation und Durchführung von videogestützten Unterrichtsqualitätsstudien werden diskutiert.

**Schlüsselwörter:** Videorating, Beurteilungsfehler, Unterrichtsqualität, Reihenfolgeeffekt, Unterrichtsforschung

*Teaching quality from the perspective of external raters—Analyses of the order effect*

#### *Abstract*

Ratings of external observers are often used for measuring instructional quality. These ratings can be influenced by the order of presentation and individual characteristics such as observers' experience and mental state (tiredness). The present study examined the influence of the presentation order of videos with varying levels of teaching quality and of individual factors such as previous experience and mental state (tiredness) on differences in observer ratings and

interactions. Therefore,  $N = 69$  students were assigned to three experimental groups. In each group, two videos of ten minutes length were presented in varying order of teaching quality: group A (low/medium), B (medium/high), and C (high/medium). Analyses showed significant differences in the ratings, depending on the sequential order, the observers' experience, and observers' tiredness. An interaction was found between experimental condition and observers' tiredness. Results suggest a bias of external judgements, especially concerning motivational quality. Consequences for interpretation and implementation of video-based studies measuring teaching quality are discussed.

*Keywords:* Order effect, Rater bias, Research on teaching quality, Teaching quality, Video rating

## **1 Einleitung**

Die Bedeutung der Qualität des unterrichtlichen Angebots für den Lernerfolg wird sowohl in theoretischen Modellen als auch in empirischen Studien betont (z.B. Lipowsky 2006). Um Aussagen über die Unterrichtsqualität treffen zu können, wird häufig auf Urteile von Externen zurückgegriffen. Gründe dafür sind die hohe Validität, die prädiktive Kraft für die Schülerleistungen, der emotionale Abstand zum Unterrichtsgeschehen sowie ein höheres Ausmaß an Vergleichsmöglichkeiten (Fauth et al. 2014a; Rakoczy 2008). Im Kontext der Raterurteile anhand von Unterrichtsvideos werden zur Sicherstellung der Objektivität Kodiermanuale entwickelt, welche je nach Inferenz des einzuschätzenden Merkmals ausdifferenziert werden, um Abweichungen in Beurteilungen zur selben Unterrichtseinheit zu minimieren (Seidel und Thiel 2017). Für die Sicherstellung der Validität und Reliabilität der Urteile werden Raterinnen und Rater in der Nutzung dieser Manuale geschult und gemeinsame Trainings durchgeführt (Lotz et al. 2013). Dennoch lassen sich Unterschiede in den Beurteilungen desselben Videos nicht vollständig eliminieren. Messergebnisse können durch Beurteilungsfehler verzerrt werden und somit zu einer Einschränkung der Nutzbarkeit und Interpretierbarkeit der Ergebnisse führen (Pietsch und Tosana 2008). Wenn die Raterinnen und Rater den gleichen Einflussfaktoren unterliegen, können Urteile aufgrund gleicher systematischer Verzerrung trotzdem hohe Interraterreliabilitäten aufweisen. Bisher gibt es nur wenige Erkenntnisse darüber, wie solche Verzerrungen von Einschätzungen der Unterrichtsqualität zustande kommen. Ein möglicher Erklärungsfaktor ist die Reihenfolge, in

der Videos beurteilt werden, da Informationen aus vorherigen Videos die neutrale Betrachtung eines darauffolgenden beeinträchtigen können (Mashburn et al. 2014). Verzerrungen der Urteile aufgrund der Darbietungsreihenfolge der Videos können zusätzlich durch individuelle Merkmale (z.B. geringe Vorerfahrung mit der Thematik und Müdigkeit) verstärkt werden (Gabriel-Busse et al. 2020; Webster et al. 1996). Unterrichtsbeobachtungen und Bewertungen des Lehrkraftverhaltens sowie der Unterrichtsqualität werden zunehmend in der Qualifikation sowie Weiterbildung von Lehrenden und Forschung eingesetzt (Mashburn et al. 2014). Demzufolge ist es bedeutsam, dass Verzerrungen in diesen Bewertungen so gering wie möglich gehalten werden. Die hier vorliegende Studie erweitert die bisherige Befundlage zu Beurteilungsfehlern und untersucht dabei gezielt Effekte der Darbietungsreihenfolge von Unterrichtsvideos sowie der individuellen Merkmale von Raterinnen und Ratern und mögliche Interaktionen.

## **2 Theoretischer Hintergrund**

### **2.1 Merkmale qualitätvollen Unterrichts**

Qualitätvoller Unterricht zeichnet sich unter anderem durch eine kognitiv anregende, störungsarme und unterstützende Lernumgebung aus (z. B. Praetorius et al. 2020). Um lernwirksame kognitive Prozesse zu aktivieren, ist das Anknüpfen an methodische sowie inhaltliche Vorkenntnisse der Lernenden und das Stellen herausfordernder Aufgaben bedeutsam (Lipowsky et al. 2009). Für eine erfolgreiche Auseinandersetzung mit dem Lernstoff ist ebenfalls die explizite Benennung von Lernzielen und das Vermitteln von inhaltlichen und methodischen Lernerwartungen von Relevanz (Meyer 2007).

Eine effiziente Klassenführung gilt als Voraussetzung für ein optimales Lernen in der Klasse. Ziel ist es, den Unterricht so zu strukturieren, dass möglichst viel Unterrichtszeit für den Lernstoff verwendet wird und Störungen möglichst vermieden werden (Brophy 1979; Emmer und Stough 2001). Von besonderer Bedeutung sind hierfür die Klarheit und die Einhaltung von Klassenregeln (Emmer und Stough 2001).

Als spezifische Facette der konstruktiven Unterstützung gilt die Motivierungsqualität, welche darauf abzielt, die Lernmotivation von Lernenden anzuregen und aufrecht zu erhalten (Kunter und Trautwein 2013). Für eine hohe Motivierung ist es wichtig, das Erleben von Selbstbestimmung zu fördern. Dies kann durch die Erfüllung psychologischer Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit erreicht werden

(Ryan und Deci 2020). Konkret bedeutet dies, dass den Lernenden Entscheidungsmöglichkeiten geboten werden und ihnen sowie ihren Äußerungen eine wertschätzende Haltung entgegengebracht werden (Hamre und Pianta 2010; Rakoczy 2008).

Die beschriebenen Merkmale lassen sich der Tiefenstruktur von Unterricht zuordnen, welche als besonders relevant für den Lernerfolg gilt (Pauli und Reusser 2003). Allerdings sind die Merkmale der Tiefenstruktur nur indirekt beobachtbar und können nur über die Wahrnehmung von Teilnehmenden des Unterrichts oder die Beurteilung durch Externe erfasst werden (Decristan et al. 2020). Generell übereinstimmten Einschätzungen über Unterrichtsqualitätsmerkmale von Lernenden, Lehrkräften und Externen nur gering (vgl. Clausen 2002; Fauth et al. 2014b). Eine Erklärung dafür ist, dass Unterrichtsqualitätsmerkmale in Abhängigkeit der einzuschätzenden Inhalte (Verhalten der Lehrkraft, Verhalten der Lernenden oder eine Mischung aus beidem) unterschiedlich von den Bewertungsperspektiven (Externe, Lehrkräfte und Lernenden) eingeschätzt werden. Für Unterrichtsqualitätsmerkmale mit geringer Beobachtbarkeit, zum Beispiel Skalen zur konstruktiven Unterstützung der Lehrkräfte, korrelieren die Urteile der Externen nicht stark mit denen der Lernenden- und Lehrkräfteeinschätzungen (Fauth et al. 2020). Höhere Zusammenhänge wurden zwischen den drei Perspektiven im Bereich der Klassenführung festgestellt (z.B. Fauth et al. 2014b). Letztlich sind Externe, Lehrkräfte und Lernende nicht in gleicher Weise kompetent das gesamte Spektrum von Unterrichtmerkmalen einzuschätzen (Clausen 2002).

## **2.2 Unterrichtsbeurteilungen von externen Raterinnen und Ratern**

Bei Unterrichtsbeurteilungen von Externen sind die Beobachtungen, die direkt im Klassenraum und die anhand von videografierten Unterrichtssequenzen erfolgen, zu unterscheiden. Beurteilungen auf Basis von Unterrichtsvideos ermöglichen, dass diese differenziert beurteilt und wiederholt angesehen werden und die Bewertungen auch zeitversetzt stattfinden können (Herrle et al. 2016; Mashburn et al. 2014). Urteile der Tiefenstruktur werden über hoch inferente Ratings abgegeben und setzen einen Inferenzschluss seitens der Raterinnen und Rater voraus, wobei von direkt beobachtbaren Merkmalen auf überdauernde Merkmale des Unterrichts geschlossen werden muss (Begrich et al. 2017). Obwohl in Studien zur videogestützten Unterrichtsanalyse Kodiermanuale und Ratertrainings eingesetzt werden, zeigten Untersuchungen, dass teilweise hohe Varianzanteile von bis zu 41,0% auf Beurteilungsfehler und nicht auf das einzuschätzende Merkmal

zurückzuführen sind (Hoyt und Kerns 1999; Pietsch und Tosana 2008; Praetorius 2014). Es bleibt ein Bedarf an theoretischen Begründungen und empirischen Prüfungen für diese Beurteilungsfehler und für den Einfluss von Merkmalen der Studie oder der Beurteilenden auf die Urteile.

### **2.3 Unterschiede in Unterrichtsbeurteilungen aufgrund des Reihenfolgeeffekts**

Ein klassischer Beurteilungsfehler ist der Reihenfolgeeffekt, der bei Gedächtnisprozessen auftritt und auf die Arbeitsweise des Kurzzeitgedächtnisses zurückzuführen ist (Ebbinghaus 1885; Kooken et al. 2017). Demnach hat die Darbietungsreihenfolge, in der Informationen präsentiert werden, einen Einfluss auf die Speicherung und auf die Beurteilung von Informationen (Cushman und Mele 2008). Zum Beispiel können sich Urteile über Information A systematisch danach unterscheiden, ob die Informationen in der Reihenfolge A-B oder B-A vorgelegt werden. Dieser Effekt kann damit erklärt werden, dass die kognitive Beschäftigung mit dem vorigen Bewertungsgegenstand auf die Verfügbarkeit bestimmter Gedächtnisinhalte wirkt (Sudman et al. 1996). Urteile über die Information A können positiver ausfallen, wenn eine negativ bewertete Information B vorausgeht oder negativer ausfallen, wenn eine positiv bewertete Information B vorausgeht (Miller und Campbell 1959). Folglich kann es zur Auf- oder Abwertung der Qualität eines Unterrichtsvideos kommen, je nachdem in welcher Darbietungsreihenfolge Unterrichtsvideos mit unterschiedlichen Qualitätsniveaus präsentiert werden. Empirisch konnte dies bestätigt werden, indem Ho und Kane (2013) feststellten, dass sich Raterinnen und Rater bei der Bewertung der Unterrichtsqualität eines Videos an dem vorigen Video orientierten und wiesen einen Zusammenhang zwischen einem zweiten Urteil und einem ersten Urteil auf. In einer Studie von Mashburn et al. (2014) konnte gezeigt werden, dass die Varianz, die auf die Urteilsverzerrung der Beurteilenden zurückgeführt werden konnte, minimiert wurde, wenn die Raterinnen und Rater die Reihenfolge von Videos in einer zufälligen Reihenfolge präsentiert bekamen. Neben der Darbietungsreihenfolge können individuelle Faktoren einerseits die Urteile selbst und andererseits das Auftreten des Reihenfolgeeffekts beeinflussen (Bless et al. 2004).

### **2.3.1 Effekte der Vorerfahrung**

#### **Effekte der Vorerfahrung auf Urteile**

Unterschiede in Urteilen können aufgrund der Interaktion von Vorerfahrungen mit der Wahrnehmung und Interpretation eines Reizes auftreten (Bless et al. 2004; Myford und Wolfe 2003). Neue Reize werden über eine Bedeutungszuschreibung mittels der im Gedächtnis gespeicherten Kategorien, Schemata und Skripte encodiert und interpretiert (Bless et al. 2004). Demnach unterscheiden sich Urteile in Abhängigkeit der Größe des Speichers an Informationen über und Erfahrungen für das einzuschätzende Thema. Für den Bereich der Unterrichtsforschung fanden Gabriel-Busse et al. (2020) heraus, dass sich Urteilsfindungen über die kognitive Aktivierung und Klassenführung von Unterrichtsvideos sowie Urteilsbegründungen von Lehramtsstudierenden nach einem Zuwachs an Erfahrung vermittelt durch ein Seminar veränderten. Aufgrund einer höheren Erfahrung orientierten sich Studierende bei der Beurteilung stärker an den theoretischen Konstrukten und gewichteten diese unterschiedlich in der Urteilsbegründung. Wolff et al. (2017) zeigten, dass divergierende Erfahrungszeit in der Schule von Lehrkräften zu unterschiedlicher Fokussierung von Unterrichtsqualitätsmerkmalen und zu unterschiedlicher Tiefe der Schlussfolgerungen bei der Beurteilung von Unterrichtsvideos führte. Beispielsweise zogen Lehrkräfte mit weniger Erfahrungszeit in der Schule eher oberflächliche Unterrichtsereignisse für die Urteilsfindung im Vergleich zu Erfahrenen heran.

#### **Effekte der Vorerfahrung auf das Auftreten von Beurteilungsfehlern**

Überdies kann Erfahrung einen Einfluss auf Verzerrungen in Urteilen haben und geringe Erfahrung kann das Auftreten von Beurteilungsfehlern verstärken (Praetorius 2013). Steht ein größerer Speicher an Informationen für die Verarbeitung zur Verfügung, so können akkuratere Beurteilungen über das Beobachtete getroffen und somit Beurteilungsfehler verringert werden (Bless et al. 2004; Lau und Plessner 2016). Liegt ein Mangel an relevanten Informationen zur Urteilsfindung vor, werden eher irrelevante Informationen berücksichtigt, welche zu systematisch verzerrten Bewertungen führen können (Messner und Schmid 2007). Empirisch konnten Feltz und Cokely (2011) zeigen, dass Erfahrung mit der Thematik prädiktiv für das Auftreten des Reihenfolgeeffekts ist und eine höhere Erfahrung das Auftreten eines solchen minimieren kann.

### **2.3.2 Effekte des mentalen Zustands**

#### **Effekte des mentalen Zustands auf Urteile**

Neben Vorerfahrungen kann auch der mentale Zustand einen Einfluss auf die Urteilsfindung haben. Um Urteile zu fällen, werden Situationen zunächst wahrgenommen (Bless et al. 2004). Die Wahrnehmung von Reizen kann von aktuellen Zuständen beeinflusst werden, welche zur Selektion von Informationen führen und demzufolge auch Auswirkungen auf das nachfolgende Urteil haben können (Martin und Wawrinowski 2014). Insbesondere die Müdigkeit kann im Wahrnehmungsprozess dazu führen, dass aufgrund von fehlender Konzentration nicht alle relevanten Informationen beachtet werden können (Bless und Keller 2006). Studien dazu haben gezeigt, dass eine niedrige Konzentrationsfähigkeit oder Müdigkeit zur Selektion der wahrgenommenen Information und anschließend zu einer selektiven Beurteilung führte (Schmidt-Atzert et al. 2004). Eine andere Weise, wie der mentale Zustand auf Urteile wirken kann, ist, dass sich Personen Affektheuristiken (z.B. „Wie fühle ich mich dabei“-Heuristik) bedienen, um von ihrem momentanen Zustand auf ihre Urteile zu schließen (Bless et al. 1990). So können Urteile positiver oder negativer ausfallen, da der eigene mentale Zustand als Informationsquelle dient (Stroebe 2014). In Bezug auf das Beurteilen von Unterrichtsqualität konnte festgestellt werden, dass Raterinnen und Rater mit zunehmender Müdigkeit strengere Urteile vergaben (Mashburn et al. 2014).

#### **Effekte der Müdigkeit auf das Auftreten von Beurteilungsfehlern**

Die Einschränkung der Informationsverarbeitung durch Müdigkeit kann auch ein höheres Auftreten von Beurteilungsfehlern durch eine mit der Müdigkeit einhergehender Tendenz schnellere Urteile zu fällen und häufigere Nutzung von fehleranfälligen Urteilsheuristiken bewirken (Engle-Friedman et al. 2018; Webster et al. 1996), mitunter etwa ein Auftreten des Reihenfolgeeffekts (Kruglanski und Webster 1996). Epley und Gilovich (2006) konnten zeigen, dass Beurteilungsfehler weniger häufig auftraten, wenn die Probandinnen und Probanden in einem konzentrierten mentalen Zustand waren. Auch in der Unterrichtsbeobachtung kann Müdigkeit Wahrnehmungsverzerrungen verstärken, indem für die Urteilsfindung relevante Unterrichtsereignisse nicht wahrgenommen werden, da offensichtlichere Ereignisse eine höhere Aufmerksamkeit auf sich ziehen (Schwindt 2008).

### 3 Forschungsfragen und Hypothesen

Ein substanzieller Anteil von Unterschieden in Urteilen kann auf Beurteilungsfehler zurückgeführt werden. Mögliche Erklärungen können Faktoren, wie die Darbietungsreihenfolge sowie individuelle Faktoren (z.B. Vorerfahrung oder mentaler Zustand) der Urteilenden sein. Zudem kann ein Reihenfolgeeffekt durch individuelle Faktoren verringert oder begünstigt werden.

**F1** Treten Verzerrungen in den Urteilen der Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierung bei einem Unterrichtsvideo mittleren Qualitätsniveaus aufgrund der sequenziellen Darbietungsreihenfolge auf?

Je nachdem in welcher Reihenfolge Unterrichtsvideos unterschiedlicher Qualitätsniveaus präsentiert werden, können Urteile unterschiedlich ausfallen. Es wird angenommen, dass ein Unterrichtsvideo mittleren Qualitätsniveaus nach einem vorigen Urteil eines Videos niedrigen Qualitätsniveaus positiver bewertet wird, als wenn vorher kein anderes Video bewertet wurde (H1.1). Ebenfalls wird erwartet, dass ein Unterrichtsvideo mittleren Qualitätsniveaus nach einem vorigen Urteil eines Videos eines hohen Qualitätsniveaus negativer bewertet wird, als wenn vorher kein anderes Video bewertet wurde (H1.2).

**F2** Stehen die individuellen Faktoren (a) Vorerfahrung (Erfahrungszeit in der Schule) oder (b) mentaler Zustand (Müdigkeit) und die Urteile der Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierung bei einem Unterrichtsvideo mittleren Qualitätsniveaus sowie Unterschiede im Auftreten des Reihenfolgeeffekts in systematischem Zusammenhang?

Aufgrund von unterschiedlich vorgeschichteten Informationen über das einzuschätzende Thema kommt es zu divergierenden Wahrnehmungen und schließlich auch Urteilen (Bless et al. 2004). Dementsprechend wird davon ausgegangen, dass die Urteile der Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierung bei einem Unterrichtsvideo mittleren Qualitätsniveaus durch die Vorerfahrung (Erfahrungszeit in der Schule) der Raterinnen und Rater mitbedingt werden (H2.1). Beurteilungsfehler können durch einen größeren Speicher an Informationen verringert werden (Lau und Plessner 2016). Demnach wird erwartet, dass eine höhere Erfahrungszeit in der Schule das Auftreten des Reihenfolgeeffekts verringert (H2.2). Hinzu wirkt der mentale Zustand (Müdigkeit) auf die Wahrnehmung von Reizen, was wiederum zur Selektion von Informationen führen und

Auswirkungen auf das nachfolgende Urteil haben kann (Martin und Wawrinowski 2014). Folglich wird davon ausgegangen, dass die Urteile der Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierung bei einem Unterrichtsvideo mittleren Qualitätsniveaus durch den mentalen Zustand (Müdigkeit) der Raterinnen und Rater mitbedingt werden (H2.3). Aufgrund von Müdigkeit werden häufig schnellere Urteile unter anderem mittels Urteilsheuristiken gefällt, was wiederum zu einem häufigeren Auftreten von Beurteilungsfehlern führen kann (Engle-Friedman et al. 2018). Es wird erwartet, dass eine höhere Müdigkeit das Auftreten des Reihenfolgeeffekts verstärkt (H2.4).

## 4 Methodisches Vorgehen

### 4.1 Stichprobe

Die Daten wurden im Frühjahr 2017 im Rahmen eines experimentellen Settings erhoben. Insgesamt nahmen  $N=69$  Studierende an der Studie teil. Die Probandinnen und Probanden wurden zufällig auf drei experimentelle Bedingungen verteilt, die sich in der Darbietungsreihenfolge der Unterrichtsvideos mit unterschiedlichen Qualitätsniveaus unterschieden: Gruppe A (niedrig/mittel), B (mittel/hoch) und C (hoch/ mittel) (vgl. Tab. 1). Eine vorige Poweranalyse ergab, dass bei einer Power von 0,80 und einem Effekt von  $f=0,40$   $N=64$  Personen benötigt werden. Demographische Angaben der Probandinnen und Probanden sowie die Erfahrungszeit in der Schule wurden im Voraus in einem Online-Fragebogen erhoben, damit Studierende des gleichen Studienfachs gleichmäßig auf die Gruppen aufgeteilt werden konnten. Die Probandinnen und Probanden der drei Gruppen unterschieden sich nicht hinsichtlich des Anteils des Geschlechts ( $\chi^2(2) = 0,16; p > 0,05$ ), ihres Alters ( $F[2, 65] = 1,64; p > 0,05$ ), ihrer Semesterzahl ( $F[2, 65] = 1,58; p > 0,05$ ) oder des Anteils des Lehramtsstudiums ( $\chi^2(2) = 0,46; p > 0,05$ ).

Tabelle 1.

*Deskriptiva der Gesamt- und Teilstichproben*

Stichproben	<i>n</i>	<i>M</i> <sub>Alter</sub> ( <i>SD</i> )	Weiblich in %	<i>M</i> <sub>Semester</sub> ( <i>SD</i> )	Lehramts- studium in %	Erfahrung in der Schule in %
Alle Probandinnen und Probanden	69	24.91 (4.40)	80.9	7.84 (4.66)	50.7	62.3
Experimentelle Bedingung A	23	24.87 (4.95)	81.8	7.32 (4.03)	47.8	54.5
Video 1 Niedriges Qualitätsni- veau						
Video 2 Mittleres Qualitätsniveau						
Experimentelle Bedingung B	23	23.78 (3.09)	82.6	6.96 (4.25)	47.8	60.9
Video1 Mittleres Qualitätsni- veau						
Video 2 Hohes Qualitätsniveau						
Experimentelle Bedingung C	23	26.14 (4.81)	78.3	9.22 (5.43)	56.5	73.9
Video1 Hohes Qualitätsni- veau						
Video2 Mittleres Qualitätsniveau						

*Anmerkung.* Die höheren Mittelwerte der Semesteranzahl in der Bedingung C können auf zwei einzelne Extremfälle zurückgeführt werden (Semester = 19; Semester = 20).

## 4.2 Instrumente und Durchführung

### Unterrichtsvideos

Die Videosequenzen stammten aus dem DFG geförderten Kooperationsprojekt *Entwicklung und Überprüfung von Kompetenzmodellen zur integrativen Verarbeitung von Texten und Bildern* (BiTe) (vgl. McElvany et al. 2012) und zeigten einen circa zehnminütigen Unterrichtseinstieg in vierten Klassen zur vorgegebenen Thematik „Südamerika“. Die Auswahl der Videosequenzen mit unterschiedlich hohen Qualitätsniveaus für die vorliegende Studie erfolgte in zwei Schritten. Zunächst wurden drei Videos aufgrund ihrer Qualität (niedrig [1] – hoch [4]; *niedrig*  $M = 2,14$ ,  $SD = 0,45$ , *mittel*  $M = 2,24$ ,  $SD = 0,46$  und *hoch*  $M = 2,63$ ,  $SD = 0,36$ ) und des Auftretens der zu untersuchenden Facetten in den ersten 10 min aus einem Pool von schon gerateten Videos des DFG-Projekts BiTe ausgesucht. Um die Auswahl der Videos zu bestätigen, wurden zusätzlich  $N = 8$  Expertinnen und Experten aus der

Unterrichtsforschung bezüglich der Unterrichtsqualität in den drei ausgewählten Videos sowie zum Untersuchungsdesign befragt. Die Zuordnung zu den Kategorien *niedrig*, *mittel* und *hoch* konnte durch die Expertisen unterstützt werden.

### **Ratingmanual und Ratingbogen**

Das Ratingmanual wurde in Anlehnung an Ohle und McElvany (2016) und Praetorius (2014) entwickelt und gliederte sich in einen strukturellen Teil (Informationen zu dem Beurteilungsprozess der Unterrichtssequenzen) sowie einen inhaltlichen Teil (Beschreibung der drei zu beurteilenden Unterrichtsqualitätsmerkmale). Zu den einzelnen Merkmalen *kognitive Aktivierung*, *Klassenführung* und *Motivierung* wurden zunächst im Ratingmanual eine Definition sowie zwei Facetten mit positiven und negativen Beispielindikatoren gegeben. Für die Erfassung der Unterrichtsqualität wurden im Ratingbogen drei Items eingesetzt, bei denen jeweils der Gesamteindruck der Qualität der kognitiven Aktivierung, Klassenführung oder Motivierung im gerade gesehenen Video auf einer vierstufigen Antwortskala (niedrig [1] – hoch [4]) eingeschätzt werden sollte (Item „Wir möchten Sie bitten, einen Gesamteindruck der Unterrichtsqualität einzuschätzen. Bitte tragen Sie unten ein, wie hoch Sie die Qualität der Merkmale im gerade gesehenen Video einschätzen.“).

### **Erfassung individueller Merkmale**

Zur Erfassung der Vorerfahrungen gaben die Studierenden an, wie viele Wochen Unterrichtserfahrung sie in der Schule gesammelt haben ( $M = 16,70$ ,  $SD = 31,32$ ;  $Min = 0,00$ ,  $Max = 200,00$ ). Die *Erfahrungszeit in der Schule* wurde mitunter im Rahmen eines Praktikums ( $n = 36$ ), einer Vertretungsstelle ( $n = 7$ ) oder eines Praxissemesters ( $n = 6$ ) gesammelt.  $N = 25$  Probandinnen und Probanden gaben an, keine Erfahrung in der Schule erworben zu haben. Für die Operationalisierung der *Müdigkeit* wurde die „Aktuelle Stimmungsskala“ eingesetzt (Dalbert 1992). Dabei sollte aus einer Liste mit Items, die verschiedene mentale Zustände beschreiben, für jedes Item die Zahl angekreuzt werden, welche den aktuellen Gefühlszustand am besten beschreibt (Vier Items: abgeschlafft, müde, erschöpft und entkräftet; Ratingskala überhaupt nicht [1] – sehr stark [7];  $M = 3,12$ ,  $SD = 1,25$ ;  $Min = 1,00$ ,  $Max = 6,75$ ;  $\alpha = 0,90$ ). Bis auf die Variable *Erfahrungszeit in der Schule* (2,9% fehlend) gab es keine fehlenden Werte.

### **Durchführung der Untersuchung**

Vorab wurde den Probandinnen und Probanden die Information gegeben, dass sie an einer Studie zur akkuraten Einschätzung von Unterrichtsqualität anhand von Videosequenzen und

der Relevanz des Studienfachhintergrundes teilnahmen. Um die Vergleichbarkeit der Beurteilungen sicherzustellen, wurden alle Probandinnen und Probanden im Bereich *Ratings von Unterrichtsqualität* drei Stunden standardisiert geschult (vgl. Tab. 2). Hierzu erfolgte zuerst eine theoretische Einführung in die Unterrichtsqualität und in die manualbasierte Videoanalyse. Im Anschluss wurden der Inhalt und die Struktur des Ratingmanuals sowie des Ratingbogens erklärt und eine gemeinsame Übung für das Beurteilen einer Unterrichtssequenz anhand eines Videos hohen Qualitätsniveaus angeleitet. Danach folgten zwei Einzelübungen, in denen die Studierenden eigenständig je ein Video niedrigen und mittleren Qualitätsniveaus mit Hilfe des Ratingmanuals und des Ratingbogens (siehe 4.2) einschätzten. Die Urteile wurden jeweils nach jedem Video gemeinsam besprochen, um die Urteilsübereinstimmung zu optimieren.

Tabelle 2.

*Durchführung der Studie*

Zeitpunkt	Schritte	Zeit				
Im Voraus	Onlinefragebogen vorab: <ul style="list-style-type: none"> <li>• Demographische Angaben</li> <li>• Studienfach</li> </ul>	5 Min.				
Einführung	Kurze theoretische Einführung in die Dimensionen der Unterrichtsqualität und manualbasierte Videoanalyse Einarbeiten in das Ratingmanual Gemeinsames Raten einer Sequenz „Video mit hohem Qualitätsniveau“ & Besprechung Einzelrating Trainingssequenz 1 „Video mit niedrigem Qualitätsniveau“ & Besprechung Einzelrating Trainingssequenz 2 „Video mit mittlerem Qualitätsniveau“ & Besprechung Pause Skala zur aktuellen Stimmung	190 Min.				
Hauptuntersuchung	Einzelrating der zwei Hauptuntersuchungssequenzen  <i>Experimentelle Bedingung A</i>  <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; text-align: center;">Video 1</td> <td style="width: 50%; text-align: center;">Video 2</td> </tr> <tr> <td style="text-align: center;">Niedriges Qualitätsniveau</td> <td style="text-align: center;">Mittleres Qualitätsniveau</td> </tr> </table>	Video 1	Video 2	Niedriges Qualitätsniveau	Mittleres Qualitätsniveau	40 Min.
Video 1	Video 2					
Niedriges Qualitätsniveau	Mittleres Qualitätsniveau					

*Experimentelle Bedingung B*

Video 1

Mittleres Qualitätsniveau

Video 2

Hohes Qualitätsniveau

*Experimentelle Bedingung C*

Video 1

Hohes Qualitätsniveau

Video 2

Mittleres Qualitätsniveau

Abschluss

5 Min.

---

240  
Minuten  
(4 Std.  
incl.  
Pausen)

---

Nach den zwei Trainingsvideos erreichten die Interraterreliabilitäten, Reliabilitäten des Mittelwertes aller Raterinnen und Rater, zufriedenstellende Werte von  $ICC2_{kognitive\ Aktivierung} = 0,82$ ,  $ICC2_{Klassenführung} = 0,86$  und  $ICC2_{Motivierung} = 0,63$  (Wirtz und Caspar 2002). Anschließend bearbeiteten die Probandinnen und Probanden Fragen zur aktuellen Stimmung, in denen unter anderem die Müdigkeit erfasst wurde. Zuletzt schätzten die Probandinnen und Probanden je nach experimenteller Bedingung die Unterrichtsqualität von zwei Videos mit einem niedrigen bzw. hohen und mittleren Qualitätsniveau ein (vgl. Tab. 3). Die Urteile erfolgten auch hier direkt nach jedem der zwei gesehenen Videos, wobei ein erneutes Ansehen des jeweiligen Videos für die Beurteilung der einzelnen Unterrichtsqualitätsmerkmale nicht erlaubt war. Die Gruppe B, die das Referenzvideo mittleren Qualitätsniveaus zuerst sah und daher keiner möglichen Beeinflussung durch ein vorheriges Video unterlag, stellte die Referenzgruppe dar. Für eine vergleichbar lange Durchführung des Experiments erhielten auch die Probandinnen und Probanden der Gruppe B ein zweites Video zum Einschätzen. Den gemeinsamen Abschluss der Sitzung bildete die Aufklärung der Teilnehmenden über den genauen Studienzweck.

Tabelle 3.

*Mittelwerte und Standardabweichungen der Unterrichtsvideos*

Experimentelle Bedingung	Qualitätsniveau	Kognitive Aktivierung	Klassenführung	Motivierung
		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
A	Niedrig	1.48 (0.51)	2.73 (0.75)	1.65 (0.49)
	Mittel	3.61 (0.50)	2.87 (0.63)	2.83 (0.72)
B	Mittel	3.52 (0.51)	2.61 (0.78)	3.00 (0.74)
	Hoch	3.61 (0.50)	3.26 (0.45)	2.96 (0.71)
C	Hoch	3.70 (0.47)	3.43 (0.59)	2.83 (0.58)
	Mittel	3.09 (0.51)	2.26 (0.75)	2.35 (0.49)

**4.3 Auswertungsstrategie**

Die Analysen zur Skalenreliabilität, Deskriptiva und *t*-Tests wurden mit Hilfe des Programms SPSS 24 durchgeführt. Für die Untersuchung des Reihenfolgeeffekts (F1) wurden *t*-Tests für unabhängige Stichproben gerechnet. Hierbei wurden im ersten Schritt die Unterschiede zwischen den Gruppen A (niedrig/mittel) und B (mittel/hoch) und im zweiten Schritt zwischen den Gruppen B (mittel/hoch) und C (hoch/mittel) jeweils für die Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierung überprüft. Es wurde eine Korrektur nach Bonferroni-Holm vorgenommen, um die Alpha-Fehler-Kumulierung zu berücksichtigen (Holm 1979). Für die Berechnungen des Unterschieds zwischen den Gruppen B (mittel/hoch) und C (hoch/mittel) im Bereich kognitive Aktivierung wurde der Mann-Whitney-U-Test gerechnet, da mit ungleichen Varianzen die Voraussetzung für die Berechnung von *t*-Tests nicht gegeben war. Für eine bessere Interpretierbarkeit der Ergebnisse wurde das Urteil des Videos mittleren Qualitätsniveaus der Gruppe B auf null gesetzt und für die anderen beiden Gruppen die Abweichungen der Urteile zum Mittelwert der Gruppe B bestimmt.

Die Fragestellung 2 zu den Effekten der Vorerfahrung und Müdigkeit wurde mit Hilfe von Regressions- sowie Moderationsanalysen mittels des Mplus-Softwarepakets (Muthén und Muthén 1998–2017) überprüft. In diesen Analysen stellten die Variablen zur Einschätzung der Unterrichtsqualität bei dem Video mittlerer Qualität die abhängigen Variablen dar. Es wurden vorab zwei Dummy-Variablen (Dummy 1: A = 1, B = 0, C = 0 und Dummy 2: A = 0, B = 0, C = 1) für die Gruppenzugehörigkeit der experimentellen Bedingung A und C gebildet. Insgesamt wurden drei Modelle je Unterrichtsqualitätsmerkmal gerechnet. Zur Überprüfung

der Hypothese 2.1 und 2.3 wurden in dem Modell I zur Kontrolle des Reihenfolgeeffekts die Dummy-Variablen der Bedingung A und C sowie die Erfahrungszeit oder die Müdigkeit als unabhängige Variable aufgenommen. Für die Überprüfung der Hypothesen 2.2 und 2.4 wurden Moderationsanalysen gerechnet, bei denen Interaktionsterme zwischen den zwei Dummy-Variablen und der Erfahrungszeit oder Müdigkeit definiert wurden. Die unabhängigen Variablen in dem Modell II stellten zur Kontrolle des Reihenfolgeeffekts die Dummy-Variablen der Bedingung A und C sowie die Interaktionsterme dar. Ein möglicher Einfluss der individuellen Faktoren auf Unterschiede im Auftreten des Reihenfolgeeffekts würde sich in einer signifikanten Interaktion zwischen den Gruppen A oder C und dem jeweiligen individuellen Faktor zeigen. In dem Modell III wurden schließlich als unabhängige Variablen zur Kontrolle des Reihenfolgeeffekts die Dummy-Variablen der Bedingung A und C, die Erfahrungszeit oder die Müdigkeit und die Interaktionsterme aufgenommen. In allen Modellen wurden die Variablen manifest modelliert und die Variablen Erfahrungszeit und Müdigkeit standardisiert. Da es sich um saturierte Modelle handelt, wurden keine globalen Fitindices angegeben (Geiser 2011). Außerdem wurden Korrelationen zwischen den unabhängigen Variablen zugelassen. Zusätzlich wurde in allen Modellen die in Mplus implementierten Optionen maximum-likelihood-estimator (ML) und full information maximum likelihood (FIML) verwendet.

## 5 Ergebnisse

### 5.1 Unterschiede in den Urteilen aufgrund der Darbietungsordnung

Bezüglich der Fragestellung 1, ob sich Unterschiede in der durchschnittlichen Beurteilung eines Videos mittleren Qualitätsniveaus aufgrund der Darbietungsordnung zeigen, wiesen die Ergebnisse der *t*-Tests auf Unterschiede in den Beurteilungen der Gruppen B und C hinsichtlich den Unterrichtsqualitätsmerkmalen kognitive Aktivierung und Motivierung hin (Abb. 1).<sup>1</sup>

---

<sup>1</sup> Werden für Kontrollanalysen zur Robustheit der Befunde die Semesteranzahl oder das Alter als Kovariaten in Varianzanalysen, bei denen alle drei Gruppen A, B und C verglichen werden, aufgenommen, bleiben die Ergebnisse stabil.

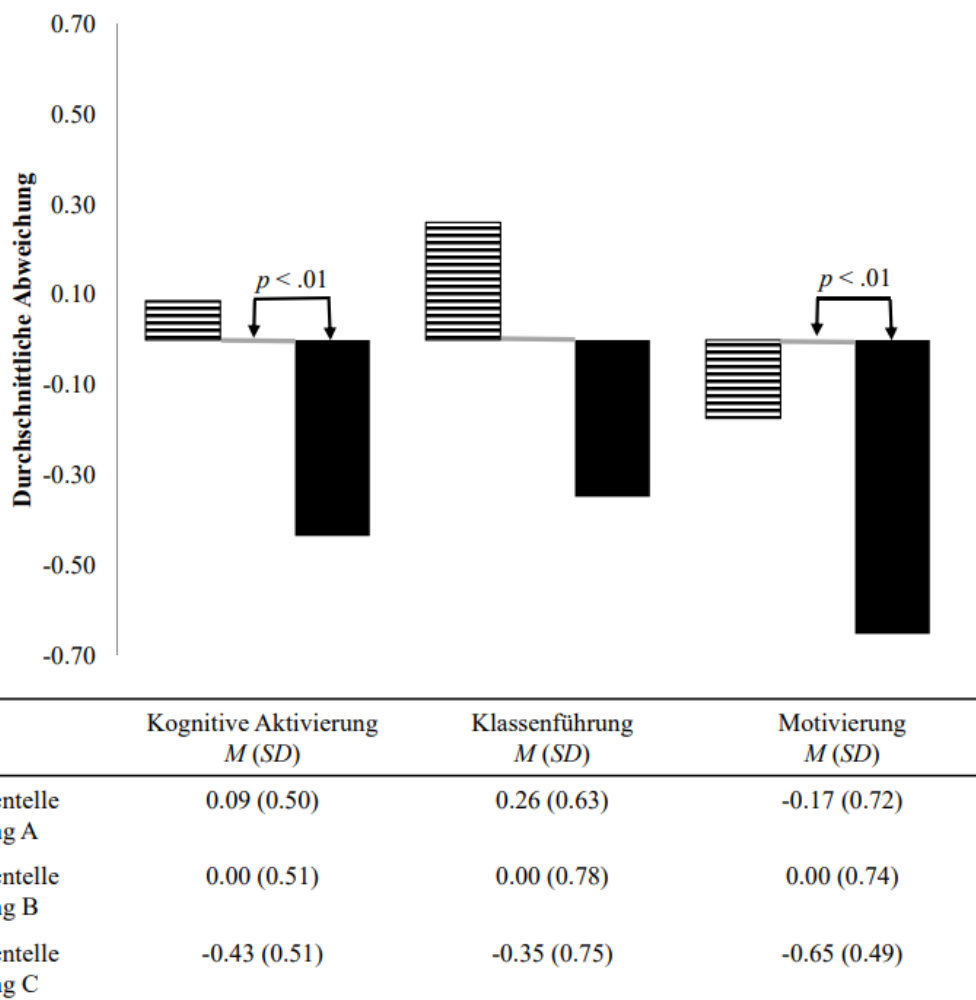


Abbildung 1. Durchschnittliche Abweichungen der Urteile des Videos mit mittlerem Qualitätsniveau der experimentellen Bedingungen A (niedrig/mittel) und C (hoch/mittel) und signifikanten Pfaden unterteilt nach den Unterrichtsqualitätsmerkmalen.

Anmerkung. Zusätzlich zu den hypothesenprüfenden Analysen ließen sich Unterschiede in den Urteilen zwischen den Gruppen A (niedrig/mittel) und C (hoch/mittel) bezüglich der kognitiven Aktivierung ( $t(44) = 3.50$ ,  $p = .009$ , Cohens  $d = 1.03$ ) und der Klassenführung ( $t(44) = 2.99$ ,  $p = .035$ , Cohens  $d = 0.88$ ) finden; jedoch nicht im Bereich Motivierung ( $t(44) = 2.65$ ,  $p = .055$ , Cohens  $d = 0.78$ ).

### Gruppe A (niedrig/mittel) vs. Gruppe B (mittel/hoch)

Zwischen den Urteilen der Gruppen A (niedrig/mittel) und B (mittel/hoch) ließen sich keine Unterschiede bezüglich der kognitiven Aktivierung ( $t(44) = 0.58$ ,  $p = 0.844$ , Cohens  $d = 0.17$ ), Klassenführung ( $t(44) = 1.25$ ,  $p = 0.654$ , Cohens  $d = 0.37$ ) und Motivierung ( $t(44) = -0.81$ ,

$p = 0,844$ , Cohens  $d = -0,24$ ) finden. Raterinnen und Rater, die vor der Beurteilung des Referenzvideos ein Video niedrigen Qualitätsniveaus sahen, beurteilten die Qualität der kognitiven Aktivierung, Klassenführung und Motivierung des Referenzvideos nicht signifikant abweichend. Dementsprechend konnte die Hypothese 1.1 zur Aufwertung der Urteile durch die Reihenfolge nicht bestätigt werden.

### **Gruppe B (mittel/hoch) vs. Gruppe C (hoch/mittel)**

Es zeigten sich signifikante Differenzen zwischen den Urteilen der Gruppen B (mittel/hoch) und C (hoch/mittel) hinsichtlich der kognitiven Aktivierung ( $U = 161,50$ ,  $Z = -2,64$ ,  $p = 0,048$ , Cohens  $d = 0,84$ ) sowie der Motivierung ( $t(44) = 3,54$ ,  $p = 0,009$ , Cohens  $d = 1,04$ ). Demzufolge schätzten die Raterinnen und Rater, die vor dem Referenzvideo ein Video hohen Qualitätsniveaus beurteilten, die Qualität der kognitiven Aktivierung und Motivierung des Referenzvideos mit mittlerer Qualität niedriger ein als die Referenzgruppe. Es wurden keine statistisch signifikanten Differenzen zwischen den Urteilen der Gruppen B und C ( $t(44) = 1,54$ ,  $p = 0,524$ , Cohens  $d = -0,14$ ) bezüglich der Klassenführung festgestellt. Raterinnen und Rater, die vor der Bewertung des Referenzvideos ein Video mit einem niedrigeren oder höheren Qualitätsniveau sahen, schätzten die Qualität der Klassenführung des Referenzvideos nicht signifikant abweichend ein als die Referenzgruppe. Die vorangegangenen Analysen verweisen darauf, dass die Unterschiede in den Bereichen kognitive Aktivierung und Motivierung auf einen Reihenfolgeeffekt der kritischeren Einschätzung im Sinne einer Abwertung zurückzuführen sind. Die Hypothese 1.2 zur Abwertung kann allerdings nur teilweise empirisch unterstützt werden, da sich nicht für alle Unterrichtsqualitätsmerkmale signifikante Unterschiede finden ließen. Insgesamt zeigten sich somit eher Abwertungs- als Aufwertungsprozesse durch einen Reihenfolgeeffekt mit mittlerer Ausgangsqualität.

## **5.2 Prädiktoren für die Urteile und das Auftreten des Reihenfolgeeffekts**

Die Ergebnisse zur Beantwortung der Fragestellung 2, ob die individuellen Faktoren Erfahrungszeit oder Müdigkeit prädiktiv für die Urteile der Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierung bei dem Unterrichtsvideo mittlerer Qualität als auch für Unterschiede im Auftreten des Reihenfolgeeffekts sind, sind in den Tab. 4 und 5 dargestellt.

**Vorerfahrung (Erfahrungszeit in der Schule)**

Es zeigte sich in allen Modellen, dass die Erfahrungszeit in der Schule nicht prädiktiv für die Urteile der kognitiven Aktivierung und Klassenführung bei dem Unterrichtsvideo mittlerer Qualität war. Hinsichtlich der Urteile der Motivierung bei dem Unterrichtsvideo mittlerer Qualität wies die Erfahrungszeit eine prädiktive Kraft auf; auch unter Berücksichtigung der Interaktionsterme. Raterinnen und Rater, die eine größere Erfahrungszeit in der Schule angaben, schätzten die Qualität der Motivierung des Referenzvideos unabhängig von der Zugehörigkeit der Bedingung niedriger ein als unerfahrene Raterinnen und Rater. Da die Prädiktion nur im Bereich Motivierung zu finden war, konnte die Hypothese 2.1 nur teilweise empirisch unterstützt werden. Die nach der Hypothese 2.2 erwarteten Prädiktionen der Interaktionen zwischen den Gruppen und der Erfahrungszeit auf die Beurteilungen der kognitiven Aktivierung, Klassenführung und Motivierung konnten nicht festgestellt werden. Somit konnte keine Verstärkung oder Verminderung des Reihenfolgeeffekts aufgrund unterschiedlicher Erfahrungszeit in der Schule nachgewiesen werden und die Hypothese 2.2 konnte demnach nicht empirisch gestützt werden.

**Mentaler Zustand (Müdigkeit)**

In Bezug auf die Urteile der kognitiven Aktivierung und Klassenführung bei dem Unterrichtsvideo mittlerer Qualität wies die Müdigkeit in allen Modellen keinen gerichteten Zusammenhang auf. Bezüglich der Motivierung war die Müdigkeit prädiktiv für die Urteile bei dem Unterrichtsvideo mittlerer Qualität. Raterinnen und Rater, die ihre Müdigkeit hoch beurteilten, schätzten unabhängig von der Zugehörigkeit der Bedingung die Qualität der Motivierung des Videos mit mittlerer Qualität höher ein als wachere Raterinnen und Rater. Dieses Ergebnis zeigte sich allerdings nicht unter Berücksichtigung der Interaktionsterme in Modell III. Da sich die Prädiktion nur für den Bereich Motivierung und auch nur in dem Modell I zeigte, konnte die Hypothese 2.3 nur teilweise empirisch unterstützt werden. Im Hinblick auf die Urteile in den Bereichen kognitive Aktivierung und Klassenführung ließen sich keine Interaktionseffekte zwischen der experimentellen Bedingung und der Müdigkeit nachweisen, jedoch bezüglich der Urteile im Bereich Motivierung. Raterinnen und Rater, die ihre Müdigkeit hoch beurteilten und der Bedingung A (niedrig/mittel) zugewiesen wurden, schätzten die Qualität der Motivierung des Videos mittlerer Qualität höher ein als wachere Raterinnen und Rater der Bedingung A. Dies bedeutet, dass der Reihenfolgeeffekt durch eine hohe Müdigkeit, also eher müderen Raterinnen und Rater, verstärkt wurde. Dies war allerdings

nur für ein positiveres Urteil nach einem Video mit geringerer Unterrichtsqualität zu beobachten und auch nur, wenn die Müdigkeit als Variable nicht einbezogen wurde. Somit konnte die Hypothese 2.4 teilweise empirisch unterstützt werden.

Tabelle 4.

Prädiktoren (Experimentelle Bedingung A und C, Erfahrungszeit und Interaktionsterme Bedingung A oder C x Erfahrungszeit) für die Urteile der Unterrichtsqualitätsmerkmale

Prädiktoren	Kognitive Aktivierung									Klassenführung						Motivierung			
	I			II			III			I			II			III			
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	
Experimentelle Bedingung A (EB A)	.08	0.13	.07	0.13	.07	0.13	.16	0.13	.17	0.13	.18	0.13	.12	0.12	-.12	0.13	-.11	0.12	
Experimentelle Bedingung C (EB C)	-.38**	0.12	-.38**	0.12	-.20	0.13	-.20	0.13	-.21	0.13	-.20	0.13	-.41**	0.12	-.43**	0.12	-.41**	0.12	
Erfahrungszeit (EZ)	.01	0.11	-	-	-.01	0.26	-.12	0.11	-	-	-.30	0.23	-.22*	0.11	-	-	-.51*	0.21	
EB A x EZ	-	-	-.05	0.11	-.05	0.14	-	-	.02	0.12	.13	0.14	-	-	-.05	0.11	.14	0.13	
EB C x EZ	-	-	.05	0.11	.06	0.21	-	-	-.07	0.12	.16	0.21	-	-	-.10	0.11	.30	0.20	
R <sup>2</sup>	.17*	0.08	.18*	0.08	.18*	0.08	.13	0.08	.12	0.07	.14	0.08	.21*	.09	.17*	.08	.23**	0.09	
Anmerkungen.	Saturierte Modelle; * $p < .05$ , ** $p < .01$ .																		

Tabelle 5.  
 Prädiktoren (Experimentelle Bedingung A und C, Müdigkeit und Interaktionstherme Bedingung A oder C x Müdigkeit) für die Urteile der Unterrichtsqualitätsmerkmale

Prädiktoren	Kognitive Aktivierung						Klassenführung						Motivierung							
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III		
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE		
Experimentelle Bedingung A (EB A)	.08	0.13	.07	0.13	.16	0.13	.16	0.13	.16	0.13	.16	0.13	.16	0.13	-.12	0.12	-.11	0.13	-.11	0.12
Experimentelle Bedingung C (EB C)	-.38**	0.12	-.38**	0.12	-.22	0.13	-.22	0.13	-.22	0.13	-.22	0.13	-.45**	0.12	-.45**	0.12	-.45**	0.12	-.45**	0.12
Müdigkeit (M)	.02	0.11	-	-	-.06	0.24	-.02	0.11	-	-	.00	0.25	.21*	0.11	-	-	.08	0.23	-	-
EB A x M	-	-	-.06	0.11	-.02	0.18	-	-	-.10	0.11	-.10	0.18	-	-	.22*	0.11	.17	0.17	-	-
EB C x M	-	-	.09	0.11	.13	0.19	-	-	.06	0.11	.06	0.20	-	-	.10	0.11	.05	0.19	-	-
R <sup>2</sup>	.17*	0.08	.19*	0.08	.19*	0.09	.11	0.07	.12	0.07	.12	0.07	.20*	.09	.21*	.09	.21**	0.09	-	-
Anmerkungen.	Saturierte Modelle; * $p < .05$ , ** $p < .01$ .																			

## 6 Diskussion

Die vorliegende Untersuchung ging der Frage nach, ob Urteile von Unterrichtsvideos von der dargebotenen Reihenfolge (F1) sowie von individuellen Merkmalen abhängig sind und welche individuellen Merkmale den Reihenfolgeeffekt verringern oder begünstigen (F2). Die Resultate bezüglich aller drei Unterrichtsqualitätsmerkmale zeigten, dass zwischen den Urteilen der Gruppe A (niedrig/mittel) und B (mittel/hoch) keine bedeutsamen Unterschiede vorlagen. Die Qualität eines Videos mittleren Qualitätsniveaus wurde nicht höher nach einem vorangestellten Urteil eines Videos niedrigen Qualitätsniveaus eingeschätzt, als wenn vorher kein anderes Video beurteilt wurde (H1.1). Die Analysen wiesen keine statistisch bedeutsamen Effekte der experimentellen Bedingung auf, allerdings ließen sich Effekte zwischen den Urteilen der Gruppe A und B von *Cohens d* = 0,17–0,37 finden.

Für die Urteile der kognitiven Aktivierung und Motivierung zeigten sich signifikante Unterschiede zwischen den Gruppen B (mittel/hoch) und C (hoch/mittel). Die kognitive Aktivierung und Motivierung wurden durchschnittlich negativer nach einer vorigen Beurteilung eines Videos hohen Qualitätsniveaus beurteilt, als wenn vorher kein anderes Video bewertet wurde (H1.2). Somit deuteten die Resultate bezüglich der kognitiven Aktivierung und Motivierung eher auf Abwertungs- als auf Aufwertungsprozesse durch einen Reihenfolgeeffekt mit mittlerer Ausgangsqualität hin.

Trotz einer hohen Interraterreliabilität im Bereich der kognitiven Aktivierung, welche für ein gelungenes Training stehen kann, weisen die Ergebnisse auf die Bedeutung der Darbietungsreihenfolge der Videos für das Urteil der kognitiven Aktivierung hin. Somit können Urteile trotz hoher Übereinstimmungen Beurteilungsfehlern unterliegen.

Für die Klassenführung ließen sich keine erwarteten Unterschiede (Gruppe A vs. B und Gruppe B vs. C) aufgrund der sequenziellen Darbietung nachweisen. Die Perspektive der Externen gilt als direktester Weg zur Erfassung der Klassenführung (Clare et al. 2001). Die Klassenführung beinhaltet im Vergleich zu den anderen Merkmalen eindeutig beobachtbare Facetten und setzt somit weniger Interpretationsaufwand voraus (Clausen 2002). Dies kann dazu führen, dass in der vorliegenden Studie keine signifikanten Unterschiede festgestellt wurden und der Grad des Interpretationsaufwandes und das Auftreten eines Reihenfolgeeffekts zusammenhängen.

In Bezug auf Urteile der Motivierung sind die negativen Abweichungen der Urteile im Bereich Motivierung der Gruppe A auffällig. Erwartet wurden positive Abweichungen zu den

Urteilen der Referenzgruppe, die sich auch für die kognitive Aktivierung und Klassenführung gezeigt haben. Eine mögliche Erklärung dafür könnte sein, dass die Motivierung als eher schwierig zu beobachtendes Konstrukt gilt und für Urteile das komplexe Zusammenwirken struktureller und inhaltlicher Unterrichtsqualitätsmerkmale sowie auch soziale und persönliche Komponenten der Lehrkräfte und Lernenden berücksichtigt werden muss (Clausen et al. 2003; Rakoczy 2008). Die Komplexität für eine Bewertung dieses Merkmals kann auch zu einer Notwendigkeit von subjektiveren Interpretationen bei hoch inferenten Ratings führen, welche wiederum häufiger von Beurteilungsfehlern betroffen sind als sichtbarere Unterrichtsqualitätsmerkmale (Praetorius 2014). Neben der Komplexität ist auch der kurze Einblick der Externen in das Unterrichtsgeschehen ein weiterer Faktor, welcher die Urteile verzerren kann. Dieser Stichprobeneffekt, eingeschränkte kurze Beobachtungsstichprobe, wird von der Beobachtbarkeit des Unterrichtsqualitätsmerkmals moderiert und so können Unterrichtsqualitätsmerkmale mit niedriger Beobachtbarkeit den Effekt verstärken (Clausen 2002). Dementsprechend konnten sich bei der Einschätzung der Motivierung eventuell zwei Faktoren gegenseitig bedingen: Komplexität des Unterrichtsqualitätsmerkmals und kurze Beobachtungsstichprobe.

Bezüglich der Fragestellung 2 zeigte sich ein Zusammenhang zwischen der Erfahrungszeit in der Schule und der Urteile des Referenzvideos. Die Studierenden mit einer höheren Erfahrungszeit in der Schule beurteilten die Motivierung des Referenzvideos negativer als Studierende mit weniger Erfahrungszeit. Eine Erklärung können implizite Theorien der erfahrenen Studierenden sein, die auf Grundlage von Erfahrung in der Schule, zum Beispiel im Rahmen von Praktika, entwickelt wurden und auch nicht durch ein Training geändert werden können (Praetorius et al. 2012). Ferner legen erfahrene Raterinnen und Rater häufig ein breiteres Spektrum für die Beurteilung an als Unerfahrene und greifen auch auf Faktoren zurück, die nicht für das Rating vorgegeben waren (Leckie und Baird 2011). Darüber hinaus zeigte sich ein gerichteter Zusammenhang zwischen der Müdigkeit und den Urteilen über die Motivierung bei dem Unterrichtsvideo mittlerer Qualität. Studierende, die ihre Müdigkeit hoch beurteilten, schätzten die Qualität der Motivierung des Videos mit mittlerer Qualität höher ein als wachere Raterinnen und Rater. Dieses Ergebnis ist gegensätzlich zu den Befunden von Mashburn et al. (2014), dass die Probandinnen und Probanden mit zunehmender Müdigkeit strengere Urteile vergaben. Die Wahrnehmung und das Urteilsvermögen werden unter anderem durch kognitive Merkmale wie die Müdigkeit bestimmt (z.B. Feltz und Cokely 2011). So kann es zum Beispiel aufgrund von Müdigkeit zu einer selektiven

Informationsaufnahme kommen und demzufolge werden Urteile auf Basis mangelnder Informationen getroffen (Schmidt-Atzert et al. 2004). Dies kann ein Grund sein, warum sich Studierende mit divergierender Ausprägung der Müdigkeit im Urteil über die Motivierungsqualität unterscheiden. Die Affektheuristik, dass Urteile auf Basis des eigenen Befindens getroffen werden (Fiske und Taylor 2017), kann an dieser Stelle keine Erklärung für die vorliegenden Resultate sein. Demnach hätten die Studierenden, die aufgrund von Müdigkeit ein schlechteres Befinden hatten, negativer die Motivierungsqualität beurteilt als wachere Studierende mit einem positiveren Befinden.

Ein Interaktionseffekt wurde zwischen der Müdigkeit und der experimentellen Bedingung A (niedrig/mittel) gefunden. Im Einklang mit bisherigen Befunden kann anhand der Ergebnisse vermutet werden, dass die Müdigkeit den Reihenfolgeeffekt verstärkt. Raterinnen und Rater, die ihren Zustand als eher müde einschätzten und zuvor ein Video niedrigen Qualitätsniveaus bewerteten, orientierten sich stärker an diesem und überschätzten die Motivierung des Videos mittlerer Qualität stärker als Personen, die ihre Müdigkeit geringer einschätzten. Auch für die verstärkte Verzerrung bei hoher Müdigkeit kann eine Begründung der Zusammenhang einer zunehmenden Müdigkeit mit sinkender Wahrnehmung der relevanten Lehrkrafthandlungen sein (Schmidt-Atzert und Amelang 2012). So haben eventuell müdere Raterinnen und Rater für die Beurteilung der Qualität bedeutsame Kriterien aufgrund von niedriger Aufmerksamkeit nicht wahrgenommen. Dass sich die Resultate der Prädiktion der Müdigkeit sowie der Interaktionseffekt nur zeigten, wenn jeweils die andere Variable nicht berücksichtigt wurde, konnte durch die Korrelation von  $r = 0,58$  ( $p = 0,000$ ) zwischen der Müdigkeit und des Interaktionsterms der Bedingung A  $\times$  Erfahrungszeit verursacht werden. Alle signifikanten gerichteten Zusammenhänge betrafen den Bereich Motivierung. Diese Resultate bestätigen, dass das Unterrichtsqualitätsmerkmal Motivierung schwierig für Externe einzuschätzen ist und externe Urteile über die Motivierung häufiger durch individuelle Merkmale beeinflusst werden.

## **6.1 Einschränkungen**

Bei der Interpretation der Ergebnisse bleibt zu beachten, dass das Training der Raterinnen und Rater mit drei Stunden vergleichsweise zu anderen Studien kurz war (z. B. Clausen et al. 2003) und besonders die gefundenen Zusammenhänge der Vorerfahrung eventuell geringer ausfallen würden, wenn zum Beispiel eine größere Spannbreite an Videos mit niedriger, mittlerer und hoher Qualität vorab gesichtet worden wäre. Hinzu sollte bei der Einordnung der Resultate

berücksichtigt werden, dass in der vorliegenden Studie nur ein Item pro Merkmal eingeschätzt wurde. Dies könnte trotz ausführlicher Beschreibung der Unterrichtsqualitätsmerkmale im Manual und Beispielfacetten zu breiten Interpretationen bei den Raterinnen und Ratern führen.

Für die Vergleiche der Urteile der Qualität des Videos mittleren Qualitätsniveaus wurde als Referenz das Urteil der Gruppe B (mittel/hoch) herangezogen, in der die Studierenden das Referenzvideo zuerst sahen und daher nicht durch ein anderes Video in ihrer Beurteilung beeinflusst wurden. Dieses Vorgehen hat einerseits den Vorteil der Möglichkeit einer Abschätzung der absoluten Größe einer möglichen Urteilsverzerrung aufgrund der wahrscheinlich eher unbeeinflussten Bewertung in Gruppe B. Andererseits können im Vergleich zu den Gruppe A (niedrig/mittel) und C (hoch/mittel), die das Referenzvideo als zweites sahen, zwischen dem zuerst einzuschätzenden Video und dem zweiten Video kognitive Prozesse auftreten (zum Beispiel kognitive Ermüdung), die wiederum zu Beurteilungsfehlern führen können.

Aufgrund des vorigen Urteils eines Videos mit einer entweder negativeren oder positiveren Unterrichtsqualität könnten die Ergebnisse nicht nur auf einen Reihenfolgeeffekt, sondern auch auf einen Kontrasteffekt hinweisen. Hierbei werden Informationen aufgrund der im Kontrast stehenden Vergleichsinformation intensiver wahrgenommen und Urteile in die entgegengesetzte Richtung verzerrt (Lenske 2016). Des Weiteren könnte auch der Ankereffekt zutragen gekommen sein, wobei vorige Informationen als Anker für zukünftige Urteile dienen (Tversky und Kahneman 1974). Es ist aber zu beachten, dass die Beurteilungsfehler nicht eindeutig trennbar sind und in einer engen Beziehung stehen (Hermann 2016). Zusätzlich führen sie alle zu dem gleichen unerwünschten Effekt – verzerrte Beurteilungen der Unterrichtsqualität.

## **6.2 Implikationen für Praxis und Forschung**

Die vorliegende Untersuchung kann aufgrund des experimentellen Designs wichtige Hinweise für die Messung der Unterrichtsqualität anhand von Videos liefern. Für weitere Studien zur Unterrichtsqualität, die auf Videoratings zurückgreifen, sollte die Reihenfolge der gezeigten Videos mit unterschiedlichen Qualitätsniveaus beachtet werden. Hierbei kann eine Randomisierung der Videos ein mögliches Auftreten eines Reihenfolgeeffekts verhindern. Mögliche Unterschiede in den Urteilen aufgrund der Vorerfahrung könnten eventuell mit einer umfänglicheren Schulung als in der vorliegenden Studie ausgeglichen werden.

Da der mentale Zustand, wie die Müdigkeit, Beurteilungsfehler verstärken und eine höhere Müdigkeit größere Verzerrungen in den Urteilen hervorrufen kann, ist es bedeutsam dies in zukünftigen Studien zu kontrollieren, zum Beispiel durch eine vorige Erfassung des mentalen Zustandes. Um Fehlattritionen aufgrund des aktuellen mentalen Zustandes zu vermeiden, ist es ebenfalls wichtig, Raterinnen und Rater zu sensibilisieren bei ihren Urteilen nicht ihren aktuellen mentalen Zustand zu berücksichtigen (Stroebe 2014). Bislang gibt es nur wenige Studien, die mögliche Einschränkungen von Urteilen Externer in der Unterrichtsforschung untersuchten. Folge sind die nicht valide Erfassung von Unterrichtsqualitätsmerkmalen und möglicherweise Fehlinterpretationen von Zusammenhängen zwischen Unterricht und beispielsweise Lernleistung. Insbesondere in dem Bereich der Motivierung sollte neben der Perspektive der Externen auch die Sicht der Lehrkräfte und Lernenden herangezogen werden. Die Komplexität dieses Unterrichtsqualitätsmerkmals, welches sich unter anderem durch die Berücksichtigung der Beziehungen der beteiligten Personen auszeichnet, ist nur eingeschränkt durch die Perspektive von Externen zu erfassen. Schließlich weist die vorliegende Studie auf mögliche Einschränkungen von Urteilen externer Raterinnen und Rater hin, indem vorige Unterrichtsszenen als mögliche Referenz für das aktuelle Urteil verwendet werden und dies insbesondere zu einer Unterschätzung der Unterrichtsqualität führen kann.

**Danksagung** Ein besonderer Dank gilt Dr. Matthias Trendtel für die Unterstützung im Bereich der statistischen Methodenwahl.

**Förderung** Das Institut für Schulentwicklungsforschung förderte und führte die Studie durch.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Interessenkonflikt** J. Iglar, A. Ohle-Peters und N. McElvany geben an, dass kein Interessenkonflikt besteht.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

## Literatur

- Begrich, L., Fauth, B., Kunter, M., & Klieme, E. (2017). Wie informativ ist der erste Eindruck? Das Thin-Slices-Verfahren zur videobasierten Erfassung des Unterrichts. *Zeitschrift für Erziehungswissenschaften*, 20(1), 23–47.
- Bless, H., & Keller, J. (2006). Urteilsheuristiken. In H.-W. Bierhoff & D. Frey (Hrsg.), *Handbuch der Psychologie: Sozialpsychologie und Kommunikationspsychologie* (S. 294–300). Göttingen: Hogrefe.
- Bless, H., Bohner, G., Schwarz, N., & Strack, F. (1990). Mood and persuasion: a cognitive response analysis. *Personality and Social Psychology Bulletin*, 16, 331–345.
- Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition: how individuals construct social reality*. Philadelphia: Psychology Press.
- Brophy, J. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71, 733–750.
- Clare, L., Valdés, R., Pascal, J., & Steinberg, J.R. (2001). *Teachers' assignments as indicators of instructional quality in elementary schools* (CSE Tech. Rep. No. 545). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.
- Cushman, F., & Mele, A. (2008). Intentional action: two-and-a-half folk concepts? In J. Knobe & S. Nichols (Hrsg.), *Experimental philosophy* (S. 171–188). Oxford: Oxford University Press.
- Dalbert, C. (1992). Subjektives Wohlbefinden junger Erwachsener: Theoretische und empirische Analysen der Struktur und Stabilität. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 13(4), 207–220.
- Decristan, J., Hess, M., Holzberger, D., & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale.

- Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift für Pädagogik*, 66, 102–116.
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Emmer, E.T., & Stough, L.M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, 36, 103–112.
- Engle-Friedman, M., Mathew, G.M., Martinova, A., Armstrong, F., & Konstantinov, V. (2018). The role of sleep deprivation and fatigue in the perception of task difficulty and use of heuristics. *Sleep Science*, 11(2), 74–84.
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014a). Student ratings of teaching quality in primary school: dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014b). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137.
- Fauth, B., Göllner, R., Lenske, L., Praetorius, A., & Wagner, W. (2020). Who sees what? Theoretical considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik*, 66, 138–155.
- Feltz, A., & Cokely, E.T. (2011). Individual differences in theory-of-mind judgments: order effects and side effects. *Philosophical Psychology*, 24(3), 343–355.
- Fiske, S.T., & Taylor, S.E. (2017). *Social cognition: from brains to culture* (3. Aufl.). London: SAGE.
- Gabriel-Busse, K., Groß-Mlynek, L., Feldhoff, T., & Harring, M. (2020). Eine Unterrichtssequenz – unterschiedliche Einschätzungen. Analyse videografiertes Unterrichtssequenzen als Bestandteil einer evidenzbasierten Lehrer/innenausbildung. In I. Gogolin, B. Hannover & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung* (ZfE-Edition, Bd. 4, S. 291–314). Wiesbaden: Springer VS.

- Geiser, C. (2011). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung*. Wiesbaden: VS.
- Hamre, B. K., & Pianta, R.C. (2010). Classroom environments and developmental processes: Conceptualization, measurement, & improvement. In J.L. Meece & J.S. Eccles (Hrsg.), *Handbook of research on schools, schooling and human development* (S. 25–41). New York: Routledge.
- Hermann, H.-P. (2016). *Tourismuspsychologie*. Heidelberg: Springer.
- Herrle, M., Rauin, U., & Engartner, T. (2016). Videos als Ressourcen zur Generierung von Wissen über Unterrichtsrealität(en). In U. Rauin & M. Herrle (Hrsg.), *Videoanalysen in der Unterrichtsforschung – Methodische Vorgehensweisen und Anwendungsbeispiele* (S. 8–28). Weinheim: Beltz Juventa.
- Ho, A.D., & Kane, T. J. (2013). *The reliability of teaching observations by school personnel*. Seattle: Bill and Melinda Gates Foundation.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Hoyt, W.T., & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: a metaanalysis. *Psychological Methods*, 4, 403–424.
- Kooken, J., Welsh, M.E., McCoach, D.B., Miller, F.G., Chafouleas, S.M., Riley-Tillmann, T. C., & Fabio, G. (2017). Test order in teacher-rated behavior assessments: is counterbalancing necessary? *Psychological Assessment*, 29(1), 98–109.
- Kruglanski, A.W., & Webster, D.M. (1996). Motivated closing of the mind: “Seizing” and “freezing”. *Psychological Review*, 103, 263–283.
- Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts*. Paderborn: Ferdinand Schöningh.
- Lau, A., & Plessner, H. (2016). *Sozialpsychologie und Sport – Ein Lehrbuch in 12 Lektionen*. Aachen: Meyer & Meyer.

- Leckie, G., & Baird, J.-A. (2011). Rater effects on essay scoring: a multilevel analysis of severity drift, central tendency, and rater experience. *Journal of Educational Measurement*, 48(4), 399–418.
- Lenske, G. (2016). *Schülerfeedback in der Grundschule. Untersuchungen zur Validität*. Münster: Waxmann.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen  
Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenz und Kompetenzentwicklung von Lehrerinnen und Lehrern*. *Zeitschrift für Pädagogik*, 51, 47–70.
- Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K., & Pauli, C. (2009). Quality of geometry instruction and its short-term impact on students? Understanding of Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537.
- Lotz, M., Gabriel, K., & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, 59(3), 357–380.
- Martin, E., & Wawrinowski, U. (2014). *Beobachtungslehre: Theorie und Praxis reflektierter Beobachtung und Beurteilung* (6. Aufl.). Weinheim: Beltz.
- Mashburn, A. J., Meyer, J.P., Allen, J.P., & Pinata, R.C. (2014). The effect of observation length and presentation order on the reliability and validity of an observational measure of teaching quality. *Educational and Psychological Measurement*, 74(3), 400–422.
- McElvany, N., Schroeder, S., Richter, T., Hachfeld, A., Baumert, J., Schnotz, W., et al. (2012). Cognitively demanding learning materials with texts and instructional pictures: teachers' diagnostic skills, pedagogical beliefs, and motivation. *European Journal of Psychology of Education*, 27, 403–420.
- Messner, C., & Schmid, B. (2007). Über die Schwierigkeit unparteiische Entscheidungen zu fällen: Schiedsrichter bevorzugen Fußballteams ihrer Kultur. *Zeitschrift für Sozialpsychologie*, 38(2), 105–110.

- Meyer, H. (2007). Zehn Merkmale guten Unterrichts. In W. Endres (Hrsg.), *Lernen lernen – wie stricken ohne Wolle? 13 Experten streiten über Konzepte und Modelle zur Lernmethodik* (S. 167–187). Weinheim: Beltz.
- Miller, N., & Campbell, D.T. (1959). Recency and primacy in persuasion as a function of the timing of speeches and measurements. *Journal of Abnormal and Social Psychology*, 59, 1–9.
- Muthén, L.K., & Muthén, B. O. (2017). *Mplus user's guide* (8. Aufl.). Los Angeles: Muthén & Muthén.
- Myford, C.M., & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Reach measurement: Part I. *Journal of Applied Measurement*, 4, 386–422.
- Ohle, A., & McElvany, N. (2016). Erfassung von Unterrichtsqualität in der Grundschule: Kognitiver Anspruch, Strukturierung und Motivierungsqualität. In N. McElvany, W. Bos, H.G. Holtappels, M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts. Dortmunder Symposium der Empirischen Bildungsforschung* (S. 117–134). Münster: Waxmann.
- Pauli, C., & Reusser, K. (2003). Unterrichtsskripts im schweizerischen und im deutschen Mathematikunterricht. *Unterrichtswissenschaft*, 31, 238–272.
- Pietsch, M., & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Praetorius, A.-K. (2013). Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter. Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 174–185.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: do they fulfill what they promise? *Learning and Instruction*, 22, 387–400.

- Praetorius, A.-K., Rogh, W., & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft, 48*, 303–318.
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht: Unterricht aus der Perspektive von Lernenden und Beobachtern*. Münster: Waxmann.
- Ryan, R.M., & Deci, E.L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary Educational Psychology*. <https://doi.org/10.1016/j.cedpsych.2020.101860>.
- Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik* (5. Aufl.). Heidelberg: Springer. Schmidt-Atzert, L., Büttner, G., & Bühner, M. (2004). Theoretische Aspekte von Aufmerksamkeits-/ Konzentrationsdiagnostik. In G. Büttner & L. Schmidt-Atzert (Hrsg.), *Diagnostik von Konzentration und Aufmerksamkeit* (S. 3–22). Göttingen: Hogrefe.
- Schwindt, K. (2008). *Lehrpersonen betrachten Unterricht: Kriterien für die kompetente Unterrichtswahrnehmung*. Münster: Waxmann.
- Seidel, T., & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaften, 32*, 1–21.
- Stroebe, W. (2014). Strategien zur Einstellungs- und Verhaltensänderung. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 231–268). Berlin: Springer.
- Sudman, S., Bradburn, N.M., & Schwarz, N. (1996). *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1131. Webster, D.M., Richter, L., & Kruglanski, A.W. (1996). On leaping to conclusions when feeling tired: mental fatigue effects on impressional primacy. *Journal of Experimental Social Psychology, 32*, 181–195.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Wolff, C.E., Jarodzka, H., & Boshuizen, H.P.A. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, 66, 295–308.

## 7 Weiterführende Analysen (Beitrag IV)

### 7.1 Students' perceived motivational support in elementary school – How accurate are teachers' judgments?

Jennifer Igler<sup>a\*</sup>, Annika Ohle-Peters<sup>b</sup>, Annika Teerling<sup>c</sup>, Olaf Köller<sup>c</sup> & Nele McElvany<sup>b</sup>

<sup>a</sup>Learning Disabilities in Rehabilitation and Education at Technical University Dortmund, Otto-Hahn-Straße 6, 44227 Dortmund, Germany; jennifer.igler@tu-dortmund.de

<sup>b</sup>Center for Research on Education and School Development (IFS) at Technical University Dortmund, Vogelpothsweg 78, 44227 Dortmund, Germany; annika.ohle-peters@tu-dortmund.de; nele.mcelvany@tu-dortmund.de

<sup>c</sup>Leibniz Institute for Science and Mathematics Education (IPN) at Kiel University, Olshausenstraße 62, 24118 Kiel, Germany; teerling@ipn.uni-kiel.de; koeller@ipn.uni-kiel.de

\*Corresponding author.

#### *Abstract*

Motivational support is a significant facet of teaching quality and relevant for students' motivation and learning. Teachers need to judge students' perceptions of motivational support accurately for adaptive teaching. To investigate the accuracy of teachers' judgments of perceived motivational support and its impact on students' intrinsic reading motivation, teachers rated how their students perceived autonomy, competence, and relatedness support on the class level ( $N = 50$  teachers) and for ten students on the individual level ( $N = 49$  teachers). Results showed low correlations between teachers' judgments and students' perceptions. This illuminates that more monitoring of motivational support is needed.

*Keywords:* Motivational support, Students' perceptions, Teacher judgment, Judgment accuracy, Elementary teachers

*Highlights:*

- Motivational support must be measured indirectly and best via students' perceptions
- Experiencing autonomy, competence, and relatedness can increase motivation
- Low correlations of teachers' judgments and students' perceptions were found
- Teachers face particular difficulties in judging autonomy support accurately
- Negative correlation of overestimating support on students' motivation was observed

*Funding:* The project was funded by the German Federal Ministry of Education and Research (funding code 01JI1504A & B).

## **1 Introduction**

Motivational support is an important indicator of teaching quality and refers to stimulating and maintaining students' motivation. In this context, students' perceptions of self-determination are essential for their motivation and learning and can be fostered by enjoying an activity or developing interest in the content (Ryan & Deci, 2020). Because students are the addressees of lessons, it is important to consider the student perspective on motivational support. Even though all learners in a class receive the same learning offer from the same teacher, studies have repeatedly shown wide variance in students' ratings of teaching quality and specifically motivational support within class (e.g., Fauth et al., 2014). In order to provide teaching that is as motivating as possible, teachers need the professional competence to support students individually in their needs and adapt their practice to students' heterogeneity. Consequently, teachers must demonstrate diagnostic competence to assess learners' perceptions of motivational support and be able to adjust their teaching accordingly. Diagnostic competence is one key aspect of teachers' professionalism, alongside classroom management, instructional competence, and subject-matter knowledge, and is defined as the ability to judge students' characteristics accurately beyond a specific situation (Herppich et al., 2018; Weinert, 2001). Teachers' judgment accuracy is determined by examining the correspondence between teachers' judgments of student characteristics and students' actual characteristics as measured

by self-reported questionnaire or standardized tests (Südkamp et al., 2018; Südkamp et al., 2012; Thiede et al., 2015). Teachers' judgments also affect instructional practice, classroom activities, students' achievement, students' motivation, and even their school careers (Bonefeld et al., 2020; Hamre & Pianta, 2001, 2005; Shavelson & Stern, 1981). Teachers who assess their students accurately are more likely to make appropriate lesson decisions (Vogt & Rogalla, 2009). Studies on judgment accuracy have examined different student characteristics central for educational success (Urhahne & Wijnia, 2021). However, existing research tends to focus on judgment accuracy regarding manifest variables (such as students' cognitive abilities, e.g., Machts et al., 2016;), even though latent variables (such as motivational characteristics, e.g., Urhahne & Zhu, 2015) also play an important role for learning at school. Promoting motivational characteristics is an important goal of school education processes, both because of their intrinsic value and because of their connection to students' achievement (Förster & Souvignier, 2014). Accordingly, it is important that teachers know how students perceive their motivational support and how they can react to encourage students in the best possible way.

### **1.1 Motivational support in the classroom and the importance of students' perceptions**

Teaching quality predicts students' outcomes and is one of the keys to students' academic success (e.g., Kunter et al., 2013; Sanusi et al., 2020; Steiner et al., 2018). Domain-independent aspects of teaching quality can be categorized into three dimensions: cognitive activation, classroom management, and student support (Praetorius et al., 2018). These dimensions are similar to the three domains of the Classroom Assessment Scoring System (instructional support, classroom organization, and emotional support; Pianta & Hamre, 2009). The dimension, student support, includes a positive climate, recognizing students' needs, paying attention to students' interests and *motivational support* by teachers (Hamre & Pianta, 2010). Self-determination theory (SDT) provides an important theoretical framework to explain how motivation can be fostered (Furtak & Kunter, 2012). Here, intrinsic motivation is particularly important and describes a certain type of motivation in which learning processes or activities are self-determined, for example based on interest in the learning content (Deci & Ryan, 2000). Intrinsically motivated students have higher performance and greater engagement (Froiland & Worrell, 2016; Taylor et al., 2014). When it comes to facilitating students' intrinsic motivation, students are most motivated to learn in class when teachers meet

their basic needs of feeling autonomous, competent, and positively related to others (Roeser et al., 2000; Ryan & Deci, 2000, 2020). Gnambs & Hanfstingl (2016) showed that lower intrinsic motivation is associated with declines in psychological need satisfaction, which means perceiving less autonomy, competence, and relatedness. *Perceived autonomy* characterizes individuals' tendency to self-organize their own behavior and to act in harmony with their integrated self. To facilitate students' perceived autonomy, it is important to allow them to choose between interesting exercises and learning subjects. Another basic need according to SDT is *perceived competence*, which refers to one's ability to influence the environment and achieve goals effectively. An optimal level of task requirements and positive feedback can contribute to students' experience of competence in the classroom. *Perceived relatedness* reflects the desire to feel connected to others and know that others care about oneself. Support by teacher to connect with class mates is relevant for creating a learning climate in which students feel integrated (Deci & Ryan, 2000, 2008; Gucciardi et al., 2020). Studies have shown that satisfying students' three basic needs is significant for school success even in elementary school (Oga-Baldwin et al., 2017, Oga-Baldwin & Fryer, 2020).

The student's perspective is particularly relevant for assessing whether students feel supported by their teacher (Fauth et al., 2020). Students' perceptions may or may not correspond with their teacher's intention to provide motivational support or their teacher's perception of how much motivational support students feel. Consequently, students' perceptions of teaching are an essential tool for evaluation, are of practical interest for teachers' reflecting on lessons, and are related to students' academic and socio-emotional outcomes (Soebari & Aldridge, 2015; Tennant et al., 2015). Inductive reasoning, which enables students to notice regularities, rules, or generalizations, already develops in elementary school, and at this age, children gain an ability to draw conclusions that allows them to observe and judge people's behaviors quite accurately (Perret, 2015). Research on students' perceptions of teaching quality has shown that students' ratings in elementary school are reliable and valid (van der Scheer, 2019), but also indicates a substantial variability in students' perceptions of motivational support in class (e.g., Fauth et al., 2014). To individually support students' motivation, teachers must meet their needs and be able to judge their students' perceived motivational support accurately.

## 1.2 Teacher judgment accuracy

Judgment accuracy is a central element of teacher professionalism and one of three components of diagnostic competence, alongside professional knowledge and diagnostic activities (Heitzmann et al., 2019; Kramer et al., 2021). Judgment accuracy describes teachers' ability to adequately assess student characteristics: for example, are teacher judgments consistent with objective assessments of students' academic achievement (Artelt & Gräsel, 2009; Hoge & Coladarci, 1989; Ready & Wright, 2011). Moreover, judgment accuracy is a requirement for teachers to effectively adapt their teaching to their students' needs (Gabriele et al., 2016; Vogt & Rogalla, 2009). Additionally, it is relevant for the teaching quality (Baumert & Kunter, 2013; Hill & Chin, 2018).

Three components indicate the accuracy of teacher judgments and allow for differentiated statements about judgment accuracy (Cronbach 1955; Helmke & Schrader, 1987; Urhahne & Zhu, 2015): level, differentiation, and rank. The *level component* captures the difference between teacher judgements and actual student characteristic, allowing for statements about over- or underestimation tendencies (e.g., Feinberg et al., 2009). The *differentiation component* is the quotient of variance of teacher judgments and variance of students' actual characteristics. This component allows for statements about over- or underestimation of the variability of the actual student characteristic (e.g., Fischer et al., 2022). In contrast to the level and differentiation components, the *rank component* is a measure of relative judgment accuracy, which means the extent to which teachers can differentiate between different relative levels of a student characteristic (Thiede et al., 2018). The rank component is the mean correlation between teacher judgments and students' characteristics. Thus, this component captures teachers' ability to predict the rank order of the student characteristic (e.g., Oudman et al., 2023). The rank component is the most frequently used criteria to assess judgment accuracy (Urhahne & Wijnia, 2021). In addition, the *global deviation* measure can be used to measure the absolute difference between a teacher judgment and actual student characteristic (Südkamp et al., 2008). Like the level and differentiation components, the global deviation measure captures absolute judgment accuracy and describes the degree to which teacher judgments match the actual levels of a student characteristic (Thiede et al., 2018).

Two different concepts are used to measure judgment accuracy: the class level and individual level. There are valid arguments for estimating teachers' judgments of students' perceived motivational support on both the class and individual level. An argument for the

former is the hierarchical structure of the educational setting. Students within one class are likely to be more similar to each other than to students in other classrooms (Dicke et al., 2012). A related problem in the context of teacher judgment accuracy concerns rating tendencies (Hoyt, 2000). Teachers' judgments when assessing many students reflect general idiosyncrasies in their ratings, such as harshness or leniency, and thus do not perfectly reflect individual students' characteristics (Dicke et al., 2012). Analyses on the class level take into account such general rating tendencies using centering strategies (Enders & Tofighi, 2007). However, it can be assumed that teacher judgments also differ intra-individually (Praetorius et al., 2013). Based on the realistic accuracy model by Funder (1999), it can be assumed that students on whom teachers pass judgments are not equivalently easy to assess. In the process of judging individual students, the perceived size of the characteristic being assessed plays an important role. Due to the variability in students' perceptions of motivational support within class, teachers may differ intra-individually in assessing students' perceptions of motivational support.

### **1.2.1 Empirical findings on teacher judgment accuracy for motivational support**

In order to individually encourage students' motivation, it is important that teachers know how supported students feel with respect to motivation. However, most prior studies have focused on teacher judgment accuracy regarding students' cognitive abilities or achievement, determining that teachers judge these characteristics quite accurately. Meta-analyses showed correlations between teachers' judgments and students' achievement higher than  $r = 0.60$  (Hoge & Coladarci, 1989; Machts et al., 2016; Südkamp et al., 2012). One explanation for teachers judging students' cognitive characteristics quite accurately is that teachers gain insight in this area by correcting their homework and tests as well as interacting with students in class. In contrast, mental states such as motivation are not always clearly reflected in student behavior, such that teachers need to engage in a higher level of inference to judge students' latent characteristics (Urhahne & Zhu, 2015). For motivational aspects, Brandmiller et al. (2020) found a low correlation between teacher perceptions of students' learning motivation and student-reported learning motivation among fourth graders ( $r = .12$ ). Kunter et al. (2008) reported moderate correlations between teachers' perceptions of cognitive autonomy support for their 9th-grade students and students' ratings of cognitive autonomy support ( $r = .26$ ) and between teachers' perceptions of social support for students and students' ratings of social support ( $r = .31$ ). Looking at the level component of judgment accuracy, teachers showed a

tendency to underestimate students' perceptions of academic self-concept, learning motivation, and learning effort, but overestimated students' perceived enjoyment (Spinath, 2005; Zhu & Urhahne, 2014). Concerning the differentiation component, a tendency to overestimate the variability in students' motivational aspects (students' enjoyment, academic self-concept, learning motivation and learning effort) has been found (Spinath, 2005; Zhu & Urhahne, 2014). Studies showed high accuracy for students' expectancy for success, with a correlation higher than  $r = .60$  (e.g., Urhahne et al., 2011). Moderate to high accuracy was also found for students' academic self-concept, engagement, and belief in their own competence, with correlations  $.17 \leq r \leq .67$  (academic self-concept: Carr & Kurtz-Costes, 1994; Givvin et al., 2001; Marsh & Craven, 1991; Praetorius et al., 2013, Zhu & Urhahne, 2014; engagement: Kaiser & Möller, 2017; conviction of competence: Karing & Artelt, 2014). Low accuracy was found for constructs like learning motivation, with correlations around  $r = .18$  (e.g., Urhahne et al., 2011), and intrinsic motivation, with correlations around  $r = .16$  (e.g., Gagné & Pèrè, 2001; Lee & Reeve, 2012). Kriegbaum et al. (2020) reported weak correlations between teachers' judgments of students' enjoyment and students' self-ratings ( $.22 \leq r \leq .34$ ). These findings are heterogeneous and indicate that teachers have difficulty judging students' motivational aspects, especially in the domains of learning motivation and intrinsic motivation. There has been hardly any empirical research on judgment accuracy regarding aspects of teaching quality, such as motivational support.

### **1.2.2 Relevance of teacher judgment accuracy**

Teachers' judgments are their first source of information on students' learning requirements (Baker et al., 2015). Teachers use their judgments to change their patterns of communication and social interaction with students and make instructional decisions that benefit students (Pielmeier et al., 2018, Urhahne & Wijnia, 2021). It is known that teachers' perceptions of students' characteristics have an influence on students' academic development (Jussim & Eccles, 1992; Rosenthal, 2002). Studies have found an influence of teachers' judgments on students' outcomes, such as achievement or motivation (e.g., Brandmiller et al., 2020; Zhu et al., 2018). The connection between teachers' judgment and students' outcomes can be explained by the mediating role of teachers' behavior and instructional practice (Hill & Chin, 2018; Zhu et al., 2018). Thus, theoretically, teachers may adjust their teaching as a result of their judgments about their students (Hill & Chin, 2018). Hill et al. (2012) demonstrated a strong connection between teachers' domain-specific knowledge of teaching and their

planning and instructional practices, which can in turn affect students' outcomes like motivation (Ekmekci & Serrano, 2022; Rice, 2003). Empirical findings also show that teachers with higher competence, including diagnostic competence, implement more challenging tasks at a high cognitive level and provide a higher degree of individual support, which might be responsible for the effects of diagnostic competence on students' outcomes (Förtsch et al., 2016; Lange et al., 2012). Fauth et al. (2019) found an effect of teacher competence on students' outcomes mediated by teaching quality. Regarding effects of under- or overestimation of students' cognitive skills, underestimated students exhibit lower expectancy for success, academic self-concept, and enjoyment, for example, than overestimated students (Urhahne, 2015; Urhahne et al., 2011). These results indicate the importance of teachers' judgments for students' outcomes such as motivation. Only a few empirical studies have investigated the effects of teachers' judgments on students' motivation, or more specifically the effects of teacher judgment accuracy on students' motivation.

### **1.3 The current study**

In order to provide adaptive motivational support, teachers need to be able to accurately judge how their students perceive the support they currently provide. Furthermore, teacher judgment accuracy is relevant for student outcomes and can predict students' motivation. Due to the general decline in students' motivational characteristics over their school careers (Musu-Gillette et al., 2015), motivation-related teaching quality and its importance for students' motivational characteristics should also be examined more intensively in elementary school. Reading motivation plays a special role in elementary school because of its connection to reading competence, which is a key interdisciplinary competence for school success (Guthrie et al., 2007; Miyamoto et al., 2018; Schröter & Bar-Kochva, 2019). A further distinction between intrinsic and extrinsic reading motivation is necessary (Deci & Ryan, 2008; Schiefele et al., 2012). In contrast to extrinsic reading motivation, in which students read because of external factors, intrinsic reading motivation refers to the decision to read as a result of internal factors, like associating the reading process with positive experiences (Conradi et al., 2014).

First, we investigate how accurately elementary school teachers the level of students' perceived motivational support in terms of autonomy, competence, and relatedness support on the class level (Research question 1). Based on the literature about teacher judgment accuracy regarding students' latent characteristics, we expected an underestimation of students' *perceived autonomy, competence, and relatedness support* (level component). Second, we

want to examine how accurately elementary school teachers judge the level, variation and rank of students' perceived motivational support in terms of autonomy, competence, and relatedness support on the individual student level (Research question 2). We hypothesized an underestimation of students' *perceived autonomy, competence, and relatedness support* (level component), an overestimation of the variance of students' *perceived autonomy, competence and relatedness support* (differentiation component), and low to moderate correlations between teachers' judgments and students' *actually perceived autonomy, competence and relatedness support* (rank component). Another aim of the study was to investigate to what extent teacher judgment accuracy regarding motivational support (global deviation measure) or overestimation or underestimation regarding motivational support (level component) predict students' intrinsic reading motivation (Research question 3a) and its development over time (Research question 3b) on the individual student level. Due to the argument that teachers with more accurate judgments can better adapt their teaching support, leading to higher intrinsic motivation among students, we expected teacher judgment accuracy to predict students' intrinsic reading motivation and its development over time.

## 2 Method

### 2.1 Participants

The study was conducted in  $N = 57$  elementary school classes in six German federal states, with a total of  $N = 945$  students and their German teachers close to the end of the school year in 2017. It represented the fourth measurement point of a larger project, (BiSS-EvalLesen) from the "Education through Language and Writing" (Bildung durch Sprache und Schrift; BiSS) program. Data were collected via questionnaires. To address the aforementioned research questions, German teachers were asked to assess how their students perceived the motivational support they provided, both with respect to the whole class as well as to ten randomly selected individual students in their class.

#### *Sample 1 (SMPI)*

For the first research question regarding judgment accuracy on the class level,  $n = 7$  teachers and their classes were eliminated due to 100% missing reports on the student ( $n = 1$ ) or teacher side ( $n = 2$ ), implausible response behavior from teachers ( $n = 3$ ), or a higher deviation than 1.12 ( $n = 1$ ). The idea behind this is that a standard deviation of  $SD \geq 1.12$  can be expected for absolutely unsystematic ratings (Gaussian distribution curve). If the ratings are systematic

(unimodal distribution curve), the standard deviation must be  $SD < 1.12$ .  $N = 825$  students (52.0% female;  $M_{years} = 10.29$ ,  $SD_{years} = 0.52$ ; 36.3% with immigrant background; 3.1% with disability) were linked to this set of 50 teachers. The  $n = 50$  German teachers (65.3% female) were between 25 and over 60 years old (6.1% 25-29 years; 30.6% 30-39 years; 18.4% 40-49 years; 32.7% 50-59 years; 12.2% over 60 years) and had taught for  $M = 18.37$  years ( $SD = 12.07$ ).

### *Sample 2 (SMP2)*

For the second research question about judgment accuracy on the individual student level,  $n = 8$  teachers and their classes were excluded from the analyses as less than 5 student judgments by teachers or information from students were available ( $n = 5$ ), or implausible response behavior by teachers ( $n = 3$ ) was shown.  $N = 421$  students (51.2% female;  $M_{years} = 10.26$ ,  $SD_{years} = 0.49$ ; 33.9% with immigrant background; 2.9% with disability) were linked to the teacher data set (out of a theoretical maximum of 490 if all ten selected students per class provided answers and teachers judged all 10 of these students). These values for the students' samples are comparable to representative samples for Germany (Stanat et al., 2017). The  $n = 49$  German teachers (65.3% female) were between 25 and over 60 years old (6.1% 25-29 years; 28.6% 30-39 years; 12.2% 40-49 years; 36.7% 50-59 years; 12.2% over 60 years) and had taught for  $M = 18.03$  years ( $SD = 11.85$ ). The  $n = 3$  classes excluded due to teachers providing implausible responses (e.g., answer pattern) were the same in sample 1 and sample 2.

## **2.2 Procedure and materials**

### **2.2.1 Procedure**

Following a pretest of all instruments, each student in class was asked to fill in a paper-pencil survey about the perceived support he or she felt with respect to the three basic needs (autonomy, competence, and relatedness) during the class's most recent German lessons. At the same time, teachers first judged the average *perceived autonomy*, *competence*, and *relatedness support* for the whole class during their most recent German lessons. Additionally, teachers were asked to judge the *perceived autonomy*, *competence*, and *relatedness support* of ten individual students in their class. In order to randomly select the students, teachers sent class lists with the students' identification numbers in advance. The ten cases were selected by picking ten random student identification numbers. Instruments for assessing motivational

support on the student and teacher side were developed based on the well-established scales for self-determination in class by Chen and colleagues (2015).

### 2.2.2 Materials

#### *Student materials*

##### *Students' perceived motivational support*

Students were asked about their *perceived autonomy support* with three items ( $\alpha_{SMP1} = .62$ ,  $\alpha_{SMP2} = .65$ ;  $ICC_{SMP1} = .25$ ,  $ICC_{SMP2} = .25$ ; e.g., “My teacher listened to the way I would like to do something”). Five items were used to assess *perceived competence support* ( $\alpha_{SMP1} = .78$ ,  $\alpha_{SMP2} = .80$ ;  $ICC_{SMP1} = .10$ ,  $ICC_{SMP2} = .11$ ; e.g., “In German class, my teacher gave me the feeling that I could also solve difficult tasks”). *Perceived relatedness support* was operationalized using three items ( $\alpha_{SMP1} = .79$ ,  $\alpha_{SMP2} = .78$ ;  $ICC_{SMP1} = .18$ ,  $ICC_{SMP2} = .18$ ; e.g., “In German class, my teacher helped me feel comfortable with my classmates”). Assessments were made on a four-point scale (1 = disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = agree).

##### *Students' intrinsic reading motivation*

To investigate research questions 3a and 3b, we used the students' intrinsic reading motivation scale by McElvany et al. (2009). Students' intrinsic reading motivation was measured with four self-report items at the beginning ( $M = 3.33$ ,  $SD = 0.79$ ,  $\alpha = .87$ ) and end of fourth grade ( $M = 3.13$ ,  $SD = 0.91$ ,  $\alpha = .91$ ) (e.g., “I enjoy reading”). Thus, the initial level of students' intrinsic reading motivation can be controlled for in the regression analysis. Responses were given on a four-point scale (1 = disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = agree). With respect to missing values, scales were not formed if more than 50.0% of items had missing values.

##### *Teacher materials*

To examine teachers' judgments of students' perceived motivational support with respect to the whole class, three items were used: *perceived autonomy support* on the class level (“The class felt that I supported them in experiencing independence”), *perceived competence support* on the class level (“The class felt that I supported them in experiencing competence”), and *perceived relatedness support* on the class level (“The class had the feeling that I supported them in experiencing belonging in class.”). Responses were given on a four-point scale (1 = disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = agree). Teachers' judgments about

the *perceived autonomy*, *competence*, and *relatedness support* of ten individual students in their class were assessed with items exactly parallel to the student items. Judgments of *perceived autonomy support* on the individual level were operationalized using three items (e.g., “The child had the feeling that I listened to the way they would like to do something”). Five items were used to assess *perceived competence support* on the individual level (e.g., “In class, I gave the child the feeling that it could also solve difficult tasks”). *Perceived relatedness support* on the individual level was assessed with three items (e.g., “In class, the child felt that I supported them so that they could feel comfortable with their classmates”). These judgments were also made on a four-point scale (1 = disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = agree).

### 2.3 Statistical analyses

To assess research questions 1 and 2, descriptive statistics, intercorrelations, the three components of judgment accuracy and the global deviation measure were computed in SPSS (Version 29). Computing the aforementioned three components of judgment accuracy and the global deviation measure are standard when determining judgment accuracy (e.g., Bonfeld et al., 2020; Zhu & Urhahne, 2014). To answer research question 1, the level component was computed as the difference between teachers’ judgments of average *perceived autonomy*, *competence*, and *relatedness support* on the class level and the aggregated mean values on the class level of students’ ratings of *perceived autonomy*, *competence*, and *relatedness support*. The global deviation measure was calculated as the absolute value of the difference between teachers’ judgments of average *perceived autonomy*, *competence*, and *relatedness support* on the class level and the aggregated mean value on the class level of students’ ratings of *perceived autonomy support*. As research question 1 involves comparing teachers’ judgments of the whole class to class mean values, the differentiating and rank components were not computed. On average, there were  $M = 16.50$  ( $SD = 4.51$ ) student responses per class. Regarding research question 2, the level component was calculated as the difference between teachers’ judgments of individual students’ *perceived autonomy*, *competence*, and *relatedness support* on the individual level and students’ individual answers on *perceived autonomy*, *competence*, and *relatedness support*. The differentiating component was determined by dividing the variance of teachers’ judgments by the variance of students’ ratings of *perceived autonomy*, *competence*, and *relatedness support* within each class. Pearson correlations between teachers’ and students’ responses were calculated for each class and then Fisher-z-

transformed. The averaged Fisher z-value was transformed back and resulted in the rank component. The global deviation measure was calculated as the averaged across students' absolute value of the difference between teachers' judgments of individual students' *perceived autonomy*, *competence*, and *relatedness support* and the individual students' own ratings of *perceived autonomy support*. The values of the level component, differentiating component, and global deviation measure were determined using *t*-tests for a random sample tested against 0 or 1. Additionally, the values of the level component, differentiating component, and global deviation measure for all three basic needs (autonomy, competence and relatedness) were compared using *t*-tests for a random sample. The rank components of *perceived autonomy*, *competence*, and *relatedness support* were compared by calculating the difference between a correlation and a fixed value (Lenhard & Lenhard, 2014). As some students were absent or did not have parental permission to participate, teachers judged an average of  $M = 8.59$  ( $SD = 1.40$ ,  $Min = 5$ ,  $Max = 10$ ) students per class.

To address research question 3a, we used structural equation models to determine to what extent teachers' absolute judgment accuracy for perceived motivational support was related to students' intrinsic reading motivation and/or whether teachers' under- or overestimation of perceived motivational support was associated with students' intrinsic reading motivation. To answer these questions, we used sample 2. Thus, analyses were conducted for the global deviation measure and for the level component. First, a series of structural equation models were run to test whether the global deviation measure or the level component of teachers' judgments with respect to each basic need predicted students' intrinsic reading motivation (models I–III & V–VII). Then, models were computed with the global deviation measure or level component of all basic needs simultaneously (autonomy, competence, and relatedness) predicting students' intrinsic reading motivation (models IV & VIII). In these two models, correlations between the global deviation measures and level components of experienced autonomy, competence, and relatedness were permitted. To answer research question 3b about whether teachers' absolute judgment accuracy for perceived motivational support and/or under- or overestimation of perceived motivational support predict intrinsic reading motivation development, structural equation models were specified controlling for the initial level of intrinsic reading motivation. Here, analyses were again conducted for the global deviation measure and for the level component. First, for each basic need, it was tested whether the global deviation measure or the level component predicted the development of students' intrinsic reading motivation over time (models VIII–

XI & XIII–XV). Afterwards, the global deviation measure or the level component of all basic needs simultaneously (autonomy, competence, and relatedness) were used to predict students' intrinsic reading motivation development (models XII & XVI). Correlations between the global deviation measures and the level components of experienced autonomy, competence, and relatedness were permitted in these two models. For these analyses, we used the Mplus software package (Muthén & Muthén, 1998 – 2017). In each model, the dependent variable, intrinsic reading motivation, was specified as a latent construct. The global deviation measure or the level component of all three basic needs – and for research question 3b, the initial level of intrinsic reading motivation – were specified as independent variables. Here, the initial level of the intrinsic reading motivation variable was specified as a latent construct. Due to the hierarchical structure of the data, we used the “cluster= idclass” option in all models, and to obtain unbiased estimates, the “type = complex” option in Mplus together with a maximum likelihood estimator (MLR). Additionally, the model-based approach implemented in Mplus (full information maximum likelihood, FIML) was used to estimate the model parameters. To answer research questions 3a and b, the individual-level data for research question 2 were also used.

## 3 Results

### 3.1 Descriptive statistics

Means, standard deviations, and correlations were calculated for the students' *perceived autonomy*, *competence*, and *relatedness support* and the teachers' judgments (see Table 1). Numerically, the teachers' judgments on the class level were on average higher than the class means of students' perceptions, with the exception of *perceived competence support*. The students' perceptions on the class level correlated strongly with each other, but did not significantly correlate with the teachers' class-level judgments. Different teachers' judgments of students' perceived motivational support on the class level also correlated highly with each other. Looking at descriptive statistics and intercorrelations for the sample related to research question 2 (see Table 1), the mean levels of teachers' judgments of individual students' *perceived competence support* were higher than the means of the students' own perceptions. In terms of intercorrelations, individual students' perceptions correlated strongly with each other, but rather low to medium, and sometimes not statistically significantly, with teachers'

judgments. In line with the correlations for research question 1, different teachers' judgments of students' perceived motivational support correlated highly with each other.

**Table 1**

*Descriptive statistics and intercorrelations of students' perceived motivational support and teachers' judgments on the class level and individual level*

Level	Perceptions and judgments	Descriptive statistics		Intercorrelations						
		<i>M</i>	<i>SD</i>	1	2	3	4	5	6	
Class level Students' perceived										
	1. autonomy support	2.78	0.42		.77**	.63**	.04	.19	.06	
	2. competence support	3.19	0.25	.77**		.73**	.10	.24	.13	
	3. relatedness support	3.08	0.41	.63**	.73**		.09	.20	.11	
Class level Teachers' judgments of students' perceived										
	4. autonomy support	3.12	0.44	.04	.10	.09		.68**	.60**	
	5. competence support	2.98	0.52	.19	.24	.20	.68**		.58**	
	6. relatedness support	3.30	0.58	.06	.13	.11	.60**	.58**		
		<i>M</i>	<i>SD</i>	7	8	9	10	11	12	
Individual level Individual students' perceived										
	7. autonomy support	2.79	0.44		.74**	.66**	.24	.11	.21	
	8. competence support	3.16	0.30	.74**		.70**	.43**	.31*	.37**	

	9. relatedness support	3.08	0.42	.66**	.70**	.40**	.23	.37*
<hr/>								
Individual Teachers' judgments								
level	of individual							
	students' perceived							
	10. autonomy support	3.05	0.45	.24	.43**	.40**	.89**	.88**
	11. competence support	3.14	0.39	.11	.31*	.23	.89**	.89**
	12. relatedness support	3.20	0.51	.21	.37**	.37*	.88**	.89**
<hr/>								

### 3.2 Teacher judgment accuracy of students' perceived motivational support on the class level

To assess teachers' judgment accuracy regarding the motivational support perceived by the whole class, we calculated the level component and the global deviation measure (see Table 2). For all three basic needs, the level component differed statistically significantly from 0 (autonomy:  $t(48) = 4.09, p < .05, d = 0.58$ ; competence:  $t(48) = -2.77, p < .05, d = -0.40$ ; relatedness:  $t(48) = 2.37, p < .05, d = 0.34$ ). Thus, the analyses showed a statistically significant and substantial underestimation of students' *perceived competence support* on the class level and overestimation of students' *perceived autonomy* and *relatedness support* on the class level. The hypothesis that teachers tend to underestimate students' perceived motivational support can only be partially confirmed empirically. The level component of *perceived autonomy support* differed statistically significantly from the level component of *perceived competence support* (autonomy vs. competence:  $t(48) = 6.55, p < .05, d = 0.94$ ), as did the level component of *perceived competence support* from the level component of *perceived relatedness support* (competence vs. relatedness:  $t(48) = -5.86, p < .05, d = -0.84$ ). However, there were no differences between the level component of *perceived autonomy support* and the level component of *perceived relatedness support* (autonomy vs. relatedness:  $t(48) = 1.39, p > .05, d = 0.20$ ). Hence, teachers differed in judging students' *perceived competence support* in comparison to *perceived autonomy and relatedness support*, but were not more accurate in judging students' *perceived autonomy support* than in judging *perceived relatedness support* on the class level.

The global deviation measure demonstrated the absolute discrepancy between students' perceptions and teachers' judgments. The lowest discrepancy was found for

*perceived competence support*, and the highest for *perceived relatedness support* on the class level. For all three basic needs, the global deviation measure differed statistically significantly from 0 (autonomy:  $t(48) = 8.61, p < .05, d = 1.23$ ; competence:  $t(48) = 9.03, p < .05, d = 1.29$ ; relatedness:  $t(48) = 10.14, p < .05, d = 1.45$ ). Thus, teachers' judgments of all aspects of perceived motivational support on the class level deviated statistically significantly from students' perceptions. There were no statistically significant differences in the global deviation measure between *perceived autonomy* and *competence* or *relatedness support* (autonomy vs. competence:  $t(48) = 1.51, p > .05, d = 0.22$ ; autonomy vs. relatedness:  $t(48) = -0.74, p > .05, d = -0.11$ ), but there was a statistically significant difference between *perceived competence support* and *perceived relatedness support* (competence vs. relatedness:  $t(48) = -2.85, p < .05, d = -0.41$ ). Hence, the results indicate that teachers judged students' *perceived competence support* more accurately than *relatedness support* on the class level.

**Table 2**

*The level component and global deviation measure on the class level*

Components and measures	Perceived autonomy support	Perceived competence support	Perceived relatedness support
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Level component	0.35 (0.60)	-0.21 (0.52)	0.23 (0.68)
Global deviation measure	0.53 (0.43)	0.44 (0.34)	0.58 (0.40)

### 3.3 Teacher judgment accuracy of students' perceived motivational support on the individual level

Next, we computed the level, differentiating, and rank components, as well as global deviation measure to examine how accurately teachers judged perceived motivational support on the individual student level (see Table 3). In terms of the level component, individual students' *perceived competence support* was judged quite accurately by teachers. In contrast, students' *perceived autonomy* and *relatedness support* tended to be overestimated by teachers. Despite the numerical differences and small to moderate effect sizes, only for *perceived autonomy support* did the level component differed statistically significantly from 0 (autonomy:  $t(48) = 3.46, p < .05, d = 0.50$ ; competence:  $t(48) = -0.10, p > .05, d = -0.01$ ; relatedness:  $t(48) = 1.93,$

$p > .05$ ,  $d = 0.28$ ). Consequently, teachers statistically significantly overestimated individual students' perceptions of autonomy support. Thus, the hypothesis that teachers tend to underestimate students' perceived motivational support cannot be confirmed empirically. Analogously to the results of research question 1, the level component of *perceived autonomy support* differed statistically significantly from the level component of *perceived competence support* (autonomy vs. competence:  $t(48) = 3.59$ ,  $p < .05$ ,  $d = 0.51$ ) and the level component of *perceived competence support* from *perceived relatedness support* (competence vs. relatedness:  $t(48) = -2.63$ ,  $p < .05$ ,  $d = -0.38$ ), but there was no difference between the level component of *perceived autonomy support* and the level component of *perceived relatedness support* (autonomy vs. relatedness:  $t(48) = 1.58$ ,  $p > .05$ ,  $d = 0.23$ ). Hence, given that only the level component of students' *perceived autonomy support* differed statistically significantly from 0, there was only a tendency of teachers overestimating individual students' *perceived autonomy support*. However, these tendencies revealed differences between *perceived competence* and *perceived autonomy support* as well as *perceived relatedness support*.

Regarding the differentiating component, the results indicated a tendency to underestimate the variability of students' perceptions of motivational support (see Table 3). The differentiating component differed statistically significantly from 1 for *perceived autonomy support* and *perceived competence support* (autonomy:  $t(48) = -2.09$ ,  $p < .05$ ,  $d = -0.30$ ; competence:  $t(48) = -4.55$ ,  $p < .05$ ,  $d = -0.65$ ; relatedness:  $t(46) = -1.53$ ,  $p > .05$ ,  $d = -0.22$ ). The analyses showed that teachers underestimated the variance of students' *perceived autonomy* and *perceived competence support*, but not *perceived relatedness support*. Hence, the hypothesis that teachers tend to overestimate the variance of students' perceived motivational support cannot be confirmed empirically. For all three basic needs, the differentiating components were not statistically significantly different from each other (autonomy vs. competence:  $t(48) = 1.66$ ,  $p > .05$ ,  $d = 0.24$ ; autonomy vs. relatedness:  $t(48) = 0.72$ ,  $p > .05$ ,  $d = 0.10$ ; competence vs. relatedness:  $t(48) = -1.10$ ,  $p > .05$ ,  $d = -0.16$ ). Thus, teachers did not differ in judging the variance of students' *perceived autonomy*, *competence*, and *relatedness support*, respectively.

The rank component captures teachers' ability to predict the rank order of students' perceptions by calculating the mean correlation. In this study, the highest correlation was found for *perceived competence support* (see Table 3). However, all correlations between students' perceived motivational support and teachers' judgments were quite low. Consequently, teachers were not successful in judging the rank order of students' *perceived*

*autonomy, competence, and relatedness support*. Thus, the hypothesis that teachers' judgements and students' *perceived autonomy, competence, and relatedness support* correlate lowly to moderately can be supported. Comparing the rank components for the three basic needs, significant differences between *perceived competence* and *perceived relatedness support* ( $z = 2.33$ ;  $p < .05$ ;  $d = 0.71$ ) were found, but not between *perceived autonomy* and *perceived competence support* ( $z = -1.50$ ;  $p > .05$ ;  $d = -0.44$ ) or between *perceived autonomy* and *perceived relatedness support* ( $z = 0.83$ ;  $p > .05$ ;  $d = 0.24$ ). Therefore, teachers were more successful in judging the rank order of students' *perceived competence* than of *perceived relatedness support*.

Turning to the global deviation measure, the lowest discrepancy was found for *perceived competence support* and the highest for *perceived autonomy support*. For all three basic needs, the global deviation measure differed statistically significantly from 0 (autonomy:  $t(48) = 19.13$ ,  $p < .05$ ,  $d = 2.73$ ; competence:  $t(48) = 25.62$ ,  $p < .05$ ,  $d = 3.66$ ; relatedness:  $t(48) = 23.420$ ,  $p < .05$ ,  $d = 3.42$ ). Accordingly, teachers' judgments deviated from students' perceptions of motivational support provided by teachers. There were no statistically significant differences between the global deviation measures for *perceived autonomy* and *perceived relatedness support* (autonomy vs. relatedness:  $t(48) = 0.61$ ,  $p > .05$ ,  $d = 0.09$ ), but the global deviation measure for *perceived competence support* differed statistically significantly from the global deviation measures for *perceived autonomy* and *perceived relatedness support* (autonomy vs. competence:  $t(48) = 3.74$ ,  $p < .05$ ,  $d = 0.53$ ; competence vs. relatedness:  $t(48) = -5.46$ ,  $p < .05$ ,  $d = -0.78$ ). Teachers judged individual students' *perceived competence support* more accurately than *perceived autonomy* and *perceived relatedness support*.

**Table 3**

*The three components of judgment accuracy and the global deviation measure on the individual level*

Components and measures	Perceived autonomy support	Perceived competence support	Perceived relatedness support
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Level component	0.28 (0.56)	-0.01 (0.41)	0.15 (0.52)
Differentiating component	0.87 (0.45)	0.76 (0.37)	0.82 (0.80)
Rank component	0.17 (0.33)	0.24 (0.44)	0.13 (0.36)
Global deviation measure	0.73 (0.27)	0.59 (0.16)	0.71 (0.21)

### **3.4 Teacher judgment accuracy as a predictor of students' intrinsic reading motivation**

The models tested showed good fit to the data (see Table 4). In models I–III, there were statistically significant paths on a 5% or 10% level from the global deviation measures of *perceived autonomy*, *competence*, and *relatedness support* to students' intrinsic reading motivation. If predictions were computed for each basic need, it appeared that students whose perceptions of autonomy, competence, or relatedness support were assessed more accurately exhibited higher intrinsic reading motivation than students who were assessed less accurately. Crucially, there were no statistically significant paths in model IV when considering all three basic needs simultaneously. The correlations between the global deviation measures of the three basic needs were between  $r = .27$  and  $r = .35$ . A structural equation model with one factor for the global deviation measure for motivational support predicting students' intrinsic reading motivation can be found in Appendix A (see Figure 1). Looking at structural equation models V–VII, only in Model VI was a significant path at the 10% level between the level component of *perceived competence support* and students' intrinsic reading motivation found. Students whose perceptions of competence support were overestimated were less motivated to read than students who were underestimated. In line with these results, model VIII showed a significant path on a 5% level between the level component of *perceived competence support* and students' intrinsic reading motivation. Looking at the data, students who had a low perception

of competence support were more likely to be overestimated. An under- or overestimation of students' *perceived autonomy* and *perceived relatedness support* was not correlated with students' intrinsic reading motivation. The correlations between the level components of the three basic needs ranged from  $r = .50$  to  $r = .62$ . Hence, the hypothesis that teacher judgment accuracy correlates with students' intrinsic motivation can be partly confirmed. In Appendix B (see Figure 2), a structural equation model with one factor for the level component of motivational support predicting students' intrinsic reading motivation can be found.

**Table 4***Predictors of students' intrinsic reading motivation*

Structural equation models		Students' intrinsic reading motivation	
	Predictors	$\beta$	<i>SE</i>
I	Global deviation measure for perceived autonomy support	-.13*	0.06
	$R^2$	.02	0.02
<i>Note.</i> CFI = 1.00, RMSEA = .00, SRMR = .01, $\chi^2 = 824.56$ , $df = 10$ , AIC = 3570.80, BIC = 3622.98; * $p < .05$ .			
II	Global deviation measure for perceived competence support	-.10*	0.05
	$R^2$	.01	0.01
<i>Note.</i> CFI = 1.00, RMSEA = .03, SRMR = .01, $\chi^2 = 834.11$ , $df = 10$ , AIC = 3603.21, BIC = 3655.51; * $p < .05$ .			
III	Global deviation measure for perceived relatedness support	-.10 <sup>†</sup>	0.06
	$R^2$	.01	0.01
<i>Note.</i> CFI = 1.00, RMSEA = .05, SRMR = .02, $\chi^2 = 853.26$ , $df = 10$ , AIC = 3415.08, BIC = 3466.67; <sup>†</sup> $p < .10$ .			
IV	Global deviation measure for perceived autonomy support	-.10	0.07
	Global deviation measure for perceived competence support	-.05	0.06
	Global deviation measure for perceived relatedness support	-.06	0.07
	$R^2$	.03	0.02

## Weiterführende Analysen (Beitrag IV)

<i>Note.</i> CFI = 1.00, RMSEA = .02, SRMR = .01, $\chi^2 = 941.98$ , $df = 18$ , AIC = 5397.86, BIC = 5494.83.			
V	Level component of perceived autonomy support	-.03	0.07
	$R^2$	.00	0.00
<i>Note.</i> CFI = 1.00, RMSEA = .00, SRMR = .01, $\chi^2 = 764.41$ , $df = 10$ , AIC = 3577.25, BIC = 3629.43.			
VI	Level component of perceived competence support	-.11 <sup>†</sup>	0.06
	$R^2$	.01	0.01
<i>Note.</i> CFI = 1.00, RMSEA = .03, SRMR = .02, $\chi^2 = 827.61$ , $df = 10$ , AIC = 3602.39, BIC = 3654.70; <sup>†</sup> $p < .10$ .			
VII	Level component of perceived relatedness support	-.01	0.07
	$R^2$	.00	0.00
<i>Note.</i> CFI = 1.00, RMSEA = .03, SRMR = .01, $\chi^2 = 851.19$ , $df = 10$ , AIC = 3410.16, BIC = 3461.72.			
VIII	Level component of perceived autonomy support	.05	0.07
	Level component of perceived competence support	-.21*	0.08
	Level component of perceived relatedness support	.11	0.09
	$R^2$	.02	0.02
<i>Note.</i> CFI = .99, RMSEA = .03, SRMR = .01, $\chi^2 = 915.60$ , $df = 18$ , AIC = 6295.87, BIC = 6392.83; * $p < .05$ .			

Models VIII–XVI showed also good fit to the data (see Table 5). In all models, the initial level of students' intrinsic reading motivation predicted students' intrinsic reading motivation. There were statistically significant paths on a 5% or 10% level from the global deviation measure for *perceived autonomy support* to students' intrinsic reading motivation when controlling for the initial level of students' intrinsic reading motivation. Hence, students whose perceptions of autonomy support were assessed more accurately exhibited more growth in intrinsic reading motivation than students who were assessed less accurately. In model XII, the correlations between the global deviation measures for the three basic needs were between  $r = .27$  and  $r = .35$ . For completeness, a structural equation model with one factor for the global deviation measure for motivational support predicting students' intrinsic reading motivation development can be found in Appendix C (see Figure 3). Regarding structural equation models XIII–XV, only in model XIII was a significant path on a 5% level between the level component of *perceived competence support* and students' intrinsic reading motivation found.

Students whose perceptions of competence support were overestimated showed less growth in their intrinsic reading motivation than students who were underestimated. In line with these results, analyses of model XVI showed a significant path on a 5% level between the level component of *perceived competence support* and students' intrinsic reading motivation development. Consequently, an under- or overestimation of students' *perceived autonomy support* and *perceived relatedness support* was not correlated with students' intrinsic reading motivation development. The correlations between the level components of the basic needs ranged from  $r = .50$  to  $r = .62$ . The hypothesis that teacher judgment accuracy correlates with students' intrinsic motivation development can also only be partly confirmed. A structural equation model with one factor for the level component of motivational support predicting students' intrinsic reading motivation development can be found in Appendix D (see Figure 4).

**Table 5***Predictors of students' intrinsic reading motivation development*

Structural equation models	Predictors	Students' intrinsic reading motivation	
		$\beta$	<i>SE</i>
VIII	Global deviation measure for perceived autonomy support	-.10*	0.04
	Initial level of students' intrinsic reading motivation	.68**	0.05
	$R^2$	.48**	0.06
<i>Note.</i> CFI = .98, RMSEA = .05, SRMR = .03, $\chi^2 = 1543.24$ , $df = 36$ , AIC = 6693.35, BIC = 6797.70; ** $p < .01$ , * $p < .05$ .			
X	Global deviation measure for perceived competence support	-.01	0.04
	Initial level of students' intrinsic reading motivation	.68**	0.05
	$R^2$	.46**	0.06
<i>Note.</i> CFI = .98, RMSEA = .06, SRMR = .05, $\chi^2 = 1590.45$ , $df = 36$ , AIC = 6795.70, BIC = 6900.31; ** $p < .01$ .			
XI	Global deviation measure for perceived relatedness support	-.06	0.05
	Initial level of students' intrinsic reading motivation	.68**	0.05
	$R^2$	.47**	0.06

Weiterführende Analysen (Beitrag IV)

<i>Note.</i> CFI = .97, RMSEA = .06, SRMR = .04, $\chi^2 = 1573.59$ , $df = 36$ , AIC = 6414.75, BIC = 6517.94; $\dagger p < .10$ .			
XII	Global deviation measure for perceived autonomy support	-.10 <sup>†</sup>	0.05
	Global deviation measure for perceived competence support	.05	0.05
	Global deviation measure for perceived relatedness support	-.05	0.05
	Initial level of students' intrinsic reading motivation	.68**	0.05
	$R^2$	.48**	0.06
<i>Note.</i> CFI = .98, RMSEA = .05, SRMR = .03, $\chi^2 = 1734.01$ , $df = 52$ , AIC = 8631.47, BIC = 8793.08; ** $p < .01$ , $\dagger p < .10$ .			
XIII	Level component of perceived autonomy support	-.06	0.04
	Initial level of students' intrinsic reading motivation	.69**	0.04
	$R^2$	.48**	0.06
<i>Note.</i> CFI = .98, RMSEA = .05, SRMR = .03, $\chi^2 = 1521.77$ , $df = 36$ , AIC = 6696.79, BIC = 6801.14; ** $p < .01$ .			
XVIII	Level component of perceived competence support	-.11*	0.04
	Initial level of students' intrinsic reading motivation	.68**	0.04
	$R^2$	.48**	0.06
<i>Note.</i> CFI = .98, RMSEA = .06, SRMR = .03, $\chi^2 = 1539.52$ , $df = 36$ , AIC = 6788.12, BIC = 6892.73; ** $p < .01$ ; * $p < .05$ .			
XV	Level component of perceived relatedness support	-.04	0.05
	Initial level of students' intrinsic reading motivation	.69**	0.05
	$R^2$	.47**	0.06
<i>Note.</i> CFI = .97, RMSEA = .06, SRMR = .03, $\chi^2 = 1579.89$ , $df = 36$ , AIC = 6382.14, BIC = 6485.26.			
XVI	Level component of perceived autonomy support	.00	0.05
	Level component of perceived competence support	-.14*	0.07
	Level component of perceived relatedness support	.05	0.07
	Initial level of students' intrinsic reading motivation	.68**	0.04
	$R^2$	.48**	0.06
<i>Note.</i> CFI = .97, RMSEA = .06, SRMR = .03, $\chi^2 = 915.60$ , $df = 18$ , AIC = 9532.54, BIC = 9694.15; ** $p < .01$ , * $p < .05$ .			

## 4 Discussion

The aim of this study was to examine teacher judgment accuracy regarding students' perceived motivational support and its impact on students' intrinsic reading motivation as well as the development of students' intrinsic reading motivation over time. Therefore, teachers were asked to make judgments about the whole class's perceived motivational support and about ten individual students on the individual level. At the same time, students assessed the motivational support they actually perceived and their intrinsic reading motivation.

### 4.1 Teacher judgment accuracy regarding motivational support (research questions 1 &2)

Looking at the findings, students' *perceived competence support* on the class level was underestimated and students' *perceived autonomy* and *perceived relatedness support* were overestimated by teachers. Regarding the global deviation measure, teachers' judgments of *perceived autonomy*, *competence*, and *relatedness support* deviated significantly from students' perceptions on the class level. Individual students' *perceived competence* and *perceived relatedness support* were not over- or underestimated. Contrary to the hypothesis, individual students' *perceived autonomy support* was significantly overestimated. In contrast to other studies, the motivational construct under study focused on teachers' offer of motivational support, and teachers were asked for their assessment of how students perceive their offer of motivational support. So, if teachers overestimated students' *perceived autonomy support*, they may have believed they act more supportively and foster students' perceived autonomy more than students see them as doing so. Support by the teacher and a constructive teacher-student relationship offer students a sense of security and belonging and are essential for learning and growth (Cui, 2022). Especially in elementary school, students who perceive better support by and a better relationship with their teachers experience more positive emotions and motivation in addition to cognitive benefits (Vandenbroucke et al., 2018). Therefore, it is particularly important for young students that their perception of teacher support is assessed accurately and supported.

Moreover, teachers underestimated the variability of students' *perceived autonomy* and *perceived competence support*, but not *perceived relatedness support*. Consequently, teachers underestimate how diversely students perceive the support they provide and perhaps also how the support they provide differs between students. From a psychological perspective, divergent

student assessments can be caused by individual differences in students' perception and interpretation. Perception and interpretation are influenced by cognitive as well as motivational and emotional processes (Greifeneder et al., 2018). This means that students not only come to school with different requirements and needs, but also perceive the teacher's learning offers in individual ways. Accordingly, when it comes to inclusion and adaptive teaching, which is about providing the best possible support to all children, it is the teacher's task to identify how students perceive their learning offers and how to individually respond to their needs. This is a complex task for which teachers should be prepared and trained. In international studies, the differentiation component is given little attention, even though it is important understanding the diversity of the assessed students' characteristics within the classroom.

In line with findings by Urhahne and Wijnia (2021) about judgment accuracy regarding motivational characteristics, our results show low correlations between teachers' judgments and individual students' perceived motivational support ( $.13 \leq r \leq .24$ ). These findings support the hypothesis that teachers face difficulties in identifying suitable indicators of students' motivation (Lee & Reeve, 2012). Hence, there is much ground for caution when using teachers' judgments of students' internal states and anything that cannot be easily observed, like motivational support (Gagné & Pèrè, 2001). Higher accuracy among teachers in judging individual students' *perceived competence support* in comparison to *perceived autonomy* or *perceived relatedness support* was found. For teachers, it is apparently easier to judge constructs that are closer to student achievement. Most likely, this is grounded in the reality that teachers receive more indicators of students' competence by correcting their homework and tests (Urhahne & Zhu, 2015). Accordingly, teachers need to apply a lower level of observer inference when judging students' competence than when judging other latent variables. The higher judgment accuracy for *perceived competence support* can be explained by the correlation between competence support and competence or achievement (Ahn et al., 2021).

Our results indicate that teachers can judge individual students' perceived motivational support more precisely than the class-average level of perceived support. A direct comparison between teachers' judgments of perceived motivational support on the class level and the individual level could not be conducted due to the different samples and ways of measuring teachers' judgments. Teachers' judgments of perceived motivational support experienced on the class level were measured indirectly, as a single item for the teacher's judgment of the

class-average level was compared to the average of multiple students' self-ratings, while the individual level was measured directly, with multiple items for teacher judgement exactly matching the students' multiple items. Zhu and Urhahne (2014) showed that single-item ratings can be a viable alternative for measuring teachers' judgments of students' motivation and that the two rating methods (single-item vs. multiple-item ratings) do not differ.

#### **4.2 Relevance of teacher judgment accuracy for students' intrinsic reading motivation and its development (research questions 3a & 3b)**

Analyses indicated that students whose perceptions of motivational support by teachers were judged more accurately exhibited higher intrinsic reading motivation. Unfortunately, these results could not be confirmed in the overall model for global deviation measures (model IV). For completeness, we checked a model with one latent factor for the global deviation measure calculated from the global deviation measures of *perceived autonomy*, *competence*, and *relatedness support* (Appendix A). These results also indicated the importance of teacher judgment accuracy regarding students' perceived motivational support for students' intrinsic reading motivation. Consequently, when it comes to fostering students' intrinsic reading motivation, it is important that teachers judge students accurately. The connection between teachers' under- or overestimation and students' intrinsic reading motivation was confirmed with respect to the level component of *perceived competence support* (model VI) as well as in the overall model with all three components (model VIII). An overestimation of *perceived competence support* leads to students having less intrinsic reading motivation. The consequences can be dramatic, as it was students who perceived low competence support in particular that were overestimated. Teachers overestimate how their motivational support is received by some students, who tend to have low intrinsic reading motivation and need to be supported in their motivation. Moreover, students can deduce teachers' expectations and perceptions from their differential behavior, which may lead them to respond to this differential behavior by changing their motivation and emotions (Babad, 1993; Brophy 1983). Urhahne (2015) also showed an indirect connection between teachers' perceptions of students' achievement and students' motivation and emotion, mediated by teachers' behavior. Thus, teachers perceive different groups of students differently and behave differently towards them. These students notice that teachers do not correctly judge their perceived motivational support, which is rated significantly higher than students perceive. This can lead to resignation, which can in turn cause poor student outcomes such as less intrinsic motivation. The model with one

latent factor for the level component does not reveal a relation between under- and overestimation and students' intrinsic reading motivation (Appendix B). Thus, this connection appears only in relation to *perceived competence support*.

Students whose *perceived autonomy support* was judged more accurately showed more growth in their intrinsic reading motivation over time (model VIII & XII). Based on the argument that teachers who assess students' *perceived autonomy support* more accurately can better adapt the support they provide, we expected a relation between teachers' supportive behavior and students' motivation. Previous studies have shown that *perceived autonomy support* in particular can increase students' intrinsic motivation. Zhou et al. (2019) showed a positive correlation between elementary school students' reports of perceived autonomy support by teachers and students' motivation. Hence, *perceived autonomy support* by the teacher, which encompasses the teacher giving students opportunities for self-initiated behavior, offering students meaningful choices, and encouraging them to participate in decision-making processes, is important for students' development in elementary school, and students at this age trust adult feedback (Cheon et al., 2019; Zhou et al., 2019). In contrast, in adolescence, peer feedback and peer motivational support become more influential for students than support by teachers or parents (Rubin et al., 1998). We also checked a model with one latent factor for the global deviation measure calculated from the global deviation measures for *perceived autonomy*, *competence*, and *relatedness support* (Appendix C). However, the results could not be confirmed. The relation between the level component of *perceived competence support* and students' intrinsic reading motivation development was shown in the single prediction model (XIII) and in the overall model with all three components (XVI). An overestimation of *perceived competence support* leads to less growth in students' intrinsic reading motivation. This also particularly affects students who assessed *perceived competence support* as low. The model with one latent factor for the level component confirms the connection between under- or overestimation and students' intrinsic reading motivation development (Appendix D). A general overestimation of perceived motivational support can cause less growth in students' intrinsic reading motivation.

### 4.3 Limitations

In addition to the strengths of the study, limitations must also be considered. The scale *perceived autonomy support* achieved only a low level of reliability ( $\alpha_{SMP1} = .62$ ,  $\alpha_{SMP2} = .65$ ). In the studies by Chen and colleagues (2015), reliability did not reach above .70 either.

Nevertheless, in line with Chen and colleagues (2015), we decided to keep experienced autonomy in the analyses in order to examine the distinct role of all three basic needs from Self-Determination Theory. In future studies, the items and prompts may need to be rewritten and adapted to elementary school students. Another possible criticism of our study is here are no longitudinal data available for judgment accuracy. Accordingly, the results for research questions 3a and 3b must be interpreted with caution. It would be interesting to examine how stable teacher judgment accuracy for motivational support is overtime. In addition, only elementary school students were examined in our study. Students' motivation and the importance of teacher support declines over the course of students' school years (Whitney & Bergin, 2018). Therefore, it would be interesting to examine not only teachers' diagnostic competence with respect to perceived motivational support in secondary school, but also how this relates to students' intrinsic motivation and its development among secondary school students.

#### **4.4 Conclusion**

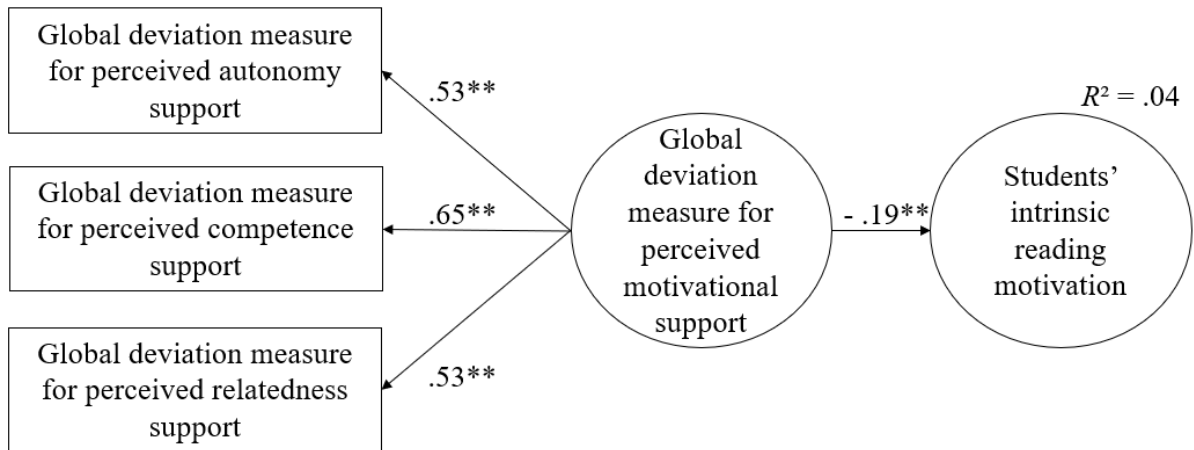
Diagnosing students' characteristics is a crucial aspect of teachers' everyday work. The teaching profession entails a high level of daily task complexity and requires the ability to multitask (Brante, 2009; Klug et al., 2013). The present study indicates that teachers face difficulties in judging students' perceived motivational support. Due to the importance and effects of teacher judgment accuracy, teachers and teacher education students should be made aware of this issue and receive training on it. As recommended by Gabriele et al. (2016), judgment accuracy could have important and useful implications for the evaluation of teacher preparation, professional development, and teacher performance in general. Teachers could more frequently compare their own perceptions to those of their students in order to see how their learning offers are received by students. In this way, they would use judgment accuracy to evaluate themselves and could adapt their teaching and support services more precisely to their students'. Hence, judgment accuracy can be a tool for improving teachers' effectiveness. However, for this to be the case, teachers must learn to recognize the valid indicators of students' motivation. This could also be a research question for future studies. Due to the significance of students' perceptions of teaching for their school success, analyses should focus on students' perceptions and teachers' assessments of students' perceptions. Students are very diverse in their perceptions as well as in their needs and prerequisites for successful learning.

Our findings have implications for teachers and researchers because they indicate that teachers tend to misestimate students' perceived motivational support, especially *perceived autonomy support*, and demonstrate the relation between judgement accuracy and students' intrinsic reading motivation.

**Appendix A**

**Figure 1**

*One-factor global deviation measure predicting students' intrinsic reading motivation*

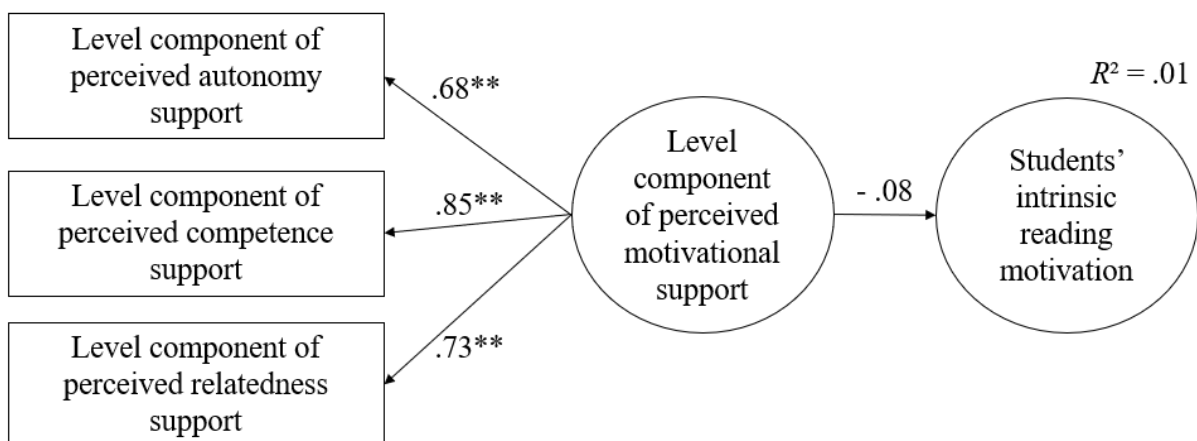


*Note.* CFI = 1.00, RMSEA = .01, SRMR = .03,  $\chi^2 = 1015.07$ ,  $df = 21$ , AIC = 5394.94, BIC = 5483.82; \*\*  $p < .01$

**Appendix B**

**Figure 2**

*One-factor level component predicting students' intrinsic reading motivation*

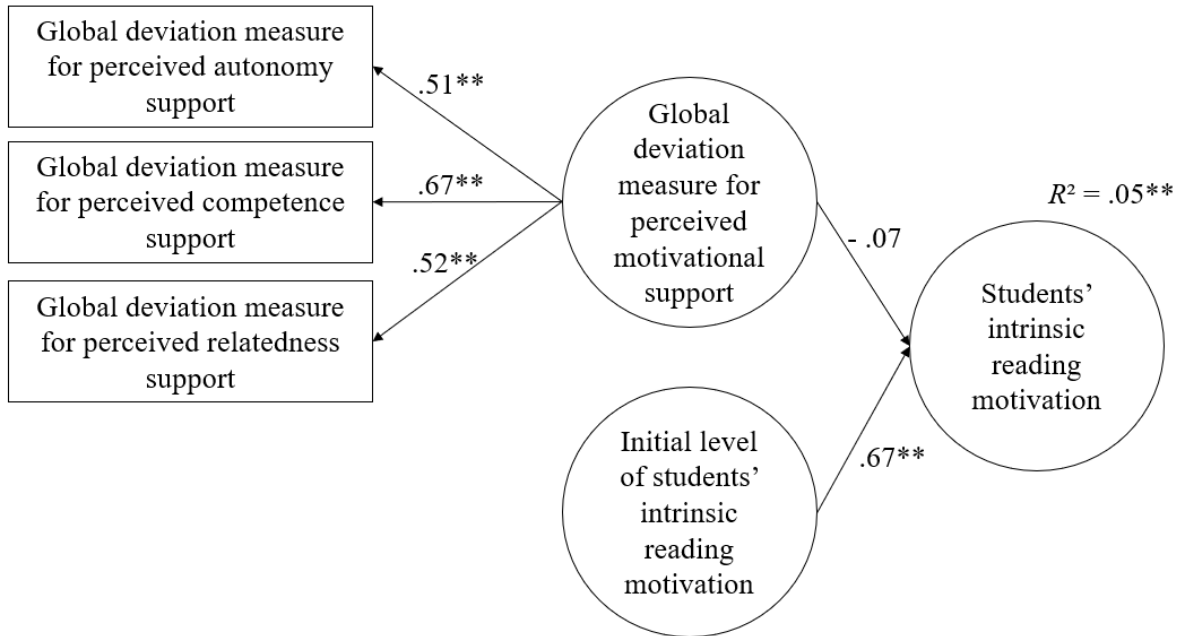


*Note.* CFI = .99, RMSEA = .04, SRMR = .03,  $\chi^2 = 1305.67$ ,  $df = 21$ , AIC = 6298.49, BIC = 6387.37; \*\*  $p < .01$

Appendix C

Figure 3

*One-factor global deviation measure predicting students' intrinsic reading motivation development*

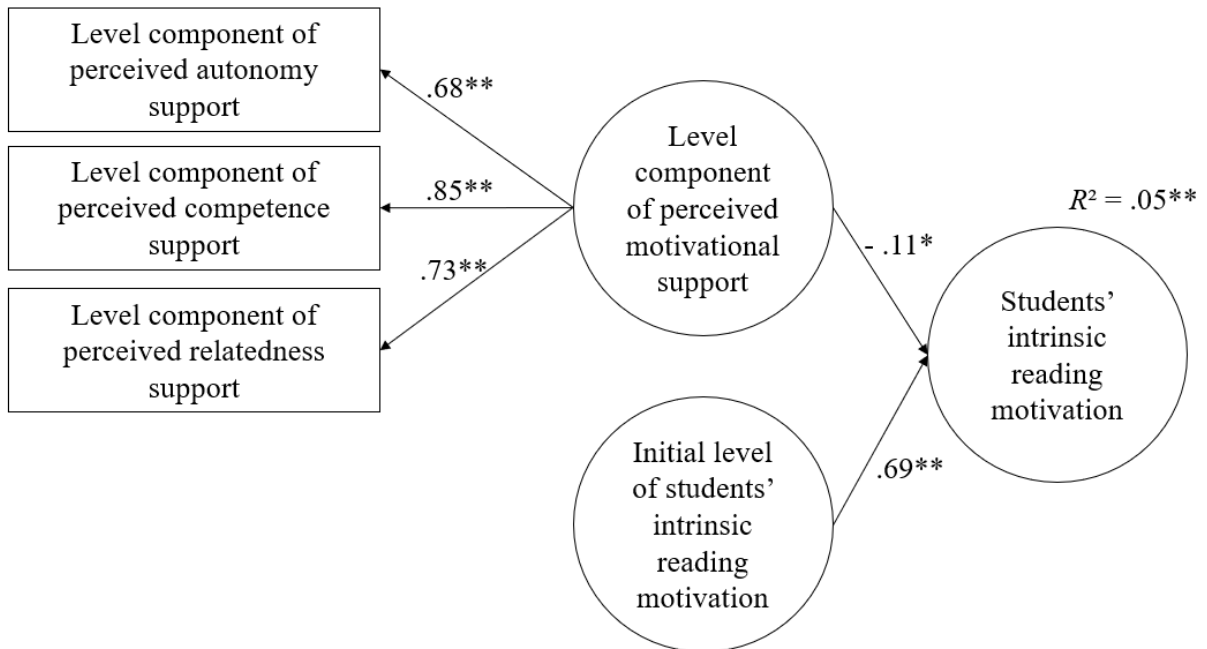


Note. CFI = .98, RMSEA = .05, SRMR = .03,  $\chi^2 = 1813.04$ ,  $df = 55$ , AIC = 8629.90, BIC = 8775.35;  $** p < .01$

## Appendix D

Figure 4

*One-factor level component predicting students' intrinsic reading motivation development*



*Note.* CFI = .98, RMSEA = .05, SRMR = .03,  $\chi^2 = 2038.88$ ,  $df = 55$ , AIC = 9531.26, BIC = 9676.71; \*\*  $p < .01$

## Notes

We would like to thank Justine Stang-Rabrig and Matthias Trendtel for their support with statistical analyses during the preparation of the manuscript.

## Funding

Funding was received for this work. This work was supported by German Federal Ministry of Education and Research (grant no. 01JI1504A & 01JI1504B).

## References

- Ahn, I., Chiu, M. M., & Patrick, H. (2021). Connecting teacher and student motivation: Student-perceived teacher need-supportive practices and student need satisfaction. *Contemporary Educational Psychology*, *64*, 101950. <https://doi.org/10.1016/j.cedpsych.2021.101950>
- Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften [Diagnostic competence of teachers]. *Zeitschrift für Pädagogische Psychologie [German Journal of Educational Psychology]*, *23*(3), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>.
- Babad, E. (1993). Teachers' differential behavior. *Educational Psychology Review*, *5*, 347e376. <http://dx.doi.org/10.1007/BF01320223>
- Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology*, *107*(3), 805–820. <https://doi.org/10.1037/edu0000008>
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Kruss, M. Neubrand, et al. (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV Project* (pp. 25–48). New York: Springer.
- Bonefeld, M., Dickhäuser, O., & Karst, K. (2020). Do preservice teachers' judgments and judgment accuracy depend on students' characteristics? The effect of gender and immigration background. *Social Psychology of Education*, *23*, 189–216. <https://doi.org/10.1007/s11218-019-09533>
- Brandmiller, C., Dumont, H., & Becker, M. (2020). Teacher perceptions of learning motivation and classroom behavior: The role of student characteristics. *Contemporary Educational Psychology*, *63*, 101893.
- Brante, G. (2009). Multitasking and synchronous work: complexities in teacher work. *Teaching and Teacher Education*, *25*, 430–436.

- Brophy, J.E. (1983). Research on self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 75, 631–661.
- Carr, M., & Kurtz-Costes, B. E. (1994). Is being smart everything? The influence of student achievement on teachers' perceptions. *British Journal of Educational Psychology*, 64(Pt 2), 263–276. <https://doi.org/10.1111/j.2044-8279.1994.tb01101.x>
- Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R. M., Sheldon, K. M., Soenens, B., Van Petegem, S., & Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, 39(2), 216–236. <https://doi.org/10.1007/s11031-014-9450-1>
- Cheon, S.H, Reeve, J., Lee, Y., Ntoumanis, N., Gillet, N., Kim, B.R., & Song, Y.G. (2019). Expanding autonomy psychological need states from two (satisfaction, frustration) to three (dissatisfaction): a classroom-based intervention study. *Journal of Educational Psychology*, 111(4), 685–702. <http://dx.doi.org/10.1037/edu0000306>
- Conradi, K., Jang, B. G., & McKenna, M. C. (2014). Motivation terminology in reading research: A conceptual review. *Educational Psychology Review*, 26, 127–164.
- Cronbach, L. J. (1955). Processes affecting scores on “understanding others” and “assumed similarity”. *Psychological Bulletin*, 52, 177–193.
- Cui, L. (2022). The role of teacher-student relationships in predicting teachers' occupational wellbeing, emotional exhaustion, and enthusiasm. *Front Psychology*, 13:896813. <https://doi.org/10.3389/fpsyg.2022.896813>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goalpursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Deci, E. L., & Ryan, R. M. (2008). Self-determination theory. A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, 49(3), 182–185.

- Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G., & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning and Individual Differences, 22*(6), 844–849.
- Ekmekci, A., & Serrano, D.M. (2022). The impact of teacher quality on student motivation, Achievement, and Persistence in Science and Mathematics. *Education Sciences, 12*(10), 649–669. <https://doi.org/10.3390/educsci12100649>
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121–138. <http://dx.doi.org/10.1037/1082-989X.12.2.121>
- Fauth, B., Decristan, J., Decker, A., Büttner, G., Klieme, E., Hardy, I., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education, 86*, 102882.
- Fauth, B., Göllner, R., Lenske, L., Praetorius, A., & Wagner, W. (2020). Who sees what? Theoretical considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik, 66*, 138–155.
- 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.
- Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102*, 453–462.
- Fischer, J., Machts, N., Bruckermann, T., Möller, J., & Harms, U. (2022). The Simulated Classroom Biology—A simulated classroom environment for capturing the action-oriented professional knowledge of pre-service teachers about evolution. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.12718>
- Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning & Instruction, 32*, 91–100. <https://doi.org/10.1016/j.learninstruc.2014.02.002>

- Förtsch, C., Werner, S., Kotzebue, L. v., & Neuhaus, B. J. (2016): Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666.
- Froiland, J. M., & Worrell, F.C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. *Psychology in the Schools*, 53(3), 321–336.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Furtak, E., & Kunter, M. (2012). Effects of Autonomy Supportive Teaching on Student Learning and Motivation. *Journal of Experimental Education*, 80(3), 284–316.
- Gabriele, A. J., Joram, E., & Park, K. H. (2016). Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes? *Learning and Instruction*, 45, 49–60.
- Gagné, F., & St Pèrè, F. (2001). When IQ is controlled, does motivation still predict achievement? *Intelligence*, 30(1), 71–100. [https://doi.org/10.1016/s0160-2896\(01\)00068-x](https://doi.org/10.1016/s0160-2896(01)00068-x)
- Givvin, K.B., Stipek, D.J., Salmon, J.M., & MacGyvers, V.L. (2001). In the eyes of the beholder: students' and teachers' judgments of students' motivation. *Teaching and Teacher Education*, 17, 321–331.
- Gnambs, T., & Hanfstingl, B. (2016). The decline of academic motivation during adolescence: An accelerated longitudinal cohort analysis on the effect of psychological need satisfaction. *Educational Psychology*, 36(9), 1691–1705.
- Greifeneder, R., Bless, H. & Fiedler, K. (2018). *Social cognition: How individuals construct social reality*. New York: Routledge.
- Gucciardi, D. F., Weixian, J. C., Gibson, W., Ntoumanis, N., & Ng, L. (2020). Motivational climate in the classroom: Factorial and convergent validity evidence of the Need-Supportive Behaviors Scale with health science students. *European Journal of Psychological Assessment*, 36(2), 324–335. <https://doi.org/10.1027/1015-5759/a000524>
- Guthrie, J.T., Hoa, A.L.W., Wigfield, A., Tonks, S.M., Humenick, N.M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years.

*Contemporary Educational Psychology*, 32, 282–313.  
<https://doi.org/10.1016/j.cedpsych.2006.05.004>

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72(2), 625–638.

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

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M.R., Ufer, S., Schmidmaier, R., Neuhaus, B. J., Siebeck, M., et al. (2019). Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. *Frontline Learning Research*, 7, 1–24.

Helmke, A., & Schrader, F. W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98.

Herppich, S., Praetorius, A.-K., Foerster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Boehmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Boehmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193.  
<https://doi.org/10.1016/j.tate.2017.12.001>

Hill, H.C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal*, 55 (5), 1076–1112. <https://doi.org/10.3102/0002831218769614>

Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education*, 118(4), 489–519. <https://doi.org/10.1086/666380>

Hoge, R.D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59, 297–313.

- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5, 64–86. <http://dx.doi.org/10.1037/1082-989X.5.1.64>
- Jussim, L., & Eccles, J. S. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63, 947–961. <http://dx.doi.org/10.1037/0022-3514.63.6.947>
- Kaiser, J., & Möller, J. (2017). Diagnostische Kompetenz von Lehramtsstudierenden [Diagnostic competence of teacher students]. In C. Gräsel, & K. Trempler (Eds.), *Entwicklung von Professionalität pädagogischen Personals* (pp. 55–74). Springer Fachmedien.
- Karing, C., & Artelt, C. (2014). Urteilsgenauigkeit von Lehrer(inne)n im emotional-motivationalen Bereich und im Leistungsbereich [Accuracy of judgment of teachers in the emotional-motivational area and in the performance area]. In M. Mudiappa & C. Artelt (Hrsg.), *BiKS - Ergebnisse aus den Längsschnittstudien. Praxisrelevante Befunde aus dem Primar- und Sekundarschulbereich [BiKS - Results from the longitudinal studies. Findings relevant to practice from primary and secondary schools]* (pp. 111–118). Bamberg: University of Bamberg Press.
- Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, 30, 38–46.
- Kramer, M., Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating pre-service biology teachers' diagnostic competences: relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. *Education Sciences*, 11(89). <https://doi.org/10.3390/educsci11030089>
- Kriegbaum, K., Steinmayr, R., & Spinath, B. (2020). Longitudinal reciprocal effects between teachers' judgments of students' aptitude, students' motivation, and grades in math. *Contemporary Educational Psychology*. <https://doi.org/10.1016/j.cedpsych.2019.101807>

- 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.
- 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.
- Lange, K., Kleickmann, T., Tröbst, S., & Möller, K. (2012). Fachdidaktisches Wissen von Lehrkräften und multiple Ziele im naturwissenschaftlichen Sachunterricht [Teachers' pedagogical content knowledge and academic outcomes in science education]. *Zeitschrift für Erziehungswissenschaft, 15*(1), 55–75.
- Lee, W., & Reeve, J. (2012). Teachers' estimates of their students' motivation and engagement: Being in synch with students. *Educational Psychology, 32*(6), 727–747. <https://doi.org/10.1080/01443410.2012.732385>
- Lenhard, W., & Lenhard, A. (2014). *Hypothesis Tests for Comparing Correlations*. available: <https://www.psychometrica.de/correlation.html>. Bibergau (Germany): Psychometrica. <https://doi.org/10.13140/RG.2.1.2954.1367>
- Machts, N., Kaiser, J., Schmidt, F.T.C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Psychology Review, 19*, 85–103. <https://doi.org/10.1016/j.edurev.2016.06.003>
- Marsh, H.W., & Craven, R.G. (1991) 'Self-other agreement on multiple dimensions of preadolescent selfconcept: inferences by teachers, mothers, and fathers'. *Journal of Educational Psychology, 83*, 393–404.
- McElvany, N., Becker, M., & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten [The importance of family characteristics for reading skills, vocabulary, reading motivation and reading behavior]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie [German Journal of Developmental and Educational Psychology], 41*(3), 121–131.
- Miyamoto, A., Pfost, M., & Artelt, C. (2018). Reciprocal relations between intrinsic reading motivation and reading competence: A comparison between native and immigrant students

in Germany. *Journal of Research in Reading*, 41, 176–196.  
<https://doi.org/10.1111/14679817.12113>

Musu-Gillette, L., Wigfield, A., Harring, J., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, 21(4), 343–370.  
<https://doi.org/10.1080/13803611.2015.1057161>

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén

Oga-Baldwin, W. L. Q., & Fryer, L. K. (2017). Exploring motivational profiles in public elementary school English classes. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in language education*. Tokyo: JALT.

Oga-Baldwin, W. L. Q., & Fryer, L.K. (2020). Girls show better quality motivation to learn languages than boys: latent profiles and their gender differences. *Heliyon*, 6(5), e04054.  
<https://doi.org/10.1016/j.heliyon.2020.e04054>

Oudman, S., Pol, J. v. d., & Gog, T. v. (2023). Effects of cue availability on primary school teachers' accuracy and confidence in their judgments of students' mathematics performance. *Teaching and Teacher Education*, 122.  
<https://doi.org/10.1016/j.tate.2022.103982>

Perret, P. (2015). Children's Inductive Reasoning: Developmental and Educational Perspectives. *Journal of Cognitive Education and Psychology*, 14(3), 389–408.

Pianta, R. C., & Hamre, B. K. (2009). Measurement and improvement of teacher-child interactions. In G. Sykes, B. Schneider, D. N. Plank & T. G. Ford (Eds.), *Handbook of education policy research* (pp. 652–660). New York: Routledge.

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

- Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging academic self-concepts. *Journal of Educational Research, 106*, 64–76.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM, 50*(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48*(2), 335–360. <https://doi.org/10.3102/0002831210374874>
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington: DC: Economic Policy Institute.
- 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. *Elementary School Journal, 100*, 443–471.
- Rosenthal, R. (2002). The Pygmalion effect and its mediating mechanisms. In J. Aronson (Ed.), *Improving academic achievement. Impact of psychological factors on education* (pp. 25–36). San Diego, CA: Academic Press.
- Rubin, K.H., Bukowski, W.M., & Parker, J.G. (1998). Peer interactions, relationships, and groups. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, pp. 619–700). New York, NY: Wiley.
- Ryan, R., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.
- Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary educational psychology, 61*.

- Sanusi, A., Sauri, S., & Nurbayan, Y. (2020). Non-native Arabic language teacher: Low teacher's professional competence: Low Quality Outcomes? *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 7, 45–60.
- Scheer, E. A. v.d., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement*, 30, 30–50. <https://doi.org/10.1080/09243453.2018.1539015>
- Scheera, v. d. E. A., Bijlsmaband, Cees, H. J. E., & Glas A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement*, 30(1), 30–50.
- Schiefele, U., Schaffner, E., Moller, J., & Wigfield, A. (2012). Dimensions of Reading Motivation and Their Relation to Reading Behavior and Competence. *Reading Research Quarterly*, 47(4), 427–463.
- Schröter, H., & Bar-Kochva, I. (2019). Keyword: Reading literacy. Reading competencies in Germany and underlying cognitive skills. *Zeitschrift für Erziehungswissenschaft [German Journal of Education]*, 22(1), 17–49.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51, 455–498. <https://doi.org/10.2307/1170362>
- Soebari, T. S., & Aldridge, J. M. (2015). Using student perceptions of the learning environment to evaluate the effectiveness of a teacher professional development programme. *Learning Environments Research*, 18, 163–178.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie [German Journal of Educational Psychology]*, 19, 85–95. <http://dx.doi.org/10.1024/1010-0652.19.12.85>
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S., & Haag, N. (2017). *IQB-Bildungstrend 2016 – Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* [IQB Trends in Student Achievement 2016 –

The Second National Assessment of German and Mathematics Proficiencies at the End of Fourth Grade]. Münster: Waxmann.

Steiner, D., Magee, J., & Jensen, B. (2018) What we teach matters. Learning First. <https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/62969/what-we-teach-matters-final-for-publication-15-nov.pdf?sequence=1>

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

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum: Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, *22*(34), 261–276.

Südkamp, A., Praetorius, A., & Spinath, B. (2018). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*, *76*, 204–213.

Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., & Koestner, R. (2014). A self-determination theory approach to predicting school achievement overtime: The unique role of intrinsic motivation. *Contemporary Educational Psychology*, *39*, 342–358.

Tennant, J. E., Demaray, M. K., Malecki, C. K., Terry, M. N., Clary, M., & Elzinga, N. (2015). Students' Ratings of Teacher Support and Academic and Social–Emotional Well-Being. *School Psychology Quarterly*, *30*(4), 494–512.

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

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., et al. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education*, *49*, 36–44.

- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education*, *45*, 73–82.
- Urhahne, D., Chao, S.-H., Florineth, M. L., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology*, *81*, 161-177.
- Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, *32*. 100374. <https://doi.org/10.1016/j.edurev.2020.100374>
- Urhahne, D., & Zhu, M. (2015). Teacher judgement and student motivation. In C. M. Rubie-Davies, J. M. Stephens & P. Watson (Eds.), *International Handbook of Social Psychology of the Classroom* (pp.304–315). London, UK: Routledge.
- Vandenbroucke, L., Spilt, J., Verschueren, K., Piccinin, C., & Baeyens, D. (2018). The classroom as a developmental context for cognitive development: a meta-analysis on the importance of teacher-student interactions for children's executive functions. *Review of Educational Research*, *88*(1), 125–164. <https://doi.org/10.3102/0034654317743200>
- Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, *25*, 1051–1060.
- Weinert, F. E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–65). Seattle, WA: Hogrefe and Huber Publishers.
- Whitney, S. D. & Bergin, D. A. (2018). Students' motivation and engagement predict reading achievement differently by ethnic group. *The Journal of Genetic Psychology*, *179*(6), 357–370. <https://doi.org/10.1080/00221325.2018.1527754>
- Zhou, L.-H., Ntoumanis, N., & Thøgersen-Ntoumani, C. (2019). Effects of perceived autonomy support from social agents on motivation and engagement of Chinese primary school students: Psychological need satisfaction as mediator. *Contemporary Educational Psychology*, *58*, 323–330. <https://doi.org/10.1016/j.cedpsych.2019.05.001>

Zhu, M., & Urhahne, D. (2014). Assessing teachers' judgements of students' academic motivation and emotions across two rating methods. *Educational Research and Evaluation*, 20(5), 411–427.

Zhu, M., Urhahne, D., & Rubie-Davies, C. M. (2018). The longitudinal effects of teacher judgement and different teacher treatment on students' academic outcomes. *Educational Psychology*, 38, 648–668. <https://doi.org/10.1080/01443410.2017.1412399>

## 8 Gesamtdiskussion

Das übergeordnete Forschungsanliegen der vorliegenden kumulativen Dissertation bestand in der Betrachtung der Wahrnehmungen von Unterrichtsqualität in der Primarstufe aus Sicht von Lernenden, Lehrkräften und externen Urteilenden. Hierbei wurden die Übereinstimmungen zwischen den Lernenden einer Klasse (Beitrag I) sowie zwischen der Perspektive der Lernenden und der Perspektive der Lehrkräfte (Beitrag IV/ weiterführende Analysen) in den Blick genommen (Fragen 1 & 2). Ein wesentlicher Schwerpunkt der Dissertation lag in der Untersuchung individueller und methodischer Prädiktoren für die Wahrnehmungen von Lernenden (Beitrag I), Lehrkräften (Beitrag II) und externen Urteilenden (Beitrag III) (Fragen 3–6; siehe Kapitel 4 Forschungsanliegen und Forschungsfragen). Im Rahmen der Einzelbeiträge wurden die Ergebnisse der Studien zu den Übereinstimmungen und Prädiktoren der Wahrnehmungen von Lernenden, Lehrkräften und externen Urteilenden gesondert diskutiert. In der Gesamtdiskussion wird sich auf die Betrachtung ausgewählter Resultate zur Beantwortung der übergeordneten Fragestellungen der Dissertation fokussiert. Ebenfalls werden Stärken und Grenzen diskutiert. Des Weiteren werden Implikationen für Forschung sowie Praxis aus den Ergebnissen abgeleitet und ein Ausblick gegeben.

### 8.1 Zusammenfassung und Diskussion der zentralen Ergebnisse

#### 8.1.1 Übereinstimmungen von Wahrnehmungen

Beitrag I (*Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität*) diente dazu, die Übereinstimmungen der Wahrnehmungen von Grundschulkindern bezogen auf die Unterrichtsqualität zu untersuchen. Bisher fokussierten hierzu die meisten internationalen und nationalen Untersuchungen die Sekundarstufe I (Burić & Kim, 2020; Lüdtke, Robitzsch, Trautwein & Kunter, 2009; Willems, 2022). Dabei konnten Fauth und Kollegen (2016) zeigen, dass Lernende im Grundschulalter durchaus in der Lage sind, valide Aussagen über die unterrichtliche Qualität zu treffen. Befunde bezüglich der Wahrnehmungen von Lernenden der Sekundarstufe I wiesen auf Divergenzen innerhalb von Klassen hin (z. B. Wagner, Göllner, Helmke, Trautwein & Lüdtke, 2013). Im Einklang mit der Vermutung, dass auch Wahrnehmungen von Grundschulkindern innerhalb von Klassen divergieren, zeigten die Analysen, dass die Wahrnehmungen von Lernenden in Grundschulklassen nur gering übereinstimmten. Mittels Vergleiche der

Intraklassenkorrelationen konnte bestätigt werden, dass die größte Übereinstimmung im Bereich des Unterrichtsqualitätsmerkmals Störungen im Unterricht, welches der Basisdimension Klassenführung zugeordnet werden kann, zu finden war (Störungen im Unterricht: ICC = .26; Motivierung durch die Lehrkraft: ICC = .19; Herausfordernde Aufgaben: ICC = .14).

Beitrag IV (*Students' perceived motivational support in elementary school – How accurate are teachers' judgments?* in weiterführende Analysen) zielte darauf ab, die Übereinstimmung zwischen zwei Perspektiven in den Blick zu nehmen: die Perspektive der Lernenden und die der Lehrkräfte. Dabei ging es aber nicht direkt um die Übereinstimmung von gleich formulierten Items zur Wahrnehmung von Unterrichtsqualität, sondern um die genaue Wahrnehmung von Lehrkräften über die wahrgenommene Motivationsunterstützung der Lernenden ihrer Klasse. Dieses Vorgehen ist aus dem Grund interessant, da die Diagnostische Kompetenz von Lehrkräften im direkten Zusammenhang mit ihrem Handeln steht. Lehrkräfte müssen die Bedürfnisse ihrer Lernenden akkurat einschätzen können, um diese bestmöglich fördern zu können und einen qualitätsvollen Unterricht anbieten zu können (Jandl, 2016; Vogelsang & Reinhold, 2013). Mittels der drei Urteilsdimensionen nach Helmke und Schrader (1987) sowie des globalen Abweichungsmaßes wurde die Beurteilungsgenauigkeit von Lehrkräften hinsichtlich der wahrgenommenen Motivationsunterstützung (Unterstützung des Autonomieerlebens, Kompetenzerlebens und Erlebens der sozialen Eingebundenheit) der Lernenden ihrer Grundschulklassen auf Klassen- und Individualebene bestimmt. Im Einklang mit den Erkenntnissen von Urhahne und Wijnia (2021) zur Übereinstimmung der Perspektiven der Lernenden und Lehrkräften in Bezug auf Motivationsmerkmale zeigten die Ergebnisse geringe Korrelationen zwischen den Wahrnehmungen von Lehrkräften und der wahrgenommenen Motivationsunterstützung einzelner Lernenden auf Individualebene ( $.13 \leq r \leq .24$ ). Sowohl auf Klassenebene als auch auf Individualebene unterschieden sich die Wahrnehmungen der Lehrkräfte von der Wahrnehmung der Lernenden hinsichtlich der Motivationsunterstützung durch die Lehrkräfte. Auf Klassenebene schätzten die Lehrkräfte die wahrgenommene Unterstützung des Kompetenzerlebens am genauesten ein, zumindest indem sie diese im Vergleich zur Unterstützung des Autonomieerlebens und Erlebens der sozialen Eingebundenheit am geringsten unterschätzten. Diese Tendenz zeigte sich auch auf Individualebene: Lehrkräfte schätzten die wahrgenommene Unterstützung des Kompetenzerlebens am genauesten ein. Die

Analysen wiesen auf Schwierigkeiten der Lehrkräfte hin, insbesondere die wahrgenommene Unterstützung des Autonomieerlebens genau wahrnehmen zu können.

### *Fazit zu Übereinstimmungen von Wahrnehmungen*

Trotz des gleichen unterrichtlichen Angebots nehmen Lernende der Primstufe die Unterrichtsqualität divergierend wahr. Auch in anderen Studien war die größte Übereinstimmung zwischen den Wahrnehmungen von Lernenden im Bereich der Klassenführung zu finden (vgl. Praetorius et al., 2017; Wagner et al., 2016, Willems, 2022). Ein Grund, warum Einschätzungen von Lernenden unterschiedlich ausfallen können, ist der Inferenzgrad der Items. Bezüglich der Einschätzungen der Unterrichtsqualitätsmerkmale kognitive Aktivierung, Klassenführung und Motivierungsqualität ist zu beachten, dass sich Items hinsichtlich ihres Inferenzgrades unterscheiden können und dies dazu führen kann, dass abhängig vom Inferenzgrad, Übereinstimmungen innerhalb von Klassen unterschiedlich hoch ausfallen können. Der Inferenzgrad lässt sich durch ein Kontinuum abbilden, bei dem niedrig- und hoch-inferente Einschätzungen nur die äußeren Pole beschreiben (Rosenshine, 1970). Für ein Urteil über vergleichbar beobachtbare Kriterien wie zum Beispiel Störungen im Unterricht muss weniger interpretiert und geschlussfolgert werden als im Vergleich zum Merkmal Nähe zur Lehrkraft. Bei diesem Unterrichtsqualitätsmerkmal ist eine höhere kognitive Schlussfolgerung und Interpretation erforderlich und wird im Vergleich zu Störung im Unterricht eher durch höher-inferente Items abgefragt. Demzufolge sind die Einschätzungen zu den unterschiedlichen Unterrichtsqualitätsmerkmalen auch unterschiedlich hoch von Beurteilungsfehlern und Verzerrungen betroffen (zum Interpretationsaufwand und zu Beurteilungsfehlern: Clausen, Reusser & Klieme, 2003; Praetorius, 2014).

Der Inferenzgrad kann auch als Erklärung für die genauere Wahrnehmung von Lehrkräften der wahrgenommenen Unterstützung des Kompetenzerlebens von Lernenden im Vergleich zur Unterstützung des Autonomieerlebens oder der sozialen Eingebundenheit in Betracht gezogen werden. In Beitrag IV wurde eine höhere Genauigkeit bei der Beurteilung der wahrgenommenen Unterstützung des Kompetenzerlebens durch Lehrkräfte im Vergleich zur wahrgenommenen Unterstützung des Autonomieerlebens oder des Erlebens der sozialen Eingebundenheit festgestellt. Für die Lehrkraft ist es offenbar einfacher, Konstrukte zu beurteilen, die näher an den Kompetenzen und Leistungen der Lernenden liegen. Mittels der Korrektur ihrer Hausaufgaben und Tests erhalten Lehrkräfte mehr Indikatoren für die Kompetenz ihrer Lernenden (Urhahne & Zhu, 2015). Dementsprechend müssen Lehrkräfte bei der Beurteilung der Kompetenz der Lernenden ein geringeres Maß an Schlussfolgerungen

anwenden als bei der Beurteilung anderer latenter Variablen. Somit kann die höhere Beurteilungsgenauigkeit für die wahrgenommene Kompetenzunterstützung durch den Zusammenhang zwischen Kompetenzunterstützung und Kompetenz bzw. Leistung erklärt werden (Ahn et al., 2021).

Ebenfalls muss beachtet werden, dass ein weiterer Grund für divergierende Wahrnehmungen von Lernenden auch ein divergierendes Verhalten von Lehrkräften gegenüber einzelnen Lernenden sein kann (Schweer, Thies & Lachner, 2017). Nach dem Regulationsmodell von Roseman (1978) haben Lehrkräfte und auch Lernende normative Erwartungen in Bezug auf fachliche und soziale Aspekte an den jeweils anderen. Je nach Erfüllung der Erwartungen folgt ein freundliches und unterstützendes oder ein distanziertes und weniger unterstützendes Verhalten (Labsch, Schmitt & Schüpbach, 2023; Schweer, Thies & Lachner, 2017). Schon Brohpy und Good (1979) wiesen Auswirkungen der Erwartungen von Lehrkräften auf die Unterrichtsqualität nach. Demnach ergaben sich je nach Leistungserwartungen Unterschiede in der Qualität des Kontakts zu einzelnen Lernenden. Des Weiteren konnte gezeigt werden, dass Lehrkräfte selbst angaben, dass sie sich auch divergent zwischen Lernenden mit und ohne Sonderpädagogischen Förderbedarf verhalten. So gaben Lehrkräfte an, aufgrund von Erwartungen einen weniger engen Kontakt zu Lernenden mit Autismus-Spektrum-Störung und Aufmerksamkeitsdefizit-Hyperaktivitätsstörung zu suchen und ein weniger konflikthafte Verhältnis mit ihnen zu führen (Zee, de Bree, Hakvoort & Koomen, 2020).

Das divergierende Verhalten von Lehrkräften könnte auch ein Argument dafür sein, dass sich in Beitrag IV die Tendenz zeigte, dass Lehrkräfte die wahrgenommene Motivierungsqualität von Lernenden auf Individual- als auf Klassenebene genauer einschätzen konnten und demnach nicht nur Lernende die Qualität des unterrichtlichen Angebots divergent wahrnahmen, sondern sich auch die Lehrkraft divergent den Lernenden gegenüber verhielt. Dies erschwert die häufig angewendete Methode Wahrnehmungen von Lernenden zu aggregieren (z. B. Lüdtke et al., 2009), um Aussagen über die Qualität des Unterrichts treffen zu können.

### **8.1.2 Individuelle und methodische Prädiktoren für die Wahrnehmungen verschiedener Perspektiven**

Die Betrachtung der Wahrnehmungen von Lernenden ist bedeutsam, da diese prädiktiv für ihren Lernerfolg sind (z. B. Kunter & Baumert, 2006; Sakiz, Pape & Hoy, 2012). Untersuchungen zeigten aber, dass die Wahrnehmungen von Unterrichtsqualität in Klassen

variiert (Wagner et al., 2016). Wenig beleuchtet wurden aber mögliche individuelle Merkmale von Lernenden, die die Wahrnehmungen bedingen und mögliche Ursachen für die Divergenz sein könnten. Mittels Mehrebenenanalysen wurde in Beitrag I (*Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität*) geprüft, welche aus der Theorie der Informationsverarbeitung abgeleiteten individuellen Merkmale divergierende Wahrnehmungen von Lernenden aus der Primarstufe bedingen. Das Vorwissen, die kognitiven Fähigkeiten, die motivationale Orientierung und der soziale Hintergrund sind bedeutsame Merkmale für die Wahrnehmungen der kognitiven Aktivierung von Lernenden. Wahrnehmungen der Klassenführung wurden durch den sozialen Hintergrund vorhergesagt. Die kognitiven Fähigkeiten und die emotionale Bindung erwiesen sich als prädiktiv für die Wahrnehmungen des unterstützenden Unterrichtsklimas. Nur das Geschlecht und der Migrationshintergrund bedingten die Wahrnehmungen der Unterrichtsqualität nicht (siehe Abbildung 5).

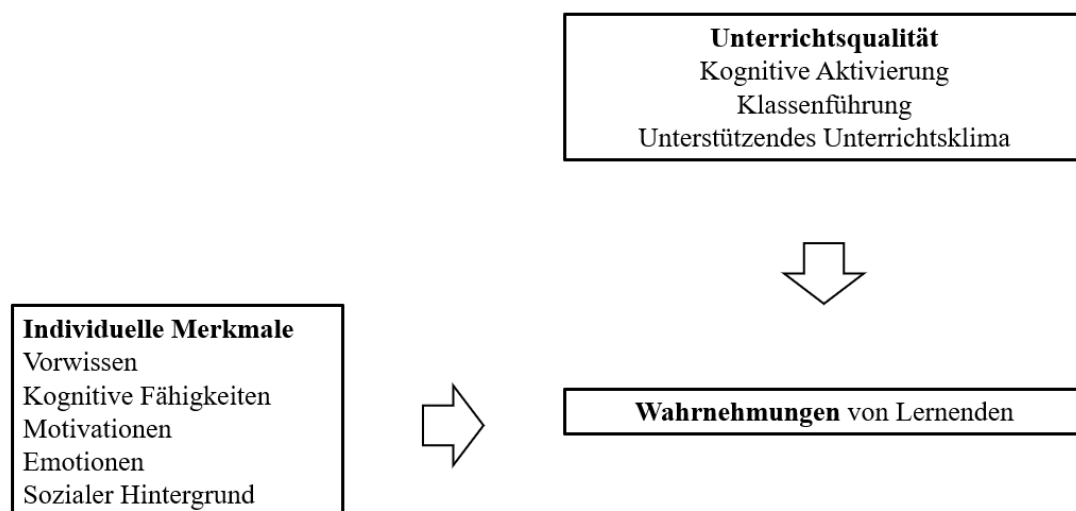


Abbildung 5. Individuelle Prädiktoren für die Wahrnehmungen von Lernenden

Welche Lehrkraftmerkmale für die Wahrnehmung der Lehrkraft über die Qualität des eigenen Unterrichts bedeutsam sind, wurde in Beitrag II (*Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht*) untersucht. Aus theoretischen Modellen kann abgeleitet werden, dass Lehrkraftmerkmale wie die eigenen Motivationen und die Selbstwirksamkeitserwartung auf die Qualität des eigenen Unterrichts wirken (Jennings & Greenberg, 2009; Helmke & Helmke, 2015; Helmke, 2022). Empirisch konnte gezeigt werden, dass die Motivationen von Lehrkräften mit der Unterrichtsqualität zusammenhängen (Roth, Assor, Kanat-Maymon &

Kaplan, 2007). Auch die Selbstwirksamkeitserwartung von Lehrkräften sagt die Qualität des unterrichtlichen Angebots vorher (Holzberger, Philipp & Kunter, 2013). Lazarides und Schiefele (2021) untersuchten die Bedeutung der Selbstwirksamkeitserwartung und des Interesses an Bildungsthemen von Lehrkräften für die Unterrichtsqualität aus Lernenden- und Lehrkraftsicht. Sie stellten fest, dass die Selbstwirksamkeitserwartung insbesondere mit der von Lehrkräften wahrgenommenen Unterrichtsqualität zusammenhing. Resultate des Beitrags II zeigten, dass die Motivation bezüglich Diagnostik die Qualität des Leseunterrichts vorhersagte. Höhere Ausprägungen in der Motivation gehen mit einer selbstwahrgenommenen häufigeren Differenzierung im Leseunterricht und einer höheren Motivierungsqualität einher. Allerdings erwies sich die Selbstwirksamkeitserwartung bezüglich Diagnostik als nicht prädiktiv für die selbstwahrgenommene Unterrichtsqualität (siehe Abbildung 6). Aufgrund der Korrelation zwischen den Variablen Motivation und Selbstwirksamkeitserwartung bezüglich Diagnostik kann auch ein Suppressoreffekt vorliegen.

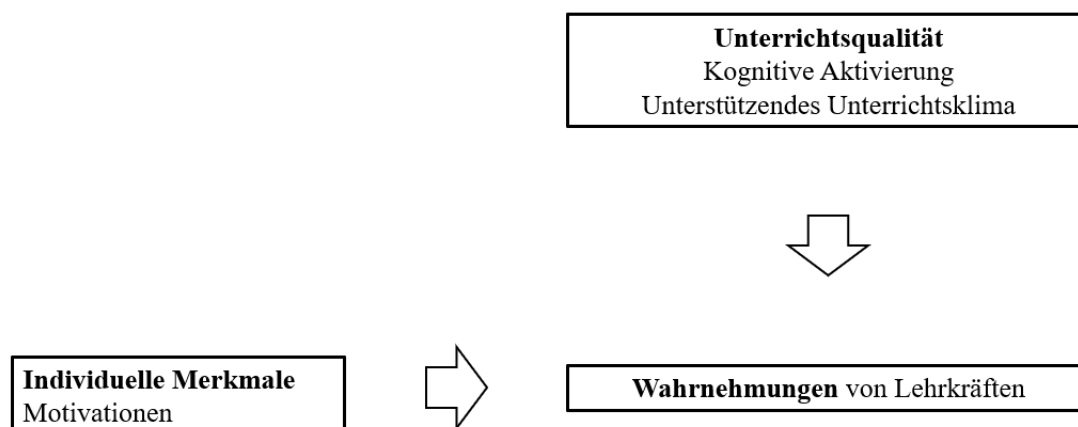


Abbildung 6. Individuelle Prädiktoren für die Wahrnehmungen von Lehrkräften

Die Perspektive der externen Urteilenden wurde lange als Königsweg bei der Erfassung von Unterrichtsqualität betitelt (Herbert & Schweig, 2021), dabei können auch die Wahrnehmungen von externen Urteilenden Einflüssen unterliegen und von Beurteilungsfehlern betroffen sein, welche die Nutzbarkeit und die Interpretierbarkeit einschränken (Pietsch & Tosana, 2008). So können individuelle Merkmale wie die Vorerfahrung und die Stimmung die Wahrnehmungen von externen Urteilenden bedingen (für Vorerfahrung: Gabriel-Busse, Groß-Mlynek, Feldhoff & Haring, 2020; für Stimmung: Schmidt-Atzert, Büttner & Bühner, 2004). Ein möglicher methodischer Einflussfaktor auf die Wahrnehmung von Unterrichtsqualität kann die Darbietungsreihenfolge der

Unterrichtsszenen sein (Mashburn et al. 2014). Unterliegen die externen Urteilenden den gleichen Einflussfaktoren, kann die Interraterreliabilität kein Indiz für ein unverzerrtes Urteil sein. Beitrag III (*Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt*) untersuchte, welche Einflussfaktoren auf die Wahrnehmungen von externen Urteilenden wirken. Resultate zeigten, dass die Vorerfahrung einen Einfluss auf die Wahrnehmungen der Motivierungsqualität hatte. Externe Urteilende mit einer höheren Vorerfahrung nahmen die Motivierungsqualität niedriger wahr als unerfahrene externe Urteilende. Auch die Müdigkeit hatte nur einen Effekt auf die Wahrnehmung der Motivierungsqualität. Externe Urteilende, die ihre Müdigkeit hoch einschätzten, nahmen die Motivierungsqualität höher wahr als wachere externe Urteilende (siehe Abbildung 7).

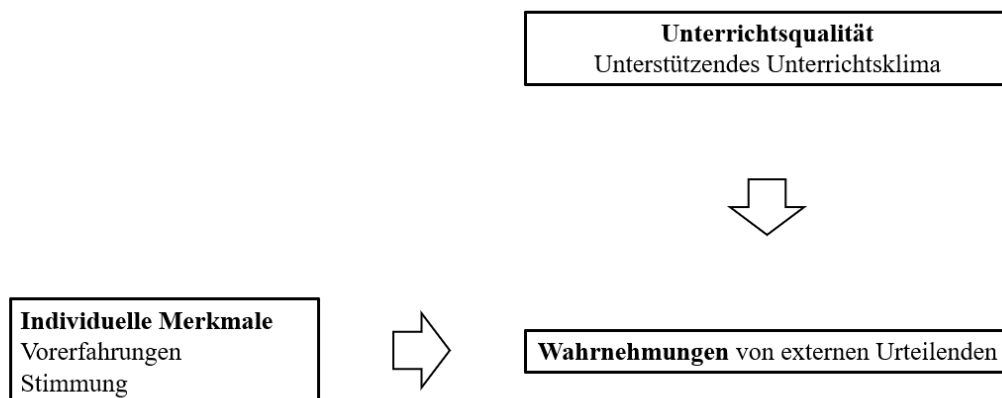


Abbildung 7. Individuelle Prädiktoren für die Wahrnehmung von externen Urteilenden

Für die Wahrnehmungen von externen Urteilenden hinsichtlich der kognitiven Aktivierung und der Motivierung erwies sich das methodische Merkmal der Darbietungsordnung von Unterrichtssequenzen als Einflussfaktor. Demnach nahmen externe Urteilende in Abhängigkeit der Reihenfolge der gesehenen Unterrichtsszenen die Qualität der kognitiven Aktivierung und Motivierung unterschiedlich wahr (siehe Abbildung 8).

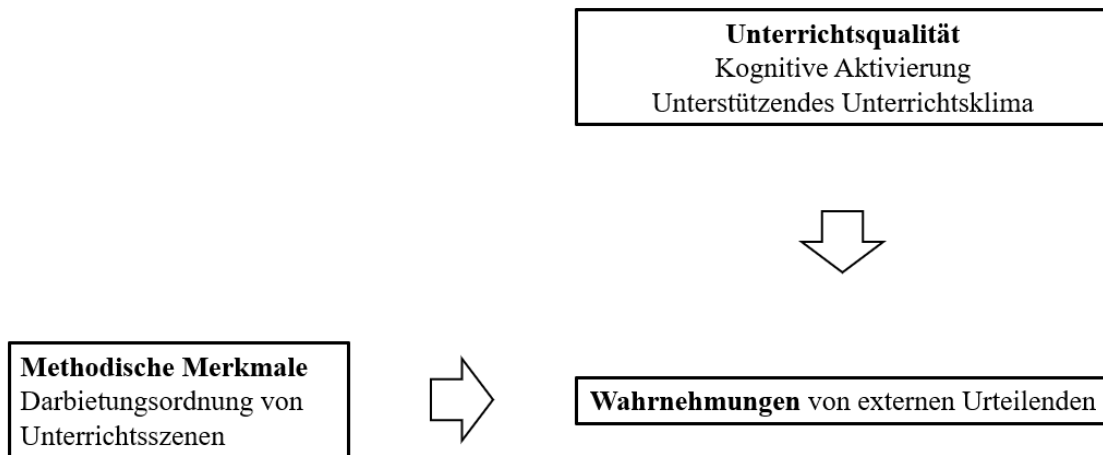


Abbildung 8. Methodische Prädiktoren für die Wahrnehmung von externen Urteilenden

#### *Fazit zu individuellen und methodischen Prädiktoren für die Wahrnehmungen*

Wie in theoretischen Modellen angenommen und in empirischen Untersuchungen schon teilweise bestätigt, haben individuelle Merkmale sowie methodische Merkmale eine Relevanz für die Wahrnehmung von Unterrichtsqualität. Einen Erklärungsansatz bieten Informationsverarbeitungsmodelle, welche den Wahrnehmungsprozess und die Wirkung individueller Einflussfaktoren näher beleuchten (z. B. Atkinson & Shiffrin, 1968; Greifeneder, Bless & Fiedler, 2018). In Bezug auf die Angaben der Untersuchungen der vorliegenden Arbeit kann nicht differenziert werden, ob die Einflussfaktoren auf die Sequenzen 1) Wahrnehmungen und Aufmerksamkeit, 2) Entschlüsselung und Interpretation oder 3) zukünftige Prozesse, Schlussfolgerungen, Urteile und Entscheidungen wirken. Die Befragten (Lernende, Lehrkräfte oder externe Urteilende) setzen in Fragebögen ihr Kreuz und der dahinterliegende Prozess wird in der empirischen Bildungsforschung eher weniger betrachtet. Der dahinterliegende Prozess bei Befragungen zur Unterrichtsqualität bleibt eine Art Black-Box (Geitz, Vater & Zimmer-Merkle, 2020). Schütte und Kolleg\*innen (2022) diskutierten hinsichtlich der frühkindlichen Bildung, dass kognitive Entwicklungsprozesse und deren Einflussfaktoren bislang noch nicht gut genug erforscht sind. Dies liegt auch daran, dass diese nur erschwert beobachtbar sind. Für eine größere Transparenz der kognitiven Prozesse beim Lernen und ihren Einflussfaktoren empfehlen Schütte und Kolleg\*innen (2022) Messungen von Blickbewegungen, welche abbilden, wohin die Lernenden ihre Aufmerksamkeit richten, und die Messung von Gehirnströmen. Hierbei können Informationen über die Informationsverarbeitung gewonnen werden. Nachteile beider Verfahren sind der große Aufwand und die Kosten für die Ausstattung (Schütte, Rose & Köller, 2022). Diese Verfahren

bieten die Chance genauer zu untersuchen, auf was Wahrnehmende im Unterricht achten und wann sie bestimmte Reize wahrnehmen. Somit können die Verfahren des Eye-Trackings und der EEG-Messung einen größeren Aufschluss über die Wahrnehmung von Unterricht geben. Auch gibt es erste Studien, welche den Ansatz verfolgen das Potenzial zur kognitiven Aktivierung mit Hilfe eines neu entwickelten Erhebungsinstruments zu erfassen: die Wahrnehmung der Unterrichtsqualität auf der gemeinsamen Auswertung aller Unterrichtsmaterialien einer Stunde durch externe Urteilende (Herbert & Schweig, 2021).

Aus den Untersuchungen geht hervor, dass für alle drei Perspektiven die Wahrnehmungen der Motivierungsqualität, als Unterrichtsmerkmal der Unterrichtsqualitätsdimension unterstützendes Unterrichtsklima, von individuellen und methodischen Merkmalen bedingt wurden. Im Gegensatz dazu wurde die Klassenführung nur in Beitrag I von einem individuellen Merkmal (sozialer Hintergrund) vorhergesagt. Demzufolge scheinen Unterrichtsqualitätsmerkmale, welche im Vergleich zu Merkmalen der Dimensionen kognitive Aktivierung und unterstützendes Unterrichtsklima mit eher niedrig-inferenten Items abgefragt werden können, weniger aufgrund von individuellen und methodischen Merkmalen verzerrt. Schließlich scheint der Inferenzgrad sowohl für die Frage der Übereinstimmung von Wahrnehmungen von Unterrichtsqualität als auch für die Frage nach individuellen und methodischen Prädiktoren für die Wahrnehmungen von Unterrichtsqualität eine Rolle zu spielen. Insbesondere ist dies bei den Wahrnehmungen von externen Urteilenden zu beachten.

Der emotionale Abstand der externen Urteilenden wird häufig als Vorteil der Perspektive genannt, dabei deuteten die Resultate darauf hin, dass externe Urteilende Schwierigkeiten aufweisen, Unterrichtsmerkmale wahrzunehmen, welche sie nicht direkt beobachten können und sich auf die Interaktion zwischen den Beteiligten am Unterricht beziehen. Hierzu zählt auch die Motivierungsqualität, die sich auf die Motivationsunterstützung der Lehrkraft für die Lernenden bezieht. Häufig wird dieses Unterrichtsqualitätsmerkmal durch Items zur Selbstbestimmungstheorie nach Ryan und Deci (2020) operationalisiert, die auf die Unterstützung des Autonomieerlebens, Kompetenzerlebens und der sozialen Einbeziehung abzielen (Schwabe et al., 2020). Die Items beziehen sich auf das Empfinden der Lernenden. Bisher liegen keine eindeutigen Befunde zu Übereinstimmungen der Perspektiven bezüglich der Wahrnehmung von Unterrichtsqualität vor. Einige Studien zeigten aber, dass die Wahrnehmungen von externen Urteilenden hinsichtlich des unterstützenden Unterrichtsklimas nicht mit den Wahrnehmungen von

Lernenden und Lehrkräften zusammenhingen (z. B. De Jong & Westerhof, 2001). Werden die Befunde des Beitrags III vor dem theoretischen Hintergrund und der bisherigen empirischen Untersuchungen diskutiert, kann festgehalten werden, dass bei der Erfassung der Motivierungsqualität nicht auf die Perspektive der externen Urteilenden zurückgegriffen werden sollte. Da in Beitrag IV (weiterführende Analysen) gezeigt wurde, dass Lehrkräfte Schwierigkeiten haben, einzuschätzen, wie ihre Lernenden die Motivierung wahrnahmen, sollte bei der Dimension des unterstützenden Unterrichtsklimas auf die Wahrnehmungen der Lernenden zurückgegriffen werden.

## **8.2 Stärken und Grenzen der vorliegenden Arbeit**

### **8.2.1 Stärken**

Eine Stärke der vorliegenden kumulativen Dissertation ist, dass die Studien einen Beitrag zur Theoriebildung leisten; insbesondere Beitrag II und Beitrag III. In den Beiträgen II und III wurden Modelle zu Lehr-Lernprozessen mit den Modellen zur Informationsverarbeitung verknüpft und um Prädiktoren des Wahrnehmungsprozesses erweitert. In Beitrag II konnte gezeigt werden, dass die Wahrnehmungen von Lernenden bezüglich der Qualität des Unterrichts mit individuellen Prädiktoren zusammenhängen. Aufgrund der Verknüpfung zu Beurteilungstendenzen in der Informationsverarbeitung konnte in Beitrag III festgestellt werden, dass die Wahrnehmungen von externen Urteilenden aufgrund des Beurteilungsfehlers des Reihenfolgeeffekts, der vorliegenden Erfahrung und der Müdigkeit divergieren können. In Untersuchungen der empirischen Bildungsforschung zur Wahrnehmung von Unterrichtsqualität werden bislang grundlegende Theorien zur Informationsverarbeitung und möglichen Einschränkungen durch individuelle und methodische Prädiktoren wenig berücksichtigt. Dies spiegelt sich auch in den Begrifflichkeiten wieder (Hierzu mehr unter 8.2.2 Grenzen. Begrifflichkeit Wahrnehmung).

Eine methodische Stärke ist, dass die Stichprobenumfänge der Beiträge I, III und IV für die durchgeführten statistischen Analysen geeignet waren. Die notwendigen Stichprobengrößen wurden teilweise vorab mittels Poweranalysen geprüft. Lediglich ist in Beitrag II die Stichprobengröße von  $N = 40$  als Einschränkung zu sehen, da kleinere Stichproben zu Verzerrungen der Teststatistik und zur Verschlechterung der Modellschätzung führen können (Algina & Olejnik, 2000; Muthén & Muthén, 2002). Weitestgehend wiesen die verwendeten Befragungsinstrumente gute bis sehr gute Reliabilitäten auf. Falls dies nicht der

Fall war, wurde dies in Beitrag transparent dargestellt und diskutiert (Beispielsweise Beitrag IV [weiterführende Analysen] Skala zur Autonomieunterstützung). Fehlende Werte wurden entweder mittels des Full Information Likelihood Algorithmus (FIML) oder auch mittels der multiplen Imputation geschätzt. So wurden beispielsweise in Beitrag I fünf Datensätze mit Hilfe der Funktion der multiplen Imputation in SPSS erstellt und in Mplus wurden für die Analysen durch die Verwendung des Befehls `type = imputation` alle fünf Datensätze berücksichtigt. Falls möglich wurden die Konstrukte latent modelliert und die geschichtete Datenstruktur wurde mittels Clusterung nach Schulen oder Klassen in den Modellen berücksichtigt.

### *Primarstufe im Fokus*

Eine konzeptionelle Stärke der Arbeit liegt darin, dass sich in allen Beiträgen die Wahrnehmungen von Unterrichtsqualität auf die Primarstufe bezieht. Viele Studien fokussieren die Untersuchung der Qualität in der Sekundarstufe (z. B. Begrich, Fauth & Kunter, 2020; Willems, 2022). Ergebnisse des Sekundarbereichs können allerdings nicht direkt übertragen werden, da sich der Primarbereich durch andere Strukturen auszeichnet. In der Primarstufe werden alle Kinder gemeinsam beschult. Es treffen alle Leistungsgruppen aufeinander; ausgenommen sind Kinder, welche auf Förderschulen beschult werden. Der Antrag auf sonderpädagogische Unterstützung kann bis auf begründete Einzelfälle, dies betrifft häufig die sonderpädagogischen Förderschwerpunkte Geistige Entwicklung sowie Körperliche und Motorische Entwicklung, erst nach Schuleintritt eingeleitet werden (Wolf & Dietze, 2022). Im Unterschied zur Sekundarstufe werden Lehrkräfte in der Primarstufe mit einer höheren Leistungsheterogenität konfrontiert. Auch werden die Kinder alle im ähnlichen Alter aus der Kindertagesstätte in die Primarstufe eingeschult, allerdings unterscheidet sich das Entwicklungsalter der Lernenden im ersten Schuljahr um bis zu drei bis vier Jahre (Largo & Beglinger, 2009; Peschel, 2012). Ein wichtiger Unterschied ist auch die intensivere Beziehung zur Lehrkraft. In der Primarstufe ist eine enge und intensive Beziehung zur Lehrkraft und ihre Unterstützung bedeutsam (Bru, Stornes, Munthe & Thuen, 2010). Die Unterstützung der Lehrkraft verliert über die Jahre an Bedeutung für die Lernenden und im Jugendalter haben das Peer-Feedback und ihre Motivationsunterstützung einen größeren Einfluss auf die Lernenden (Rubin, Bukowski & Parker, 1998; Bru et al., 2010). Dementsprechend ist eine große Stärke, dass in allen vier Beiträgen der kumulativen Dissertation die Motivierungsqualität durch die Lehrkraft in der Primarstufe untersucht wurde. Hervorzuheben ist Beitrag IV, in welchem in Bezug auf die Unterrichtsqualität nur das

Unterrichtsmerkmal Motivierungsqualität in den Blick genommen und analysiert wurde, wie genau die Lehrkräfte wahrnehmen konnten, wie ihre Lernenden in der Primarstufe ihre Motivationsunterstützung wahrnahmen. Bisher wurde die Motivierungsqualität des Unterrichts in der Primarstufe selten untersucht (Lepper, 2022). Ebenfalls werden in der kumulativen Dissertation drei Perspektiven zur Erfassung der Unterrichtsqualität berücksichtigt. Allerdings beziehen sich die Wahrnehmungen aller drei Perspektiven nicht auf den gleichen Unterricht in der Primarstufe. Demzufolge können ihre Wahrnehmungen nicht direkt miteinander verglichen werden.

### *Unterrichtsfach Deutsch und Unterrichtseinheit zur Bild-Textintegration*

In der empirischen Bildungsforschung zur Unterrichtsqualität beziehen sich Studien häufig auf die Wahrnehmungen zum Unterrichtsfach Mathematik (z. B. Klieme, Lipowsky, Rakoczy & Ratzka, 2006; Lepper, Stang, & McElvany, 2021; Rakoczy, 2008; Willems, 2022). Die Beiträge I und III beziehen sich auf Unterrichtseinheiten zur Bild-Textintegration, welche entweder im Rahmen des Deutschunterrichts oder des Sachunterrichts durchgeführt wurden. In Beitrag II bezogen sich die Wahrnehmungen der Lehrkräfte auf die Qualität des Leseunterrichts. In Beitrag IV gaben die Lernenden und die Lehrkräfte an, wie sie die Qualität des letzten Deutschunterrichts wahrnahmen. Zwar gelten die drei Unterrichtsqualitätsdimensionen als Unterrichtsfach übergreifend und bestehend aus Formulierungen generischer Unterrichtsqualitätsmerkmale, jedoch bezieht sich auch die Mehrzahl an Untersuchungen auf die Unterrichtsqualität im naturwissenschaftlichen und mathematischen Unterricht (Praetorius & Gräsel, 2021). Im Hinblick auf die Wahrnehmungen von Unterrichtsqualität können beispielweise dahingehend Unterschiede aufgrund des Unterrichtsfaches entstehen, da der naturwissenschaftliche und mathematische Unterricht häufig als männliche Domäne beschrieben wird, hingegen der sprachliche Bereich und Leseunterricht eher als weibliche Domäne gilt (Espinoza & Strasser, 2020; Nowicki & Lopata, 2017). Somit sind Ergebnisse zur Wahrnehmung der Unterrichtsqualität im Unterrichtsfach Mathematik nicht unbedingt direkt übertragbar und müssen auch für unterschiedliche Fächer überprüft werden. Eine Stärke der vorliegenden Dissertation ist, dass die Befunde zum häufig untersuchten Unterrichtsfach Mathematik um Befunde zum Unterrichtsfach Deutsch, zum Leseunterricht und zum Bereich Bild-Text-Integration ergänzt werden. In allen in dieser Dissertation untersuchten Unterrichtseinheiten stand der sprachliche Bereich im Mittelpunkt der Einheit.

### 8.2.2 Grenzen

Abgesehen von Stärken müssen auch Grenzen der vorliegenden Arbeit berichtet und diskutiert werden. Nicht in allen Beiträgen wurden die Wahrnehmungen aller drei Unterrichtsqualitätsdimensionen eingeholt. So wurden nur in den Studien der Beiträge I und III drei Unterrichtsqualitätsmerkmale erhoben, die den drei Unterrichtsqualitätsdimensionen zuzuordnen sind. In Beitrag II wurden Lehrkräfte nur zu ihrer Wahrnehmung der Unterrichtsqualitätsdimensionen kognitive Aktivierung und Motivierungsqualität befragt. Im Rahmen der Befragung des Beitrags 4 wurden Lehrkräfte befragt, wie ihre Schüler\*innen nur ihre angebotene Motivierungsqualität wahrnahmen. Es wurden keine Wahrnehmungen bezüglich der Dimension kognitive Aktivierung oder Klassenführung erhoben. Somit bleibt ein Bedarf, Lehrkräftewahrnehmungen bezüglich der Klassenführung im Hinblick auf mögliche Prädiktoren sowie die diagnostische Kompetenz von Lehrkräften in Bezug auf die Wahrnehmung von Lernenden der kognitiven Aktivierung oder Klassenführung zu untersuchen.

#### *Begrifflichkeit Wahrnehmung*

Wie schon in Kapitel 2.1 beschrieben, wird bislang in der empirischen Bildungsforschung nicht eindeutig zwischen den Begrifflichkeiten Wahrnehmung, Beobachtung, Urteil und Einschätzung klar differenziert. Dieses Problem zeigt sich deutlich, wenn es darum geht, eine Begrifflichkeit für alle Perspektiven zu wählen. Nach Kloss (2013) verläuft die Wahrnehmung eher beiläufig und unbeabsichtigt und eine Beobachtung erfolgt zielgerichtet und theoriegeleitet. Übertragen auf die Perspektiven zur Erfassung von Unterrichtsqualität bedeutet dies, dass auch immer das Studiendesign berücksichtigt werden muss, bevor der Begriff Wahrnehmung oder Beobachtung verwendet werden kann. Orientiert am Studiendesign wären Beispiele für Wahrnehmungen von Unterrichtsqualität, wenn Lernende oder Lehrkräfte ohne vorige Informationen direkt im Anschluss an eine Unterrichtsstunde zu ihrer Wahrnehmung von Unterrichtsqualitätsmerkmalen befragt werden. Wissen Lehrkräfte aber vorab, welche Themen bei der Studie abgefragt werden, können sie bis zur Erhebung beispielsweise auf Störungen im Unterricht achten und gezielt beobachten, bevor sie ein Urteil darüber abgeben müssen. Ähnlich ist es bei Befragungen von externen Urteilenden. Meistens sind externe Urteilende vorab über die Unterrichtsqualitätsmerkmale informiert und sie sollen diese gezielt und theoriegeleitet beobachten und beurteilen. Demzufolge passt oftmals bei externen Urteilenden die Bezeichnung Beobachtung statt Wahrnehmung von Unterrichtsqualität besser. Für weitere Arbeiten in der Bildungsforschung sollte die

Verwendung von Begrifflichkeiten weiter geklärt werden. Dafür sollten unter anderem Modelle von Lehr-Lernprozessen mit Aspekten des Wahrnehmungs-/Urteilsprozesses verbunden werden. Es wird deutlich, dass bislang Modelle von Lehr-Lernprozessen und der Prozess der Informationsverarbeitung nicht genug miteinander verbunden wurden, um konkrete Begrifflichkeiten in der empirischen Bildungsforschung zu wählen.

### *Itemformulierungen*

Nach Fauth und Kolleg\*innen (2018) sollten Items für die Erfassung von Unterrichtsqualität nach dem beobachtbaren Verhalten der Lernenden fragen. In den Befragungen der vorliegenden kumulativen Dissertation sind unterschiedliche Itembezüge zu finden. Die Items in den Beiträgen beziehen sich nicht immer nur auf das beobachtbare Verhalten der Lernenden, sondern auch auf das Lehrkraftverhalten sowie eine Kombination aus Lernenden- und Lehrkraftverhalten. In Beitrag I fällt auf, dass der Itembezug zwischen den einzelnen Items für die Unterrichtsqualitätsmerkmale variiert. Für das Unterrichtsqualitätsmerkmal herausfordernde Aufgaben wurden die Lernenden nach dem Lernendenverhalten gefragt („Wir Schüler mussten ganz schön denken, bis wir den Text und das Bild verstanden haben.“). Für die Erfassung der Störungen im Unterricht sollten die Lernenden ihre Wahrnehmungen zur Kombination aus Lernenden- und Lehrkraftverhalten schildern („Im Unterricht zu Südamerika wurde der Unterricht oft gestört“). Die Items der Motivierungsqualität bezogen sich auf das Lehrkraftverhalten, beziehungsweise ein Item auch auf das Lernendenverhalten („Im Unterricht gerade hat unser / e Lehrer / in das Lesen der Texte mit Bildern sehr spannend gestaltet.“). Des Weiteren beziehen sich die Items zum Merkmal herausfordernde Aufgaben auf die Wahrnehmung der ganzen Klasse. Für die Beantwortung dieser Items müssen die Lernenden die Wahrnehmungen der ganzen Klasse mitteln können und ihnen müssen auch die Informationen über die Wahrnehmungen aller Mitschüler\*innen vorliegen. Insbesondere bei jüngeren Lernenden wird deswegen empfohlen sich nur auf Aussagen über die eigenen Wahrnehmungen zum Unterricht zu beschränken (Lenske, 2016).

In Beitrag IV ist kritisch zu sehen, dass sich die Itemformulierungen für alle drei Unterrichtsqualitätsdimensionen bei den Lernenden und Lehrkräften auf das Lehrkraftverhalten beziehen (Beispiel Lernende: „Ich hatte das Gefühl, dass meine Lehrerin/mein Lehrer mir Entscheidungsmöglichkeiten geboten hat.“ Beispiel Lehrkräfte: Das Kind hatte das Gefühl, dass ich ihm Entscheidungsmöglichkeiten geboten habe.). Die Ergebnisse des Beitrags wiesen auf geringe Übereinstimmungen und divergierende Wahrnehmungen von Lernenden und Lehrkräften hin. Ein Grund hierfür kann aber auch der

Itembezug sein. Untersuchungen zeigten, dass die Übereinstimmungen der Wahrnehmungen von Lernenden und Lehrkräften höher sind, wenn sich die Items auf das Verhalten der Lernenden bezieht (z. B. Wettstein, Ramseier, Scherzinger & Gasser, 2016). Demnach sollten sich zukünftige Untersuchungen, welche die diagnostische Kompetenz und Wahrnehmungsübereinstimmung untersuchen, auf das Lernendenverhalten beziehen. Ebenfalls könnten die Resultate aus Beitrag IV mittels Befragungen mit einem Itembezug auf das Lernendenverhalten überprüft werden und die Untersuchung repliziert werden. Positiv ist aber, dass sich die Items zur Erfassung der Unterrichtsqualität aus Lernendensicht des Beitrags IV auf die zuletzt erlebte Unterrichtsstunde beziehen und der zugrundeliegende Beurteilungszeitraum eng abgesteckt ist. Allerdings erfolgten alle Befragungen der Beiträge I-IV retrospektiv und alle Befragten trafen erst nach dem erlebten Unterricht Aussagen zu ihrer Wahrnehmung der Unterrichtsqualität. Auch in der experimentellen Studie des Beitrags III wurden die externen Urteilenden zu einer Unterrichtseinheit befragt, welche sie kurz zuvor gesehen haben. Dabei können insbesondere retrospektive Angaben von Verzerrungen, wie von Erinnerungseffekten, betroffen sein (Pieper, 2023).

#### *Variabilität von Wahrnehmungen der Unterrichtsqualitätsmerkmale*

In Studien wird bislang die Variabilität von Einschätzungen hinsichtlich Unterrichtsqualitätsmerkmalen kaum berücksichtigt. Dabei können Lernende in ihren Einschätzungen zu der Qualität des unterrichtlichen Angebots erheblich schwanken (Götz, Lüdtke, Nett, Keller & Lipnevich, 2013; Rakoczy, Frick, Weiß-Wittstadt, Tallon & Wagner 2022). In einer Studie von Rakoczy und Kolleg\*innen (2022) zeigte sich, dass die größte Varianz in den Unterrichtswahrnehmung (72.0 %) aufgrund von intraindividuellen Schwankungen zustande kam. Die Befragung von Lernenden bezog sich auf zwei Doppelstunden, die sich auf zwei verschiedene Tage verteilten. Insgesamt wurden 10 Messungen durchgeführt (Rakoczy et al., 2022). Im Hinblick auf die Resultate kann für weitere Studien methodisch in Frage gestellt werden, ob eine einmalige Einschätzung des Unterrichts genügt, um valide Aussagen über die Unterrichtsqualität treffen zu können. Darüber hinaus wurde bisher die Frage nach Prädiktoren für die intraindividuellen Schwankungen nicht hinreichend geklärt. Auch die Frage, wie retrospektive Einschätzungen getroffen werden, wurde bisher weniger beleuchtet. Einerseits könnte eine retrospektive Einschätzung das Mittel verschiedener Einschätzungen abbilden, andererseits könnte sich eine Einschätzung auch nur auf einen letzten Urteilsprozess beziehen (Rakoczy et al., 2022). Auch bezüglich der Förderung von Lernenden ist die Variabilität in den Wahrnehmungen zu

beachten. Für Lernende bedeuten die Ergebnisse zur Variabilität von Unterrichtsqualitätsmerkmalen über verschiedene Zeitpunkte hinweg, dass zukünftig untersucht werden muss, welche Faktoren individuelle Wahrnehmungen bestimmen und wie individuelle Verläufe positiv beeinflusst werden können (Rakoczy et al., 2022). Insbesondere variieren Wahrnehmungen von Lernenden im Bereich Motivierungsqualität über den Tag hinweg (Patall et al., 2018; Martin et al., 2015). Da die wahrgenommene Motivierungsqualität aber relevant für Lernendenmerkmale wie die Motivation ist (Lepper et al., 2021), bleibt zu klären, wie die wahrgenommene Motivierungsqualität aufrechterhalten werden kann.

Bezüglich der Einschätzungen von externen Urteilenden über die Unterrichtsqualität via Videoratings stellten Praetorius und Kolleg\*innen (2014) fest, dass insbesondere das Unterrichtsqualitätsmerkmal kognitive Aktivierung nicht stabil über mehrere Unterrichtsstunden eingeschätzt werden konnte und insgesamt neun Unterrichtsstunden pro Lehrkraft eingeschätzt wurden, um reliable Einschätzungen von den Externen über die Qualität der kognitiven Aktivierung zu erhalten. Empfehlungen, die aus den Ergebnissen abgeleitet wurden, waren, dass wenn nicht mehrere Unterrichtsstunden pro Lehrkraft zur Verfügung stehen und eingeschätzt werden können, sich in Befragungen einzelner Unterrichtsstunden auf Unterrichtsqualitätsmerkmale bezogen wird, welche deutlich stabiler über Unterrichtsstunden hinweg eingeschätzt werden können; zum Beispiel das Merkmal Klassenführung (Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2014). Konträr dazu zeigten Gabriel-Busse und Lipowsky (2021), dass neben der kognitiven Aktivierung auch für die Klassenführung mindestens drei Unterrichtsstunden für eine reliable Einschätzung der Qualität notwendig sind. Schließlich ist nicht hinreichend geklärt, inwiefern die Variabilität von Einschätzungen über Unterrichtsqualitätsmerkmale auch die anderen Perspektiven zur Erfassung von Unterrichtsqualität betrifft und welche Faktoren dies bedingen.

### **8.3 Implikationen**

Aus den Resultaten der Beiträge können Implikationen für Theorie, Forschung und Praxis abgeleitet werden. Einige Implikationen wurden bereits in den vorigen Kapiteln genannt. Weitere Schlussfolgerungen werden für die Forschung und Praxis abgeleitet. Aufgrund der Ergebnisse des Beitrags I kann das Angebots-Nutzungs-Modell von Unterricht nach Helmke (2022) erweitert werden. Das Modell setzt die Merkmale von Unterrichtsqualität in Verbindung mit Eigenschaften der Lehrperson und den Wirkungsweisen und Zielkriterien des

Unterrichts. Hierbei sieht Helmke den Unterricht als Angebot, welches nicht notwendigerweise zu den Wirkungen des Unterrichts führt. Um der Komplexität von Lehr-Lern-Prozessen gerecht zu werden, müssen vermittelnde und interpretative Prozesse auf Lernendenseite berücksichtigt werden. Die Wirksamkeit des Lernens wird von zwei Mediationsprozessen auf Lernendenseite begleitet. Im ersten Schritt werden die Prozesse von den wahrgenommenen Erwartungen der Lehrkraft, dem Unterrichtsangebot und den darauffolgenden Interpretationen mediiert. Der zweite parallel verlaufende Vorgang beinhaltet die Auswirkungen der unterrichtlichen Maßnahmen auf die motivationalen, emotionalen und volitionalen Prozesse auf Schüler\*innenseite (Hemlke, 2022). Helmke sieht zunächst keine Verbindung zwischen den Merkmalen von Lernenden, zu denen die Faktoren auf Familienseite und das Lernpotenzial der Lernenden zählen, und der Wahrnehmung sowie Interpretation der unterrichtlichen Angebote. Aufgrund der Resultate des Beitrags I kann der Pfad dieses Modells, die Verbindung zwischen Wahrnehmung sowie Interpretation der unterrichtlichen Angebote und individuellen Merkmale der Lernenden (Faktoren auf Familienseite und Lernpotenzial der Lernenden), ergänzt werden. Weitere Implikationen für die Forschung und Praxis, die mittels der Resultate der Beiträge I, II, III und IV gefolgert werden können, werden vorgestellt.

### **8.3.1 Implikationen für die Forschung**

Die Ergebnisse beider übergeordneten Forschungsfragen verweisen darauf, dass Wahrnehmungen über ein und denselben Unterricht divergieren und bei Befragungen zu Unterrichtsqualitätsmerkmalen individuelle und methodische Merkmale bedingen, wie die Wahrnehmungen von Lernenden, Lehrkräften und externen Urteilenden ausfallen. Hinter den Unterrichtsqualitätsmerkmalen, die erhoben werden, stecken somit auch immer individuelle Merkmale der Befragten sowie methodische Merkmale der Befragung. In den Angaben zur Unterrichtsqualität verschwimmen somit viele Faktoren und erschweren es, von einem zugrundeliegenden Konstrukt zu sprechen. Aufgrund der Bedeutsamkeit von Unterrichtsqualität für den Lernerfolg und für den Abbau von Ungleichheiten (Klieme, 2020), sollte weiterhin versucht werden die Unterrichtsqualität zu erfassen, um Lernenden bestmögliche Förderung anzubieten. Mögliche Einflussfaktoren sollten aber bei der Befragung oder spätestens bei der Diskussion der Ergebnisse zu Wahrnehmungen von Unterrichtsqualität berücksichtigt werden. Um den Einfluss von individuellen Prädiktoren zu berücksichtigen, können Faktoren, welche die Informationsverarbeitung bedingen, beispielsweise das

Vorwissen, Motivationen und Emotionen (Greifeneder et al., 2018), in Befragungen mit erhoben und in Analysen miteinbezogen werden. Insbesondere in Bezug auf Forschung zu möglichen Prädiktoren für die Unterrichtsqualität aus Sicht von Lehrkräften wird zu wenig differenziert, ob die abhängige Variable, die Unterrichtsqualität, über Wahrnehmungen von Lernenden oder Lehrkräften erfasst wurde. Methodische Einflussvariablen sollten weitestgehend vermieden werden und vorab mitgedacht werden. So sollte beispielsweise bei der Erfassung der Wahrnehmungen externer Urteilender hinsichtlich der Unterrichtsqualität auf die Reihenfolge geachtet werden, in der Unterrichtsvideos präsentiert werden.

Neue Ansätze wie Erhebungen mittels EEG, Eye-Tracking oder Analysen des Unterrichtsmaterials bieten Chancen für die Erfassung von Unterrichtsqualität. Allerdings müssen auch für diese neuen Methoden mögliche Einflussfaktoren erst untersucht werden.

Bislang wird in theoretischen Modellen und auch hier im Untersuchungsmodell der Arbeit die Unterrichtsqualität als eigener Faktor dargestellt, welcher beispielweise von Lernenden wahrgenommen sowie genutzt werden muss und im Zusammenhang mit Lehrkraftmerkmalen steht (z. B. Angebots-Nutzungs-Modell nach Helmke, 2022; Wirkungsgeflecht der Klassenführung nach Helmke & Helmke, 2015; Grunddimensionen der Unterrichtsqualität und deren vermutliche Wirkung nach Klieme & Rakoczy, 2008; Untersuchungsmodell von Lepper, 2022). Das unterrichtliche Angebot mit seiner Qualität steht im Zentrum der Modelle und wird als gegeben angenommen (auch Vieluf, Praetorius, Rakoczy, Kleinknecht & Pietsch, 2020). Wenn die Übereinstimmungen zwischen Perspektiven hinsichtlich der Unterrichtsqualitätsdimensionen gering bis moderat ausfallen, insbesondere für die Dimensionen kognitive Aktivierung und unterstützendes Unterrichtsklima, und auch Lernende, Lehrkräfte und externe Urteilende aufgrund von individuellen Dispositionen in ihren Wahrnehmungen divergieren, ist es fragwürdig, warum immer von dem einen qualitativ guten Unterricht ausgegangen wird. Hinsichtlich der Unterrichtsqualitätsdimension Klassenführung wurden bislang die höchsten Übereinstimmungen und auch die wenigsten Einflussfaktoren für die Wahrnehmung gefunden. In dem vereinfachten Angebots-Nutzungsmodell von Lipowsky (2006) wird die Unterrichtsqualität verstanden als Zusammenspiel vom Lehrkrafthandeln, welches eine gewisse Quantität und Qualität der Lernangebote darbietet, und dem Lernendenhandeln, welches daraus besteht, das Lernangebot wahrzunehmen und zu nutzen. Somit kann nach Lipowsky (2006) die Unterrichtsqualität für einzelne Kinder auch anders ausfallen, da sie entweder auch eine andere Qualität und Quantität des Lernangebots durch die Lehrkraft

erhalten und diese auch individuell wahrnehmen und nutzen. In zukünftigen Modellen sollte die Darstellung der Unterrichtsqualität mitbedacht werden.

Eine pädagogische Antwort auf die wachsende Heterogenität von Schulklassen ist seit einigen Jahren die Forderung nach einer stärkeren Anpassung des Unterrichts an die individuellen Voraussetzungen und Interessen der Lernenden (Lipowsky & Lotz, 2015). Strategien im Umgang mit der steigenden Heterogenität sind die Innere Differenzierung, Individualisierung, Adaptives Unterrichten, Offene Unterrichtsformen und Entdeckendes Lernen (ebd.). Der gemeinsame Ausgangspunkt aller Strategien ist, dass das Individuum im Fokus steht und das einzelne Kind durch angepasste Förderung bestmögliche Lernfortschritte erreicht sowie auch die Persönlichkeitsentwicklung und motivational-emotionale Merkmale gefördert werden sollen. Beispielsweise verfolgt der Ansatz des Offenen Unterrichts die Idee, dass die lehrkraftgelenkten Unterrichtsphasen durch mehr geöffnete Unterrichtsphasen ersetzt werden sollen und Lernende somit die Möglichkeit erhalten, die Selbst-, Sach- und Sozialkompetenz mittels eigenen Handelns und Konstruieren zu entwickeln (Peschel, 2002, 2020). Demnach sollte die Unterrichtsqualitätsforschung stärker die Passung zwischen den Bedürfnissen der Lernenden und dem individualisierten Angebot von Lehrkräften in den Blick nehmen. Bisher zielten die Itemformulierungen bei der Befragung von Lernenden, Lehrkräften und externen Urteilenden eher noch auf lehrkraftgelenkte Unterrichtsphasen ab (Beispiel für Lernende: „Schülerinnen und Schüler auffordern, laut vorzulesen.“ IGLU, McElvany et al., 2023). Werden zunehmend offenere Unterrichtskonzepte eingesetzt, müssen die Itemformulierungen für die Befragungen für die Wahrnehmung von Unterrichtsqualität angepasst werden; sie sollten sich entweder konkret auf einzelne Unterrichtsmethoden beziehen oder allgemeingültig und methodenunabhängig formuliert sein. Bei aggregierten Wahrnehmungen von Lernenden sollte zukünftig stärker diskutiert werden, welche Aussagen aufgrund von den individuellen Abweichungen der Lernenden überhaupt getroffen werden können. Mit dem Trend der immer offeneren Unterrichtskonzepte und dem Fokus auf individuelle Förderung sowie Berücksichtigung individueller Bedürfnisse, werden auch Aussagen von Lehrkräften und externen Urteilenden für einen guten Unterricht für alle Kinder der Klasse schwieriger zu treffen sein.

### **8.3.2 Implikationen für die Praxis**

Der Beitrag IV (weiterführende Analysen) zeigte, dass Lehrkräfte ihr Angebot an Motivierungsunterstützung abweichend von den Wahrnehmungen ihrer Lernenden

wahrnahmen. Lehrkräfte sollten häufiger ein Feedback einholen, damit sie die Sicht der Lernenden zukünftig genauer einschätzen können. Hierfür ist die Voraussetzung, dass Lehrkräfte das Feedback von Lernenden als bedeutsam erachten und mit Lernenden ins Gespräch kommen. Dies ist für die Weiterentwicklung der Professionalität von Lehrkräften bedeutsam (Zierer, Wisniewski, Schatz, Weckend & Helmke, 2019). Positive Motivation und Haltung der Lehrkräfte können dazu führen, dass Lehrkräfte regelmäßige Dialoge mit ihren Lernenden führen und diese als Rückmeldung zur Gestaltung des eigenen Unterrichts nutzen. Sie sehen Lernende als wichtige Ansprechpartner\*innen für die Spiegelung ihrer eigenen Wirksamkeit (ebd.).

Feedback von Lernenden bezüglich des Lernstandes einzuholen ist bedeutsam, um den Unterricht individuell an die Bedürfnisse der Lernenden anzupassen. Einen Ansatz für den an den individuellen Bedürfnissen angepassten Unterricht bietet das Response-to-Intervention-Modell. Auf der Basis von Lern-/Entwicklungsverlaufsdiagnostik werden geeignete Fördermaßnahmen ausgewählt, um Lernschwierigkeiten bei Lernenden zu verhindern und die Lernenden bestmöglich zu fördern. Das Modell basiert darauf, dass Lernprobleme durch die Veränderung der äußeren Bedingungen verringert werden können. Lernschwierigkeiten werden nicht als individuelle Störung angesehen und werden als Nichtpassung zwischen den individuellen Lernvoraussetzungen sowie Bedürfnissen und dem unterrichtlichen Angebot verstanden. Eine Forderung des Modells ist, dass Unterricht kontinuierlich und professionell an die Voraussetzungen und die aktuellen Leistungen der Lernenden angepasst werden muss. Bedeutsam für die Förderung und die Anpassung des Unterrichts ist das Antwortverhalten der Lernenden (Response-to-Intervention) (Huber & Grosche, 2012; Kuhl & Hecht, 2014). Demnach legt das Modell einen Fokus auf die effektive Förderung aller und zielt damit auch auf einen qualitätsvollen Unterricht. Zu kritisieren ist aber die nach der Leistung vorgenommenen Hierarchisierung im Modell. Da die Beiträge der vorliegenden kumulativen Dissertation auch zeigten, dass die Motivierungsqualität und die diagnostische Kompetenz für den Lernerfolg von Lernenden wichtig sind, sollte das Modell um die Betrachtung von motivational-emotionalen Merkmalen von Lernenden erweitert und übertragen werden.

Nach den Ergebnissen der Arbeit und im Hinblick auf das Response-to-Intervention Modell sollte die Rolle der Lehrkraft beinhalten, den Unterricht an individuelle Bedürfnisse anzupassen und zu überprüfen, wie ihre Lernenden ihr unterrichtliches Angebot wahrnehmen, um ihren Unterricht und ihr Förderangebot stetig an die Lernenden anzupassen. Für eine erfolgreiche Umsetzung, so konnten es auch die Ergebnisse des Beitrags II bestätigen, sind

die Motivationen von Lehrkräften von besonderer Bedeutung. Die Motivation bezüglich der Diagnostik hängt positiv mit der Häufigkeit der Differenzierung im Unterricht und mit der Motivierung der Lernenden zusammen. Doch als Lehrkraft motiviert zu bleiben, ist angesichts der besonderen, sich stets verändernden Anforderungen nicht einfach. Die Bildungspraktiken haben sich in den letzten Jahren erheblich verändert und fordern eine stetige Anpassung der Lehrkräfte; von stark strukturierten Klassenräumen und Unterrichtsmethoden hin zu mehr Flexibilität sowie Wahlmöglichkeiten (Carpenter, Witherby & Tauber, 2020) und zugleich eine wachsende Heterogenität der Klassen. Allerdings gaben (angehende) Lehrkräfte an, ungenügend auf diese Herausforderungen vorbereitet zu sein (Gold, Pfirrmann & Holodynski, 2021; McMahon, Forde & Dickson, 2015; O'Neill & Stephenson, 2012; Wolff, Jarodzka & Boshuizen, 2021). Demzufolge muss die Aus- und Weiterbildung der (angehenden) Lehrkräfte fortwährend an die geforderten Kompetenzen der Lehrkräfte angepasst werden, damit diese einen qualitativ wertvollen und für ihre Lernenden abgestimmten Unterricht anbieten können.

## **8.4 Ausblick**

Schule, Bildung und damit Unterricht sind stetig im Wandel. Die Entwicklung orientiert sich dabei an politischen Bestimmungen sowie an den gesellschaftlichen Entwicklungen (hierzu schon Herzog, 1994; Petko, Döbeli Honegger & Prasse, 2018). Das unterrichtliche Angebot hat sich immer wieder neu an die Bedürfnisse der Schüler\*innenschaft anzupassen. Somit ergeben sich auch neue Merkmale, welche bei der Einschätzung der Unterrichtsqualität berücksichtigt werden müssen. Schlagworte, welche gerade die Schulentwicklung bestimmen sind die Inklusion und Zuwanderung (Zusammengefasst als Umgang mit Heterogenität), Digitalisierung und Fernunterricht bzw. Blended Learning. Wird sich die Qualität des Unterrichts auch unter solchen Gesichtspunkten angeschaut, dann bleibt das Gerüst, welches die Angebots-Nutzungs-Modelle liefern. Auch in Bezug auf Umgang mit Heterogenität, Digitalisierung und Fernunterricht bleibt das unterrichtliche Angebot seitens der Lehrkraft, welches durch die Wahrnehmung und Nutzung von Lernenden zum Lernerfolg führt. Hierbei beeinflussen sowohl Merkmale der Gesellschaft, Schule, Lehrkräfte, Lernenden, Familien und Peers den Lehr-Lern-Prozess. Allerdings verändert sich das unterrichtliche Angebot und auch Einflussmerkmale, welche Lernenden zum Lernerfolg verhelfen. Zum Beispiel rückten national durch die Ratifizierung der UN-Behinderterrechtskonvention im Jahre 2009 die Adaptivität und Individualisierung als wichtige Unterrichtsqualitätsmerkmale in den Fokus

(Schröer, 2013). Aufgrund der Anpassung von Unterricht an die gesellschaftlichen Bedürfnisse bleibt die Forschungsfrage nach gutem Unterricht bestehen und muss immer wieder neu geprüft werden. Damit verbunden sind die Wahrnehmungen des unterrichtlichen Angebots und die Frage nach (individuellen und methodischen) Merkmalen, welche die Wahrnehmung der sich wandelnden Unterrichtsqualitätsmerkmale bedingen. Klieme (2020) beschreibt, dass alle drei generischen Unterrichtsqualitätsdimensionen (Kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima) unabhängig vom methodischen Arrangement sowie vom gewählten Inhalt bedeutsam sind und demnach auch beispielsweise im Fernunterricht gelten. Die Ausgestaltung der Unterrichtsqualitätsdimensionen ist allerdings abhängig vom Arrangement sowie vom Inhalt und muss angepasst werden.

Forschungsfragen, welche in Zukunft näher beleuchtet werden müssen, sind beispielsweise welche Unterrichtsqualitätsmerkmale relevant für verschiedene Situationen sind, wie Lernende mit Einschränkungen oder Migrationshintergrund den Unterricht wahrnehmen und welche individuellen Merkmale ihre Wahrnehmung und somit auch ihre Nutzung von Lernangeboten positiv beeinflussen. Erste Studien gibt es zwar schon in diesen Bereichen (z. B. Unterrichtsqualität und Heterogenität: Moser Opitz, Maag Merki, Pfaffhauser, Stöckli & Garrote, 2021; Finkelstein, Sharma & Furlonger, 2021; Unterrichtsqualität und Digitalisierung: Perrin & Wang, 2021; Quast, Rubach & Lazarides, 2020; Unterrichtsqualität und Fernunterricht: Klieme, 2020; Wittwer & Voss, 2022). Allerdings bleibt ein Bedarf an der Weiterentwicklung von theoretischen Modellen und Belegung durch empirische Untersuchungen.

## Literaturverzeichnis II

- Ahn, I., Chiu, M. M. & Patrick, H. (2021). Connecting teacher and student motivation: Student-perceived teacher need-supportive practices and student need satisfaction. *Contemporary Educational Psychology*, 64, 101950. <https://doi.org/10.1016/j.cedpsych.2021.101950>
- Algina, J. & Olejnik, S. (2000). Determining Sample Size for Accurate Estimation of the Squared Multiple Correlation Coefficient. *Multivariate Behavioral Research*, 35, 119–136.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.
- Begrich, L., Fauth, B. & Kunter, M. (2020). Who sees the most? Differences in students' and educational research experts' first impressions of classroom instruction. *Social Psychology of Education*, 23, 673–699. <http://doi.org/10.1007/s11218-020-09554-2>
- Brophy, J. E. & Good, T. L. (1976). *Die Lehrer-Schüler-Interaktion*. München: Urban & Schwarzenberg.
- Bru, E., Stornes, T., Munthe, E. & Thuen, E. (2010). Children's perceptions of teacher support across the transition from primary to secondary school. *Scandinavian Journal of Educational Research*, 54(6), 519–533. <https://doi.org/10.1080/00313831.2010.522842>
- Burić, I. & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning Instruction*, 66, 101302. <https://doi.org/10.1016/j.learninstruc.2019.101302>
- Carpenter, S. K., Witherby, A. E. & Tauber, S. K. (2020). On students' (mis)judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition*, 9(2), 137–151. <https://doi.org/10.1016/j.jarmac.2019.12.009>
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.

- De Jong, R. & Westerhof, K. J. (2001). The quality of student ratings of teacher behavior. *Learning Environments Research*, 4, 51–85.
- Espinoza, A. M. & Strasser, K. (2020). Is reading a feminine domain? The role of gender identity and stereotypes in reading motivation in Chile. *Social Psychology of Education*, 23(4), 861–890.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2016). Urteile zur Unterrichtsqualität von Grundschüler\_innen: Mehr als „Ich mag meine Lehrerin“? In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler & R. Schmidt (Hrsg.), *Facetten grundschulpädagogischer und grundschuldidaktischer Forschung*. Wiesbaden: Springer.
- Fauth, B., Göllner, R., Lenske, G., Praetorius, A.-K. & Wagner, W. (2018). *Warum (und inwiefern) ist Unterrichtsqualität eine Frage der Perspektive? Auf dem Weg zu einem besseren Verständnis perspektivenspezifischer Unterrichtsbeurteilungen (Symposiumsbeitrag)*. Basel, GEBF. Verfügbar unter <https://unterrichtsforschung.dipf.de/de/arbeitsergebnisse/tagungsbeitraege/beitraege-als-pdf/geb-2018-fauth-gollner-lenske-praetorius-wagner.pdf>
- Finkelstein, S., Sharma, U. & Furlonger, B. (2021) The inclusive practices of classroom teachers: a scoping review and thematic analysis. *International Journal of Inclusive Education*, 25(6), 735–762. <https://doi.org/10.1080/13603116.2019.1572232>
- Gabriel-Busse, K., Groß-Mlynek, L., Feldhoff, T. & Haring, M. (2020). Eine Unterrichtssequenz – unterschiedliche Einschätzungen. Analyse videografiert Unterrichtssequenzen als Bestandteil einer evidenzbasierten Lehrer/innenausbildung. In I. Gogolin, B. Hannover & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung (ZfE-Edition, Bd. 4, S. 291–314)*. Wiesbaden: Springer.
- Gabriel-Busse, K. & Lipowsky, F. (2021). 90 Minuten Mathematikunterricht bei gleichbleibender Unterrichtsqualität? Analysen zur zeitlichen Stabilität und Generalisierbarkeit von Ratings zur Unterrichtsqualität im 2. Schuljahr. *Unterrichtswissenschaft*, 49, 137–163. <https://doi.org/10.1007/s42010-020-00086-4>
- Geitz, E., Vater, C. & Zimmer-Merkle, S. (2020). *Black Boxes–Versiegelungskontexte und Öffnungsversuche Interdisziplinäre Perspektiven* (p. 418). Berlin: De Gruyter. <https://doi.org/10.1515/9783110701319>

- Gold, B., Pfirrmann, C. & Holodynski, M. (2021). Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *Journal of Teacher Education*, 72(4), 31–447.
- Götz, T., Lüdtke, O., Nett, U. E., Keller, M. M. & Lipnevich, A. A. (2013). Characteristics of teaching and students' emotions in the classroom: investigating differences across domains. *Contemporary Educational Psychology*, 38, 383–394. <https://doi.org/10.1016/j.cedpsych.2013.08.001>
- Greifeneder, R., Bless, H. & Fiedler, K. (2018). *Social cognition: How individuals construct social reality*. New York: Routledge.
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung. Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Seelze: Klett-Kallmeyer.
- Helmke, A. & Helmke, T. (2015). Wie wirksam ist gute Klassenführung? Effiziente Klassenführung ist nicht alles, aber ohne sie geht alles andere gar nicht. *Pädagogik-Leben*, 7–11. Verfügbar unter [https://bildung-rp.de/fileadmin/user\\_upload/p\\_files/Materialien/PL\\_Publikationen/15\\_2\\_PL/P\\_L\\_2-2015\\_S.6-11.pdf](https://bildung-rp.de/fileadmin/user_upload/p_files/Materialien/PL_Publikationen/15_2_PL/P_L_2-2015_S.6-11.pdf)
- Helmke, A. & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98.
- Herbert, B. & Schweig, J. (2021). Erfassung des Potenzials zur kognitiven Aktivierung über Unterrichtsmaterialien im Mathematikunterricht. *Zeitschrift für Erziehungswissenschaft*, 24(4), 955–983. <https://doi.org/10.25656/01:25456>
- Herzog, W. (1994). Gesellschaftlicher Wandel und schulische Autonomie - Erneuerung der Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerbildung*, 12(2), 206–223. <https://doi.org/10.25656/01:13285>
- Holzberger, D., Philipp, A. & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality. A longitudinal analysis. *Journal of Educational Psychology*, 105(3), 774–786.
- Huber, C. & Grosche, M. (2012). Das Response-to-Intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 8, 312–322.

- Jandl, S. (2016). *Mathematikspezifisches Wissen von Sonderpädagoginnen und Sonderpädagogen. Eine empirische Studie zur Entwicklung, Evaluation und dem Einsatz eines Befragungsinstruments*. [Dissertation Universität Zürich]. Verfügbar unter <https://www.zora.uzh.ch/id/eprint/133521/1/Jandl.pdf>
- Jennings, P. A. & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, 79(1), 491–525. <https://doi.org/10.3102/0034654308325693>
- Klieme, E. (2020). Guter Unterricht – auch und besonders unter Einschränkungen der Pandemie? – In D: Fickermann & B. Edelstein [Hrsg.], *Langsam vermisse ich die Schule ...". Schule während und nach der Corona-Pandemie* (S. 117–135). Münster; New York: Waxmann. <https://doi.org/10.25656/01:20233>
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule, Abschlussbericht des DFG-Schwerpunktprogramms* (S. 128–146). Münster: Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Kloss, J. (2013). Grundschüler als Experten für Unterricht – Empirische Überprüfung der Validität von Unterrichtsbeurteilungen durch Schüler der dritten und vierten Jahrgangsstufe. In E. Witruk, *Beiträge zur Pädagogischen und Rehabilitationspsychologie* (Volume 6). Frankfurt am Main: Peter Lang.
- Kuhl, J. & Hecht, T. (2014). Prävention von Lernschwierigkeiten durch die Implementierung von Diagnostik und Förderung - Ein Praxisbeispiel für das erste Schuljahr. *Zeitschrift für Heilpädagogik*, 65, 406–415.
- 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.

- Labsch, A., Schmitt, M. & Schüpbach, M. (2023). Nehmen Schüler\* innen ohne sonderpädagogische Förderbedarfe in inklusiven Klassen die Beziehung zu ihren Lehrkräften anders wahr als ihre Peers in nicht-inklusive Klassen? *Zeitschrift für Pädagogik*, 6, 811–831.
- Largo, R. H. & Beglinger, M. (2009). *Schülerjahre: Wie Kinder besser lernen* (4th ed.). München: Piper.
- Lazarides, R. & Schiefele, U. (2021). The relative strength of relations between different facets of teacher motivation and core dimensions of teaching quality in mathematics—a multilevel analysis. *Learning and Instruction*, 76, 101489.
- Lenske, G. (2016). *Schülerfeedback in der Grundschule. Untersuchungen zur Validität*. Münster: Waxmann.
- Lepper, C. (2022). *Motivationsunterstützende Unterrichtsgestaltung in der Grundschule – Unterrichtsqualität und Textmerkmale im Fokus*. Veröffentlichte Dissertation. Verfügbar unter [https://eldorado.tu-dortmund.de/bitstream/2003/40746/1/Dissertation\\_Lepper.pdf](https://eldorado.tu-dortmund.de/bitstream/2003/40746/1/Dissertation_Lepper.pdf)
- Lepper, C., Stang, J. & McElvany, N. (2021). Bedeutung der wahrgenommenen Motivierungsqualität für intrinsische Motivation und Selbstkonzept von Grundschulkindern. *Unterrichtswissenschaft*, 50, 125–147. <https://doi.org/10.1007/s42010-021-00121-y>
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenz und Kompetenzentwicklung von Lehrerinnen und Lehrern*. *Zeitschrift für Pädagogik*, 51, Beiheft (S. 47–70). Weinheim: Beltz.
- Lipowsky, F. & Lotz, M. (2015). Ist Individualisierung der Königsweg zum erfolgreichen Lernen? Eine Auseinandersetzung mit Theorien, Konzepten und empirischen Befunden. In G. Mehlhorn, K. Schöppe & F. Schulz (Hrsg.), *Begabungen entwickeln & Kreativität fördern* (S. 155–219). München: kopaed.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings in multilevel modelling. *Contemporary Educational Psychology*, 34, 120–131.

- Martin, A. J., Papworth, B., Ginns, P., Malmberg, L. E., Collie, R. J. & Calvo, R. A. (2015). Realtime motivation and engagement during a month at school: Every moment of every day for every student matters. *Learning and Individual Difference*, 38, 26–35. <https://doi.org/10.1016/j.lindif.2015.01.014>
- Mashburn, A. J., Meyer, J. P., Allen, J. P. & Pinata, R. C. (2014). The Effect of Observation Length and Presentation Order on the Reliability and Validity of an Observational Measure of Teaching Quality. *Educational and Psychological Measurement*, 74(3), 400–422.
- McElvany, N., Lorenz, R., Frey A., Goldhammer, F., Schilcher, A. & Stubbe, T. C. (Hrsg.) (2023). *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre*. Münster: Waxmann.
- McMahon, M., Forde, C. & Dickson, B. (2015). Reshaping teacher education through the professional continuum. *Educational Review*, 67(2), 158–178. <https://doi.org/10.1080/00131911.2013.846298>
- Moser Opitz, E., Maag Merki, K., Pfaffhauser, R., Stöckli, M. & Garrote, A. (2021). Die Wirkung von unterschiedlichen Formen von co-teaching auf die von den Schülerinnen und Schülern wahrgenommene Unterrichtsqualität in inklusiven Klassen. *Unterrichtswissenschaft*, 49(3), 443–466. <https://doi.org/10.1007/s42010-021-00103-0>
- Muthén, L. K. & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620.
- Nowicki, E. A. & Lopata, J. (2017). Children’s implicit and explicit gender stereotypes about mathematics and reading ability. *Social Psychology of Education*, 20, 329–345. <https://doi.org/10.1007/s11218-015-9313-y>
- O’Neill, S. & Stephenson, J. (2012). Does classroom management coursework influence pre-service teachers’ perceived preparedness or confidence? *Teaching and Teacher Education*, 28(8), 1131–1143.
- Patall, E. A., Steingut, R. R., Vasquez, A. C., Trimble, S. S., Pituch, K. A. & Freeman, J. L. (2018). Daily autonomy supporting or thwarting and students’ motivation and engagement in the high school science classroom. *Journal of Educational Psychology*, 110(2), 269–288. <https://doi.org/10.1037/edu0000214>

- Perrin, S. & Wang, L. (2021). COVID-19 and rapid digitalization of learning and teaching: Quality assurance issues and solutions in a Sino-foreign higher education institution. *Quality assurance issues and solutions*, 29, 463–476.
- Peschel, F. (2002). Qualitätsmaßstäbe - Hilfen zur Beurteilung der Offenheit von Unterricht. In U. Drews & W. Wallrabenstein, *Freiarbeit in der Grundschule. offener Unterricht in Theorie, Forschung und Praxis* (S. 160–171). Frankfurt am Main: Grundschulverband - Arbeitskreis Grundschule e.V. <https://doi.org/10.25656/01:17637>
- Peschel, F. (2012). Individualisierung, Inklusion und Offener Unterricht. Missverständnisse, Fallstricke und Teufelskreise. In M. Balliet & U. W. Kliebisch (Hrsg.), *LehrerHandeln. Kompetent, effizient, kongruent* (S. 124–139). Baltmannsweiler: Schneider Verlag.
- Peschel, F. (2021). Offener Unterricht. Idee - Realität - Perspektive und ein praxiserprobtes Konzept zur Diskussion. Teil I: Allgemeindidaktische Überlegungen (11., unveränd. Neuaufl.). Baltmannsweiler: Schneider Verlag
- Petko, D., Döbeli Honegger, B. & Prasse, D. (2018). Digitale Transformation in Bildung und Schule: Facetten, Entwicklungslinien und Herausforderungen für die Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 36(2), 157–174 <https://doi.org/10.25656/01:17094>
- Pieper, V. (2023). *(Lehrkraft-)Feedback im Klassenraum – Eine qualitative Videostudie zu Interaktionen von Lehrkräften und Schüler\*innen im Mathematikunterricht vierter Grundschulklassen in Niedersachsen*. Veröffentlichte Dissertation. Verfügbar unter <https://voado.uni-vechta.de/bitstream/handle/21.11106/469/Vanessa%20Pieper%20Dissertation%202023.pdf?sequence=2&isAllowed=y>
- Pietsch, M. & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Praetorius, A.-K. & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? *Unterrichtswissenschaft*, 49, 167–188. <https://doi.org/10.1007/s42010-021-00119-6>

- Praetorius, A.-K., Lauer mann, F., Klassen, R. M., Drexler, K., Nitsche, S., Janke, S., Dickhäuser, O. & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education*, 65, 241–254.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Quast, J., Rubach, C. & Lazarides, R. (2021). Lehrkräfteeinschätzungen zu Unterrichtsqualität mit digitalen Medien: Zusammenhänge zur wahrgenommenen technischen Schulausstattung, Medienunterstützung, digitalen Kompetenzselbsteinschätzungen und Wertüberzeugungen. *Zeitschrift für Bildungsforschung*, 11(2), 309–341. <http://dx.doi.org/10.1007/s35834-021-00313-7>
- Rakoczy, K. (2008). Motivationsunterstützung im Mathematikunterricht. Münster: Waxmann
- Rakoczy, K., Frick, U., Weiß-Wittstadt, S., Tallon, M. & Wagner, E. (2022). Einmal begeistert, immer begeistert? Eine Experience-Sampling Studie zur wahrgenommenen Unterrichtsqualität und Motivation von Schülerinnen und Schülern im Kunstunterricht. *Unterrichtswissenschaft*, 50, 211–236. <https://doi.org/10.1007/s42010-022-00147-w>
- Rosemann, B. (1978). Bedingungsvariablen der Lehrer-Schüler-Beziehung: Erwartungskonkordanz und das Verhalten von Lehrern und Schülern. *Psychologie in Erziehung und Unterricht*, 25, 39–49.
- Rosenshine, B. (1970). Evaluation of Classroom Instruction. *Review of Educational Research*, 40, 279–300.
- Roth, G., Assor, A., Kanat-Maymon, Y. & Kaplan, H. (2007). Autonomous motivation for teaching. How self-determined teaching may lead to self-determined learning. *Journal of Educational Psychology*, 99(4), 761–774.
- Rubin, K.H., Bukowski, W.M. & Parker, J.G. (1998). Peer interactions, relationships, and groups. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, pp. 619–700). New York, NY: Wiley.

- Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Sakiz, G., Pape, S. J. & Hoy, A. W. (2012). Does perceived teacher affective support matter for middle school students in mathematics classrooms? *Journal of School Psychology*, 50, 235–255. <https://doi.org/10.1016/j.jsp.2011.10.005>
- Schmidt-Atzert, L., Büttner, G. & Bühner, M. (2004). Theoretische Aspekte von Aufmerksamkeits-/ Konzentrationsdiagnostik. In G. Büttner & L. Schmidt-Atzert (Hrsg.), *Diagnostik von Konzentration und Aufmerksamkeit* (S. 3–22). Göttingen: Hogrefe.
- Schröer, H. (2013). Inklusion versus Integration – *Zauberformel oder neues Paradigma?* *Migration und soziale Arbeit*, 35, 249–255.
- Schütte, K., Rose, H. & Köller, O. (2022). Frühkindliche Entwicklung stärken. Eine Zukunftsallianz aus Fachpraxis, Wissenschaft, Verwaltung, Politik und Zivilgesellschaft. Münster: Waxmann. <https://doi.org/10.25656/01:24576>
- Schwabe, F., Schlitter, T., Iglar, J., Ohle-Peters, A., Teerling, A., Köller, O. & McElvany, N. (2020). Lesemotivation, Leseselbstkonzept und Leseverhalten am Ende der Grundschulzeit – Wirksamkeit und differenzielle Effekte der schulischen Teilnahme an einer bundesweiten Förderinitiative. *Zeitschrift für pädagogische Psychologie*, 1, 1–18.
- Schweer, M. K. W., Thies, B. & Lachner, R. P. (2017). Soziale Wahrnehmungsprozesse und unterrichtliches Handeln. Eine dynamisch transaktionale Perspektive. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 121–145). Wiesbaden: Springer VS.
- Urhahne, D. & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, 32. 100374. <https://doi.org/10.1016/j.edurev.2020.100374>
- Urhahne, D. & Zhu, M. (2015). Teacher judgement and student motivation. In C. M. Rubie-Davies, J. M. Stephens & P. Watson (Eds.), *International Handbook of Social Psychology of the Classroom* (pp. 304–315). London, UK: Routledge.

- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts – Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift für Pädagogik, Beiheft*, 66, 63–80.
- Vogelsang, C. & Reinhold, P. (2013). Gemessene Kompetenz und Unterrichtsqualität. Überprüfung der Validität eines Kompetenztests mit Hilfe der Unterrichtsvideographie. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken. Fachdidaktische Forschungen* (Band 4) (pp. 319–334). Münster: Waxmann.
- 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(5), 705–721.
- Wettstein, A., Ramseier, E., Scherzinger, M. & Gasser, L. (2016). Unterrichtsstörungen aus Lehrer- und Schülersicht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48, 171–183.
- Willems, A. S. (2022). Individuelle Schüler\*innenprofile des situationalen und dispositionalen Interesses und ihre Bedeutung für die Wahrnehmung der Unterrichtsqualität im Fach Mathematik. *Zeitschrift für Erziehungswissenschaft*, 25(2), 377–404.
- Wittwer, J. & Voss, T. (2022). Kognitive Aktivierung im Unterricht unter Pandemiebedingungen aus instruktionspsychologischer Sicht. In M. Schambeck & W. Verbarg (Hrsg.), *Wie Religion für Krisen taugt – Zum Beitrag religiöser Bildung in Krisenzeiten* (S.116–135). Vandenhoeck Ruprecht. <https://doi.org/10.13109/9783666703294.116>
- Wolf, L. M. & Dietze, T. (2022). Ein Überblick über die Organisation der Feststellung von sonderpädagogischen Förderbedarfen in Deutschland. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 325–344). Regensburg: Universitätsbibliothek.

- Wolff, C. E., Jarodzka, H. & Boshuizen, H. P. (2021). Classroom management scripts: A theoretical model contrasting expert and novice teachers' knowledge and awareness of classroom events. *Educational Psychology Review*, 33, 131–148.
- Zee, M., Bree, E. de, Hakvoort, B. & Koomen, H. (2020). Exploring relationships between teachers and students with diagnosed disabilities: A multi-informant approach. *Journal of Applied Developmental Psychology*, 66, 101101. <https://doi.org/10.1016/j.appdev.2019.101101>
- Zierer, K., Wisniewski, B., Schatz, C., Weckend, D. & Helmke, A. (2021). Wie kann Feedback der Lernenden die Unterrichtsqualität verbessern? *Journal für LehrerInnenbildung*, 19(1), 26–40. <https://doi.org/10.25656/01:18108>

## 9 Anhang

### 9.1 Liste der Einzelbeiträge

Veröffentlichte Beiträge der kumulativen Dissertation

Beitrag I. **Igler, J.**, Ohle-Peters, A. & McElvany, N. (2019). Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität. *Zeitschrift für Pädagogische Psychologie*, 33, 191–205.  
<https://doi.org/10.1024/1010-0652/a000243>

Beitrag II. **Igler, J.**, Ohle-Peters, A., Schlitter, T., Teerling, A., Asseburg, R., McElvany, N. & Köller, O. (2018). Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht. In F. Schwabe, N. McElvany, W. Bos & H. G. Holtappels (Hrsg.), *Jahrbuch der Schulentwicklung Band 20. Schule und Unterricht in gesellschaftlicher Heterogenität* (S. 58–81). Weinheim: Beltz Juventa.

Beitrag III. **Igler, J.**, Ohle-Peters, A. & McElvany, N. (2022). Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt. *Zeitschrift für Erziehungswissenschaft*, 25, 1431–1453. <https://doi.org/10.1007/s11618-022-01121-z>

Beiträge zu weiterführenden Analysen

Beitrag IV. **Igler, J.**, Ohle-Peters, A., Teerling, A., Köller, O. & McElvany, N. (unpublished). Students' perceived motivational support in elementary school – How accurate are teachers' judgments?

### 9.2 Eigenanteile der Doktorandin bei den Einzelarbeiten

In Bezug auf die Stellungnahme der DGfE zu Publikationsbasierten Dissertationen in der Erziehungswissenschaft (Thole et al., 2011<sup>2</sup>) werden für jeden einzelnen Beitrag die

---

<sup>2</sup> Thole, W., Lohmann, I., Aufenanger, S., Breuer, K., Hascher, T., Müller, H.-R. & Reh, S. (2011). Publikationsbasierte Dissertationen in der Erziehungswissenschaft Stellungnahme der DGfE. Verfügbar unter: [https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2011\\_Publikationsbasierte\\_Promotionen.pdf](https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2011_Publikationsbasierte_Promotionen.pdf)

Eigenanteile der Doktorandin dargestellt. Hierbei werden die Aspekte (i) Konzeption der schriftlichen Abfassung des Beitrags, (ii) Statistische Auswertungen, (iii) Schriftliche Abfassung des Beitrags, (iv) Revision nach Rückmeldung der Co-Autorinnen und (v) Revision nach Rückmeldung der Gutachter\*innen beschrieben.

## **Beitrag I**

**Igler, J.,** Ohle-Peters, A. & McElvany, N. (2019). Mit den Augen eines Grundschulkindes. Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität. *Zeitschrift für Pädagogische Psychologie*, 33(3-4), 191–205. <https://doi.org/10.1024/1010-0652/a000243>

### *Konzeption der schriftlichen Abfassung des Beitrags*

Jennifer Igler hat die schriftliche Abfassung hauptverantwortlich konzipiert. Hierbei waren die Co-Autorinnen beratend tätig.

### *Statistische Auswertungen*

Die statistischen Auswertungen wurden hauptverantwortlich von Jennifer Igler durchgeführt, wobei Dr. Matthias Trendtel beratend tätig war.

### *Schriftliche Abfassung des Beitrags*

Der Text wurde vollständig von Jennifer Igler verfasst.

### *Revision nach Rückmeldung der Co-Autorinnen*

Anhand von Anmerkungen beider Co-Autorinnen hat Jennifer Igler den Text überarbeitet und finalisiert.

### *Revision nach Rückmeldung der Gutachter\*innen*

Der Text wurde nach Erhalt der Anmerkungen und Änderungsvorschläge der Gutachter\*innen hauptverantwortlich von Jennifer Igler überarbeitet. Die Co-Autorinnen waren beratend tätig und gaben Feedback zu den Überarbeitungen.

## Beitrag II

**Igler, J.,** Ohle-Peters, A., Schlitter, T., Teerling, A., Asseburg, R., McElvany, N. & Köller, O. (2018). Bedeutung motivationaler Lehrkraftmerkmale und der Beteiligung an innovativen Programmen für die Qualität von Leseunterricht. In F. Schwabe, N. McElvany, W. Bos & H. G. Holtappels (Hrsg.), *Jahrbuch der Schulentwicklung Band 20. Schule und Unterricht in gesellschaftlicher Heterogenität* (S. 58–81). Weinheim: Beltz Juventa.

### *Konzeption der schriftlichen Abfassung des Beitrags*

Die schriftliche Abfassung wurde hauptverantwortlich von Jennifer Igler konzipiert. Die Co-Autor\*innen waren dabei beratend tätig.

### *Statistische Auswertungen*

Die statistischen Auswertungen wurden hauptverantwortlich von Jennifer Igler durchgeführt. Die Co-Autor\*innen fungierten als Berater\*innen.

### *Schriftliche Abfassung des Beitrags*

Jennifer Igler verfasste den Text des Beitrages hauptverantwortlich. Sie erhielt dazu Rückmeldungen von den Co-Autor\*innen.

### *Revision nach Rückmeldung der Co-Autor\*innen*

Nach Erhalt der Rückmeldungen der Co-Autor\*innen wurde der Beitrag von Jennifer Igler überarbeitet und finalisiert.

### *Revision nach Rückmeldung der Gutachter\*innen*

Nach Erhalt der Anmerkungen und Änderungsvorschläge der Gutachter\*innen wurde der Beitrag hauptverantwortlich von Jennifer Igler überarbeitet. Die Co-Autor\*innen gaben ihr Feedback zu den Überarbeitungen.

## Beitrag III

**Igler, J.,** Ohle-Peters, A. & McElvany, N. (2022). Unterrichtsqualität aus der Sicht von externen Raterinnen und Ratern – Analysen zum Reihenfolgeeffekt. *Zeitschrift für Erziehungswissenschaft*, 25, 1431–1453. <https://doi.org/10.1007/s11618-022-01121-z>

*Konzeption der schriftlichen Abfassung des Beitrags*

Die schriftliche Abfassung wurde von Jennifer Iglar konzipiert. Die Co-Autorinnen gaben dazu Rückmeldungen.

*Statistische Auswertungen*

Die statistischen Auswertungen wurden hauptverantwortlich von Jennifer Iglar vorgenommen, wobei die Co-Autor\*innen beratend tätig waren.

*Schriftliche Abfassung des Beitrags*

Der Text wurde vollständig von Jennifer Iglar verfasst.

*Revision nach Rückmeldung der Co-Autorinnen*

Anhand von Anmerkungen beider Co-Autorinnen hat Jennifer Iglar den Text überarbeitet und finalisiert.

*Revision nach Rückmeldung der Gutachter\*innen*

Der Text wurde nach Erhalt der Anmerkungen und Änderungsvorschläge der Gutachter\*innen hauptverantwortlich von Jennifer Iglar überarbeitet. Die Co-Autorinnen waren beratend tätig und gaben Feedback zu den Überarbeitungen.

**Beitrag IV (weiterführende Analysen)**

**Iglar, J.,** Ohle-Peters, A., Teerling, A., Köller, O. & McElvany, N. (unpublished).  
Students' perceived motivational support in elementary school – How accurate are teachers' judgments?

*Konzeption der schriftlichen Abfassung des Beitrags*

Die schriftliche Abfassung wurde hauptverantwortlich von Jennifer Iglar konzeptualisiert, wobei die Co-Autor\*innen Beratung leisteten.

*Statistische Auswertungen*

Die statistischen Auswertungen wurden hauptverantwortlich von Jennifer Iglar durchgeführt. Die Co-Autor\*innen sowie Dr. Justine Stang-Rabrig und Dr. Matthias Trendtel fungierten als Berater\*innen.

*Schriftliche Abfassung des Beitrags*

Die Hauptverantwortung der Verschriftlichung des Beitrages lag bei Jennifer Iglar. Die Co-Autor\*innen gaben Rückmeldungen.

*Revision nach Rückmeldung der Co-Autor\*innen*

Der Beitrag wurde nach Erhalt der Rückmeldungen der Co-Autor\*innen von Jennifer Iglar überarbeitet und finalisiert.

*Revision nach Rückmeldung der Gutachter\*innen*

Der Text wurde nach Erhalt der Anmerkungen und Änderungsvorschläge der Gutachter\*innen hauptverantwortlich von Jennifer Iglar überarbeitet. Die Co-Autor\*innen waren beratend tätig und haben ihr Feedback zu den Überarbeitungen gegeben.

### 9.3 Eidesstattliche Erklärung

Hiermit versichere ich schriftlich und eidesstattlich gemäß § 11 Abs. 2 PromO v.08.02.2011/08.05.2013:

1. Die von mir vorgelegte Dissertation ist selbstständig verfasst und alle in Anspruch genommenen Quellen und Hilfen sind in der Dissertation vermerkt worden.
2. Die von mir eingereichte Dissertation ist weder in der gegenwärtigen noch in einer anderen Fassung an der Technischen Universität Dortmund oder an einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegt worden.

---

Ort, Datum

---

Unterschrift

3. Weiterhin erkläre ich schriftlich und eidesstattlich, dass mir der „Ratgeber zur Verhinderung von Plagiaten“ und die „Regeln guter wissenschaftlicher Praxis der Technischen Universität Dortmund“ bekannt sind und von mir in der vorgelegten Dissertation befolgt worden sind.

---

Ort, Datum

---

Unterschrift