

Holger Gärtner\*, Felicitas Thiel\*, Christopher Kellermann & Max Nachbauer

## Ein theoriebasierter Schülerfragebogen für Unterrichtsevaluation

### Zusammenfassung

*Der vorliegende Beitrag beschreibt die theoriegeleitete Entwicklung und psychometrische Überprüfung eines Schülerfragebogens zur Erfassung von Unterrichtsqualität, der vorrangig als formatives Evaluationsinstrument in der schulischen Praxis eingesetzt werden soll. Grundlage des Schülerfragebogens stellt ein theoretisch begründetes Modell von Unterrichtsqualität dar, welches die drei Anforderungsbereiche Unterstützung des Wissenserwerbs, Motivierung und Klassenmanagement sowie die querliegende Dimension Individualisierung umfasst. Der Fragebogen wurde an einer Stichprobe von Schüler:innen an Berufsschulen eingesetzt. Am ersten Messzeitpunkt wurde eine vorläufige Version des Schülerfragebogens verwendet. Basierend auf den Ergebnissen wurde der Schülerfragebogen optimiert und an einem zweiten Messzeitpunkt überprüft. Konfirmatorische Faktorenanalysen belegen eine gute Passung zwischen empirischen Daten und theoretischem Modell. Die Skalen sind empirisch klar voneinander trennbar, mit der Ausnahme einer relativ hohen Faktorkorrelation auf Klassenebene. Auch die Reliabilitäten der Skalen erweisen sich mit einer Ausnahme als gut bis sehr gut. Die Ergebnisse sprechen dafür, dass Lehrkräfte mithilfe des Schülerfragebogens konkrete Ansatzpunkte für die Weiterentwicklung ihres Unterrichts ermitteln können.*

### Schlagworte

*Unterrichtsevaluation, Feedback, Unterrichtsqualität, Schülerfragebogen*

---

\* geteilte Erstautor:innenschaft

Prof. Dr. Holger Gärtner (Korrespondenzautor), ORCID: 0000-0001-8814-9341, Freie Universität Berlin, Arbeitsbereich Unterrichts- und Schulevaluation, Habelschwerdter Allee 45, 14195 Berlin  
E-Mail: [holger.gaertner@fu-berlin.de](mailto:holger.gaertner@fu-berlin.de)

Prof. Dr. Felicitas Thiel · M.A. Christopher Kellermann · Dipl.-Päd. Max Nachbauer, Freie Universität Berlin, Arbeitsbereich Schulpädagogik/Schulentwicklungsforschung, Habelschwerdter Allee 45, 14195 Berlin  
E-Mail: [felicitas.thiel@fu-berlin.de](mailto:felicitas.thiel@fu-berlin.de)  
[christopher.kellermann@fu-berlin.de](mailto:christopher.kellermann@fu-berlin.de)  
[max.nachbauer@fu-berlin.de](mailto:max.nachbauer@fu-berlin.de)

## Development of a Theory-Driven Student Questionnaire for Teacher Evaluation

### Abstract

*This article describes the theory-driven development and psychometric evaluation of a student questionnaire for the assessment of teaching quality. The student questionnaire is supposed to be used in the context of formative evaluation within schools. It is based on a model of teaching quality that comprises the domains supporting knowledge acquisition, motivating, and classroom management as well as the lateral dimension individualization. The questionnaire was used on a sample of students in vocational schools. At the first measurement time point, a preliminary version of the student questionnaire was applied. Based on the results, the questionnaire was revised and tested at a second measurement time point. Confirmatory factor analyses yielded a good fit between empirical data and theoretical model. All scales are empirically distinguishable from each other with the exception of one relatively high factor correlation on the class level. Further, reliabilities of the scales are good to very good with one exception. The results suggest that teachers can use the student questionnaire for identifying leverage points for improving their teaching.*

### Keywords

*teacher evaluation, feedback, teaching quality, student questionnaire*

## 1. Einleitung

Der im vorliegenden Beitrag vorgestellte Schülerfragebogen zur Messung von fachübergreifenden Aspekten der Unterrichtsqualität wurde im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts „Unterrichtsbeurteilung und -feedback durch Schulleitungen“ entwickelt. Für eine Interventionsstudie wurde hier zunächst ein Ratingbogen für Schulleitungen, die Unterricht in ganz unterschiedlichen Fächern (auch fachfremd) beurteilen müssen und auf dieser Grundlage ein Feedback geben, entwickelt und validiert (Gärtner, Thiel, Nachbauer & Kellermann, 2021). Der komplementär zu diesem Ratingfragebogen entwickelte Schülerfragebogen wird im vorliegenden Beitrag vorgestellt. Ein auf das Ratinginstrument abgestimmter Schülerfragebogen erlaubt nicht nur eine Evaluation des Unterrichts in kürzeren Abständen zwischen den regelmäßigen, aber häufig nur maximal einmal jährlich stattfindenden Unterrichtsbeurteilungen und -feedbacks durch die Schulleitung, er eröffnet auch eine weitere Perspektive auf den Unterricht (Clausen, 2002; Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Gruehn, 2000). Eine kombinierte Einschätzung der Unterrichtsqualität aus unterschiedlichen Perspektiven ist deshalb für die Unterrichtsentwicklung von hohem Wert.

Im Folgenden wird zunächst die Konzeptualisierung von Unterrichtsqualität in einem Modell, das sowohl dem Ratingbogen als auch dem Schülerfragebogen zugrunde liegt, dargestellt, mit dem Ziel, die theoretischen Dimensionen und Unterrichtsmerkmale zu identifizieren, die wesentliche Ansatzpunkte für die Unterrichtsentwicklung darstellen. Die theoretische Konstruktion geht von drei basalen Anforderungen des Unterrichtens aus: (a) der Unterstützung der schülerseitigen Verarbeitung von Informationen zu bedeutungsvollem Wissen, (b) der Motivierung zur Aneignung eines Lernangebots sowie zur Aufrechterhaltung von Anstrengung auch im Falle des Misserfolgs, (c) der Gestaltung und Steuerung der Interaktion in der Schulklasse mit dem Ziel der Maximierung der aktiven Lernzeit. Für alle drei Dimensionen sowie für das Unterrichtsmerkmal Individualisierung werden auf der Grundlage des nachfolgend dargestellten theoretischen Modells Subskalen ausdifferenziert. Es wird davon ausgegangen, dass die entsprechenden Einzelmerkmale eine konkrete Rückmeldung zu unterschiedlichen Ansatzpunkten der Unterrichtsentwicklung liefern.

## 2. Unterrichtsevaluation und Konzeptualisierung von Unterrichtsqualität

Im Zuge der Implementation umfassender Qualitätssicherungs- und -entwicklungsverfahren in den letzten 20 Jahren wurden in fast allen deutschen Bundesländern Verfahren der externen und internen Evaluation implementiert (Thiel et al., 2019). Der Evaluation des Unterrichts kommt hier eine besondere Bedeutung zu. Allerdings kommen bislang nur in wenigen Ländern validierte Evaluationsinstrumente zum Einsatz. Das gilt für Ratingbögen von Schulinspektionen in ähnlicher Weise wie für Schülerfragebögen, die im Rahmen der internen Evaluation eingesetzt werden (Helmke et al., 2018; Thiel et al., 2019). Auf der anderen Seite wurden – ausgehend von den TIMSS-Videostudien (Klieme & Rakoczy, 2003) und dem in den USA entwickelten Instrument Classroom Assessment Scoring System (CLASS; Pianta, La Paro & Hamre, 2008) – in der Unterrichtsforschung in den letzten Jahren mehrere Instrumente zur Unterrichtsbeurteilung entwickelt, die drei Basisdimensionen der Unterrichtsqualität erfassen (Praetorius, Klieme, Herbert & Pinger, 2018). Die Forschung zu den drei Basisdimensionen prägt die gegenwärtige Diskussion um die Operationalisierung von Unterrichtsqualität. Obwohl sich die Operationalisierung dieser Basisdimensionen in den unterschiedlichen Instrumenten teilweise deutlich unterscheidet, haben sich als gemeinsame Bezeichnungen der Dimensionen inzwischen die drei Begriffe effiziente Klassenführung, kognitive Aktivierung und schülerorientiertes Unterrichtsklima (Klieme & Rakoczy, 2003) bzw. classroom management, cognitive activation und student support durchgesetzt (Praetorius et al., 2018). Die Identifikation der Dimensionen erfolgte zunächst im Rahmen empirischer Analysen, später wurden die drei Dimensionen mit Bezug auf konstruktivistische Lerntheorien, Motivationstheorien, Unterrichtsklimaforschung

und Ansätze zum Klassenmanagement theoretisch unterlegt. Durch die Zusammenführung mehrerer Unterrichtsmerkmale in wenige, inhaltlich relevante Dimensionen genügen die Instrumente forschungsrelevanten Anforderungen wie Sparsamkeit, Vermeidung von Multikollinearität oder hoher Erklärkraft für den Lernerfolg von Schüler:innen in hohem Maß (Praetorius et al., 2018). Ursprünglich für das Fach Mathematik entwickelt, wurde infolge zahlreicher Studien in anderen Fächern der Anspruch auf einen generischen Charakter der drei Basisdimensionen erhoben. Von Vertreter:innen der Fachdidaktik wird allerdings immer wieder der Bedarf an einer fachdidaktischen Spezifikation, insbesondere der Items zur kognitiven Aktivierung, artikuliert. So wird gegenwärtig z. B. eine fachdidaktische Weiterentwicklung des Modells der Basisdimensionen in der Mathematikdidaktik diskutiert (Prediger & Neugebauer, 2021), eine ähnliche Diskussion findet in den Didaktiken der Naturwissenschaft statt (Steffensky & Neuhaus, 2018).

Unabhängig von einer fachlichen Spezifizierung stellt die Beurteilung der Dimension kognitive Aktivierung hohe Ansprüche an die Beurteilenden, werden hier doch auf der Grundlage einer konstruktivistischen Lerntheorie hoch-inferente Indikatoren zur Unterrichtsqualität operationalisiert. Außerdem scheint es erforderlich zu sein, Unterricht über einen längeren Zeitraum zu beobachten, um zu einer zuverlässigen Einschätzung der Qualität kognitiver Aktivierung zu kommen. Praetorius et al. (2014) konnten in einer Generalisierbarkeitsstudie zeigen, dass für eine zuverlässige Einschätzung dieser Qualitätsdimension wenigstens neun Unterrichtsstunden erforderlich sind. Die Items der Dimension kognitive Aktivierung erfassen höherwertige Verstehensprozesse, die für den Erwerb von elaborierten Kompetenzen im Fachunterricht unverzichtbar sind (Praetorius et al., 2018). Dagegen werden empirisch gut abgesicherte Merkmale guten Unterrichts wie Klarheit und Strukturiertheit (Helmke et al., 2018), die für die Unterstützung kognitiver Prozesse ebenfalls relevant sind, sowohl in den Modellannahmen als auch in den gegenwärtigen Instrumenten oft nicht operationalisiert.

Vor dem Hintergrund der fachdidaktischen Diskussion auf der einen Seite und der Schwierigkeit der Operationalisierung der Dimension kognitive Aktivierung auf der anderen Seite ist die Frage zu stellen, ob für Evaluationsinstrumente, die eine Beurteilung von Unterrichtsqualität sowie ein Unterrichtsfeedback aus fachübergreifender Perspektive (Ratingbogen für Schulleitungen) oder Schülerperspektive (Schülerfragebogen) unterstützen und deren primärer Zweck die Unterrichtsentwicklung ist, nicht ein Ansatz fruchtbar sein könnte, der sich stärker an allgemeinen kognitionspsychologischen und motivationspsychologischen Ansätzen orientiert. Ein Vorteil dieses Ansatzes ist, dass mit dem Instrument zugleich ein leicht verständliches theoretisches Modell vorliegt, das keine fachdidaktischen Kenntnisse voraussetzt. Ein solches Modell ist für die Gestaltung der Qualifizierung von Beurteilenden aus der Praxis, hier Schulleitungen, wie für die Führung eines bedeutungsvollen Feedbackgesprächs gleichermaßen hilfreich (Heneman & Milanowski, 2003). An diesen Überlegungen ansetzend, haben wir eine an allgemeinen kognitionspsychologischen, motivationspsychologischen sowie mikrosoziologischen Ansätzen orientierte Operationalisierung von drei Basisdimensionen vorgenommen

(Ophardt & Thiel, 2013). Die hier vorgeschlagene Operationalisierung der drei Basisdimensionen unterscheidet sich in vier Punkten von den vorliegenden Fragebögen, die im Rahmen unterschiedlicher Forschungsprojekte entwickelt worden sind:

1. Anstatt kognitiver Aktivierung wird der gesamte Prozess der Unterstützung des Wissenserwerbs auf der Grundlage eines Lehr-Lern-Prozessmodells (Klauer & Leutner, 2012) operationalisiert; das heißt, es werden auch basale Merkmale der Unterrichtsqualität wie Klarheit und Strukturiertheit oder Üben einbezogen.
2. An die Stelle der Dimension Schülerorientierung/konstruktive Unterstützung, die sowohl motivationspsychologische als auch kognitionspsychologische Konstrukte beinhaltet, tritt die Dimension Motivierung. Auf der Grundlage des Erwartung-mal-Wert-Modells (Eccles & Wigfield, 2002; Rheinberg, 2008) werden unterschiedliche Ansatzpunkte für die Motivierung der Schüler:innen identifiziert.
3. Die Dimension Klassenmanagement umfasst alle Aspekte, die sich auf die Gestaltung einer lerndienlichen Interaktion in der Schulklasse beziehen; also auch das Klassen-/Unterrichtsklima, das anders als Motivierungsstrategien auf die grundlegende Beziehung zwischen Lehrkraft und Schüler:innen abhebt, die wiederum die Grundlage für die Gestaltung und Steuerung der sozialen Interaktionen in der Schulklasse darstellt und damit theoretisch dem Klassenmanagement zuzuordnen ist.
4. Es werden Merkmale der Individualisierung des Unterrichts für die kognitiven und motivationalen Aspekte des Lernens operationalisiert, weil diese Merkmale für die Unterrichtsentwicklung in inklusiven Schulen eine große Rolle spielen.

Für alle drei Dimensionen sowie für das Unterrichtsmerkmal Individualisierung werden auf der Grundlage des nachfolgend dargestellten theoretischen Modells Subskalen ausdifferenziert. Die entsprechenden Einzelmerkmale erlauben eine konkrete Rückmeldung zu unterschiedlichen Ansatzpunkten der Unterrichtsentwicklung. Die theoretische Konstruktion geht von drei basalen Anforderungen des Unterrichtens aus: (a) der Unterstützung der schülerseitigen Verarbeitung von Informationen zu bedeutungsvollem Wissen, (b) der Motivierung zur Aneignung eines Lernangebots sowie zur Aufrechterhaltung von Anstrengung auch im Falle des Misserfolgs, (c) der Gestaltung und Steuerung der Interaktion in der Schulklasse mit dem Ziel der Maximierung der aktiven Lernzeit.

## 2.1 Unterstützung des Wissenserwerbs

Die erste Dimension, die die kognitiven Funktionen des Lernens betrifft, wird als Unterstützung des Wissenserwerbs bezeichnet. Ihre Konzeption beruht auf dem Lehr-Lern-Prozessmodell von Klauer und Leutner (2012), das wiederum unterschiedliche Ansätze der Informationsverarbeitung und des Gedächtnisses integriert. Das Modell unterscheidet vier kognitive Komponenten des Lernens: (a) Informierung, (b) Informationsverarbeitung, (c) Speicherung und Abruf von In-

formationen und (d) Transfer. Informierung wird in unserem Modell als Subskala *Erklärung und Präsentation* operationalisiert. Die Herausforderung bei der Einführung neuer Konzepte oder der Darstellung neuen Wissens besteht darin, dass im sensorischen Register viele Reize aufgenommen werden, aber die Kapazität des Arbeitsgedächtnisses für die Weiterverarbeitung der Reize sehr gering ist. Informationen sollten daher so aufbereitet werden, dass die extrinsische kognitive Belastung des Arbeitsgedächtnisses, im Unterschied zur lernbezogenen Belastung (*germane load*), möglichst gering ist. Das bedeutet: Die Gestaltung der Lernmaterialien muss so erfolgen, dass keine unnötige Beanspruchung des Arbeitsgedächtnisses der Lernenden entsteht (Klauer & Leutner, 2012; Sweller, 1988). Eine gute Informationsdarbietung ist entsprechend durch eine strukturierte Präsentation sowie verständliche Erklärungen und Veranschaulichungen durch Beispiele gekennzeichnet. Im Anschluss an die Informationsaufnahme erfolgt die Informationsverarbeitung. Lehrkräfte sollten in dieser Phase die Verstehensprozesse der Schüler:innen unterstützen. Diesbezüglich unterscheiden Klauer und Leutner (2012) zwei Aspekte: die Elaboration von Informationen und ihre anschließende Reduktion. Elaborieren umfasst Prozesse wie Klärung von Zusammenhängen, die Kontextualisierung von Informationen oder die Entfaltung unterschiedlicher Detailspekte. Angeregt werden können Elaborationsprozesse durch Vergleiche und kognitiv aktivierende Fragen, die auf Voraussetzungen, Bedingungen und Folgen abzielen. Reduktive Prozesse haben die Funktion, Informationen zu komprimieren. Lehrkräfte können diese anregen, indem sie behandelte Inhalte noch einmal zusammenfassen. Neben elaborativen und reduktiven Strategien ist der Einsatz von metakognitiven Strategien hilfreich für die Schüler:innen. Metakognitive Strategien umfassen die Planung, Überwachung und Anpassung des Lernprozesses (Friedrich & Mandl, 1992). Lehrkräfte können den Einsatz metakognitiver Strategien beispielsweise dadurch unterstützen, dass sie die Schüler:innen dazu auffordern, ihre Vorgehensweisen zu planen und ihre Lösungen selbst zu bewerten. Die langfristige Verfügbarkeit von Wissen ist von einer nachhaltigen Speicherung im Langzeitgedächtnis durch ausreichende Übung abhängig. Neben einer ausreichenden Häufigkeit von Übungsphasen ist es relevant, wie die Übungen konkret ausgestaltet werden. Wichtig ist, dass besonders schwierige Aspekte intensiv geübt werden und dass die Resultate des Übens überprüft und gegebenenfalls korrigiert werden (Anderson, 1982). Die vierte kognitive Funktion des Lernens ist der Transfer. Hier geht es darum, erworbenes Wissen so zu flexibilisieren, dass es auf neue Probleme übertragen werden kann bzw. darum, gelernte Prinzipien oder Strategien in anderen Zusammenhängen anzuwenden. Die Schüler:innen müssen dazu die Tiefenstruktur eines Sachverhalts erfassen (Klauer & Leutner, 2012). Der Transfer kann angeregt werden, indem die Schüler:innen ihr theoretisches Wissen auf praktische Probleme anwenden oder indem die Schüler:innen zu einem Thema mehrere Aufgaben bearbeiten, die den Kontext variieren.

## 2.2 Motivierung

Neben den kognitiven Funktionen sind die motivationalen Funktionen für das Lernen zentral. Motivation sorgt dafür, dass Lernaktivitäten begonnen und aufrechterhalten werden. Grundlage für die Konzeptualisierung der Unterrichtsmerkmale zur Erfassung der Motivierungsqualität ist das Erwartung-mal-Wert-Modell (Eccles & Wigfield, 2002; Rheinberg, 2008). Vier Aspekte der Motivierung, die für den Unterricht eine zentrale Bedeutung haben, lassen sich hieraus ableiten: Durch (a) das Wecken von Interesse, (b) die Förderung von Lernfreude und (c) die Verdeutlichung des Nutzens können Anreize für Lernaktivitäten geschaffen werden, während (d) die Stärkung der Selbstwirksamkeitserwartung eine günstige Einschätzung der eigenen Erfolgswahrscheinlichkeit beim Lernen fördert (Eccles & Wigfield, 2002).

Das Wecken von Interesse ist ein wichtiger Ansatzpunkt für die Motivierung der Schüler:innen. Interesse bezieht sich auf die jeweiligen Lerninhalte und stellt damit einen gegenstandsbezogenen Anreiz dar. Es setzt sich aus wertbezogenen Überzeugungen und Gefühlsüberzeugungen zusammen (Schiefele, 2009). Lehrkräfte können Interesse wecken, indem sie die Schüler:innen mit neuen Einsichten überraschen oder indem sie im Unterricht mit Enthusiasmus auftreten und so selbst Interesse an den Lerninhalten demonstrieren. Im Gegensatz zum Interesse als gegenstandsbezogenem Anreiz bezieht sich Lernfreude auf die Art der Lernaktivitäten und wird daher auch als tätigkeitsbezogener Anreiz bezeichnet (Rheinberg, 2008). Die Lehrkräfte sollten demnach darauf achten, dass die Schüler:innen im Unterricht Tätigkeiten nachgehen können, bei denen sie Freude und Spaß erleben. Hierzu zählen beispielsweise explorierende, gestalterische oder spielerische Aktivitäten, die die Lernenden entsprechend ihrer Fähigkeiten herausfordern und ein Aufgehen im Tun ermöglichen (Csikszentmihalyi & Rathunde, 1993). Ein weiterer Anreiz bezieht sich darauf, welchen Nutzen die Lerninhalte für die Schüler:innen haben (Hulleman & Harackiewicz, 2009). Für die entsprechende Motivierungsstrategie ist es entscheidend, dass Lehrkräfte verdeutlichen, warum bestimmte Lerninhalte für die Schüler:innen relevant sind und dass sie Beispiele geben, in welchen Lebenssituationen bestimmte Kenntnisse oder Kompetenzen benötigt werden. Anreize alleine motivieren allerdings nicht zum Lernen. Die Schüler:innen müssen gleichzeitig auch die Erwartung haben, das Lernziel mit ihren eigenen Kompetenzen erreichen zu können. Aus diesem Grund stellt die Förderung von Selbstwirksamkeitserwartungen (Bandura, 1976) einen wichtigen Ansatzpunkt für Motivierung dar. Relevant ist hierbei, dass die Lehrkraft den Schüler:innen zeigt, dass sie ihnen etwas zutraut und dass sie die Schüler:innen spezifisch und aufgabenbezogen lobt (Brophy, 1981). Weiterhin sollte die Lehrkraft Herausforderungen formulieren, deren Bewältigung möglich ist. Bei Schwierigkeiten sollte die Lehrkraft die Schüler:innen ermutigen, aber die Lösung nicht vorwegnehmen, sodass die Schüler:innen Erfolgserfahrungen machen können (Bandura, 1993).

## 2.3 Klassenmanagement

Schulklassen sind dynamische Interaktionssysteme, in denen Lernen nur dann gelingt, wenn ein gewisses Maß an sozialer Ordnung sichergestellt ist (Doyle, 2006; Ophardt & Thiel, 2013; Piowar, 2013). Diese sozialen Faktoren des Lernens werden im Konstrukt des Klassenmanagements konzeptualisiert. Im Vordergrund steht die Maximierung der aktiven Lernzeit für alle Schüler:innen der Klasse. Fünf zentrale Funktionen können im Hinblick auf das Klassenmanagement der Lehrkraft unterschieden werden (Thiel, Richter & Ophardt, 2012): (a) Unterrichtsklima, (b) Monitoring, (c) Steuerung des Unterrichtsflusses, (d) Gruppenaktivierung und (e) der Umgang mit Störungen.

Für ein lernförderliches Unterrichtsklima muss zunächst eine Interaktionsordnung durch die Einführung von Regeln etabliert werden. Die Beachtung von Regeln durch die Schüler:innen in der Form von gegenseitigem Respekt, Rücksichtnahme und die Bereitschaft zur Unterstützung findet ebenso Ausdruck im Unterrichtsklima wie entsprechende Verhaltensweisen der Lehrkraft, in denen das Arbeitsbündnis mit der Klasse zum Ausdruck kommt. Ein effektives Klassenmanagement setzt voraus, dass die Lehrkraft das Geschehen in der ganzen Klasse wahrnimmt (Kounin, 2006). Dies betrifft das Monitoring. Damit ist gemeint, dass die Lehrkraft über die gesamte Unterrichtsstunde hinweg die Aktivitäten aller Schüler:innen im Blick behält und bemerkt, wenn sich einzelne Schüler:innen mit unterrichtsfremden Dingen beschäftigen. Für die Maximierung aktiver Lernzeit ist weiterhin eine effektive Steuerung des Unterrichtsflusses wichtig, einerseits, weil bei unnötigen Pausen Lernzeit verloren geht, andererseits, weil das Handlungsprogramm durch Leerläufe oder Nachfragen unterbrochen wird und dann mühevoll wiederaufgebaut werden muss (Thiel et al., 2012). Für die Aufrechterhaltung des Unterrichtsflusses sind nahtlose Übergänge zwischen Unterrichtsphasen und die Vermeidung von inhaltlichen Abschweifungen wichtig. Auf die Bedeutung der Gruppenaktivierung hat bereits Kounin (2006) mit seiner Videostudie aufmerksam gemacht. Gruppenaktivierung meint, dass die Lehrkraft alle Schüler:innen aktiv in den Unterricht einbindet und sicherstellt, dass kein Leerlauf für einzelne Schüler:innen entsteht. Störungen des Unterrichts entstehen, weil die Schüler:innen die Aufmerksamkeit nicht über längere Zeit aufrechterhalten können, weil sie von ihren Mitschüler:innen abgelenkt werden, diese beindrucken wollen oder weil sie sich ungerecht behandelt fühlen (Thiel, 2016). Wenn Störungen auftreten, sollte die Lehrkraft sich klar verhalten und das Störverhalten möglichst unmittelbar beenden. Ein angemessener Umgang mit Störungen berücksichtigt die Schwere des Störverhaltens. Bei kleineren Störungen sind beiläufige nonverbale Signale häufig ausreichend. Bei schwereren Störungen können dagegen verbale Ermahnungen und gegebenenfalls auch die Ankündigung von Sanktionen notwendig sein (Thiel, 2016). Für jede Störungsintervention gilt, dass sie das Arbeitsbündnis zwischen Lehrkraft und Schüler:innen nicht beschädigen sollte. Zurechtweisungen müssen aus diesem Grund konstruktiv erfolgen.



## 2.4 Individualisierung

Da sich Schüler:innen hinsichtlich ihrer Lernvoraussetzungen unterscheiden, stehen Lehrkräfte schließlich vor der Herausforderung der Individualisierung des Unterrichts (Klieme & Warwas, 2011). Individualisierung wird im vorliegenden Modell als querliegende Dimension verstanden, das heißt, Individualisierung stellt keinen eigenständigen Anforderungsbereich dar, sondern bezieht sich inhaltlich auf die zuvor dargestellten Bereiche. Individualisierende Maßnahmen sind primär in den Bereichen Unterstützung des Wissenserwerbs und Motivierung erforderlich, da die Schüler:innen sowohl unterschiedliche kognitive (z. B. Vorwissen) als auch unterschiedliche motivationale Lernvoraussetzungen (z. B. Interessen) aufweisen. Im Bereich des Klassenmanagements ist Individualisierung weniger relevant, da sich Klassenmanagement per definitionem auf die gesamte Klasse bezieht. Dementsprechend werden die zwei Aspekte individualisierter Wissenserwerb und individualisierte Motivierung unterschieden.

Auf der Grundlage der dargestellten Subdimensionen des theoretischen Modells wurden Skalen sowohl für den Unterrichtsbeurteilungsbogen für Schulleitungen (Gärtner et al., 2021) als auch für den in diesem Beitrag vorgestellten Schülerfragebogen operationalisiert.

## 3. Entwicklung eines fachübergreifenden Schülerfragebogens

Mehrere Studien haben gezeigt, dass Schüler:innen in der Lage sind, eine bedeutungsvolle, valide und reliable Einschätzung zur Qualität von Unterricht zu geben (Gärtner, 2010; Gärtner & Brunner, 2018; Göllner et al., 2016; Gruehn, 2000; Kunter et al., 2005; Wagner, Göllner, Helmke, Trautwein & Lüdtke, 2013). Schülerbefragungen zum Unterricht haben im Vergleich zu Unterrichtsratings durch trainierte Beobachter:innen eine Reihe von Vorteilen (vgl. auch Clausen & Göbel, 2020): Erstens legen Schüler:innen ihrer Urteilsbildung eine größere Zahl von Ereignissen bzw. einen längeren Beobachtungszeitraum zugrunde als externe Beobachter:innen, zweitens haben sie viele Möglichkeiten, unterschiedliche Lehrkräfte zu vergleichen und damit eine soziale Bezugsnorm zu entwickeln, drittens ist die subjektive Perspektive der Lernenden auf die Unterstützung des Lernprozesses oder die Beziehung zur Lehrkraft von besonderem Interesse und viertens kann ein Vergleich der Perspektive der Schüler:innen mit der Perspektive der Lehrkraft zur Überprüfung der eigenen Wahrnehmung führen (Helmke et al., 2018).

Die Entwicklung (und der Einsatz) von Schülerfragebögen ist allerdings mit besonderen Herausforderungen verbunden. Dass die Schüler:innen- und die Lehrkraftperspektive in vielen Studien vergleichsweise weit auseinanderliegen (Clausen, 2002; Fauth, et al., 2014), kann durchaus auch als Validitätsproblem interpretiert werden. In jüngster Zeit ist die bereits von Clausen (2002) angeregte Diskussion

um den Einfluss eines „affektiven Gesamteindrucks“ auf die Wahrnehmung einzelner Merkmale wieder aufgenommen und in der wahrnehmungspsychologischen Perspektive des *Halo-Effekts* untersucht worden (Röhl & Rollett, 2020; Wagner, 2008). Röhl und Rollett (2020) konnten einen *Communion-Bias*, also die dominierende Bedeutung der subjektiven Einschätzung der Beziehung zur Lehrkraft für die Beurteilung der Unterrichtsqualität durch Schüler:innen nachweisen. Ein anderer Untersuchungsstrang beschäftigt sich damit, ob Schüler:innen die Bedeutung einzelner Items überhaupt verstehen. So untersucht Lenske (2016) das Verständnis der Items bei Grundschüler:innen und kommt zu dem Befund, dass die Bedeutung einiger Items von den Schüler:innen nicht korrekt erschlossen wird. Dies gilt insbesondere für die hoch-inferenten Items zur kognitiven Aktivierung. In einer aktuellen Studie von Lenske und Praetorius (2020) wurde der Befund der unzureichenden oder falschen semantischen Decodierung der Items zur kognitiven Aktivierung an einer Stichprobe von Sekundarschüler:innen bestätigt. Die Befunde stehen im Einklang mit der – im Vergleich mit der Dimension des Klassenmanagements – schwachen Korrelation von Schülerurteilen zur kognitiven Aktivierung und zu den Leistungen (Hanisch, 2018). Neben Fragen der Urteilsverzerrung durch einen Sympathie-Bias und den Untersuchungen zum Verständnis von Items ist die Frage der Trennschärfe zwischen den einzelnen Dimensionen relevant. Hier zeigt eine Untersuchung von Wagner (2008), dass Schüler:innen die Differenzierung zwischen unterschiedlichen theoretischen Dimensionen häufig schwerfällt.

Die Vorteile einer Nutzung von Schüler:inneneinschätzungen der Unterrichtsqualität für Unterrichtsentwicklung sind, so lassen sich die vorliegenden Befunde interpretieren, an bestimmte Voraussetzungen der Fragebogenkonstruktion und Itemformulierung gebunden:

- Items sollten durchgängig auf einen Bezug auf Schüler:innenverhalten verzichten, um eine mögliche selbstdienliche Urteilsverzerrung zu vermeiden.
- Items sollten möglichst konkrete Verhaltensweisen der Lehrkraft beschreiben, deren Einschätzung für die Schüler:innen nicht an das Verständnis von didaktischen Kernkonzepten (z. B. Fehlerfreundlichkeit) geknüpft ist.
- Items sollten grundsätzlich einfach formuliert sein.
- Die theoretische Trennschärfe der unterschiedlichen Dimensionen sollte auch für Schüler:innen prinzipiell erkennbar sein.

## 4. Ziele der Studie

Auf der Grundlage des dargestellten Modells von Unterrichtsqualität wurde, als Ergänzung zum bereits etablierten Beobachtungsbogen, ein Schülerfragebogen entwickelt. Das Ziel der vorliegenden Studie ist die psychometrische Überprüfung dieses Fragebogens anhand unterschiedlicher Gütekriterien. Ausgangspunkt der Prüfung ist die Fragestellung, ob Lehrkräfte mittels des vorliegenden Instrumentes reliable und valide Rückmeldungen über verschiedene Aspekte des Unterrichts erhalten

können. Zur Prüfung gehören dementsprechend verschiedene Maße der Reliabilität der entwickelten Skalen sowie Maße der Konstruktvalidität. Der letztgenannte Aspekt bezieht sich auf die Frage, ob die einzelnen Skalen eindimensionale Konstrukte abbilden und ob sich die Konstrukte empirisch voneinander trennen lassen.

## 5. Methode

### 5.1 Entwicklung des Schülerfragebogens

Ausgehend von dem beschriebenen Modell der Unterrichtsqualität wurden Items für die zuvor beschriebenen Aspekte von Unterrichtsqualität entwickelt. Hierbei wurde, wo dies sinnvoll erschien, auf bereits vorliegende Skalen zurückgegriffen (Baumert et al., 2008; Baumert, Gruehn, Heyn, Köller & Schnabel, 1997; Gärtner, 2010; Helmke, 2017; Jäger & Helmke, 2008; Nationales Bildungspanel, 2013; Rakoczy, Buff & Lipowsky, 2005; Seidel, Prenzel, Duit & Lehrke, 2003; Thiel, Ophardt & Piwowar, 2013). Die Itemformulierungen des Schülerfragebogens orientieren sich an den in Kapitel 2 auf der Grundlage des Forschungsstands herausgearbeiteten Kriterien:

- Um soziale Erwünschtheit zu minimieren, wurden alle Items auf das Lehrkraftverhalten bezogen. Mit der Ausnahme des Unterrichtsklimas zwischen den Schüler:innen enthält der Fragebogen keine Items, die eine Selbstbewertung der Schüler:innen nahelegen (z. B. Clausen, 2002).
- Auch zur Erfassung der Unterstützung kognitiver Aktivitäten (z. B. Vergleiche) werden konkrete Aktivitäten der Lehrkraft abgefragt (z. B. „Unsere Lehrkraft lässt unterschiedliche Lösungswege von Aufgaben vergleichen und bewerten.“). Es wird vermieden, dass Schüler:innen ihre eigenen kognitiven Prozesse einschätzen müssen. Dadurch wird das Verständnis der Items erleichtert (z. B. Lenske & Praetorius, 2020).
- Alle Items sind sprachlich einfach formuliert. Auf komplexe Formulierungen wurde verzichtet (z. B. Lenske, 2016).
- Die drei Dimensionen sind trennscharf entsprechend der theoretisch hergeleiteten Unterrichtsmerkmale (siehe Kapitel 1) operationalisiert, um für die Schüler:innen die Abgrenzung von Fragen zur Unterstützung des Verstehensprozesses, zur Motivierung und zum Klassenmanagement nachvollziehbar zu machen (z. B. Wagner, 2008).

Als Format wurden 4-stufige Likert-Skalen gewählt. Beispielitems finden sich in Tabelle 1.

Tabelle 1: Unterrichtsmerkmale nach Anforderungsbereichen, Beispielimens und Reliabilitäten zu Messzeitpunkt 1 und 2

Unterrichtsmerkmal	Beispielimens	Messzeitpunkt 1			Messzeitpunkt 2				
		Items	Ω	ICC1	ICC2	Items	Ω	ICC1	ICC2
Unterstützung des Wissenserwerbs Erklärung und Präsentation	Unsere Lehrkraft gibt anschauliche Beispiele, die uns das Verstehen der Lerninhalte erleichtern.	3	.75	.29	.87	4	.84	.32	.85
Unterstützung des Verstehens	Unsere Lehrkraft lässt uns unsere Aufgabenlösungen selbst korrigieren.	4	.70	.11	.67	5	.76	.12	.63
Speicherung durch Üben	Wir erhalten gezielt Übungsaufgaben für besonders schwierige Themen.	5	.73	.20	.80	5	.80	.24	.79
Anwendung und Transfer	Unsere Lehrkraft stellt uns häufig Aufgaben, bei denen wir unser Wissen auf neue Situationen übertragen müssen.	5	.75	.21	.81	5	.80	.17	.72
Motivierung									
Weckung von Interesse	Auch bei bekannten Themen überrascht uns unsere Lehrkraft oft mit neuem Wissen.	6	.90	.28	.86	6	.91	.31	.85
Förderung von Freude	Die Bearbeitung der Aufgaben, die unsere Lehrkraft stellt, macht mir Spaß.	4	.89	.29	.87	4	.88	.19	.74
Verdeutlichung des Nutzens	Unsere Lehrkraft erklärt uns, warum es für das spätere Leben wichtig ist, das Unterrichtsthema zu lernen.	4	.82	.28	.86	4	.87	.24	.79
Förderung von Selbstwirksamkeit	Unsere Lehrkraft muntert uns auf, damit wir auch bei schwierigen Aufgaben nicht den Mut verlieren.	4	.88	.24	.84	4	.88	.28	.83
Klassenmanagement									
Lerndienliches Unterrichtsklima SuS	Im Unterricht gehen die Schülerinnen und Schüler respektvoll miteinander um.	2	.87	.32	.88	4	.91	.23	.79
Lerndienliches Unterrichtsklima LK	Unsere Lehrkraft ist immer freundlich zu den Schülerinnen und Schülern.					4	.88	.26	.81
Monitoring	Unsere Lehrkraft bekommt mit, was die Schülerinnen und Schüler im Unterricht machen.	4	.87	.16	.76	4	.90	.22	.77
Steuerung des Unterrichtsflusses	Unsere Lehrkraft schweift häufig vom Thema ab. <sup>1</sup>	4	.78	.22	.82	4	.77	.07	.49
Gruppenaktivierung	Im Unterricht unserer Lehrkraft beteiligen sich meist alle Schülerinnen und Schüler.	2	.75	.21	.81	4	.88	.16	.70
Umgang mit Störungen	Unsere Lehrkraft reagiert angemessen auf Störungen (also nicht zu streng und nicht zu mild).	4	.83	.21	.81	4	.84	.25	.80
Individualisierung	Unsere Lehrkraft gibt uns je nach Leistungsstärke unterschiedliche Aufgaben.	3	.75	.20	.80	6	.89	.11	.60

Anmerkungen. ICC = Intraklassenkorrelation; SuS = Schüler:innen; LK = Lehrkraft. <sup>1</sup> = Das Item wurde invertiert, sodass hohe Werte für eine gute Steuerung des Unterrichtsflusses stehen.

## 5.2 Stichprobe

Eingesetzt wurde der Schülerfragebogen im Rahmen des DFG-Projektes „Unterrichtsbeurteilungen und -feedback durch Schulleitungen“. Hierbei handelt es sich um ein Forschungsprojekt zur Überprüfung der Wirksamkeit von Unterrichtsfeedback. Entsprechend erfolgte die Rekrutierung der Schülerstichprobe über die am Projekt beteiligten Schulen und Lehrkräfte. Die vorliegende Studie verwendet Daten von zwei Messzeitpunkten. Messzeitpunkt 1 diente der Pilotierung der ersten Version des Fragebogens sowie der weiteren Optimierung. Entsprechend wurden für Messzeitpunkt 1 mehr Items entwickelt als für die Endversion intendiert. Auf der Grundlage der psychometrischen Analysen wurde eine überarbeitete Version des Fragebogens erstellt und für Messzeitpunkt 2 verwendet.

Die Erhebungen fanden im Oktober 2017 (Messzeitpunkt 1) und im Juni 2019 (Messzeitpunkt 2) an beruflichen Schulen in Berlin statt. Zu Messzeitpunkt 1 beteiligten sich 6 berufliche Schulen mit insgesamt 81 Klassen und 1299 Schüler:innen an der Ausgangsbefragung zur Wahrnehmung des Unterrichts (durchschnittlich 16 Schüler:innen pro Klasse). Die Befragung erfolgte in unterschiedlichen Jahrgangsstufen (13.8 % Jahrgangsstufe 10, 32.3 % Jahrgangsstufe 11, 13.2 % Jahrgangsstufe 12, 40.6 % Missing) und unterschiedlichen Fächern (33.3 % MINT<sup>1</sup>-Fächer, 13.7 % sprachliche Fächer, 8.2 % sozialwissenschaftliche Fächer, 4.1 % sonstige Fächer, 40.6 % Missing). Zu Messzeitpunkt 2 beteiligten sich an den 6 beruflichen Schulen noch 24 Klassen und 294 Schüler:innen an der Befragung. Wiederum wurden durch die Befragung unterschiedliche Jahrgangsstufen (6 % Jahrgangsstufe 11, 54.8 % Jahrgangsstufe 12, 14.4 % Jahrgangsstufe 13, 24.7 % Missing) und unterschiedliche Fächer (56.9 % MINT-Fächer, 8.7 % sprachliche Fächer, 7.7 % sozialwissenschaftliche Fächer, 2 % sonstige Fächer, 24.7 % Missing) abgedeckt. Alle Befragungen wurden vor Ort von geschulten Testleiter:innen mit Papierfragebogen durchgeführt.

## 5.3 Analysen

Die aus dem theoretischen Modell abgeleiteten Skalen wurden zunächst einer Itemanalyse unterzogen, um besonders leichte bzw. schwierige sowie wenig trennscharfe Items zu identifizieren. Innerhalb der Itemanalyse wurden explorative Faktorenanalysen (EFA) durchgeführt, da es sich um ein neues Instrument handelt, über dessen empirische Faktorenstruktur bisher noch nichts bekannt ist. Zur Optimierung der einzelnen Skalen wurden Items mit geringen Faktorladungen bzw. Parallelladungen ausgeschlossen. Die explorativ ermittelten Skalen wurden anschließend mithilfe von konfirmatorischen Faktorenanalysen (CFA) im Rahmen des Gesamtmodells überprüft. Da ein Gesamtmodell mit allen Unterrichtsmerkmalen auf zwei Ebenen aufgrund zu vieler zu schätzender Parameter zu komplex für die

1 MINT = Mathematik, Informatik, Naturwissenschaften und Technik.

vorliegende Stichprobe war, wurde zunächst eine CFA auf Individualebene mit allen Unterrichtsmerkmalen und anschließend getrennte Multilevel-CFAs für die drei Bereiche Unterstützung des Wissenserwerbs, Motivierung und Klassenmanagement durchgeführt. Das Unterrichtsmerkmal Individualisierung bleibt als querliegende Anforderung alleinstehend. Die Güte der Modellpassung wird anhand der folgenden Fit-Indizes geprüft: CFI, RMSEA,  $\chi^2/df$  (Hu & Bentler, 1999).

Weiterhin wurden Reliabilitätsmaße berechnet, einmal als interne Konsistenz in der Form des Koeffizienten  $\Omega$  (McDonald, 1999) sowie für die jeweils aggregierten Skalenmittelwerte in Form der Intraklassenkorrelation (ICC<sub>2</sub>; Gärtner, 2010; Lüdtke, Trautwein, Kunter & Baumert, 2006). Diese Reliabilität gilt als Voraussetzung für die Berechnung von Klassenmittelwerten für die Unterrichtsevaluation. Ergänzt werden diese Kennwerte um die Angabe der ICC<sub>1</sub>, welche ausweist, ob sich die gemessenen Aspekte der Unterrichtsqualität zwischen den einzelnen Klassen unterscheiden (Lüdtke et al., 2006). Die Analysen zur faktoriellen Validität und internen Konsistenz erfolgten mit Mplus 8 (Muthén & Muthén, 1998–2017). Fehlende Werte wurden durch die Full-Information-Maximum-Likelihood-Schätzmethode behandelt, sodass jeweils alle Fälle eingeschlossen werden können (Lüdtke, Robitzsch, Trautwein & Köller, 2007). Der Anteil fehlender Werte lag an Messzeitpunkt 1 je nach Item zwischen 0.9 % und 7.1 %, an Messzeitpunkt 2 zwischen 0.0 % und 1.7 %.

## 6. Ergebnisse

### 6.1 Deskriptive Statistiken

Die Items des Schülerfragebogens wurden zunächst deskriptiv anhand der Mittelwerte und Standardabweichungen analysiert. An beiden Messzeitpunkten zeigte sich, dass mit Ausnahme der Items zur Individualisierung alle Itemmittelwerte über dem theoretischen Mittelwert liegen. Demnach bewerten die Schüler:innen bei der großen Mehrheit der Merkmale den Unterricht als gut.

### 6.2 Psychometrische Eigenschaften der ursprünglichen Version des Schülerfragebogens zu Messzeitpunkt 1

Von den ursprünglich 98 Items bezogen auf 15 Skalen wurden im Rahmen der EFA 44 Items ausgeschlossen, sodass sich die nachfolgenden Analysen zunächst auf 14 Skalen bestehend aus 54 Items (2 bis 6 Items pro Skala) beziehen. Gründe für den Ausschluss von Items waren geringe Faktorladungen und Parallelladungen auf zwei unterschiedlichen Unterrichtsmerkmalen.

Aufgrund der Ergebnisse der EFA wurden zwei inhaltliche Änderungen an der Gesamtkonstruktion des Fragebogens vorgenommen. Erstens umfasste die

Skala zum Unterrichtsklima zwei Faktoren, von denen einer das Verhalten der Schüler:innen untereinander beschreibt und der andere das Verhalten der Lehrkraft. Der Lehrkraft-Faktor konnte aufgrund der geringen Itemanzahl zu Messzeitpunkt 1 zunächst nicht abgebildet werden, dies wurde bei der Überarbeitung des Schülerfragebogens aufgegriffen. Zweitens konnte die zunächst angenommene Unterscheidung von zwei unterschiedlichen Skalen zur Erfassung von Individualisierung, einmal im Bereich Unterstützung des Wissenserwerbs und einmal im Bereich Motivierung, empirisch nicht bestätigt werden. Daher wurden die Items zu einer bereichsübergreifenden Individualisierungs-Skala zusammengefasst. Diese zwei Änderungen sind im Einklang mit empirischen Analysen zum Beobachtungsbogen, bei denen sich die gleiche Faktorenstruktur zeigte (Gärtner et al., 2021).

Die Annahme von 14 inhaltlich separaten Aspekten von Unterrichtsqualität konnte durch die CFA auf Individualebene bestätigt werden. Die Fit-Indizes sprechen für eine gute Modellpassung: CFI = .95, RMSEA = .03,  $\chi^2/df = 1.99$ . Alle Items laden substantziell auf den intendierten Skalen ( $.52 < \lambda < .86$ ; s. Tabelle A1). Die Faktorkorrelationen variieren zwischen  $r = -.25$  und  $r = .89$  und zeigen, dass sich die theoretischen Konstrukte empirisch voneinander trennen lassen (s. Tabelle A2).

Die Ergebnisse der separaten Multilevel-CFAs zeigen, dass die theoretisch angenommene Faktorenstruktur von wenigen Ausnahmen abgesehen auch auf der Klassenebene Gültigkeit hat. Sowohl im Bereich Unterstützung des Wissenserwerbs als auch im Bereich Klassenmanagement sind gute Modellpassungen und theoriekonforme Faktorladungen und -korrelationen auf Klassenebene gegeben (Unterstützung des Wissenserwerbs: CFI = .97, RMSEA = .02,  $\chi^2/df = 1.69$ ,  $.45 < \lambda < 1.00$ ,  $.45 < r < .84$ ; Klassenmanagement: CFI = .96, RMSEA = .04,  $\chi^2/df = 2.78$ ,  $.77 < \lambda < 1.00$ ,  $.39 < r < .86$ ). Im Bereich Motivierung liegen ebenfalls eine gute Modellpassung (CFI = .96, RMSEA = .04,  $\chi^2/df = 2.77$ ) und substantielle Faktorladungen ( $.84 < \lambda < 1.00$ ) vor, allerdings weisen die Faktoren Weckung von Interesse und Förderung von Freude eine sehr hohe Korrelation von  $r = .97$  auf (restliche Faktorkorrelationen  $.82 < r < .90$ ; s. Tabellen A1 und A3). Faktorkorrelationen über  $r > .90$  legen nahe, dass zwei Skalen im Wesentlichen das gleiche Konstrukt erfassen (Kline, 2011). Somit können die Schülerwahrnehmung der beiden Unterrichtsmerkmale Weckung von Interesse und Förderung von Freude auf der Klassenebene empirisch nicht klar getrennt werden. Angesichts der inhaltlichen Nähe der beiden Konstrukte ist dieser Befund theoretisch plausibel.

Schließlich wurden die Reliabilitäten der Skalen überprüft. Sowohl die internen Konsistenzen der Skalen ( $.70 < \Omega < .90$ ) als auch die Reliabilitäten der aggregierten Klassenwerte ( $.67 < ICC2 < .88$ ) sind als gut zu bewerten. Zudem unterscheiden sich die aggregierten Klassenwerte substantziell zwischen den Klassen. Mehrheitlich werden zwischen 20 % und 30 % der Varianz der Schülerurteile durch die Klassenzugehörigkeit bestimmt ( $.11 < ICC1 < .32$ ; s. Tabelle 1).

### **6.3 Psychometrische Eigenschaften der überarbeiteten Version des Schülerfragebogens zu Messzeitpunkt 2**

Zur Optimierung des Fragebogens wurden neue Items generiert, um Skalen mit wenigen Items inhaltlich zu erweitern. Insgesamt wurde der Fragebogen um 13 neue Items ergänzt, sodass alle Skalen nun vier bis sechs Items umfassen (siehe Tabelle 1). Von besonderem Interesse war hierbei, dass der auf die Lehrkraft bezogene Aspekt des Unterrichtsklimas mit der überarbeiteten Version angemessen erfasst werden kann. Der Fragebogen wurde zunächst mit einer CFA überprüft. Aufgrund der deutlich kleineren Stichprobe wurde hierbei nur eine CFA auf Individualebene berechnet. Die Ergebnisse zeigen eine zufriedenstellende Passung zwischen der theoretisch angenommenen Faktorenstruktur und den empirischen Daten (CFI = .91, RMSEA = .04,  $\chi^2/df = 1.51$ ). Alle Items laden substantiell auf den zugewiesenen Konstrukten ( $.53 < \lambda < .90$ ; s. Tabelle A2). Somit erweisen sich die neu formulierten Items als faktoriell valide. Erfreulicherweise zeigte sich, dass das Unterrichtsklima in Bezug auf das Verhalten der Lehrkraft nun basierend auf einer höheren Anzahl an Items empirisch abgebildet werden kann. Die Faktorkorrelationen fallen teilweise etwas höher aus und bewegen sich jetzt bei zwei Korrelationen an der Obergrenze einer empirischen Trennbarkeit (s. Tabelle A2). Die Korrelation zwischen Weckung von Interesse und Förderung von Freude beträgt  $r = .90$ , die Korrelation zwischen Erklärung und Präsentation und Speicherung durch Üben beträgt  $r = .89$  (restliche Korrelationen  $.17 < r < .81$ ). Schließlich wurden die Reliabilitäten der Skalen des überarbeiteten Fragebogens analysiert (s. Tabelle 1). Die internen Konsistenzen der Einzelskalen fallen weiterhin gut aus ( $.76 < \Omega < .91$ ). Die ICC2 fällt gut bis akzeptabel aus ( $.60 < ICC2 < .85$ ), mit Ausnahme der Skala zur Steuerung des Unterrichtsflusses ( $ICC2 = .49$ ). Dies ist insofern überraschend, als dass diese Skala nicht verändert wurde. Die ICC1 verweist mehrheitlich auf bedeutende Varianzanteile zwischen den Klassen ( $.07 < ICC1 < .32$ ). Insgesamt gesehen sprechen die empirische Abbildung aller Unterrichtsmerkmale, die valide Faktorenstruktur und die mit einer Ausnahme zufriedenstellenden Reliabilitäten dafür, dass die Überarbeitung des Fragebogens zu einer Verbesserung des Instruments geführt hat.

## **7. Diskussion**

Ausgangspunkt dieses Beitrags ist ein Ansatz für Unterrichtsevaluation, welcher auf einem theoretisch begründeten Modell von Unterrichtsqualität basiert. Dieses Modell ist insbesondere für Unterrichtsevaluation und -feedback geeignet, weil es eine klare theoretische Konzeptualisierung der Unterrichtsmerkmale liefert, den Fokus auf veränderbare Aspekte des Unterrichts legt, an konkreten Verhaltensweisen der Lehrkraft im Unterricht ansetzt und mit empirischen Befunden der Lehr-Lernforschung vereinbar ist. Auf der Grundlage dieses Modells wurde in früheren



Arbeiten ein beobachtungsbasiertes Feedbackverfahren entwickelt und erfolgreich empirisch überprüft. Die vorliegende Studie erweitert den Ansatz nun um einen Schülerfragebogen. Dieser kann entweder als Ergänzung oder als Alternative zum Beobachtungsbogen verwendet werden.

Vor dem Hintergrund des Forschungsstands zu Schülerurteilen zum Unterricht wurde bei der Itemformulierung darauf geachtet, dass keine Selbstbewertung der Schüler:innen nahegelegt wird, dass Prozesse der Unterstützung komplexer kognitiver Aktivitäten als konkrete Aktivitäten der Lehrkraft abgefragt werden und die Items sprachlich leicht verständlich sind.

Die Ergebnisse der durchgeführten Analysen zeigen insgesamt gute Kennwerte, sodass den Lehrkräften mithilfe des Schülerfragebogens ein belastbares Klassenprofil zur Wahrnehmung des Unterrichts rückgemeldet werden kann. Dieses Profil unterscheidet inhaltlich eng gefasste Unterrichtsmerkmale als Stellschrauben für die Entwicklung der Unterrichtsqualität. Die konfirmatorische Prüfung konnte mehrheitlich eine gute Passung der empirischen Daten an das theoretische Modell bestätigen. Ausnahmen betreffen die beiden begründbaren Änderungen hinsichtlich des Unterrichtsklimas (als Trennung von Lehrkraft- und Schüler:innenverhalten) und der Individualisierung (als Zusammenfassung von Aspekten der individualisierten Unterstützung des Wissenserwerbs und der Motivierung). Die CFAs belegen die empirische Trennbarkeit der Unterrichtsmerkmale. Die einzige Ausnahme bildet hierbei die hohe Korrelation zwischen Weckung von Interesse und Förderung von Freude auf der Klassenebene. Die hohe Korrelation verweist darauf, dass gegenstandsbezogene Motivation (Interesse) und tätigkeitsbezogene Motivation (Spaß am Lernen) im Unterricht häufig nahe beieinanderliegen oder – darauf verweist auch die Theorie der organismischen Integration der Motivation von Deci und Ryan (1985) – ineinander übergehen können.

Auch die Reliabilitäten der Skalen fallen gut aus, wiederum mit einer Ausnahme (Steuerung des Unterrichtsflusses). Zusammenfassend ist festzuhalten, dass sowohl faktorielle Validität als auch Reliabilität jeweils bei fast allen Skalen gewährleistet sind. Somit genügt eine Unterrichtsevaluation basierend auf dem Schülerfragebogen wichtigen psychometrischen Anforderungen.

Wir betrachten die Befunde als Bestätigung der eingangs formulierten Annahme, dass eine fächerübergreifende Modellierung von Unterrichtsqualität auf Basis klassischer Bezugstheorien trennscharf operationalisiert werden kann. Die Operationalisierung von Items auf der Grundlage der zentralen theoretischen Ansätze der Informationsverarbeitung, der Motivation und der sozialen Interaktion scheint hinsichtlich der empirischen Trennbarkeit von Anforderungsbereichen des Unterrichts ein vielversprechender Ansatz zu sein. Die wahrgenommene Nützlichkeit der Operationalisierung für die Ableitung von konkreten Ansatzpunkten der Unterrichtsentwicklung aus Sicht von Lehrkräften muss weiterführend untersucht werden.

Einschränkend ist festzuhalten, dass die vorliegenden Analysen auf einer Stichprobe im berufsbildenden Bereich beruhen. Die Stichprobe ist somit hinsichtlich der beteiligten Fächer sehr heterogen, was im vorliegenden Fall die Überprüfung

der Messinvarianz über verschiedene Fächer hinweg nicht ermöglicht hat. Zukünftige Forschung mit dem Schülerfragebogen sollte somit folgende Fragen in den Blick nehmen:

- Replikation der Ergebnisse an einer Stichprobe aus allgemeinbildenden Schulen,
- Prüfung der Messinvarianz über Fächer, Jahrgangsstufen und Zeit,
- Validierung anhand von Außenkriterien, wie z. B. der Leistungs- oder Interessenentwicklung,
- Prüfung des Einsatzes des Schülerfragebogens in der Primarstufe (Lenke, 2016),
- Perspektivenvergleich zwischen Beobachtung und Schülerwahrnehmung.

Neben weiteren psychometrischen Analysen sollte außerdem explizit die Nützlichkeit von Unterrichtsevaluationen mithilfe des Schülerfragebogens überprüft werden. Nützlichkeit bezieht sich darauf, ob ein Unterrichtsfeedback basierend auf dem Schülerfragebogen zu einer Verbesserung der Unterrichtsgestaltung führt (Gärtner & Vogt, 2013). Im Rahmen der Studie zum beobachtungs-basierten Unterrichtsfeedback durch Schulleitungen hat sich der Vorteil der gut verständlichen Kategorisierung der drei Basisdimensionen sowie der Kommunikation der Einzelskalenergebnisse als Stellschrauben für die Unterrichtsentwicklung bestätigt (Kellermann et al., 2022). Entsprechend gehen wir davon aus, dass auch ein alleiniger Einsatz des Schülerfragebogens mit entsprechenden Erläuterungen des theoretischen Modells an die Lehrkräfte zu einer Verbesserung der Unterrichtsgestaltung führen kann (Gärtner & Vogt, 2013).

## Literatur

- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369–406. <https://doi.org/10.1037/0033-295X.89.4.369>
- Bandura, A. (1976). *Lernen am Modell*. Klett.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148. [https://doi.org/10.1207/s15326985ep2802\\_3](https://doi.org/10.1207/s15326985ep2802_3)
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., ... Tsai, Y.-M. (2008). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*. Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O. & Schnabel, K.-U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation – Band 1. Skalen Längsschnitt I, Wellen 1–4*. Max-Planck-Institut für Bildungsforschung.
- Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, 51(1), 5–32. <https://doi.org/10.3102/00346543051001005>
- Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive?* Waxmann.
- Clausen, M. & Göbel, K. (Hrsg.). (2020). Unterrichtsrückmeldungen durch Schüler\*innen [Themenheft]. *Empirische Pädagogik*, 34(1).
- Csikszentmihalyi, M. & Rathunde, K. (1993). The measurement of flow in everyday life: Toward a theory of emergent motivation. In R. A. Dienstbier & J. E. Jacobs

- (Hrsg.), *Nebraska symposium on motivation 1992. Developmental perspectives on motivation* (S. 57–97). University of Nebraska Press.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. Plenum. <https://doi.org/10.1007/978-1-4899-2271-7>
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson & C. S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice, and contemporary issues* (S. 97–125). Erlbaum.
- Eccles, J. S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Friedrich, H. F. & Mandl, H. (1992). Lern- und Denkstrategien – ein Problemaufriß. In H. F. Friedrich & H. Mandl (Hrsg.), *Lern- und Denkstrategien. Analyse und Intervention* (S. 3–54). Hogrefe.
- Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen: Ein Vergleich verschiedener Maße zur Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 24(2), 111–122. <https://doi.org/10.1024/1010-0652/a000009>
- Gärtner, H. & Brunner, M. (2018). Once good teaching, always good teaching? The differential stability of student perceptions of teaching quality. *Educational Assessment, Evaluation and Accountability*, 30(2), 159–182. <https://doi.org/10.1007/s11092-018-9277-5>
- Gärtner, H., Thiel, F., Nachbauer, M. & Kellermann, C. (2021). Unterrichtsentwicklung durch Unterrichtsfeedback – Erste Entwicklungsschritte eines Beobachtungsbogens für Schulleitungen. *Zeitschrift für Bildungsforschung*, 11(3), 529–547. <https://doi.org/10.1007/s35834-021-00321-7>
- Gärtner, H. & Vogt, A. (2013). Wie Lehrkräfte Ergebnisse eines Schülerfeedbacks verarbeiten und nutzen. *Unterrichtswissenschaft*, 41(3), 252–267.
- Göllner, R., Wagner, W., Klieme, E., Lüdtke, O., Nagengast, B. & Trautwein, U. (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. In Bundesministerium für Bildung und Forschung (Hrsg.), *Bildungsforschung Bd. 44: Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments* (S. 63–82). BMBF.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Waxmann.
- Hanisch, A.-K. (2018). *Kognitive Aktivierung im Rechtschreibunterricht: Eine Interventionsstudie in der Grundschule*. Waxmann.
- Helmke, A. (2017). *Schülerfragebogen zum Unterricht*. Verfügbar unter [www.unterrichtsdiagnostik.de/downloads/fragebogen/](http://www.unterrichtsdiagnostik.de/downloads/fragebogen/)
- Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. & Ade-Thurow, M. (2018). *Unterrichtsdiagnostik mit EMU. Evidenzbasierte Methoden der Unterrichtsdiagnostik Version 7.0*. Verfügbar unter [www.unterrichtsdiagnostik.de/media/files/Broschuere\\_Version\\_7.0.pdf](http://www.unterrichtsdiagnostik.de/media/files/Broschuere_Version_7.0.pdf)
- Heneman, H. G. & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(3), 171–195. <https://doi.org/10.1023/B:PEEV.0000032427.99952.02>
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

- Hulleman, C. S. & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326(5958), 1410–1412. <https://doi.org/10.1126/science.1177067>
- Jäger, R. & Helmke, A. (2008). *Projekt MARKUS Skalendokumentation*. Verfügbar unter [www.iqb.hu-berlin.de/fdz/studies/MARKUS/MARKUSSkalendoku.pdf](http://www.iqb.hu-berlin.de/fdz/studies/MARKUS/MARKUSSkalendoku.pdf)
- Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F. (2022). The effect of school leaders' feedback on teaching quality perceived by students – Results of an intervention study. *School Effectiveness and School Improvement. Advance online publication*. <https://doi.org/10.1080/09243453.2022.2075018>
- Klauer, K. J. & Leutner, D. (2012). *Lehren und lernen: Einführung in die Instruktionspsychologie*. Beltz.
- Klieme, E. & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 333–359). Springer. [https://doi.org/10.1007/978-3-322-97590-4\\_12](https://doi.org/10.1007/978-3-322-97590-4_12)
- Klieme, E. & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805–818.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford.
- Kounin, J. S. (2006). *Techniken der Klassenführung*. Waxmann.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., ... Neubrand, M. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. *Zeitschrift für Erziehungswissenschaft*, 8(4), 502–520. <https://doi.org/10.1007/s11618-005-0156-8>
- Lenske, G. (2016). *Schülerfeedback in der Grundschule*. Waxmann.
- Lenske, G. & Praetorius, A.-K. (2020). Schülerfeedback – was steckt hinter dem Kreuz auf dem Fragebogen? *Empirische Pädagogik*, 34(1), 11–29.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- 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(1/2), 85–96. <https://doi.org/10.1024/1010-0652.20.12.85>
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Erlbaum.
- MET Project. (2012). *Gathering feedback for teaching. Combining high-quality observations with student surveys and achievement gains*. Bill & Melinda Gates Foundation.
- Muthén, L. K. & Muthén, B. O. (1998–2017). *Mplus user's guide. Eighth edition*. Muthén & Muthén.
- Nationales Bildungspanel. (2013). *Startkohorte 3: Klasse 5 (SC3). Welle 2. Erhebungsinstrumente (Feldversion)*. Verfügbar unter [www.neps-data.de/Portals/o/NEPS/Datenzentrum/Forschungsdaten/SC3/5-0-0/SC3\\_Q\\_w2.pdf](http://www.neps-data.de/Portals/o/NEPS/Datenzentrum/Forschungsdaten/SC3/5-0-0/SC3_Q_w2.pdf)
- Ophardt, D. & Thiel, F. (2013). *Klassenmanagement. Ein Handbuch für Studium und Praxis*. Kohlhammer.
- Pianta, R. C., La Paro, K. & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS)*. Paul H. Brookes.
- Piowar, V. (2013). Multidimensionale Erfassung von Kompetenzen im Klassenmanagement: Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe. *Zeitschrift für Pädagogische Psychologie*, 27(4), 215–228. <https://doi.org/10.1024/1010-0652/a000108>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM –*

- Mathematics Education*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- 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. <https://doi.org/10.1016/j.learninstruc.2013.12.002>
- Prediger, S. & Neugebauer, P. (2021). Capturing teaching practices in language-responsive mathematics classrooms. Extending the TRU framework “teaching for robust understanding” to L-TRU. *ZDM – Mathematics Education*, 53(2), 289–304. <https://doi.org/10.1007/s11858-020-01187-1>
- Rakoczy, K., Buff, A. & Lipowsky, F. (2005). Teil 1: Befragungsinstrumente. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. GFFP.
- Rheinberg, F. (2008). *Motivation*. Kohlhammer.
- Röhl, S. & Rollett, W. (2020). Alles nur die Nettigkeit der Lehrkraft? Die Communion der Lehrkräfte als Erklärung für den Halo-Bias in Schülerbefragungen zur Unterrichtsqualität. *Empirische Pädagogik*, 34(1), 30–45.
- Schiefele, U. (2009). Situational and individual interest. In K. R. Wenzel & A. Wigfield (Hrsg.), *Educational psychology handbook series. Handbook of motivation at school* (S. 197–222). Routledge/Taylor & Francis Group.
- Seidel, T., Prenzel, M., Duit, R. & Lehrke, M. (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Leibniz-Institut für die Pädagogik der Naturwissenschaften.
- Steffensky M. & Neuhaus B. J. (2018). Unterrichtsqualität im naturwissenschaftlichen Unterricht. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 299–313). Springer. [https://doi.org/10.1007/978-3-662-56320-5\\_18](https://doi.org/10.1007/978-3-662-56320-5_18)
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- Thiel, F. (2016). *Interaktion im Unterricht*. Barbara Budrich. <https://doi.org/10.36198/9783838545714>
- Thiel, F., Ophardt, D. & Piowar, V. (2013). *Kompetenzen des Klassenmanagements. Schülerfragebogen*. Verfügbar unter [www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/forschung/abgelaufeneprojekte/kodek/kodek/Buch\\_Dokumente/KODEK-S\\_Lehrerin.pdf](http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/forschung/abgelaufeneprojekte/kodek/kodek/Buch_Dokumente/KODEK-S_Lehrerin.pdf)
- Thiel, F., Richter, S. G. & Ophardt, D. (2012). Steuerung von Übergängen im Unterricht. Eine Experten-Novizen-Studie zum Klassenmanagement. *Zeitschrift für Erziehungswissenschaft*, 15(4), 727–752. <https://doi.org/10.1007/s11618-012-0325-5>
- Thiel, F., Tarkian, J., Lankes, E.-M., Maritzen, N., Ricke-Baulecke, T. & Kroupa, A. (Hrsg.). (2019). *Datenbasierte Qualitätssicherung und -entwicklung in Schulen. Eine Bestandsaufnahme in den Ländern der Bundesrepublik Deutschland*. Springer. <https://doi.org/10.1007/978-3-658-23240-5>
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht – am Beispiel der Studie DESI der Kultusministerkonferenz*. Universität Koblenz-Landau.
- 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>

## Appendix

Tabelle A1: Itemstatistiken (Mittelwert, Standardabweichung und Faktorladung  $\lambda$  aus CFA und CFA<sub>2level</sub>) zu MZP 1 und MZP 2

Item	MZP 1				MZP 2						
	N	M	SD	$\lambda_{CFA}$	$\lambda_{CFA-level_1}$	$\lambda_{CFA-level_2}$	N	M	SD	$\lambda_{CFA}$	
Unterstützung des Wissenserwerbs – Erklärung und Präsentation											
Unsere Lehrkraft präsentiert neues Wissen klar und strukturiert.	1287	3.21	0.70	.73	.64	.99	293	3.27	0.67	.79	
Unsere Lehrkraft kann schwierige Themen gut erklären.	1271	3.12	0.74	.75	.69	.98	294	3.24	0.72	.82	
Unsere Lehrkraft bringt anschauliche Beispiele, die uns das Verstehen der Lernerhalte erleichtern.	1269	3.21	0.75	.66	.59	.88	294	3.36	0.64	.67	
Wenn unsere Lehrkraft uns etwas erklärt, wissen wir immer ganz genau, was sie meint.							292	2.98	0.67	.73	
Unterstützung des Wissenserwerbs – Unterstützung des Verstehens											
Unsere Lehrkraft lässt uns unsere Aufgabenlösungen selbst korrigieren.	1245	2.64	0.90	.52	.47	.79	293	3.02	0.75	.65	
Unsere Lehrkraft stellt Aufgaben, bei denen mehrere Lösungswege möglich sind.	1248	2.79	0.84	.66	.63	.86	294	3.14	0.65	.65	
Unsere Lehrkraft lässt unterschiedliche Lösungswege von Aufgaben vergleichen und bewerten.	1254	2.88	0.82	.71	.68	1.00	292	3.14	0.74	.64	
Bei unserer Lehrkraft kann ich zum Lösen schwieriger Aufgaben meine eigenen Strategien einsetzen.	1242	2.98	0.80	.56	.59	.45	292	3.16	0.75	.67	
Unsere Lehrkraft stellt Aufgaben, für deren Lösung man sich Zeit zum Nachdenken nehmen muss.							293	2.81	0.88	.53	
Unterstützung des Wissenserwerbs – Speicherung durch Üben											
Wir erhalten gezielt Übungsaufgaben für besonders schwierige Themen.	1255	2.89	0.81	.58	.53	.90	294	3.09	0.80	.65	
Im Unterricht gibt es regelmäßig Phasen, in denen wir Übungsaufgaben bearbeiten.	1264	3.19	0.76	.54	.51	.86	294	3.25	0.79	.71	
Zu Übungsaufgaben oder Hausaufgaben erhalten wir immer eine Rückmeldung, was an unseren Lösungen richtig oder falsch ist.	1260	3.07	0.86	.53	.49	.84	293	3.11	0.85	.64	
Unter den Übungsaufgaben, die uns unsere Lehrkraft gibt, sind oft Aufgaben, bei denen man sieht, ob man etwas wirklich verstanden hat.	1254	2.98	0.78	.68	.63	1.00	293	3.15	0.70	.70	
Die Übungsaufgaben im Unterricht unserer Lehrkraft sind ähnlich, aber doch immer wieder anders, so dass man genau aufpassen muss.	1248	2.96	0.77	.62	.55	.98	290	3.06	0.71	.61	
Unterstützung des Wissenserwerbs – Anwendung und Transfer											
Bei unserer Lehrkraft nutzen wir unser Wissen zur Bearbeitung von Alltagsproblemen.	1270	2.63	0.87	.61	.57	.67	293	2.85	0.86	.70	
Unsere Lehrkraft gibt uns Problemaufgaben, in denen wir unser Wissen anwenden müssen.	1261	2.94	0.79	.57	.57	.90	294	3.22	0.69	.66	

Tabelle A1 wird fortgesetzt

Fortsetzung Tabelle A1

Item	MZIP 1					MZIP 2				
	N	M	SD	$\lambda_{CFA}$	$\lambda_{CFA-level_1}$	$\lambda_{CFA-level_2}$	N	M	SD	$\lambda_{CFA}$
Wir machen im Unterricht kleine Projekte, bei denen wir das, was wir gelernt haben, anwenden können.	1256	2.74	0.95	.56	.51	.67	294	2.89	0.87	.55
Unsere Lehrkraft stellt uns häufig Aufgaben, bei denen wir unser Wissen auf neue Situationen übertragen müssen.	1258	2.83	0.79	.70	.67	.99	294	3.09	0.72	.81
Wenn wir im Unterricht unserer Lehrkraft üben, wenden wir das Gelehrte oft auf andere Dinge an.	1244	2.83	0.77	.62	.54	.93	293	2.99	0.75	.67
Motivierung – Weckung von Interesse										
Unsere Lehrkraft kann Schülerinnen und Schüler manchmal richtig begeistern.	1263	2.91	0.86	.76	.70	.94	293	3.02	0.82	.79
Unsere Lehrkraft kann auch trockene Themen wirklich interessant machen.	1267	2.80	0.85	.80	.73	1.00	293	2.97	0.81	.79
Auch bei bekannten Themen überrascht uns unsere Lehrkraft oft mit neuem Wissen.	1256	2.89	0.82	.72	.65	.95	292	3.05	0.77	.78
Unsere Lehrkraft steckt uns mit ihrer Begeisterung an.	1257	2.74	0.87	.79	.73	.98	293	2.92	0.87	.85
Unsere Lehrkraft gestaltet den Unterricht oft sehr spannend.	1263	2.76	0.88	.82	.77	.98	292	2.90	0.79	.80
Die von unserer Lehrkraft behandelten Inhalte sind richtig spannend.	1266	2.64	0.88	.79	.74	.93	292	2.74	0.81	.78
Motivierung – Förderung von Freude										
Die Bearbeitung der Aufgaben, die unsere Lehrkraft stellt, macht mir Spaß.	1261	2.76	0.85	.83	.78	.98	292	2.80	0.79	.83
Bei unserer Lehrkraft ist man oft so mit Aufgaben beschäftigt, dass man gar nicht merkt, wie die Zeit vergeht.	1257	2.82	0.91	.72	.64	.96	291	2.75	0.85	.73
Im Unterricht unserer Lehrkraft bin ich mit Freude dabei.	1261	2.94	0.84	.85	.81	1.00	289	2.98	0.80	.83
Der Unterricht unserer Lehrkraft macht mir Spaß.	1259	2.98	0.87	.86	.82	.98	291	3.01	0.79	.86
Motivierung – Verdeutlichung des Nutzens										
Unsere Lehrkraft sagt uns, wofür die Lerninhalte nützlich sind.	1264	3.15	0.77	.73	.69	.93	293	3.35	0.71	.80
Unsere Lehrkraft informiert uns darüber, dass die Lerninhalte wichtig für andere Themen sind.	1261	3.12	0.77	.69	.62	.94	293	3.38	0.66	.74
Unsere Lehrkraft erklärt uns, warum es für das spätere Leben wichtig ist, das Unterrichtsthema zu lernen.	1255	2.96	0.91	.75	.69	.88	292	3.19	0.83	.81
Unsere Lehrkraft macht deutlich, dass das Unterrichtsthema auch im Alltag wichtig ist.	1252	2.86	0.91	.73	.67	.84	291	3.09	0.83	.77
Motivierung – Förderung von Selbstwirksamkeit										
Unsere Lehrkraft muntert uns auf, damit wir auch bei schwierigen Aufgaben nicht den Mut verlieren.	1251	2.90	0.84	.78	.71	.97	291	3.01	0.81	.78
Unsere Lehrkraft lobt unsere Fortschritte.	1252	2.99	0.83	.74	.70	.97	291	3.19	0.84	.80

Tabelle A1 wird fortgesetzt

Fortsetzung Tabelle A1

Item	MZIP 1					MZIP 2				
	N	M	SD	$\lambda_{CFA}$	$\lambda_{CFA-level_1}$	$\lambda_{CFA-level_2}$	N	M	SD	$\lambda_{CFA}$
Unsere Lehrkraft zeigt uns, welche Fortschritte wir durch Anstrengung erreichen können.	1241	2.85	0.84	.73	.69	.94	292	3.07	0.84	.82
Unsere Lehrkraft traut uns etwas zu.	1254	3.17	0.74	.67	.64	.90	291	3.34	0.74	.78
Klassenmanagement – Lerndienliches Unterrichtsklima SuS										
Im Unterricht gehen die Schülerinnen und Schüler respektvoll miteinander um.	1260	3.14	0.85	.74	.81	.98	291	3.16	0.80	.83
Im Klassenzimmer nehmen alle aufeinander Rücksicht.	1251	3.00	0.92	.81	.86	1.00	291	3.00	0.86	.90
Im Unterricht beachten die Schülerinnen und Schüler die Regeln für einen guten Umgang miteinander.							292	3.08	0.77	.90
Wenn es im Unterricht zu Konflikten kommt, versuchen die Schülerinnen und Schüler, gemeinsam eine Lösung zu finden.							291	3.03	0.85	.75
Klassenmanagement – Lerndienliches Unterrichtsklima LK <sup>1</sup>										
Unsere Lehrkraft hat eine gute Beziehung zu den Schülerinnen und Schülern.	293	3.35	0.67				293	3.35	0.67	.72
Unsere Lehrkraft verhält sich uns gegenüber fair.	293	3.42	0.70				293	3.42	0.70	.83
Wir vertrauen unserer Lehrkraft.	291	3.40	0.74				291	3.40	0.74	.86
Unsere Lehrkraft ist immer freundlich zu den Schülerinnen und Schülern.	290	3.43	0.65				290	3.43	0.65	.81
Klassenmanagement – Monitoring										
Unsere Lehrkraft bekommt mit, was die Schülerinnen und Schüler im Klassenzimmer machen.	1254	3.20	0.79	.74	.72	.97	293	3.27	0.77	.86
Unsere Lehrkraft merkt, was beim Rest der Klasse vorgeht, auch wenn sie mit einzelnen Schülerinnen und Schülern beschäftigt ist.	1248	3.04	0.79	.81	.77	.99	293	3.17	0.76	.85
Unsere Lehrkraft merkt sofort, wenn Schülerinnen und Schüler beginnen, sich mit etwas anderem zu beschäftigen.	1245	2.96	0.84	.79	.77	.97	293	3.16	0.77	.78
Unsere Lehrkraft hat immer die ganze Klasse im Blick.	1248	3.07	0.82	.81	.78	1.00	291	3.15	0.79	.86
Klassenmanagement – Steuerung des Unterrichtsflusses										
Bei unserer Lehrkraft kommt es im Unterricht häufig zu Verzögerungen und Wartezeiten. (umgepolt)	1251	2.82	0.88	.69	.67	.82	292	2.14	0.80	.71
Unsere Lehrkraft schweift häufig vom Thema ab. (umgepolt)	1241	2.98	0.89	.68	.66	.77	291	1.93	0.81	.61
Bei unserer Lehrkraft nehmen Übergänge von einem Unterrichtsabschnitt zum nächsten (z. B. vom Klassengespräch zur Stillarbeit) oft viel Zeit in Anspruch. (umgepolt)	1240	2.69	0.85	.60	.58	.89	293	1.93	0.75	.62
Bei unserer Lehrkraft dauert es zu Beginn der Stunde sehr lange, bis die Schülerinnen und Schüler ruhig werden und zu arbeiten beginnen. (umgepolt)	1260	3.02	0.94	.75	.67	.99	291	1.70	0.79	.73

Tabelle A1 wird fortgesetzt



Fortsetzung Tabelle A1

Item	MZIP 1					MZIP 2					
	N	M	SD	$\lambda_{CFA}$	$\lambda_{CFA-level_1}$	$\lambda_{CFA-level_2}$	N	M	SD	$\lambda_{CFA}$	
<b>Klassenmanagement – Gruppenaktivierung</b>											
Im Unterricht unserer Lehrkraft beteiligen sich meist alle Schülerinnen und Schüler.	1253	2.74	0.77	.74	.68	.96	293	2.75	0.73	.81	
Bei unserer Lehrkraft nimmt meist die gesamte Klasse aufmerksam am Unterrichtsgeschehen teil.	1248	2.76	0.81	.81	.78	1.00	294	2.97	0.73	.85	
Unsere Lehrkraft schafft es, dass sich meist alle Schülerinnen und Schüler mit dem Unterrichtsthema beschäftigen.							290	2.85	0.76	.86	
Bei unserer Lehrkraft werden alle Schülerinnen und Schüler in den Unterricht mit einbezogen.							292	3.10	0.72	.67	
<b>Klassenmanagement – Umgang mit Störungen</b>											
Bei unserer Lehrkraft können wir im Unterricht ungestört arbeiten.	1258	3.06	0.80	.67	.61	.95	294	3.10	0.74	.60	
Unsere Lehrkraft reagiert angemessen auf Störungen (also nicht zu streng und nicht zu mild).	1242	3.11	0.85	.65	.61	.77	291	3.24	0.74	.67	
Unsere Lehrkraft gelingt es, Störungen schnell zu beenden.	1250	3.15	0.81	.84	.81	.99	293	3.29	0.73	.88	
Unsere Lehrkraft reagiert so auf Störungen, dass die Schülerinnen und Schüler aufhören zu stören.	1250	3.11	0.81	.81	.78	.97	292	3.28	0.73	.87	
<b>Individualisierung</b>											
Unsere Lehrkraft gibt uns je nach Leistungsstärke unterschiedliche Aufgaben.	1253	1.96	0.99	.69			294	2.21	0.92	.76	
Unsere Lehrkraft gibt uns je nach Interesse unterschiedliche Aufgaben.	1244	2.07	0.94	.83			293	2.41	0.96	.69	
Unsere Lehrkraft gibt uns Inhalte und Themen zur Auswahl.	1231	2.20	0.95	.70			294	2.57	0.94	.72	
Unsere Lehrkraft gibt guten Schülerinnen und Schülern Extraaufgaben, die schwieriger sind.							293	2.22	0.94	.84	
Wenn Schülerinnen und Schüler etwas nicht verstehen, erhalten sie zusätzliche Aufgaben, um noch einmal zu üben.							292	2.43	0.94	.78	
Unsere Lehrkraft stellt verschiedene Aufgaben, so dass alle Schülerinnen und Schüler Spaß am Lernen haben.							294	2.68	0.89	.74	

Anmerkungen: MZIP = Messzeitpunkt; CFA = konfirmatorische Faktorenanalyse;  $\lambda_{CFA}$  = Faktorladungen aus Gesamtmodellprüfung auf Schülerebene;  $\lambda_{CFA-level_1}$  = Faktorladungen auf Schülerebene aus Modellprüfung pro Anforderungsdimension mit 2 Ebenen;  $\lambda_{CFA-level_2}$  = Faktorladungen auf Klassenebene aus Modellprüfung pro Anforderungsdimension mit 2 Ebenen.

Tabelle A2: Faktorinterkorrelationen zu MZP 1 (unterhalb der Diagonale) und MZP 2 (über der Diagonale) aus CFA auf Schülerebene

Unterrichtsmerkmal	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Unterstützung des Wissenserwerbs															
F1 Erklärung und Präsentation		.71	.74	.67	.79	.74	.64	.81	.39	.71	.64	.51	.62	.63	.39
F2 Unterstützung des Verstehens	.49		.89	.79	.62	.58	.64	.68	.24	.64	.56	.37	.46	.49	.48
F3 Speicherung durch Üben	.72	.65		.80	.63	.61	.75	.76	.23	.72	.56	.45	.55	.61	.53
F4 Anwendung und Transfer	.63	.59	.74		.68	.68	.74	.65	.33	.60	.50	.38	.51	.51	.49
Motivierung															
F5 Weckung von Interesse	.76	.54	.65	.70		.90	.64	.80	.33	.65	.59	.40	.70	.59	.53
F6 Förderung von Freude	.71	.46	.59	.59	.89		.60	.73	.40	.61	.52	.44	.64	.54	.46
F7 Verdeutlichung des Nutzens	.60	.44	.63	.71	.69	.64		.72	.30	.63	.54	.25	.49	.47	.45
F8 Förderung von Selbstwirksamkeit	.74	.55	.67	.61	.79	.75	.68		.29	.80	.67	.36	.63	.63	.55
Klassenmanagement															
F9 Lerndienliches Unterrichtsklima SuS	.30	.19	.24	.22	.28	.28	.19	.29		.32	.39	.43	.45	.45	.25
F10 Lerndienliches Unterrichtsklima LK											.64	.35	.46	.58	.35
F11 Monitoring	.61	.39	.53	.49	.61	.58	.56	.61	.33			.39	.53	.69	.30
F12 Steuerung des Unterrichtsflusses	.30	.02	.17	.02	.10	.12	-.02	.12	.23	.23	.23		.42	.54	.17
F13 Gruppenaktivierung	.58	.36	.46	.48	.63	.64	.51	.62	.50	.57	.57	.17		.62	.46
F14 Umgang mit Störungen	.62	.36	.54	.45	.53	.52	.44	.59	.49	.64	.64	.36	.68		.35
F15 Individualisierung	.33	.48	.40	.54	.54	.48	.50	.45	.17	.31	-.25	.46	.24		

Anmerkungen. MZP = Messzeitpunkt; CFA = konfirmatorische Faktorenanalyse; SuS = Schüler:innen; LK = Lehrkraft.

Tabelle A3: Faktorinterkorrelationen zu MZP 1 auf Klassenebene

Unterrichtsmerkmal	F1	F2	F3	F4	F5	F6	F7	F8	F9	F11	F12	F13	F14
Unterstützung des Wissenserwerbs													
F1 Erklärung und Präsentation													
F2 Unterstützung des Verstehens	.59												
F3 Speicherung durch Üben	.84	.69											
F4 Anwendung und Transfer	.85	.45	.80										
Motivierung													
F5 Weckung von Interesse													
F6 Förderung von Freude					.97								
F7 Verdeutlichung des Nutzens					.90	.90							
F8 Förderung von Selbstwirksamkeit					.90	.87	.82						
Klassenmanagement													
F9 Lerndienliches Unterrichtsklima SuS													
F11 Monitoring									.40				
F12 Steuerung des Unterrichtsflusses									.61	.39			
F13 Gruppenaktivierung									.72	.80	.50		
F14 Umgang mit Störungen									.70	.83	.75	.86	

Anmerkungen. MZP = Messzeitpunkt; SuS = Schüler:innen.