

Unterrichtsqualität aus der Perspektive von Schülerinnen und Schülern

Von der Fakultät Bildung
der Leuphana Universität Lüneburg zur Erlangung des Grades

Doktor der Mathematikdidaktik
- Dr. phil. -

genehmigte Dissertation von
Stefan Blanck

geboren am 22. November 1981 in Staaken (Berlin)

Eingereicht am: 21. November 2019

Mündliche Verteidigung (Disputation) am: 12. Juni 2020

Erstbetreuer und Erstgutachter:

Professor Dominik Leiß

Zweitgutachter:

Professor Olaf Köller

Drittgutachter:

Professor Thilo Kleickmann

Druckjahr: 2020

Danksagung

Diese Dissertation wäre nicht ohne die wohlwollende, inspirierende und zielführende Unterstützung von Olaf Köller und Thilo Kleickmann möglich gewesen. Sie waren von Anfang an zwei großartige Mentoren, die mich von den ersten Schritten an vertrauensvoll durch die Täler voller Zweifel getragen haben und mich vorbehaltlos in allen Belangen geduldig beraten haben. Ohne festen Anschluss an eine Arbeitsgruppe waren Sie mir geduldige Sparringspartner für wissenschaftliche Diskussionen und meinen eigenen Erkenntnisgewinn. Besser hätte eine Betreuung nicht sein können - vielen herzlichen Dank dafür! Genauso wohlwollend und fördernd hat mich Dominik Leiss unterstützt und meine Promotionsbetreuung von der ersten Minute an vorbehaltlos übernommen. Gemeinsam mit Michael Besser führte er mich durch die formalen Anforderungen der Leuphana Universität Lüneburg. Eine Dissertation neben einer Vollzeitarbeit als Lehrer und als junger Familienvater kann ohne den Pragmatismus und den Blick für das Wesentliche, wie ihn diese vier in meinem Fall gezeigt haben, nicht angefertigt werden! Diese persönlichen Erfahrungen mit diesen vier Lehrenden versöhnen mich mit zum Teil tiefen Wunden aus meinen frühen Studienjahren.

Um meine Studie durchführen zu können, haben mir meine Kollegen Tobias und Christian sowie meine Kollegin Lena einen vertrauensvollen Einblick in ihren Unterricht gewährt. Sie haben sich noch intensiver und enger miteinander abgestimmt und diverse Stunden ihrer Unterrichtszeit meiner Forschung zur Verfügung gestellt. Ich danke euch dafür! Die Durchführung der Studie an meiner Schule ermöglichte darüberhinaus das Wohlwollen des Leitungsteams, das durch seinen hohen Innovationsanspruch einen guten Nährboden für die Feldforschung schuf und mir so manches Mal mit einer flexiblen Arbeitszeitgestaltung entgegen kam. Ich bin mir sicher, mir war das Schreiben der Arbeit nur an dieser Schule möglich. Dafür bin ich dankbar! Für redaktionelle und handwerkliche Fragen haben mir Sina und Luis viele Stunden ihrer Zeit geschenkt und mir damit lästige Sorgen abgenommen. Ich danke euch!

Schließlich möchte ich mich bei all denjenigen bedanken, die mich während der letzten Jahre begleitet und auch ohne professionelles Interesse an meiner Arbeit unterstützt

haben: meinen guten Freunden und meiner Familie, die in Telefonaten oder bei Treffen all die Details meiner Arbeit immer wieder anhören mussten und mich aufgemuntert haben, wenn ich mit der einen oder anderen Herausforderung konfrontiert war. Euer Zuhören hat mir eine Arbeitsgruppe ersetzt und mir oft geholfen, meine Gedanken zu sortieren.

Und dann sind da noch meine liebe Heike und mein lieber Anton, ihr zwei habt oft den Preis meiner nächtlichen Schreibtischarbeit bezahlt oder mir die wertvollen Sonntagnachmittage zur Verfügung gestellt, um diese Dissertation fertigzustellen. Ich weiß, dass ich diese Arbeit nicht für euch anfertigen musste. Ihr habt mich unterstützt, weil ihr mich liebt und mir das Projekt so wichtig war. Ihr verdient dafür nicht nur meinen größten Dank, sondern auch eine Entschuldigung für die ganze Mühe. Ich liebe euch!

Hamburg, November 2019

Stefan Blanck

Inhaltsverzeichnis

1	Zusammenfassung	1
2	Theoretischer Hintergrund	5
2.1	Einführung	5
2.2	Einordnung der eigenen Arbeit	6
2.3	Unterrichtsqualität	8
2.3.1	Basisdimension 1 Kognitive Aktivierung	12
2.3.2	Basisdimension 2 Instruktionale bzw. Kognitive Unterstützung	13
2.3.3	Basisdimension 3 Emotionale Unterstützung	13
2.3.4	Basisdimension 4 Effizienz der Klassenführung	14
2.3.5	Zusammenfassung	16
2.4	Unterrichtswahrnehmungen	17
2.4.1	Erfassung von Unterrichtsqualität	17
2.4.2	Die Perspektive von Schülerinnen und Schülern	20
2.4.3	Stabilität der Unterrichtswahrnehmung	21
2.5	Einfluss von Lehrkräftekompetenzen auf Lernleistungen	24
2.6	Unterrichtswahrnehmung und Leistung	27
3	Ableitung der Fragestellung und Hypothesen	29
3.1	Stabilität der Unterrichtswahrnehmung	29
3.2	Unterrichtswahrnehmung und Lernleistungen	30
4	Methode	33
4.1	Überblick über das Vorgehen	33
4.2	Beschreibung der Stichprobe	37
4.3	Beschreibung der Instrumente	39
4.3.1	Der Fragebogen	39
4.3.2	Reliabilität	44
4.3.3	Leistungsmessung	46
4.3.4	Ergebnisse der Leistungsmessung deskriptiv	48

4.3.5	Modellfits allgemein	49
5	Ergebnisse	51
5.1	Stabilität der Unterrichtswahrnehmung	52
5.1.1	Stabilität des Qualitätsniveaus	53
5.1.2	Stabilität der Rangfolge	62
5.1.3	Stabilität zwischen und innerhalb der Unterrichtsthemen	69
5.2	Unterrichtswahrnehmung und Leistung	71
5.2.1	Lineare Strukturgleichungsmodelle mit Kovariaten	71
5.2.2	Lineare Strukturgleichungsmodelle und Korrelation zur Leistung	75
5.2.3	Regressionsanalysen zur Leistung und Unterrichtsqualität	76
6	Diskussion der Ergebnisse	87
6.1	Diskussion der Stabilität der Unterrichtswahrnehmung	87
6.1.1	Stabilität des Qualitätsniveaus	87
6.1.2	Stabilität der Rangfolge	89
6.1.3	Stabilität zwischen und innerhalb der Unterrichtsthemen	89
6.2	Diskussion Unterrichtswahrnehmung und Leistung	90
6.3	Diskussion des methodischen Vorgehens	91
6.4	Bilanz und Ausblick	92
A	Literaturverzeichnis	99
B	Itemanalyse	109
C	Fragebogen	135

Abbildungsverzeichnis

2.1	Basisdimensionen guten Unterrichts und deren Wirkungen	16
2.2	Vollständiges Vermittlungsmodell für die Kompetenz von Lehrkräften . .	26
4.1	Darstellung des zeitlichen Ablaufs der Studie im Schuljahr 2016/17 . . .	35
4.2	Darstellung der Reliabilitätskonstante α zu den 8 Messzeitpunkten	45
5.1	Lineares Strukturgleichungsmodell der Kognitiven Aktivierung	57
5.2	Lineares Strukturgleichungsmodell der Instrukionalen Unterstützung . .	58
5.3	Lineares Strukturgleichungsmodell der Emotionalen Unterstützung	59
5.4	Lineares Strukturgleichungsmodell der Klassenführung	60
5.5	Mittelwerte der Kognitiven Aktivierung der 3 Klassen	67
5.6	Mittelwerte der Instrukionalen Unterstützung der 3 Klassen	68
5.7	Mittelwerte der Emotionalen Unterstützung der 3 Klassen	68
5.8	Mittelwerte der Klassenführung der 3 Klassen	69
5.9	Pfaddiagramm zur Kognitiven Aktivierung	78
5.10	Pfaddiagramm zur Instrukionalen Unterstützung	80
5.11	Pfaddiagramm zur Emotionalen Unterstützung	82
5.12	Pfaddiagramm zur Klassenführung	84

Tabellenverzeichnis

2.1	Spanne der verwertbaren Koeffizienten aus verschiedenen Studien	19
4.1	Skala 1 - Kognitive Aktivierung	40
4.2	Skala 2 - Instruktionale Unterstützung	41
4.3	Skala 3 - Emotionale Unterstützung	42
4.4	Skala 4 - Effizienz der Klassenführung	43
4.5	Skala - Mathematikbezogenes Selbstkonzept	43
4.6	Interne Konsistenz zu jedem Messzeitpunkt für die Gesamtstichprobe . .	44
4.7	Reliabilität der Selbstkonzeptskala	45
4.8	Deskriptive Darstellung der Leistung	48
4.9	Korrelationen der jeweiligen Leistungsmessungen	49
5.1	Korrelationen zwischen den Qualitätsdimensionen über die Messzeiten . .	52
5.2	Deskriptive Darstellung der Kognitiven Aktivierung	53
5.3	Deskriptive Darstellung der Instruktionalen Unterstützung	53
5.4	Deskriptive Darstellung der Emotionalen Unterstützung	54
5.5	Deskriptive Darstellung der Klassenführung	54
5.6	Zentrierte Korrelationen der Qualitätsdimension Kognitive Aktivierung .	62
5.7	Zentrierte Korrelationen der Qualitätsdimension Instruktionale Unter- stützung	63
5.8	Zentrierte Korrelationen der Qualitätsdimension Emotionale Unterstützung	64
5.9	Zentrierte Korrelationen der Qualitätsdimension Klassenführung	65

1 Zusammenfassung

Die Beurteilung von Unterrichtsqualität stellt in der schulischen Praxis eine Schwierigkeit dar, weil sie eng mit der Frage danach, wer den Unterricht bewertet, verknüpft ist. Üblicherweise schätzen Lehrkräfte ihren Unterricht selbst ein. Seltener wird Unterrichtsqualität von geschulten, externen Beobachtern beurteilt. Eine weitere relevante Perspektive auf die Qualität des gehaltenen Unterrichts stellt die der Schülerinnen und Schüler dar. Die Qualität dieser Perspektive steht im Fokus dieser Arbeit.

Der Begriff Unterrichtsqualität gliedert sich im deutschsprachigen Raum in drei Qualitätsdimensionen auf: die Kognitive Aktivierung, die Konstruktive Unterstützung und die Klassenführung. In dieser Arbeit wird die Unterstützungsdimension aufgefächert in zwei Qualitätsdimensionen, in die Instruktionale Unterstützung und die Emotionale Unterstützung. So ergeben sich vier Basisdimensionen von Unterrichtsqualität, die aus der Perspektive von Schülerinnen und Schülern in dieser Arbeit untersucht werden sollen.

Die überwiegende Mehrheit von Studien zur Unterrichtsqualität geht davon aus, dass die Lehrqualität von Lehrerinnen und Lehrern ein stabiles Verhaltensmuster der Lehrperson ist (Wagner et al., 2015). Die Anzahl der beobachteten Stunden, die für die reliable und valide Feststellung der Lehrqualität nötig ist, wird in den verschiedenen Studien unterschiedlich eingeschätzt. Nach Brophy (2006) sind es mindestens 20 bis 30 Unterrichtsstunden. Praetorius (2014) kommt zu dem Ergebnis, dass die unterrichtliche Basisdimension der Klassenführung sehr stabil ist und es zur Einschätzung nur einer einzigen Unterrichtsstunde bedarf. Die Basisdimension der Kognitiven Aktivierung hingegen benötigt mindestens 9 beobachtete Unterrichtsstunden, um ein verlässliches Reliabilitätsniveau zu erhalten. Die Validität solcher Aussagen wird dann nicht selten an ihrem Zusammenhang mit Leistungen der Schülerinnen und Schüler festgestellt. In aller Regel beziehen sich die gewonnenen Erkenntnisse auf die Klassenebene.

Der bisherige Wissensstand zur Stabilität des Qualitätsniveaus von Unterricht in den einzelnen Qualitätsdimensionen ist noch unzureichend und nicht in jeder Qualitätsdimension umfassend erforscht. Weiterer Forschungsbedarf besteht in der Untersuchung

der Stabilität der Rangfolge interindividueller Unterschiede zwischen den Qualitätsdimensionen bei der Beurteilung durch Schülerinnen und Schüler. Bisherige Studien haben bisher nur unzureichend untersucht, ob es einen Zusammenhang zwischen den Unterrichtsthemen und der Beurteilung der Qualitätsdimensionen durch Schülerinnen und Schüler gibt. Die meisten Studien in dem Forschungsfeld stützen ihre Ergebnisse auf Erkenntnisse, die auf Durchschnitten der gesamten Klasse beruhen, nicht auf Individualergebnissen. Die hier vorliegende Arbeit knüpft mit ihren Fragestellungen an diese Wissenslücken an. Es werden dabei zwei Hauptfragestellungen untersucht. Zum einen wird die Stabilität der Unterrichtswahrnehmung in folgenden drei Teilfragestellungen untersucht:

- Wie stabil ist das beobachtete Qualitätsniveau in den einzelnen Dimensionen?
- Wie stabil sind die Einschätzungen der Unterrichtsqualität in den vier Dimensionen Kognitive Aktivierung, Instruktionale Unterstützung, Emotionale Unterstützung und Klassenführung über die Zeit?
- Sind die Beurteilungen der Unterrichtsqualität themenunabhängig?

Zum anderen wird der Frage

- Welchen Zusammenhang zeigen Mathematikleistungen und das mathematikbezogene Selbstkonzept mit den Unterrichtsbeurteilungen der Schülerinnen und Schüler?

nachgegangen.

Die vorliegende Arbeit wurde in 8 längsschnittlichen Erhebungen mit einer Gruppe von in drei Klassen parallel unterrichteten Fünftklässlerinnen und Fünftklässlern eines Hamburger Gymnasiums durchgeführt ($N = 85$). Das Instrument zur Erfassung der Unterrichtsqualität aus der Perspektive von Schülerinnen und Schülern in den vier Basisdimensionen Kognitive Aktivierung (7 Items), Instruktionale und Emotionale Unterstützung (9 Items und 5 Items) sowie Klassenführung (5 Items) wurde auf der Grundlage der Skalen aus Fauth, Decristan, Rieser, Klieme und Büttner (2014b) und Kauertz et al. (2011) entwickelt. Das Instrument kam zu acht Messzeitpunkten zum Einsatz. Die Skala zur Erfassung des mathematikbezogenen Selbstkonzeptes (4 Items) der Schülerinnen und Schüler stammt aus Bos, Dudas, Gröhlich, Guill und Scharenberg (2010) und wurde zu Beginn und am Ende der Messreihe einmal verwendet. Die Reliabilitäten der jeweiligen Skalen zu den jeweiligen Messzeitpunkten nimmt immer akzeptable (Cronbachs $\alpha > .70$), oft sogar gute Werte an (Cronbachs $\alpha > .80$). Die Leistungsfähigkeit in Mathematik wurde im Rahmen des standardisierten Tests KERMIT5 erfasst und auf

der Grundlage von 4 Klassenarbeiten und Schulnoten im Laufe des Schuljahres durch die Lehrkraft eingeschätzt. Es wurden lineare Strukturgleichungsmodelle (LGM) zur Messung der Veränderungen der eingeschätzten Unterrichtsdimensionen über die acht Messzeitpunkte mit der Software Mplus (L. K. Muthén & Muthén, 2014) berechnet. In weiteren Schritten wurden dann die Mathematikleistung und das mathematikbezogene Selbstkonzept als Prädiktorvariablen eingefügt und ihr Effekt auf die Unterrichtsbeurteilungen untersucht. Es wurde die Korrelation zwischen der Mathematikleistung und dem mathematikbezogenen Selbstkonzept zu den linearen Strukturgleichungsmodellen hin untersucht. Der Zusammenhang zwischen Unterrichtswahrnehmung und Lernleistung wurde für jede Qualitätsdimension als Regressionsanalyse berechnet und jeweils als Pfaddiagramm dargestellt. Klassenspezifische Tendenzeffekte bei der Beantwortung der Items durch Schülerinnen und Schüler wurden herausgerechnet.

Die Modellfits der berechneten linearen Strukturgleichungsmodelle zur Untersuchung der Beobachtungsstabilität weisen akzeptable Werte für χ^2 , *CFI* und *TLI* auf. Die Wachstumsanalysen zeigen, dass das Niveau der Kognitiven Aktivierung über die Zeit stabil bleibt. Beide Unterstützungsdimensionen werden mit zunehmender Zeit etwas niedriger beurteilt, gleiches gilt für die Klassenführung. Weiterhin zeigen Stabilitätsanalysen, dass die Unterrichtsbeurteilungen über die Zeit eine relativ hohe interindividuelle Stabilität aufweisen (Korrelationen zwischen den Beurteilungen zu den verschiedenen Messzeitpunkten $> .50$).

Höhere mathematische Leistungen (im Test) führen zu signifikant niedrigeren Beurteilungen der Kognitiven Aktivierung und der Emotionalen Unterstützung des Unterrichts. Bessere Noten gehen mit höherer wahrgenommener Unterstützung einher. Ein besseres mathematikbezogenes Selbstkonzept führt zu einer signifikant höheren Beurteilung der Kognitiven Aktivierung und zu einer signifikant niedrigeren Beurteilung der Klassenführung des Unterrichts. Insgesamt belegen die Ergebnisse, dass alle Basisdimensionen der Unterrichtsqualität relativ stabil von den Schülerinnen und Schülern und unabhängig von den unterrichteten Themen eingeschätzt werden. Die häufigere Erhebung erlaubt aber die bessere Modellierung von Niveauveränderungen über die Zeit.

2 Theoretischer Hintergrund

In diesem Abschnitt wird der theoretische Hintergrund, auf dem die durchgeführte Studie basiert, beschrieben. Dabei wird zunächst eine Einordnung der eigenen Arbeit in das genaue Forschungsfeld vorgenommen und kurz dargelegt, worin der Mehrwert dieser Arbeit in Abgrenzung zu anderen Arbeiten besteht. Der Begriff der Unterrichtsqualität wird daran anschließend in Abschnitt 2.3 erklärt. Es werden die verschiedenen Facetten des Begriffes und die Genese des Forschungsstandes dargelegt. Die vier Basisdimensionen guten Unterrichts, auf denen diese Arbeit fußt, werden genauer eingeordnet und beschrieben. Im Anschluss daran wird zusammengetragen, was der aktuelle Forschungsstand zur Erfassung der Unterrichtsqualität ist. Dabei wird speziell der Wissensstand zur Perspektive der Schülerinnen und Schüler dargelegt. Der besondere Fokus soll dabei auf die Erkenntnisse zur Stabilität der Unterrichtswahrnehmung von Schülerinnen und Schülern und auf den Wissensstand zu den Zusammenhängen von Unterrichtswahrnehmung und Lernleistung gelegt werden.

2.1 Einführung

Im schulischen Alltag verantworten Lehrerinnen und Lehrer den Unterricht für ihre Klassen weitgehend alleine. In der Regel erhalten sie keine regelmäßige Rückmeldung zu ihrer Arbeit im Unterricht von Kolleginnen und Kollegen oder durch ihre Schulleitungen. Traditionell ist der einzige Leitstrahl, an dem sich die Lehrpersonen bei der Gestaltung des Unterrichts orientieren, ihre eigene Einschätzung der Lernsituation. Dieser Leitstrahl hängt nicht nur von dem Professionswissen der Lehrperson ab, sondern unter anderem auch von deren Motivation, Urteilsvermögen und Leistungsanspruch, um nur einige Einflussfaktoren zu benennen. Selbst bei enger planerischer Zusammenarbeit in den Kollegien und günstigen Schulstrukturen werden wichtige Qualitätsmerkmale guten Unterrichts, wie zum Beispiel die Klassenführung oder die individuellen Lernkontakte mit den Schülerinnen und Schülern, davon wenig beeinflusst. Eine kontinuierliche

lehrpersonunabhängige Qualitätsrückmeldung ist auf diese Weise nicht gegeben, eine Eichung des Leitstrahls kann so nicht erfolgen.

Auch wenn heutzutage mehr Pädagoginnen und Pädagogen regelmäßig Feedback ihrer Schülerinnen und Schüler in ihren Unterricht implementiert haben, als das wahrscheinlich vor 20 Jahren der Fall war, und es zunehmend behördliche Programme wie *Frischer Wind für die Qualitätsentwicklung - Dokumentation der Veranstaltung zur Qualitätsentwicklung an allgemeinbildenden Schulen am 25. Oktober 2018* (2018) gibt, die diese Problematik in den Fokus nehmen, ist zu bezweifeln, dass die regelmäßige, institutionalisierte und selbstverständliche Befragung von Schülerinnen und Schülern eine hohe Anzahl von Schulen und Klassenzimmern erreicht hat. Es existieren immer noch viele Vorbehalte hinsichtlich der Einschätzungsqualität und Einschätzungslegitimation von Unterricht durch Schülerinnen und Schüler. Eine andere Schwierigkeit besteht in der Praxis darin, dass selbst gutwillige Lehrerinnen und Lehrer, die ihren Leitstrahl mit den Einschätzungen von Schülerinnen und Schülern abgleichen wollen, daran scheitern, dass sie zu wenig über die Aussagenqualität von Schülerinnen und Schülern wissen oder darin wie eine gewinnbringende Rückmeldung einzuholen ist, unterstützt werden müssen. Schulen hinken hierbei deutlich hinter dem Wissenstand der Forschung hinterher. Es besteht eine Verbindungslücke zwischen dem Wissen der theoriegeleiteten Forschenden und den Praktizierenden vor Ort in den Schulen.

Die Motivation dieser Arbeit besteht nicht nur darin, neues Wissen zu heben, sondern relevantes Wissen in der Praxis nutzbar zu machen. Die Motivation zu dieser Arbeit entstand in der Praxis des Schulalltags aus der täglichen Beobachtung des beschriebenen Missstandes heraus. Schülerinnen und Schüler sollen nicht nur zufällig guten Unterricht erleben, sondern durch eine gestärkte Partizipation Argumente erhalten, ihr Unterrichtschicksal mitzugestalten.

2.2 Einordnung der eigenen Arbeit

„The notion of how the student experiences the lesson is critical to engagement and success in participating in learning - more so for adolescents than for elementary students.“ (Hattie, 2012, S. 140)

Das vorderste Forschungsinteresse dieser Arbeit gilt der Beurteilungsqualität von schulischem Unterricht. Untersucht wird die Qualität der Beurteilung von Unterricht durch Schülerinnen und Schüler über den Zeitraum eines Schuljahres. Es werden die Einflüsse

des mathematikbezogenen Selbstkonzeptes und der Mathematikleistung in Abhängigkeit des Unterrichtsinhaltes auf die Einschätzung der Unterrichtsqualität der Schülerinnen und Schüler hin empirisch untersucht. Diese Arbeit ordnet sich als Teil der empirischen Bildungsforschung in den Bereich der Lehr-Lernforschung ein.

Mitte der 1990er Jahre fand die deutsche Erweiterung der TIMSS-Video-Studie statt. In dieser Studie wurde der Mathematikunterricht achter Klassen in Deutschland, den USA und Japan analysiert (Stigler & Hiebert, 1999). Hier wurden unter anderem die Aktivitätsstruktur, der Unterrichtszweck und Typen mathematischer Probleme, die Unterrichtsgegenstand waren, unterschieden. Speziell für die Anwendung in Deutschland wurde das Design der TIMSS-Video-Studie von 1995 angepasst und verändert (Baumert et al., 1997; Kunter, 2005). Die Autoren greifen dabei auf 21 bereits existierende Skalen zurück, um zusätzlich die Wahrnehmung des Klassenraumes durch Lehrpersonen und Schülerinnen und Schüler zu bewerten. Diese 21 Skalen sind unter anderem in Gruehn (2000) beschrieben. Um Vergleiche zwischen den Beobachtungsperspektiven von Lehrpersonen, Lernenden und externen Beobachtern herstellen zu können, wurden diese 21 Skalen von allen drei Personengruppen bewertet (Clausen, 2002). Durch eine explorative Faktorenanalyse erstellten Klieme, Schümer und Knoll (2001) aus diesen 21 Skalen die drei Basisdimensionen der Unterrichtsqualität: Klassenraummanagement, Unterstützung der Studierenden und Kognitive Aktivierung. Diese drei Qualitätsdimensionen sind in der Bildungsforschung seitdem mehr oder weniger üblich im deutschsprachigen Raum, um die Unterrichtsqualität zu systematisieren. Diese Übertragung von Unterrichtskriterien aus einem internationalen Kontext nennt Douven (2017) Abduction. Praetorius, Klieme, Herbert und Pinger (2018) verweisen darauf, dass solche Übertragungen wissenschaftlich streng zu überprüfen sind, bevor man sie als gültig erachtet.

Die vorliegende Arbeit erweitert den Qualitätsrahmen aus diesen drei Qualitätsdimensionen auf vier, durch die Auffächerung der Unterstützungsdimension in zwei eigenständige, von einander abzugrenzende Qualitätsdimensionen. Dabei werden Quellen aus dem angloamerikanischen Raum als Vorlage genutzt. Der so neu übertragene Qualitätsrahmen wird auf seine psychometrische Güte hin untersucht und getestet und mit seiner Hilfe werden aktuelle Fragestellungen der Unterrichtsforschung bearbeitet. Dabei besteht, in Abgrenzung zu anderen Arbeiten in diesem Forschungsfeld, die Besonderheit dieser Arbeit darin, dass der Unterricht aller untersuchten Klassen harmonisiert wurde, das heißt, alle 3 Klassen wurden im gleichen Zeitraum, mit ähnlichen Methoden und in enger Absprache zwischen den unterrichtenden Kolleginnen und Kollegen zum gleichen Thema unterrichtet. Diese Voraussetzung findet man in keiner anderen vorliegenden Studie. Des Weiteren beziehen sich die meisten anderen Studien zu Schüler- und Schülerinnen-

ratings vorrangig auf die Klassenebene. In dieser Studie hingegen steht die individuelle Wahrnehmung der einzelnen Schülerinnen und Schüler im Vordergrund.

2.3 Unterrichtsqualität

Was genau guten Unterricht ausmacht und welche Kriterien die Qualität von Unterricht messen, wird von vielen Autoren in zahlreichen Listen und Katalogen von Schlüsselmerkmalen seit langem beschrieben und zusammengestellt. So formuliert Brunnhuber (1977) schon vor 1980 „Prinzipien guter Unterrichtsgestaltung“, die es einzuhalten galt, um eine hohe Unterrichtsqualität zu erreichen. Auf Seite 14 benennt er die Faktoren Zielorientierung, Motivierung, Strukturierung, Aktivierung, Angemessenheit und Leistungssicherung als wirksam. Eine Einigkeit in Bezug auf die optimalen Kriterien zur Einschätzung von Unterrichtsqualität herrscht indes in der Forschungsgemeinde nicht. Weinert, Schrader und Helmke (1989, S. 899) definieren Unterrichtsqualität als „jedes stabile Muster von Instruktionsverhalten, das als Ganzes oder durch einzelne Komponenten die substantielle Vorhersage oder Erklärung von Schulleistung ermöglicht.“ Helmke (2015) gibt einen Überblick über bedeutende Neuveröffentlichungen. Der Autor bezieht sich dabei auf eine Zusammenstellung von Scheltwort (2006). Darunter finden sich unter anderem Arbeiten von Brophy (2000), Walberg und Paik (2000) und Meyer (2004, 2007).

Helmke (2015) führt ab Seite 168ff. folgende zehn Qualitätsbereiche auf, die hier kurz zusammenfassend dargestellt werden sollen:

Klassenführung. Er unterscheidet vier Ansätze und Denkrichtungen dessen, was Klassenführung darstellt:

- Klassenlehrer sein, also Ansprechpartner für Kolleginnen und Kollegen, Eltern und Schülerinnen und Schüler in allen Belangen der Klassenorganisation, bei Fragen und Problemen.
- Klassenführung als Inbegriff erfolgreichen Unterrichtens und Führens. Klassenführung bedeutet in diesem Sinne, die Führungsaufgabe bewusst wahrzunehmen und Konflikte zu bearbeiten, institutionelle und interkulturelle Einbindung vorzunehmen, ein lernfreundliches Klima zu schaffen und dergleichen mehr.
- Klassenführung als Reaktion auf Disziplinstörungen ist die verbreitetste nationale Sichtweise der Arbeitsaufgabe. Wobei die Herstellung von Ruhe, Ord-

nung und Drill sehr veraltete und traditionsbehaftete Assoziationen dieser Rolle darstellen.

- Integrativer Ansatz, bei dem vorbeugende Maßnahmen und ein proaktives Handeln im Mittelpunkt stehen.

Klarheit und Strukturiertheit. Es geht darum, dass Informationen so klar und verständlich wie möglich präsentiert und den Schülerinnen und Schülern strukturiert dargeboten werden, um Lernprozesse zu initiieren. Klarheit meint dabei akustische Verständlichkeit, sprachliche Präzision, inhaltliche Kohärenz und fachliche Korrektheit. Als senderbezogenes Qualitätsmerkmal werden Modulation und Lautstärke der Sprache genauso in den Blick genommen wie Mimik und Gestik. Helmke (2015) führt an, dass die Effektstärke von Klarheit und Verständlichkeit der Sprache bei $d = 0.75$ liegt und beruft sich damit auf Hattie (2009).

Strukturiertheit bedeutet zum einen Unterricht, der optimal geplant und sequenziert ist, und zum anderen Unterricht so zu gestalten, dass die Wissensbasis der Schülerinnen und Schüler genutzt wird.

Konsolidierung, Sicherung. Lernen stellt keine bloße Informationsaufnahme dar. Es muss eine Herstellung von Verknüpfung zu anderem, bereits vorhandenem Wissen, ein Anwenden folgen. Ein Üben, ein selbstreflektiertes Auseinandersetzen mit dem Lerngegenstand, ist ein Qualitätsmerkmal guten Unterrichts.

Aktivierung. Unterricht muss die Schülerinnen und Schüler aktivieren. Aktivierung als Konzept umfasst vier Facetten:

- Kognitive Aktivierung, im Sinne des selbstgesteuerten Lernens mit individuellen Lernstrategien und Methoden.
- Soziale Aktivierung, zum Beispiel mit kooperativen Lernformen.
- Aktivierung im Sinne einer aktiven Teilhabe am Unterricht und bei der Planung und Durchführung dessen.
- Körperliche Aktivierung, im Gegensatz zur passiv-sitzenden Lernhaltung.

Motivierung. Helmke (2015) führt zahlreiche Quellen an, die die Bedeutung dieses Qualitätsmerkmals stützen. Motivation ist unabdingbar, um Lernprozesse zu befördern und müssen fortwährend im Blick behalten werden. Dieser Aspekt ist aber nicht so zu verstehen, dass eine Motivierung des Lernens ausschließlich von der Lehrperson ausgeht. Lernende sollen zunehmend selbst Motive finden, sich mit dem Lehrgegenstand auseinanderzusetzen und ihren Lernprozess selbst zu steuern. Diese Anreize

können in Leistung, im Finden von Anschluss oder im Gewinnen von Macht liegen. Eine extrinsische Motivation ist dabei keinesfalls als schädliches Mittel anzusehen, um Lernprozesse wahrzunehmen.

Lernförderliches Klima. Mit diesem Begriff ist eine Lernumgebung gemeint, die es Schülerinnen und Schülern erleichtert, Lernprozesse aufzunehmen. Dies beinhaltet Aspekte des Wohlbefindens, der Zufriedenheit, der positiven Grundeinstellung dem Unterrichtsgegenstand und der Leistungsbereitschaft gegenüber, aber auch der Regelklarheit der Gerechtigkeit und Fürsorge durch die Lehrperson. Besonders wichtig in diesem Zusammenhang ist der Umgang der Lehrperson mit Fehlern der Schülerinnen und Schüler, ein angemessenes Unterrichtstempo und eine angemessene Wartezeit sowie der Abbau von Ängsten bei Schülerinnen und Schülern.

Schülerorientierung. Hier geht es vor allem darum, dass Schülerinnen und Schüler unabhängig ihres Lern- und Leistungsvermögens als Personen ernst genommen und wertgeschätzt werden. Die Schülerinnen und Schüler müssen sich durch die Lehrperson respektiert fühlen. Die Lehrperson muss sich angemessen Zeit für persönliche Anliegen und bei Schwierigkeiten nehmen. Die Lehrperson muss persönliches Interesse an Stärken und Schwächen der Lernenden glaubhaft bekunden. Es muss angemessen auf Anregungen und Vorschläge der Schülerinnen und Schüler eingegangen werden. Und dergleichen mehr.

Kompetenzorientierung. Ein Ziel guten Unterrichts besteht darin, den Kompetenzerwerb von Schülerinnen und Schülern zu fördern. Eine Grundlage kompetenzorientierten Unterrichts ist eine an messbaren Ergebnissen ausgerichtete, empirische Orientierung. Dazu müssen Curricula, Jahrespläne und ähnliches so gestaltet sein, dass der kumulative Prozess des Kompetenzerwerbs von distalen in proximale Kompetenzen zerlegt wird (Helmke, 2015; Lange, 2005). In dieses Teilkriterium gehört auch die Leistungsmessung, um die Unterrichtsergebnisse zu überprüfen.

Passung. Hier geht es um den adaptiven Umgang mit Heterogenität im Gruppenunterricht.

Angebotsvielfalt. Verschiedene Unterrichtsformen mit variierenden Vermittlungsstilen und verschiedenen Graden der Lenkung durch die Lehrperson sollten sich abwechseln und ohne einem Dogma zu folgen, optimal zum Erreichen der Lernziele gewählt werden.

Auf Seite 169 fasst Helmke (2015) einige seiner Qualitätsbereiche so zusammen, dass sie letztlich vier grundlegenden Dimensionen von Unterrichtsqualität zuordenbar sind:

erstens der Klassenführung, zweitens der Förderung der Informationsverarbeitung, drittens primär der Förderung der Lernbereitschaft und indirekt dem Lernerfolg und viertens dem Sachverhalt, der Unterschiedlichkeit von Bildungszielen Rechnung zutragen. Diese vier Basisdimensionen finden sich im wesentlichen gleich bei anderen Autoren wieder.

Kunter, Baumert, Blum, Klusmann und Krauss (2011, S. 87ff.) unterscheiden Kunter und Voss die drei Dimensionen der Unterrichtsqualität in COACTIV¹ in

- Effizienz der Klassenführung,
- Potenzial zur Kognitiven Aktivierung und
- Konstruktive Unterstützung.

Kunter et al. (2011) unterscheiden beim Merkmal der Konstruktiven Unterstützung zwei Teilkonstruktionen, nämlich die *Strukturierung* (oder anders die instruktionale bzw. konstruktive Unterstützung) und die *Qualität* (emotionale Unterstützung) der Beziehung zwischen Lehrenden und Lernenden. Diese benannten Basisdimensionen ordnen Sie den sogenannten „Tiefenstrukturen“ zu (Oser & Baeriswyl, 2001; Seidel, 2003). Im Gegensatz zu „Sichtstrukturen“, die eher sichtbare übergeordnete Organisationsmerkmale und Rahmenbedingungen beschreiben, sind Tiefenstrukturen Kennzeichen des direkten Lehr-Lern-Prozesses. Tiefen- und Sichtstrukturen bestehen unabhängig voneinander und sind klar voneinander abgrenzbar (Lipowsky, 2002).

Laut Kunter et al. (2011) hat die empirische Unterrichtsforschung bereits eindeutig gezeigt, dass Kompetenzzuwächse bei Schülerinnen und Schülern überwiegend durch Tiefenstrukturen erklärt werden können. Die Autoren zitieren dazu Seidel und Shavelson (2007) und Hattie (2009).

Unter dem Begriff *Konstruktive Unterstützung* oder *Unterstützendes Klima* fassen viele, vor allem deutschsprachige Autoren, die *Instruktionale/Kognitive Unterstützung* und die *Emotionale Unterstützung* zusammen. Im amerikanischen Sprachraum werden diese beiden Dimension oft getrennt und als eigenständige Basisdimensionen betrachtet.

So auch in R. Pianta, Hamre und Allen (2012). Ab Seite 365ff. stellen die Autoren die Basisdimensionen dar, die dem in Amerika sehr verbreiteten Classroom Assessment Scoring System² (R. C. Pianta, La Paro & Hamre, 2004) zugrunde liegen. Hier gibt

¹Die COACTIV-Studie *Kurzform für Cognitive Activation in the Classroom: The Orchestration of Learning Opportunities for the Enhancement of Insightful Learning in Mathematics* oder zu deutsch: *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz* befragte und testete 2003/2004 die Mathematiklehrkräfte der PISA-Klassen zu ihrem fachdidaktischem und mathematischem Wissen.

²kurz CLASS

es drei Basisdimensionen: Classroom Organization Domain (Effizienz der Klassenführung), Instructional Interaction Domain (Instruktionale Unterstützung) und Emotional Interaction Domain (Emotionale Unterstützung). Bei CLASS ist die Dimension Kognitive Aktivierung in der Dimension Instructional Interaktion Domain enthalten (in der Facette Concept Development).

Auch in dieser Arbeit werden die Facetten Instruktionale bzw. Kognitive Unterstützung und Emotionale Unterstützung zu zwei eigenständigen Basisdimensionen erhoben. Damit liegen dieser Studie folgende vier Basisdimensionen der Unterrichtsqualität, die erfasst und untersucht werden, zugrunde.

2.3.1 Basisdimension 1 Kognitive Aktivierung

Die Unterrichtsdimension der Kognitiven Aktivierung stellt, im Gegensatz zu anderen Qualitätsdimensionen, ein junges Konstrukt dar (Baumert et al., 2004; Klieme, Lipowsky, Rakoczy & Ratzka, 2006).

Lipowsky (2015), aber auch die Autoren Kunter et al. (2011) verstehen unter Kognitiver Aktivierung, dass Lernende zur vertieften Auseinandersetzen mit dem Unterrichtsgegenstand angeregt werden. Es soll ein aktiver, mentaler Auseinandersetzungsprozess angestoßen werden. Kunter et al. (2011, S.88) fassen weiter zusammen: „Aus lernpsychologischer Sicht bedeutet dies, dass bestehende Wissensstrukturen verändert, erweitert, vernetzt, umstrukturiert oder neu gebildet werden sollten.“

Dabei ist die didaktische Weise wie Wissen aufgebaut und mit vorhandenen Kompetenzen und Fähigkeiten verknüpft wird, entscheidend (Schneider & Stern, 2010; Stern, Schalk & Schumacher, 2016). Verschiedene Quellen verweisen auf die Bedeutung des didaktischen Niveaus der Lehrperson für die Kognitive Aktivierung der Schülerinnen und Schüler (Brophy, 2000; for Mathematics for Teaching Project, 2010).

Die Kognitive Aktivierung muss beispielsweise durch geeignete, anregende Aufgaben erreicht werden. Diese Basisdimension ist also nicht an formale Sichtstrukturen gekoppelt und im Unterricht einfach zu beobachten. Es sind verschiedene Indikatoren notwendig, um die Kognitive Aktivierung der Lernenden approximativ zu beobachten. Lipowsky (2015) führt auf, welche Möglichkeiten eine Lehrperson hat, einen kognitiv aktivierenden Lernprozess zu fördern. In das Zentrum rückt er dabei folgende Schwerpunkte:

- Konfrontation der Lernenden mit kognitiv herausfordernden Aufgaben.
- Initiierung von kognitiven Konflikten.

- Verweisung auf Unterschiede in inhaltsbezogenen Ideen, Konzepten, Positionen, Interpretationen und Lösungen.

um nur drei zu nennen.

Weiterführende Aspekte zur Aktivierung finden sich auch in Helmke (2015). Der Autor sieht ein Ziel der Kognitiven Aktivierung in der Selbststeuerung des Lernens. Neben anderen Erkenntnissen zur Bedeutung dieser mentalen Aktivierung führt der Autor an, dass laut Hattie (2009) die Wirksamkeit von metakognitiven Strategien bei $d = 0.69$ bzw. für *study skills* bei $d = 0.59$ für den Lernenden liegen. Zu gleichen Ergebnissen gelangen auch Schumacher, Barth, Lipscher und Hänger-Surer (2017).

2.3.2 Basisdimension 2 Instruktionale bzw. Kognitive Unterstützung

Hier ist die Struktur der Beziehung zwischen Lehrenden und Lernenden gemeint. In der englischsprachigen Literatur wird diese Dimension als *scaffolding* bezeichnet (Pea, 2004). Komplexe Sachverhalte werden im Unterricht gegliedert und die Anforderungen an die Lernenden angepasst. Den Lehrenden gelingt eine strukturierte, an die Schwierigkeiten der Lernenden angepasste Hilfestellung. Um die Anforderungen des Lerninhaltes für Lernende bewältigbar zu machen, muss eine geeignete Strukturierung der Lerninhalte vorgenommen werden und es müssen geeignete Maßnahmen ergriffen werden (Reiser, 2004). Kunter et al. (2011, S. 90) schätzen ein, dass diese durch „eine Dekompensation komplexer Aufgaben in überschaubare Schritte, durch die Begleitung der Lernprozesse oder das strukturierende Eingreifen bei Verständnisschwierigkeiten und Fehlern erreicht werden.“

2.3.3 Basisdimension 3 Emotionale Unterstützung

Unter der Emotionalen Unterstützung verstehen R. Pianta et al. (2012) die Bemühungen Lehrender, ihre Schülerinnen und Schüler sozial und emotional beim Lernen zu unterstützen. Die Lehrperson berücksichtigt die Interessen und die Motivation seiner Schülerinnen und Schüler (Hamre & Pianta, 2010). Eine angenehme Lernatmosphäre ist durch die Lehrenden zu erzeugen und eine gute Lehrer-Schüler-Beziehung forcieren. Gespür für Verständnisprobleme, Geduld bei individuellen Schwierigkeiten sowie ein konstruktiver Umgang mit Fehlern der Lernenden durch konstruktive Rückmeldungen werden durch diese Dimension erfasst. Aber auch die Förderung der sozialen und emotionalen Interaktion zwischen den Schülerinnen und Schülern gehört in diese Dimension. Es geht um

die Lernatmosphäre, die eine sozio-emotionale Auswirkung auf das Wohlbefinden der Schülerinnen und Schüler hat.

In der deutschen Forschung werden die Facetten dieser Dimension unter dem Begriff des Unterrichtsklimas zusammengefasst (Den Brok, Brekelmans & Wubbels, 2004). Kunter et al. (2011) bewerten den Klimabegriff allerdings als ein unscharfes Konstrukt in der Unterrichtsforschung, weshalb er auch in der COACTIV-Studie vermieden wird. Stattdessen unterscheidet man dort emotionale und motivationale Unterstützung und bezieht sich dabei auf Gruehn (2000) und Lipowsky (2015).

Durch Ryan und Deci (2000) wurde gezeigt, dass die wahrgenommene Unterstützung im Kompetenzerleben, autonomes Lernen und gelungene soziale Beziehungen Voraussetzungen für Motivation sind und Grundbedürfnisse darstellen. Diese emotionalen Aspekte schulischen Lernens stellen einen starken Prädiktor für selbstbestimmtes Lernen dar (Ryan & Deci, 2000). Rakoczy (2008) sieht die emotionale Unterstützung als ein Kernelement der Unterstützung Lernender an und sieht darin eine Qualitätsdimension guten Unterrichts.

2.3.4 Basisdimension 4 Effizienz der Klassenführung

In Schulen findet Unterricht fast ausschließlich im Klassenverband statt. Einen sehr umfangreichen Überblick zur Forschung auf diesem Gebiet, seit der grundlegenden Arbeit von Kounin (1970) in den 1970er Jahren, findet man in Seidel (2015). Darin heißt es, auf Seite 109: „Klassenführung ist alles, was Lehrpersonen mittels Aktivitäten und Haltungen zur Steuerung der Interaktionen in der Klasse beitragen [...]“. Die Klassenführung wird in manchen Quellen als Schlüsselkomponente erfolgreichen Unterrichtens bezeichnet (Emmer & Stough, 2001; Hattie, 2009). Hochweber, Hosenfeld und Klieme (2014) identifizieren zwei Kernkomponenten, zum einen Vermeidung unerwünschter Verhaltensweisen und zum anderen die Stärkung des gewünschten Lernverhaltens.

Auch Helmke (2015) gibt einen umfassenden Überblick zu dieser Basisdimension. Im Wesentlichen unterscheidet er dabei vier Ansätze dessen, was Klassenführung ausmacht, nämlich:

- Klassenlehrer sein,
- Klassenführung als Inbegriff erfolgreichen Unterrichtens und Führens,
- Klassenführung als Reaktion auf Störungen und den
- Integrativen Ansatz.

Diese Ansätze werden ausführlich beschrieben und im Kontext aktueller Forschungen diskutiert. Er formuliert auch entsprechende Kriterien, die die Basisdimension Klassenführung erfassen sollen und erörtert welcher theoretische Überbau diese Dimension beschreibt. Aber letztlich geht es darum, dass die Lehrperson durch geschickte Orchestrierung die zur Verfügung stehende Zeit in echte, effiziente und störungsfreie Lernzeit für die Lernenden verwandelt (Kuger, 2016).

In Helmke (2015) wird die Bedeutung und Relevanz dieses Merkmals für den Leistungsfortschritt und das Leistungsniveau der Schülerinnen und Schüler beschrieben. Demnach gibt es keinen eindeutigeren und konsistenteren Bedingungsfaktor für schulische Leistungen. Der Autor zitiert unter anderem Hattie (2009), wonach das *classroom management* einen starken Effekt von $d = 0.52$ auf den Lernerfolg haben soll. Zu diesem Schluss kommen auch Kunter et al. (2011, S. 88): „Je störungsärmer der Unterricht ist und je mehr effektive Lernzeit zur Verfügung steht, umso höher sind die Leistungen der Schülerinnen und Schüler.“

2.3.5 Zusammenfassung

Verschiedene Studien haben gezeigt, dass die Basisdimensionen *Effizienz der Klassenführung* und *Instruktionale Unterstützung* positiv mit Lernleistungen und Lernverhalten von Schülerinnen und Schülern verbunden sind (Camron, Connor & Morrison, 2005). Die Basisdimension *Emotionale Unterstützung* hat eher einen Einfluss auf indirekte, lernförderliche Faktoren wie auf die Motivation oder die Selbstachtung der Schülerinnen und Schüler (Kunter et al., 2013). Das Maß, in dem Lehrerinnen und Lehrer auf die Bedürfnisse ihrer Schülerinnen und Schüler eingehen, fördert deren Motivation und deren akademisches Selbstkonzept, so die Autoren.

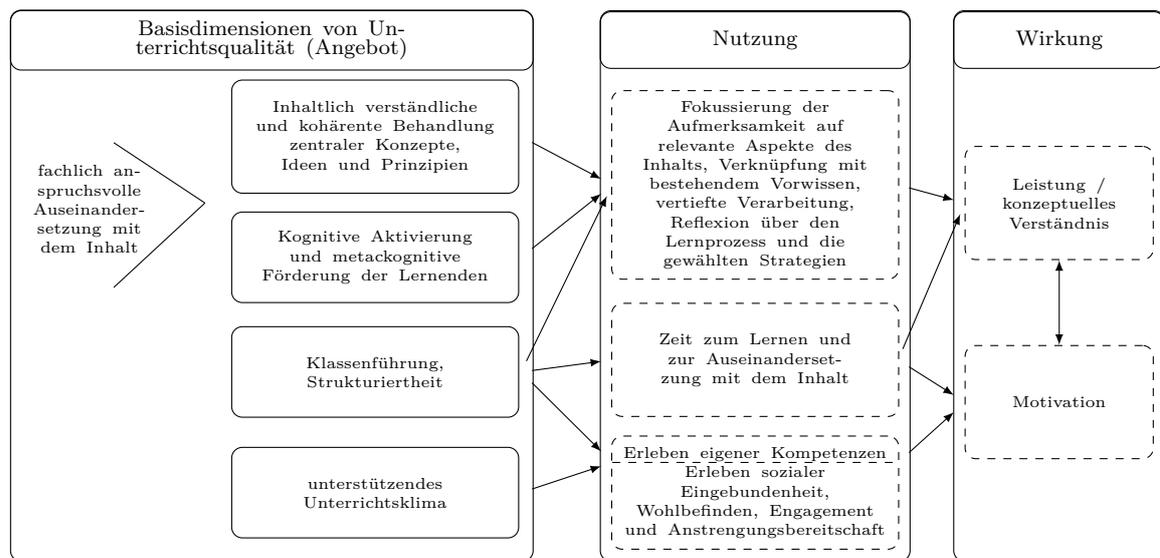


Abbildung 2.1: Basisdimensionen guten Unterrichts und deren angenommene Wirkungen; aus Lipowsky (2015, S. 97) entnommen.

Eine gelungene Zusammenfassung der Basisdimensionen, deren Nutzen und Wirken stellt die Abbildung 2.1 dar. Sie stammt aus Lipowsky (2015) und wurde von dem Autor nach Klieme et al. (2006) und Klieme, Pauli und Reusser (2009) modifiziert. Hier werden ähnliche vier Basisdimensionen verwendet wie in der vorliegenden Arbeit. Die Abbildung zeigt deren angenommenen Wirkungen. Das Modell zeigt die zentrale Rolle der Klassenführung. Wenig Störungen und ein angenehmes Unterrichtsklima führen zu einem lerneffizienten, motivierenden Unterricht. Dabei wirkt eine gelungene Kognitive Aktivierung vor allem auf die fachliche, inhaltsbezogene Ebene ein und die Unterstützungsdimensionen auf ein lernförderliches, zwischenmenschliches Miteinander in sozialer Eingebundenheit.

2.4 Unterrichtswahrnehmungen

2.4.1 Erfassung von Unterrichtsqualität

Unterricht stellt einen Prozess dar, dessen Qualitätsbeurteilung anhand von subjektiven Einschätzungen entweder von Selbsteinschätzungen oder anhand von Fremdeinschätzungen bewertbar ist. Die Qualität ist dabei nicht objektiv messbar, sondern hängt von der Perspektive ab. Grundsätzlich bieten sich drei Beobachtungsperspektiven des Unterrichtsprozesses an. Zum einen die Selbsteinschätzung der Qualität durch die Lehrperson, zum anderen die Beurteilung durch einen geschulten, sachkundigen (externen) Beobachter und natürlich eine Beurteilung durch die zu unterrichtenden Lernenden. Beurteilung von Unterrichtsqualität findet in der Fachliteratur durch diese drei Personengruppen statt. Es werden in der Literatur die verschiedenen Vor- und Nachteile der jeweiligen Perspektive diskutiert (Clausen, 2002; Desimone, Smith & Frisvold, 2010; Helmke, 2015). Geeignet scheint die jeweilige Perspektive dann zu sein, wenn die Beobachtungsergebnisse mit denen einer der beiden anderen Perspektiven übereinstimmt, man spricht von *Inter-Rater-Übereinstimmung* oder es einen hohen Zusammenhang zu zugeschriebenen Erfolgsmerkmalen von Unterricht, wie etwa dem Kompetenzzuwachs gibt. Jede Perspektive bringt spezifische Beobachterfehler mit sich. Der Beobachter verfälscht das Beobachtungsergebnis, diese Effekte werden mit dem Begriff des Rater-Bias oder des Rater-Fehlers bezeichnet. Die Interpretation der fehlenden Übereinstimmung der drei verschiedenen Perspektiven und die Rater-Bias sind unter anderem Gegenstand der Arbeit von Praetorius (2014).

Die Selbsteinschätzung des gehaltenen Unterrichts durch die Lehrpersonen hat, wegen ihres höher eingeschätzten Professionswissens, wieder an Bedeutung in der Forschung zugenommen (Baumert & Kunter, 2006; Blömeke, 2004). Andere Gründe liegen auch in dem wesentlich geringeren Zeitaufwand im Vergleich zu anderen Beobachtungsperspektiven (Clausen, 2002; Desimone et al., 2010). Andererseits zeigen empirische Befunde auch diverse Schwierigkeiten dieser Perspektive auf (Clausen, 2002). So finden sich laut Clausen (2002) zu wenig Übereinstimmungen zwischen den Selbstbeurteilungen und anderen Beurteilungsdaten. Die prädiktive Validität der Beobachtung für relevante Qualitätskriterien für Unterrichtsqualität ist zu schlecht. Die Ursachen dafür werden darin gesehen, dass die Lehrperson Teil des zu beurteilenden Prozesses ist und deshalb auf der einen Seite zum Beispiel selbstwertschätzende Tendenzen zeigt (Clausen, 2002), auf der anderen Seite ihr auch nicht alle Aspekte des eigenen Verhaltens einer Selbstbeschreibung zugänglich sind (Vazire & Solomon, 2015).

Eine Unterrichtsbeobachtung zur Erfassung der Unterrichtsqualität durch externe Beobachter, entweder direkt im Klassenraum oder anhand von Videoaufzeichnungen, um Unterrichtsmaterial, -konzepte oder andere Aspekte zu beurteilen, gilt weithin als der Idealfall in der Lehr-Lern-Forschung. Allerdings sind diese mit erheblichem zeitlichen und finanziellen Aufwand verbunden. Die beurteilenden Experten müssen sehr aufwendig geschult werden, insbesondere dann, wenn hoch inferente Beobachtungen gewünscht werden, bei denen nicht nur die reine Beobachtung vorgenommen wird, sondern auch noch auf abstrakte Sachverhalte geschlossen werden muss. Die Auswertung auf diese Weise erfasster Unterrichtsstunden erfordert in der Regel erheblich mehr Zeit als die eigentliche Unterrichtsstunde selbst (Fauth, Decristan, Rieser, Klieme & Büttner, 2014a). Die entstehende Güte dieses aufwendigen Verfahrens lässt in aller Regel dennoch zu wünschen übrig und erfordert eine Anpassung der untersuchten Unterrichtsmerkmale (Derry et al., 2010).

Zu den verschiedenen Beobachtungsperspektiven stellen zum Beispiel Clausen (2002) oder Helmke (2015) weiterführende Aussagen vor. In der Arbeit von Praetorius et al. (2018) kann man mehrere, nach verschiedenen Aspekten hin bewertete, sehr genaue tabellarische Auflistungen aller bis 2018 erfolgten Studien zur Erfassung der im deutschsprachigen Raum üblichen drei Grunddimensionen guten Unterrichts mit Blick auf die Beobachtungsperspektive hin finden. Hier wird auch die Güte der jeweiligen Beobachtung detailliert mit angegeben. Die jeweiligen Zuverlässigkeitsmaße dieser Studien wurden in Praetorius et al. (2018) nach Beobachtungsperspektive geordnet in einer Tabelle zusammengefasst. Die Tabelle 2.1 ist eine Kopie dieser Zusammenstellung und aus Praetorius et al. (2018, S. 10) entnommen.

Tabelle 2.1: Spanne der verwertbaren Koeffizienten aus den verschiedenen Studien, die sich auf die drei Basisdimensionen konzentrierten; aus Praetorius, Klieme, Herbert und Pinger (2018, S. 10), (dort als Table 3 deklariert) entnommen.

	Cronbach's α	ICC_1/ICC_2	G coefficient p^2
Teacher ratings			
Classroom management	[0.86; 0.94]	-/-	-
Student support	[0.79; 0.86]	-/-	-
Cognitive activation	[0.67; 0.86]	-/-	-
Students ratings			
Classroom management	[0.73; 0.91]	-/-	-
Student support	[0.78; 0.95]	[0.25; 0.29]/[0.85; 0.93]	-
Cognitive activation	[0.73; 0.91]	[0.12; 0.23]/[0.66; 0.87]	-
Observers ratings			
Classroom management	[0.75; 1.00]	-/[0.68; 0.98]	[0.68; 0.98] ^a
Student support	[0.62; 0.98]	-/[0.59; 0.98]	[0.44; 0.94]
Cognitive activation	[0.74; 0.98]	-/[0.65; 0.97]	[0.58; 0.63]

G coefficients are based on designs including raters as a facet (as well as other facets, such as measurement points and items, depending on the study)

- no data available as the indicator does not make sense for the respective ratings or has not been published

^a We excluded the measure of classroom management in the study of Taut und Rakoczy (2016), as the G coefficient of 0.05 indicates a severe problem with raters, scales, or the analyses

Praetorius et al. (2018) schreiben dazu, dass die Zuverlässigkeit der drei Basisdimensionen in den meisten, wenn auch nicht in allen Studien zufriedenstellend war. Insgesamt fanden Praetorius et al. (2018) 6 Studien, die die faktorielle Gültigkeit der drei Basisdimensionen untersuchten. Davon führten laut Praetorius et al. (2018) drei eine Konformitätsfaktoranalyse durch (Fauth et al., 2014a, 2014b; Künsting, Neuber & Lipowsky, 2016; Kunter & Voss, 2013), die alle die Drei-Faktoren-Struktur belegten. Davon wurden hier auch bereits einzelne im Vorfeld erwähnt. Des Weiteren schreiben Praetorius et al. (2018), dass eine von drei Studien, die die explorative Faktoranalyse durchführten, ebenfalls die Drei-Faktoren-Struktur bestätigte und von Lipowsky et al. (2009) stammt. Die beiden übrigen haben in ihren Analysen auch andere Qualitätsdimensionen betrachtet (Kunter, Brunner et al., o.D.; Taut & Rakoczy, 2016).

2.4.2 Die Perspektive von Schülerinnen und Schülern

Beobachtungen durch Schülerinnen und Schüler sind nicht aufwendig und daher kostengünstig. Schülerinnen und Schüler haben über einen langen Zeitraum Erfahrungen mit ihren Lehrenden, was eine genauere Einschätzung zulässt; ihr Urteil kann sich auf einen langen Beobachtungszeitraum beziehen. Schülerinnen und Schüler sind in der Lage, Vergleiche zwischen verschiedenen Lehrpersonen zu ziehen (Lüdke, Trautwein, Kunter & Baumert, 2006). Bei der Untersuchung stellt sich die Frage, wie aussagekräftig die Aussagen von Lernenden über die Unterrichtsqualität sind. Allgemeiner könnte man fragen, wer eigentlich am besten Unterrichtsqualität einschätzen kann, welche Beobachtungsperspektive also die zuverlässigste Aussage bietet.

Neben anderen Autoren werfen Fauth et al. (2014b) und Kunter und Baumert (2006) die These auf, dass die identische Formulierung der Items der Bewertungsskalen bei Lehrenden und Schülerinnen und Schülern nicht garantiert, dass auch die gleichen Assoziationen bei beiden Gruppen ausgelöst werden. Die psychometrische Güte, also Reliabilität und Validität der Untersuchungen, ist somit nur schwer einschätzbar.

Mit der Interpretation dieser mangelnden Übereinstimmungen zwischen den Perspektiven beschäftigt sich die Arbeit von Praetorius (2014). Hier kommt die Autorin zu dem Schluss, dass die fehlenden Zusammenhänge zwischen den verschiedenen Perspektiven auch auf Messprobleme zurückzuführen sind. Auch hier ist das Fazit, dass in Untersuchungen die Reliabilität und Validität der einzelnen Perspektiven sicherzustellen sind.

Bei der Erfassung von Unterrichtsqualität durch Schülerinnen und Schüler kommt es zu dyadischen Tendenzeffekten zwischen Schülerinnen und Schülern auf der einen Seite und deren Lehrpersonen auf der anderen Seite. Diese beeinflussen die Schülerinnen und Schüler bei der Beantwortung der Items einer Skala eines Fragebogens (Göllner, Wagner, Eccles & Trautwein, 2018). Häufig werden die Forschungsergebnisse der Unterrichtsbeurteilung durch Schülerinnen und Schüler als gemeinsame Wahrnehmungen auf der Klassenebene präsentiert. Es gibt nur wenige Studien, die individuelle Wahrnehmungen von Unterrichtsqualität untersuchen, dann aber werden idiosynkratisch Wahrnehmungen von den von allen Schülerinnen und Schülern innerhalb einer Klasse geteilten Wahrnehmungen nicht getrennt. In der Studie von Göllner et al. (2018) werden diese klassenspezifischen Tendenzen bei der Unterrichtswahrnehmung herausgerechnet, indem letztlich, vereinfacht dargestellt, die Differenz zwischen individuellem Mittelwert und Klassenmittelwert berechnet wird.

2.4.3 Stabilität der Unterrichtswahrnehmung

Da sich die Untersuchungen dieser Arbeit auf die Beurteilung von Unterrichtsqualität auf der Grundlage von Beobachtungen durch Schülerinnen und Schüler stützen, sollen hier entscheidende Aussagen zur Stabilität der Unterrichtswahrnehmung und zur Aussagegüte der Beurteilungen durch Schülerinnen und Schüler zusammengestellt werden. Die allermeisten Aussagen der im Folgenden angeführten Studien zur Perspektive von Schülerinnen und Schülern beziehen sich auf die Aussagen der Klasse. Aussagen, die die Individualebene von Schülerinnen und Schülern berücksichtigen, sind dabei nur vereinzelt zu finden. Diese werden im Folgenden extra ausgewiesen, sonst beziehen sich die Aussagen auf die Klassenebene.

In der Literatur findet man verschiedene Aussagen über die Qualität bestimmter Beobachtungsperspektiven. Kritisch gesehen wird immer wieder die fehlende didaktisch-pädagogische Ausbildung von Schülerinnen und Schülern, die auch durch ihre umfangreiche Unterrichtserfahrung nicht ausgeglichen werden kann.

In Clausen (2002) kam der Autor zu dem Ergebnis, dass die Bewertung von Unterrichtsqualität zwischen Lehrenden, Lernenden und externen Beobachtern stark von einander abweichen. Clausen (2002) berichtet von relativen Übereinstimmungen für 12 verschiedene Aspekte der Unterrichtsqualität von einer Bandbreite im Bereich von $-.28 \leq r \leq .42$.

Die psychometrische Güte von Schülerinnen- und Schülerurteilen, aber auch Fragen danach, wie gut und zuverlässig Schülerinnen und Schüler theoretisch distinkte Aspekte des Unterrichtsgeschehens beurteilen können, in wieweit Schülerinnen- und Schülerurteile über Unterrichtsfächer und Klassen hinweg vergleichbar sind, werden in dem BMBF Projekt „Erfassung der Unterrichtsqualität in Large-Scale-Studien: Optimierung der Modellierung und Itemauswahl“ (Trautwein, Lüdtke, Klieme, Nagengast & Wagner, 2011) untersucht.

Eine Darstellung der Ergebnisse dieser Studie liefern Göllner et al. (2016). Um der Frage der Validität der Beurteilungen des Unterrichts durch Schülerinnen und Schüler nachzugehen, wird hier vor allem die durch Wagner, Göllner, Helmke, Trautwein und Lüdtke (2013) mittels Mehrebenen-Strukturgleichungsmodelle ausgewertete Studie Deutsch-Englisch-Schülerleistungen-International (DESI) („DESI“, 2018; Klieme, 2018) angeführt. Im Rahmen von DESI wurden zwischen 2001 und 2006 über 400 Schulklassen des Jahrganges 9 hinsichtlich ihres Leistungsstandes in der Beherrschung der englischen und der deutschen Sprache untersucht. Laut Göllner et al. (2016) wurden in Wagner et al. (2013) fünf Unterrichtsdimensionen untersucht: Schülermotivierung, Verständlichkeit, Schülerorientierung, Strukturiertheit und Klassenführung. Im Ergebnis zitiert Göll-

ner et al. (2016) Wagner et al. (2013) so, dass im Ergebnis eine dimensionale Struktur der Beurteilungen durch Schülerinnen und Schüler vorliegt. Innerhalb von Klassen und auch zwischen Klassen konnten empirische Dimensionen getrennt werden: „Das Urteil von Schülerinnen und Schülern spiegelt nicht nur primär stabile oder Transkontingente Anteile der Beurteiler wieder, sondern beschreibt die spezifische Qualität einer Lehrkraft entlang verschiedener Qualitätsdimensionen“, (Göllner et al., 2016, S. 68). Um die Validität der Aussagen von Schülerinnen und Schülern abzusichern, wurde in Wagner et al. (2013) die Vergleichbarkeit der Beurteilungen von Schülerinnen und Schüler über verschiedene Kontexte, also Fächer, verschiedene Klassen und Unterrichtsinhalte hinweg geprüft. Göllner et al. (2016) bewertet die Ergebnisse aus Wagner et al. (2013) so, dass die Kontextunabhängigkeit nur eingeschränkt gegeben ist. Aspekte, die einen hohen Schüler- und Schülerinnenbezug haben, können nur eingeschränkt lehrkraftunabhängig verglichen werden. Ursachen dafür werden in Schüler- und Schülerinnenmerkmalen gesehen, die eine erhebliche Rolle spielen und sich in verschiedenen Klassen stark von einander unterscheiden.

Insgesamt resümieren die Autoren Fauth et al. (2014b), dass die Einschätzungen von Grundschulschülerinnen und -schülern dennoch brauchbare Ergebnisse zur Bewertung von Unterrichtsqualität liefern. Dem stimmen auch andere Autoren (Clausen, 2002; Gruehn, 2000; Helmke, Piskol, Pikowsky & Wagner, 2009) zu. Es hängt von den Unterrichtsaspekten ab, zu denen man eine Aussage haben möchte.

Auch Kunter und Baumert (2006) erhalten bei ihrem Vergleich der Bewertung von Unterrichtsqualität zwischen Lernenden und Lehrenden eine hohe Bandbreite der Korrelation im Bereich von $.09 \leq r \leq .64$. Kunter et al. (2011) gehen aufgrund verschiedener Analysen, die sie zur COACTIV-Studie ausgewertet haben, wie der in Kunter und Baumert (2006), davon aus, dass Lehrende und Lernende Aspekte der Klassenführung gleichermaßen gut beurteilen können (Kunter et al., 2011, S. 92). Schülerinnen und Schüler erfassen die Unterstützungsdimensionen am besten, Lehrerinnen und Lehrer können die intendierte Anlage des Unterrichts am besten einschätzen. Der Einschätzung der Autoren nach können sowohl Lernende, als auch Lehrende das Potenzial zur Kognitiven Aktivierung des Unterrichts nur sehr begrenzt einschätzen. Hier sind Expertenanalysen notwendig, um die eingesetzten Aufgaben zu beurteilen.

Eine Zusammenfassung des Wissens zur Stabilität der Unterrichtswahrnehmung aus der Perspektive von Schülerinnen und Schülern findet man unter anderem in Wagner et al. (2015, S. 3) und in Praetorius, Pauli, Reusser, Rakoczy und Klieme (2014, S. 3). Im Folgenden werden Teile dieser Zusammenfassungen wiedergegeben.

Die überwiegende Mehrheit der Studien geht davon aus, dass die Lehrqualität von Lehrerinnen und Lehrern ein stabiles Verhaltensmuster der Person ist (Brown, Jones, LaRusso & Aber, 2010). In Praetorius et al. (2014) findet sich eine tabellarische Übersicht über die wichtigsten 14 Videostudien zur Unterrichtsqualität. Dieser Übersicht kann man die zum Teil sehr kurzen Beobachtungszeiträume entnehmen. In manchen, wie in TIMSS 1999, liegt der Beobachtung nur eine Unterrichtsstunde zugrunde. Neuere empirische Erkenntnisse stellen die Beobachtungsqualität durch nur wenige Beobachtungsstunden aber in Frage. Das Hauptargument gegen kurze Beobachtungszeiträume ist, dass verschiedene Facetten der Lehrqualität unterschiedlich stabil sind, da von Stunde zu Stunde Inhalte, Interaktions- und Lehrmethoden wechseln. Vermutlich wird jede Form des Verhaltens Schwankungen unterliegen, die auf die Eigenheiten einer Situation und der Umstände zurückzuführen sind (Göllner et al., 2016). Auf Seite 4 zitieren die Autoren Praetorius et al. (2014) den Beitrag Brophy (2006), der zum Beispiel 20 bis 30 Stunden Beobachtung für notwendig erachtet, um Unterrichtsqualität einschätzen zu können.

Die Autoren Wagner et al. (2015, S. 3, Eigene Übersetzung) fassen die Analysen der Variabilität dennoch damit zusammen, dass „es vernünftig scheint anzunehmen, dass Lehrer- und Schülerbewertungen von Lehrqualitäten über die Zeit hinweg konsistent sind.“ Allerdings schränken sie auch ein, dass die Beurteilung über mehrere Messzeitpunkte genauer als zu einem einzelnen Messzeitpunkt sein sollte.

In Praetorius et al. (2014) werden bestehende Studien und Metastudien zur Qualitätsstabilität diskutiert. Die Studie selbst kommt zu dem Ergebnis, dass die Basisdimension Klassenführung (hier als *Classroom management* geführt) über mehrere Beobachtungen sehr stabil war. Nach Einschätzung der Autoren reicht daher eine Unterrichtsstunde aus, um ein Reliabilitätsniveau von .70 zu erreichen. Zu einer sehr ähnlichen Einschätzung kommen die Autoren bei der Basisdimension Emotionale Unterstützung (hier als *personal learning support* geführt). Um aber die Kognitive Aktivierung (hier als *cognitive activation* geführt) richtig einschätzen zu können und ein verlässliches Reliabilitätsniveau zu erhalten, waren in der Studie mindestens 9 Unterrichtsstunden Beobachtungen nötig.

Je nach Qualitätsaspekt sind also verschiedene Anforderungen an die Beobachtungsdauer zu richten.

Viele Befragungsinstrumente geben darüber hinaus nicht den Beurteilungszeitraum für die Unterrichtsqualität an. Gerade bei Befragungen von Schülerinnen und Schülern ist der zeitliche Bezug, den diese beim Ankreuzen der Fragebögen wählen, wichtig. Wählen einige als Bezugszeitraum das vergangene Schuljahr, andere die letzten Unterrichtsstun-

den, könnten, ein variierendes Verhaltensmuster der Lehrperson vorausgesetzt, sehr verschiedene Ergebnisse herauskommen. Der Vergleich, auch mit anderen Datenquellen, ist dann nicht möglich (Göllner et al., 2016). Göllner et al. (2016) führen an, dass in dem BMBF Projekt „Erfassung der Unterrichtsqualität in Large-Scale-Studien: Optimierung der Modellierung und Itemauswahl“ (Trautwein et al., 2011) analysierte Daten aus dem ebenfalls BMBF geförderten Projekt „Lernen mit Plan“ (Orgin, Silber, Friedrich, Trautwein & Schmitz, 2017) zeigen, dass für Hauptschüler und Hauptschülerinnen im Alter zwischen 11 und 12 Jahren für das Fach Mathematik eine geringere Zeitspezifität besteht als für die Urteile der Lehrpersonen, mit Ausnahme der Qualitätsdimension Klassenführung, hier variieren die Urteile der Lehrkräfte stärker. Demnach erklären sich etwa drei Viertel der Unterschiede zwischen den Beurteilungen durch zeitlich stabile Komponenten. Insgesamt, fasst Göllner et al. (2016) zusammen, beinhaltet die Beurteilung von Unterricht bedeutende Anteile von Zeitbezug. Aus der Perspektive von Schülerinnen und Schülern zeigt sich jedoch eine höhere zeitliche Konsistenz als aus der Perspektive von Lehrerinnen und Lehrern.

Nicht unerwähnt soll bleiben, dass in der schon zuvor angeführten Arbeit von Praetorius et al. (2018) vier Studien gefunden wurden, die die zeitliche Konsistenz der Qualitätsbeurteilung der drei in Deutschland üblichen Basisdimensionen aus der Perspektive von Schülerinnen und Schülern und zwei aus der Perspektive von Lehrerinnen und Lehrern untersucht haben. Die Ergebnisse gehen in den bereits angeführten Quellen auf.

Die Studie der Autoren Fauth et al. (2014b) basiert auf der Befragung von Grundschulschülern und Grundschulschülerinnen. Die Autoren weisen darin nach, dass die Beurteilungen der Qualitätsdimension Kognitive Aktivierung mit den Interessen der Lernenden zum Unterrichtsthema korrelieren.

2.5 Einfluss von Lehrkräftekompetenzen auf Lernleistungen

Lehrkräfte haben einen zentralen Einfluss auf die Gestaltung von Unterricht und damit auch auf die Unterrichtsqualität. Eine dabei naheliegende Frage besteht darin, wie eine Lehrperson sein muss, welche Eigenschaften sie haben muss, damit Schülerinnen und Schüler Lernerfolge erzielen können. Welchen Einfluss hat dabei die Kompetenz von Lehrerinnen und Lehrern auf die Unterrichtsqualität und die Entwicklung von Schülerinnen und Schülern? Am bekanntesten ist der Ansatz zur Klassifikation von unterrichtsrelevanten Merkmalen von Lehrpersonen nach Shulman (1986). Hier werden folgende Arten des Lehrerwissens unterschieden (Baumert & Kunter, 2006; Haag & Lohrmann, 2006;

Helmke, 2015):

1. Fachwissen in dem unterrichteten Fach,
2. fachübergreifendes pädagogisches Wissen,
3. fachdidaktisches Wissen,
4. curriculares Wissen und
5. Philosophie des entsprechenden Schulfaches.

Andere bedeutende Klassifikationen stammen unter anderem von Weinert (2001) oder Bromme und Haag (2004).

Die bedeutendste nationale Studie zum Professionswissen von Lehrkräften im Fach Mathematik ist die COACTIV-Hauptstudie, die als Teiluntersuchung der nationalen PISA-Studie 2003/2004 empirische Daten aufgenommen hat. Dabei wurden Schülerinnen und Schüler zehnter Klassen im Abstand von einem Jahr befragt. Daten dieser Studie werden auch in Kunter et al. (2013) ausgewertet. In dieser Studie werden im Wesentlichen drei Hypothesen dazu, welche lehrerbezogenen Merkmale die besten Lernerfolge erzielen, untersucht:

Die eine Hypothese geht auf Kennedy, Ahn und Choi (2008) zurück und wird als „Bright-Person-Hypothese“ (BPH) bezeichnet. Diese Hypothese unterstellt, dass die kognitiven Fähigkeiten einer Lehrperson ausschlaggebend und entscheidend für die Qualität ihres Unterrichtes ist. Nach Feldon (2007) besteht die Ursache dafür in der Idee, dass Unterricht eine sehr anspruchsvolle komplexe Tätigkeit darstellt und die damit verbundenen Aufgaben eine hohe Flexibilität bedürfen. Kunter et al. (2013) führen an, dass diese Hypothese viele Unterstützer hat, unter anderem die OECD, wie aus for Economic Cooperation and Development (2005) klar wird.

Die andere, als „Knowledgeable-Teacher-Hypothese“ (KTH) bezeichnete Hypothese geht auf die bereits erwähnte Klassifikation von Shulman (1986) zurück, bei der erlerntes und profundes Fachwissen die Schlüsselkompetenz zu gutem Unterricht darstellt. Dieser Hypothese folgend gewinnt die Lehrerbildung an erheblicher Bedeutung.

Als dritte Hypothese wird das Konzept der professionellen Kompetenz eingeführt. Es verbindet verschiedene Erklärungsansätze und wird aus anderen hochkomplexen Professionen des Arbeitslebens auf den Lehrberuf übertragen. Es definiert Kompetenz als Summe aus Fähigkeiten, Fachwissen, Einstellung und anderen motivationsabhängigen Faktoren, die in diesem hochspezialisierten Berufsfeld zum Erfolg führen (Epstein & Hundert, 2002; Klieme, Hartig & Rauch, 2008). Mehr dazu in Kunter et al. (2013).

Im Ergebnis werden die in der Abbildung 2.2 dargestellten Effekte der verschiedenen Einflussgrößen in Kunter et al. (2013) nachgewiesen. Betrachtet wird dabei die Klassebene, aber auch die individuelle Ebene einzelner Schülerinnen und Schüler.

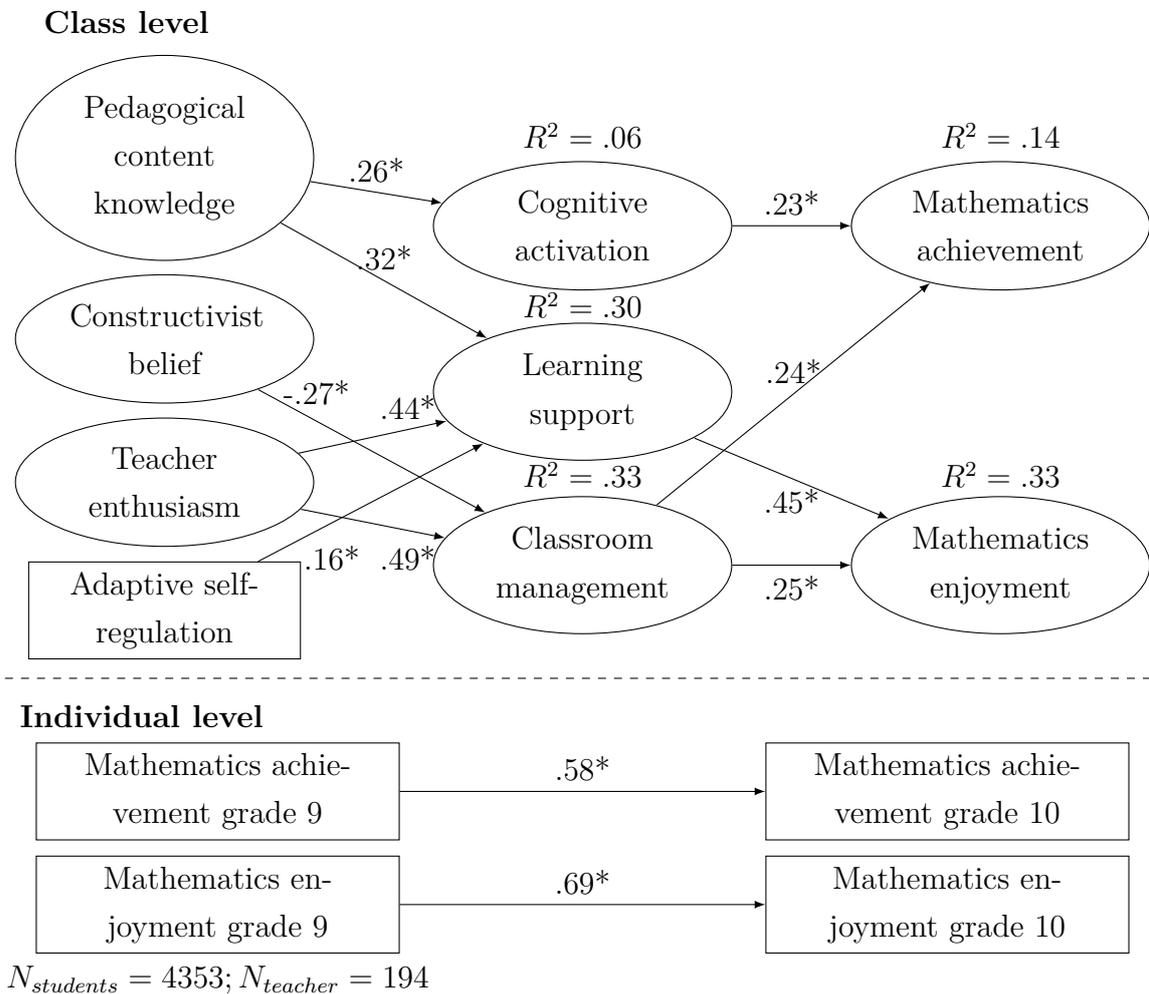


Abbildung 2.2: Vollständiges Vermittlungsmodell für die professionelle Kompetenz von Lehrkräften, das die Vorhersagekraft für Unterrichtsqualität und Schülerergebnisse zeigt. Ellipsen stellen Variablen dar, die auf mehreren Indikatoren basieren. Rechtecke stellen Variablen dar, die auf einzelnen Indikatoren basieren, * $p < .05$; aus Kunter et al. (2013, S. 814) entnommen, dort als *Figure 1* deklariert.

Zusammenfassend stellen Kunter et al. (2013) fest, dass weder die BPH noch die KTH alleine ausreichen, um Lernerfolge für Schülerinnen und Schüler zu erklären. Es konnte gezeigt werden, dass allgemeine kognitive Fähigkeiten der Lehrperson keinen Zusammenhang zu ihrem Lehrverhalten liefern. Hier zählen eher die berufsbezogenen Kompe-

tenzen, wie sie in der KTH beschrieben sind. Bestimmte Aspekte der allgemeinen kognitiven Fähigkeiten der Lehrpersonen, von denen vermutet werden darf, einen Einfluss auf Schülerinnen- und Schüler-Lernleistungen zu haben, wurden allerdings in dieser Studie nicht untersucht. Motivationale Aspekte der Lehrpersonen spielen eine bedeutende Rolle und haben Effekte auf das Lernverhalten der Schülerinnen und Schüler. Letztlich stellt die formulierte dritte Hypothese den besten Kompromiss als Erklärungszusammenhang zwischen Merkmalen der Lehrperson und Schülerinnen- und Schülerleistungen dar. So das Fazit von Kunter et al. (2013).

Vor dem Hintergrund der vorliegenden Arbeit sind diese Erkenntnisse bedeutend, da über die Kenntnis der Unterrichtsqualität einer Lehrperson Maßnahmen zur Verbesserung der individuellen Fortbildungsnotwendigkeit im Bereich des Professionswissens Effekte für die Unterrichtsqualität und damit für die Lernleistungen von Schülerinnen und Schülern erzielt werden könnten. Würde die Qualität der Lernleistungen von Schülerinnen und Schülern einzig durch die BPH erklärt werden, dann wären fortbildende Maßnahmen ineffektiv und eine Messung der Unterrichtsqualität könnte lediglich als Detektor, nicht aber als Steuerungsinstrument dienen. Ungenügende Aspekte der Unterrichtsqualität könnten nicht verbessert werden. Dem ist nachgewiesenermaßen nicht der Fall.

2.6 Unterrichtswahrnehmung und Leistung

Schülerinnen und Schüler sind im Unterricht ein Teil des zu beurteilenden Prozesses und damit in besonderem Maße involvierte Akteure. Göllner et al. (2016) verweisen auf Gigliotti und Buchtel (1990), Greenwald und Gillmore (1997) und Marsh und Roche (1997), deren empirische Studien schon in den 1990er Jahren zu dem Ergebnis kommen, dass die Beurteilungen durch Schülerinnen und Schüler von beispielsweise dem Geschlecht, dem Interesse, der individuellen Urteilstendenz (Milde-Strenge-Effekt) und auch von ihrem Leistungsstand abhängen. Auch Clausen (2002) weist der Schülerinnen-Schülerperspektive Halo-Effekte nach, die zu einer schlechteren Beschreibung verschiedener Qualitätskriterien führen können.

Diese Aussagen werden auch durch Fauth et al. (2014b) gestützt. Auf den Seiten 6ff. kann man zum Beispiel nachlesen, dass die Einschätzung der Schüler und Schülerinnen zur Kognitiven Aktivierung nicht mit deren Lernleistungen korrelieren, sie können, so die Interpretation der Autoren, diese Unterrichtsdimension nicht gut beurteilen. Diese

Ergebnisse beziehen sich nicht auf die Individualebene, sondern auf die Klassenebene. Auf die Individualebene hingegen bezieht sich die Erkenntnis, dass die Beurteilung keines Unterrichtsmerkmals prädiktiv für die Leistungen der Schülerinnen und Schüler ist (Fauth et al., 2014b).

3 Ableitung der Fragestellung und Hypothesen

Im vorangegangenen Kapitel 2 wurden die bedeutendsten Erkenntnisse bisheriger Forschungen rund um das Thema *Unterrichtswahrnehmungen durch Schülerinnen und Schüler* in Bezug auf Stabilität und Leistungsabhängigkeit angeführt und es wurde auf das Einflussvermögen auf bestimmte Aspekte der Lehrkräfte eingegangen. Nun sollen die der Studie zugrundeliegenden Fragestellungen und Ausgangshypothesen formuliert werden.

3.1 Stabilität der Unterrichtswahrnehmung

Die übergeordnete erste Hauptfragestellung der vorliegenden Arbeit, „Wie stabil ist die Unterrichtswahrnehmung?“ unterteilt sich in folgende drei Teilfragestellungen:

(a) Stabilität des Qualitätsniveaus

Wie stabil ist das beobachtete Qualitätsniveau in den einzelnen Dimensionen?

Diese Frage betrachtet die intraindividuelle Veränderung der beobachteten Qualitätsniveaus. Die Fähigkeit, Unterricht kognitiv für Lernende aufzubereiten und ein Thema gut strukturell anleiten zu können, hängt, so die hier zugrundeliegende These, von den fachlichen Kompetenzen der Lehrperson ab, diese ist sicher themenabhängig. Es lässt sich daher vermuten, dass hier keine Stabilität über die Zeit zu erzielen ist, im Gegensatz zu den Qualitätsmerkmalen Emotionale Unterstützung und Klassenführung. Hier wird ein hohes Maß an Stabilität im Qualitätsniveau erwartet.

(b) Stabilität der Rangfolge

Wie stabil sind die Einschätzungen der Unterrichtsqualität in den vier Dimensionen Kognitive Aktivierung, Instruktionale Unterstützung, Emotionale Unterstützung und

Klassenführung über die Zeit?

Hier geht es um die Stabilität der Rangfolge interindividueller Unterschiede. Aus der Theorie heraus lässt sich vermuten, dass die Beurteilung der Kognitiven Aktivierung durch Schülerinnen und Schüler schlechter gelingt als beispielsweise die Beurteilung der Emotionalen Unterstützung. Daher wird an dieser Stelle vermutet, dass die Kognitive Aktivierung keinen hohen Grad an Stabilität im Vergleich zu den anderen Qualitätsdimensionen erzielen wird.

(c) Stabilität zwischen und innerhalb der Unterrichtsthemen

Sind die Beurteilungen der Unterrichtsqualität themenunabhängig?

Gestützt auf den im Abschnitt 2.5 beschriebenen Wissensstand wird hier vermutet, dass die Beurteilungen der Kognitiven Aktivierung und der Instrukionalen Unterstützung von dem didaktischen Vermögen der Lehrkraft abhängig sind. Die Fähigkeit empathisch auf die Bedürfnisse der Schülerinnen und Schüler einzugehen oder im Klassenraum eine angemessene Lernatmosphäre zu schaffen, so die These, sind eher in der Persönlichkeit der Lehrkraft verhaftet. Daher besteht die Hypothese, dass die Beurteilung der Unterrichtsqualität in den Dimensionen der Kognitiven Aktivierung und der Instrukionalen Unterstützung themenabhängig sind, die der Emotionalen Unterstützung und der Klassenführung keine Themenabhängigkeit zeigen, also themenunabhängig stabil sind.

3.2 Unterrichtswahrnehmung und Lernleistungen

Die zweite Hauptfragestellung der vorliegenden Arbeit lautet:

„Welchen Zusammenhang zeigen die Mathematikleistungen und das mathematikbezogene Selbstkonzept mit den Unterrichtsbeurteilungen der Schülerinnen und Schüler?“

Der Anspruch, jedem Schüler und jeder Schülerin im Unterricht gerecht zu werden, erfordert eine leistungsgerechte Binnendifferenzierung. Wenn diese im Idealfall auf die Unterrichteten zugeschnitten ist, wäre hier zu erwarten, dass alle Schülerinnen und Schüler, unabhängig von ihren Leistungsniveaus, gleichermaßen gut kognitiv aktiviert würden. Diesem hohen Anspruch gerecht zu werden, ist das große Ziel eines jeden Unterrichts. Zu erwarten ist hier, dass Schülerinnen und Schüler mit hoher mathematischer Leistungsfähigkeit die Qualitätsdimension der Kognitiven Aktivierung tendenziell anders

beurteilen als Schülerinnen und Schüler mit einer nur geringen mathematischen Leistungsfähigkeit. Ähnlich verhält es sich mit der Instrukionalen Unterstützung. Sich gut unterstütztühlende Schülerinnen und Schüler sollten gemäß Cornelius-White (2007) eine höhere Mathematikleistung erzielen. In diesen beiden Qualitätsdimensionen wird eine leistungsabhängige Einschätzung des erlebten Unterrichts vermutet. In den Dimensionen Emotionale Unterstützung und Klassenführung hingegen wird davon ausgegangen, dass die Leistung keinen geeigneten Prädiktor darstellt.

Anders sollte es sich mit dem mathematikbezogenen Selbstkonzept der Schülerinnen und Schüler verhalten. Insbesondere Lernende mit einem niedrigen Selbstkonzept dürften sensibel auf die geleistete Emotionale Unterstützung und die Klassenführung reagieren. Vermutet wird an dieser Stelle, dass das mathematikbezogene Selbstkonzept ein geeigneter Prädiktor für die Emotionale Unterstützung und die Klassenführung darstellt.

In der vorliegenden Arbeit soll der Fokus auf der Individualebene der Schülerinnen und Schüler liegen, das heißt, es sollen individuelle Einschätzungen und Wahrnehmungen der Schülerinnen und Schüler der Unterrichtsqualität betrachtet werden. Die unter Abschnitt 2.4 zusammengetragenen Forschungen und Befunde stellen zumeist die Klassenebene, das heißt die Unterrichtsqualität in Klassen oder bei einer Lehrkraft, in den Mittelpunkt. Die Analysen dieser Studien stützen sich im Wesentlichen auf Klassenmittelwerte. Darum soll es hier nicht, beziehungsweise nur bedingt gehen. Die in Abschnitt 2.4 angeführten Ergebnisse lassen sich also nicht ohne weiteres übertragen. Darin sollen *das Neue* und *der Mehrwert* dieser Arbeit liegen.

4 Methode

In diesem Kapitel wird das methodische Vorgehen der Studie dargestellt. Es werden die Testklassen und die Stichprobe beschrieben, es wird auch auf die verwendeten Testinstrumente und deren jeweilige Güte eingegangen.

4.1 Überblick über das Vorgehen

Die Messung der Unterrichtsbeobachtungen mit den erstellten Fragebögen erfolgte über das Schuljahr 2016/17 verteilt zu 8 Messzeitpunkten. Zum ersten und zum achten Messzeitpunkt wurde über die Messung der Unterrichtsqualität hinaus die Messung der individuellen mathematikbezogenen Selbstkonzepte vorgenommen. Ursprünglich war bei der zeitlichen Wahl der Messzeitpunkte eine engere Kopplung an konkrete unterrichtsinhaltliche Schwerpunkte vorgesehen. In der konkreten Umsetzung gelang das nicht mehr in aller Konsequenz. Da die Testklassen mit dem Schuljahr neu an der Testschule waren und neu zusammengestellt wurden, erschien ein Beginn der Beobachtungsserie zu einem Zeitpunkt im fortgeschrittenen Schuljahr sinnvoll. Feiertage, Krankheit der Kollegen und Kolleginnen in den Testklassen, klasseninterne Wandertage und andere Störungen der Regelmässigkeit waren der Grund für die zeitlich nicht äquidistante Verteilung der Messzeitpunkte. Da die individuellen Mathematikstunden in den drei Testklassen innerhalb einer Unterrichtswoche nicht parallel lagen, sind auch die genauen Messzeitpunkte innerhalb der entsprechenden Wochen nicht unbedingt an gleichen Tagen. Wann innerhalb der entsprechenden Woche die Messung vorgenommen wurde, ist also in den Klassen verschieden. Es wurde aber darauf geachtet, dass der Abstand bei enger liegenden Messzeiträumen möglichst gestreckt wurde, d.h. die Messungen fanden je nach konkreter Situation eher zu Beginn der Woche oder eher am Ende der Woche statt.

Neben den neu erstellten Fragebögen zur Erfassung der Unterrichtsqualität wurden die mathematischen Kompetenzen der Schülerinnen und Schüler durch vier Klassenarbeiten, die Halbjahresnote, einen standardisierten Test (KERMIT 5) zu Beginn des Schuljahres durch das IfBQ und eine Testung am Ende der Messreihe erfasst. Die Klassenarbeiten

wurden von den unterrichtenden Lehrpersonen entwickelt und waren in allen Testklassen gleich. Der Leistungstest KERMIT 5 wurde durch das IfBQ Hamburg durchgeführt, hier werden lediglich die dort ermittelten Ergebnisse genutzt. Der Messzeitpunkt konnte nicht beeinflusst werden, stellte sich allerdings als ein geeigneter Zeitpunkt heraus, um das Leistungsniveau zu Beginn der Messserie zu erhalten. Die Leistungstestung am Ende der Messreihe speist sich aus Items der LAU-Testungen. Trotzdem es sich ausdrücklich nicht um die LAU-Testung handelte und auch keine entsprechende Skalierung der Items vorlag, wird diese Testung im Folgenden mit LAU bezeichnet.

Die Abbildung 4.1 stellt die zeitliche Einordnung der Messungen in das Schuljahr graphisch dar. Die acht Messungen der Unterrichtsqualität sind darin rechts durch **M1** bis **M8** abgekürzt. Links daneben befinden sich die Messzeitpunkte der Leistungsmessungen, die Klassenarbeiten eins bis vier werden durch **K1** bis **K4**, sowie die Halbjahresnoten des ersten und zweiten Schulhalbjahres durch **H1** und **H2** abgekürzt. Links in der Grafik sind kurz die Überschriften der jeweiligen Unterrichtsinhalte dargestellt.

Unterrichtsinhalte	Woche	Messungen	
Kein Unterricht	1		
Kennenlernwoche	2		
Statistik	3		
	4	KERMIT 5	
	5		
Klassenreise	6	K 1	
	7		
Ferien	8		
	9		
Natürliche Zahlen	10		
	11		
	12		
Anti-Mobbing-Projekt	13		
Natürliche Zahlen	14		
	15		M 1
	16		
Ferien	17	K 2	
	18		
Ferien	19		
Forscher-Projekt	20		
Figuren	21		
	22	H 1	M 2
und Körper	23		
	24		
	25		M 3
	26		
Ferien	27	K 3	M 4
	28		
Flächen- und Rauminhalt	29		
	30		
	31		
	32		M 5
	33		
	34		
	35		
36		M 6	
Steinzeit-Projekt	37		
Ferien	38	K 4	
	39		
Brüche und Dezimalzahlen	40		
	41		
	42		M 7
	43		
	44		
	45		
Brüche und Dezimalzahlen	46	LAU	M 8
	47	H 2	

1. Schulhalbjahr

2. Schulhalbjahr

Abbildung 4.1: Darstellung des zeitlichen Ablaufs der Studie im Schuljahr 2016/17

Die Datenerhebung selbst wurde mittels der Fragebögen durch die Fachlehrer und Fachlehrerinnen durchgeführt und dauerte jeweils zwischen 30 und 45 Minuten.

Datenauswertung

Alle nachfolgenden Berechnungen wurden mit der Software SPSS (Corp., 2015) und mit der Software *Mplus* (L. K. Muthén & Muthén, 2014) durchgeführt. Bei der Analyse der erhobenen Daten wurde wie folgt vorgegangen:

Zunächst muss die Güte des neu zusammengestellten Fragebogens mit seinen Skalen ermittelt werden.

- Dazu wurden die Skalen zur Messung der Unterrichtsqualität hinsichtlich ihrer Reliabilität hin untersucht. Das bedeutet, für die Skalen wurden zu jedem Messzeitpunkt ihre internen Konsistenzen Cronbachs α berechnet und eine Itemanalyse durchgeführt, also die Homogenität und Trennschärfe berechnet.
- Da jede Skala zu jedem Messzeitpunkt gute Qualität nachwies, wurden die individuellen Mittelwerte der Schülerinnen und Schüler über die einzelnen Skalen zu jedem Messzeitpunkt berechnet.

Um die Fragestellungen zur Stabilität der Unterrichtswahrnehmung aus Abschnitt 3 zu untersuchen, wurden folgende Analysen und Berechnungen vorgenommen:

- Es wurden für jede Skala zur Untersuchung der Unterrichtsqualität die individuellen Mittelwerte eines jeden Schülers, einer jeden Schülerin über alle 8 Messzeitpunkte berechnet. Für diesen Wert wird im Folgenden die Symbolik \bar{M} verwendet.
- Im Anschluss wurde für jede der 4 Qualitätsdimensionen die Korrelationen zwischen den jeweiligen Skalenmittelwerten zu jedem Messzeitpunkt berechnet. Dabei wurden die Daten, nach dem Vorbild der Studie von Göllner et al. (2018), über die jeweiligen Klassen hinweg zentriert, um klassenspezifische Tendenzeffekte bei der Beantwortung der Items durch die Schülerinnen und Schüler herauszurechnen.
- Um die Qualität der Leistungsdaten abzusichern, wurden die Korrelationen der Leistungsmessungen untereinander berechnet und dabei gute Werte nachgewiesen.
- Damit wurden dann für jede Qualitätsskala, für jeden Messzeitpunkt und den jeweiligen Mittelwert \bar{M} die Korrelationen zu den individuellen Leistungsdaten berechnet, um Hinweise auf konvergente Validität zu finden.

- Danach wurde die Signifikanz der Mittelwertveränderungen über die 8 Messzeitpunkte mittels einer Varianzanalyse berechnet.
- Im Anschluss daran wurden Interaktionseffekte (mixed ANOVA) zwischen der Zeit und der Klasse berechnet, um Haupteffekte zu untersuchen.
- Dann wurden lineare Strukturgleichungsmodelle (LGM) zur Messung der Veränderung der eingeschätzten Unterrichtsdimensionen über die acht Messzeitpunkte mit der Software *Mplus* (L. K. Muthén & Muthén, 2014) berechnet. *Mplus* wurde auch für die folgenden Berechnungen verwendet.

Zur Untersuchung der Zusammenhänge zwischen der Unterrichtswahrnehmung und der Lernleistung der Schülerinnen und Schüler, wie sie in Fragestellung 2, Kapitel 3 formuliert wurden, kam es zu folgenden Berechnungen:

- Es wurden die Korrelationen zwischen der Mathematikleistung und dem mathematikbezogenen Selbstkonzept zu den linearen Strukturgleichungsmodellen hin untersucht.
- Der Einfluss der Prädiktorvariablen Mathematikleistungen und des mathematikbezogenen Selbstkonzepts auf das lineare Strukturgleichungsmodell wurde dann als Modell mit Kovariaten berechnet.
- Der Zusammenhang zwischen der Unterrichtswahrnehmung und der Lernleistung wurde als Regressionsanalyse berechnet und als dazu passende Pfaddiagramme, auch mit höherer autoregressiver Ordnung, dargestellt. Wie schon zuvor wurden hier ebenfalls nach dem Vorbild von Göllner et al. (2018) die Daten der Unterrichtswahrnehmung auf der einen Seite und die Leistungsdaten auf der anderen Seite zentriert, um einerseits klassenspezifische Tendenzeffekte bei der Beantwortung der Items durch die Schülerinnen und Schüler und andererseits klassenspezifische Beurteilungseffekte bei der Notengebung herauszurechnen.

4.2 Beschreibung der Stichprobe

Insgesamt liegen für die 8 Erhebungswellen Daten von $N = 85$ Teilnehmenden vor. Davon waren $N = 53$ (62,4%) Schülerinnen und 32 (37,6%) Schüler. Die acht Erhebungen fanden als Teilkohorte in drei parallel (allerdings von drei verschiedenen, sich eng abstimmanden Lehrpersonen) unterrichteten fünften Klassen eines vierzügigen hamburg

Gymnasiums statt. Darin liegt eine besondere Qualität der Stichprobe, in der sich diese Studie von den meisten anderen Studien absetzt.

Klasse 1 setzt sich zusammen aus 28 Teilnehmenden, davon $N = 16$ (57,1%) Schülerinnen und $N = 12$ (42,9%) Schülern.

In Klasse 2 waren $N = 18$ (64,3%) Mädchen und $N = 10$ (35,7%) Jungen, also insgesamt $N = 28$ Teilnehmende.

Aus Klasse 3 waren $N = 29$ Teilnehmende, sie setzt sich aus $N = 19$ (65,5%) Mädchen und $N = 10$ (34,5%) Jungen zusammen.

Alle Schülerinnen und Schüler sind 2005 oder 2006 geboren. Die überwiegende Mehrzahl der Schülerinnen und Schüler haben einen Altersabstand von weniger als einem Jahr.

Die jeweiligen Klassen wurden zu Beginn des Schuljahres, in dem die Längsschnittstudie stattfand, neu zusammengesetzt. Dabei kamen alle Schülerinnen und Schüler von verschiedenen Grundschulen neu an dieser weiterführenden Schule zusammen. Die Zusammenstellung der Klassen erfolgte nicht unter Leistungsgesichtspunkten, sondern auch unter dem Aspekt des sozialen Klimas. In Hamburg entscheidet allein der Elternwille und die Entfernung zwischen Wohnort und Schule, auf welche Schule ein Kind nach der Grundschule gehen wird. Das führt zu eingangs sehr leistungsheterogenen Klassen. Die Klassen 1 und 2 werden von ihrem Klassenlehrer beziehungsweise ihrer Klassenlehrerin im Fach Mathematik unterrichtet. Klasse 3 erhält den Fachunterricht nicht bei der Klassenlehrerin beziehungsweise dem Klassenlehrer.

Das Gymnasium, an dem die Erhebungen stattfanden, hatte zu dem Erhebungszeitpunkt den hamburger Sozialindex¹ 5, das geht aus einer kleinen Anfrage der hamburger Bürgerschaft vom 28.02.2013 und der entsprechenden Dokumentation des IfBQ dazu (*Drucksache 20/7094*, 2013, S. 27) hervor. Die Testschule hat demnach eher günstige soziale Rahmenbedingungen. Es ist also davon auszugehen, dass die untersuchte Kohorte eben dem entspricht.

Zu Beginn des Schuljahres waren die Schülerinnen und Schüler aller drei Testklassen im Mittel von gleicher Leistungsstärke, das zeigen die in Tabelle 4.8, auf Seite 48 aufgeführten KERMIT Ergebnisse zu Beginn der fünften Klasse. Darin zu sehen ist auch, dass die Testkohorte insgesamt zu Beginn der fünften Klasse leicht vom hamburger Leistungs-

¹Der Sozialindex beschreibt die sozialen Rahmenbedingungen der Schule und hat eine Skalierung von 1 bis 6, wobei 1 für ein sehr schwieriges Sozialgefüge steht und 6 für sehr günstige soziale Rahmenbedingungen. Weiterführend siehe dazu für Bildungsmonitoring und Qualitätsentwicklung (o.D. a) und Schulte, Hartig und Pietsch (2014).

mittel von 500 Skalenpunkten, (Thonke & Lücken, 2014) auf der KERMIT-Skala nach oben abweicht. Die Schülerinnen und Schüler erreichen im Mittel 521,78 Punkte auf der KERMIT-Skala.

Detaillierte Darstellungen, zu welchem Erhebungszeitpunkt wieviele Schülerinnen und Schüler bei der Berechnung von Skalenmittelwerten berücksichtigt werden konnten, finden sich im Anhang B, ab Seite 109. Entscheidende Abweichungen zu dem hier geschilderten gibt es dabei nicht, deshalb die Auslagerung in den Anhang.

Missing Data

Nicht alle Schülerinnen und Schüler waren zu jedem Zeitpunkt der Erhebung in der Schule. So kam es durch vereinzelte Abwesenheiten, vereinzeltes Auslassen eines Items oder vereinzeltes ungenaues Ankreuzen zu fehlenden Datenpunkten. Die Auslassungen sind dabei ganz klar nicht systemisch begangen oder auf Panelmortalität zurückzuführen. Von den insgesamt 18360² Datenpunkten fehlten am Ende 803³. Das entspricht einem Anteil von rund 4,4%⁴ Missings. Insbesondere bei Analysen über Skalenmittelwerte ist dieser geringe Anteil zu vernachlässigen und nicht weiter durch Imputationsverfahren oder ähnliches (Enders, 2010; Lüdtke, Robitzsch, Trautwein & Köller, 2007) zu behandeln. Skalenmittelwerte einer Person wurden nur bei ausreichend hoher Rücklaufquote gebildet.

4.3 Beschreibung der Instrumente

In diesem Abschnitt sollen die verwendeten Testinstrumente vorgestellt werden. Dabei werden nicht nur selbst zusammengestellte Instrumente verwendet, sondern auch auf etablierte Instrumente, etwa der Leistungsmessung zurückgegriffen.

4.3.1 Der Fragebogen

In diesem Abschnitt soll die Zusammenstellung der Items zu Skalen der entsprechend vorgestellten Dimensionen der Unterrichtsqualität beschrieben werden und der daraus

²26 Items zu 8 Erhebungen entsprechen 208 Datenpunkten, dazu kommen 4 Items zu 2 Datenpunkten zum Selbstkonzept, das bedeutet 216 Datenpunkte je Teilnehmenden, bei 85 Teilnehmenden sind das 18360 Datenpunkte. Davon entfallen auf Klasse 1 und 2 je 6048 und 6264 auf Klasse 3.

³Davon entfallen auf Klasse 1 295, auf Klasse 2 208 und auf Klasse 3 300 Missings.

⁴Klasse 1: 4,9%, Klasse 2 3,3% und Klasse 3 4,8%

resultierende, eingesetzte Fragebogen vorgestellt werden. Der fertige Fragebogen erfasst auch eine anonymisierte Kodierung der Befragten, um eine längsschnittliche Zuordnung zu ermöglichen. Der Fragebogen ist im Anhang in seinem Ausgabelayout dargestellt.

Dimension 1 - Skala zur Kognitive Aktivierung

Diese Skala wurde in Fauth et al. (2014b, S. 8) unter *Cognitive activation* veröffentlicht und für diese Arbeit aus dem Englischen übersetzt zu:

Tabelle 4.1: Skala 1 - Kognitive Aktivierung

	Text
1	Wir bearbeiten im Mathematikunterricht Aufgaben, über die ich sehr gründlich nachdenken muss.
2	Unsere Mathematiklehrerin / unser Mathematiklehrer fragt mich, was ich verstanden habe und was nicht.
3	Unsere Mathematiklehrerin/ unser Mathematiklehrer stellt Fragen, über die ich sehr gründlich nachdenken muss.
4	Unsere Mathematiklehrerin/ unser Mathematiklehrer stellt uns Aufgaben, die auf den ersten Blick schwierig wirken.
5	Unsere Mathematiklehrerin/ unser Mathematiklehrer fragt uns bei einem neuen Thema, was wir schon wissen.
6	Unsere Mathematiklehrerin/ unser Mathematiklehrer stellt uns Aufgaben, über die ich gerne nachdenke.
7	Unsere Mathematiklehrerin/ unser Mathematiklehrer möchte, dass ich meine Antworten erkläre.

Dimension 2 - Skala zur Instruktionen Unterstützung

Aus der in Fauth et al. (2014b, S. 8) unter *Supportive climate* veröffentlichten Skala sind die beiden folgenden Items 15 und 16 übernommen und für diese Studie ins Deutsche übersetzt worden.

	Text
15	Wenn ich einen Fehler mache, sagt unsere Mathematiklehrerin / unser Mathematiklehrer mir, wie ich es besser machen kann.

und

	Text
16	Unsere Mathematiklehrerin / unser Mathematiklehrer sagt mir, worin ich schon gut bin und was ich noch lernen muss.

Das Item 11 ist keiner Literaturquelle entnommen, sondern selbst ergänzt.

Die restlichen Items dieser Skala stammen aus dem Projekt PLUS⁵ und sind in Kauertz et al. (2011) zu finden. Das Projekt PLUS untersucht nicht primär den Mathematikunterricht in Klasse 5. Die Items wurden an diese Untersuchung angepasst und dazu leicht abgewandelt. Item 8 findet sich in Kauertz et al. (2011, S. 8) mit der Item-ID *Sc1*, Item 9 und Item 10 auf Seite 51 mit den Item-IDs *Sc5*, beziehungsweise *Sc6*, Item 12 auf Seite 57 mit der Item-ID *Tm1*, Item 13 auf Seite 51 mit der Item-ID *Tm2* und Item 14 auf Seite 83 mit der Item-ID *Is4*. So ergibt sich die dieser Studie zugrundeliegende Skala zu:

Tabelle 4.2: Skala 2 - Instruktionale Unterstützung

	Text
8	Unsere Mathematiklehrerin/ unser Mathematiklehrer weist darauf hin, was wir uns merken sollen.
9	Im Mathematikunterricht geht es oft um zu viele Fragen gleichzeitig.
10	Im Mathematikunterricht weiß ich oft nicht, worüber wir gerade sprechen.
11	Beim Arbeiten im Mathematikunterricht weiß ich genau, was ich tun soll.
12	Unsere Mathematiklehrerin/ unser Mathematiklehrer interessiert sich für unsere Erklärungen.
13	Unsere Mathematiklehrerin/ unser Mathematiklehrer erklärt oft mit Fremdwörtern, die wir nicht verstehen.
14	Unsere Mathematiklehrerin/ unser Mathematiklehrer hilft mir nur so viel, dass ich auch alleine eine Erklärung finden kann.
15	Wenn ich einen Fehler mache, sagt unsere Mathematiklehrerin/ unser Mathematiklehrer mir, wie ich es besser machen kann.
16	Unsere Mathematiklehrerin/ unser Mathematiklehrer sagt mir, worin ich schon gut bin und was ich noch lernen muss.

⁵PLUS ist die Kurzform für Professionswissen von Lehrkräften, naturwissenschaftlicher Unterricht und Zielerreichung im Übergang von der Primar- zur Sekundarstufe

Dimension 3 - Skala zur Emotionalen Unterstützung

In Fauth et al. (2014b, S. 8) unter *Supportive Climate* sind die folgenden Items, allerdings um die beiden Items 15 und 16 ergänzt, veröffentlicht. Diese beiden Items wurden in Fauth et al. (2014b) der Basisdimension 2 - Instruktionale/Kognitive Unterstützung zugeordnet. Für die Erfassung der Basisdimension 3 sind sie ungeeignet und werden deshalb in dieser Untersuchung nicht verwendet. Es ergibt sich also danach die nun kürzere Skala zu:

Tabelle 4.3: Skala 3 - Emotionale Unterstützung

	Text
17	Unsere Mathematiklehrerin/ unser Mathematiklehrer ist auch dann nett zu mir, wenn ich einen Fehler mache.
18	Ich bin unserer Mathematiklehrerin/ unserem Mathematiklehrer wichtig.
19	Unsere Mathematiklehrerin/ unser Mathematiklehrer ermutigt mich, wenn ich eine Aufgabe schwierig finde.
20	Unsere Mathematiklehrerin/ unser Mathematiklehrer lobt mich, wenn ich etwas gut gemacht habe.
21	Unsere Mathematiklehrerin/ unser Mathematiklehrer glaubt, dass ich schwierige Aufgaben lösen kann.

Dimension 4 - Skala zur Effizienz der Klassenführung

Diese Skala wurde in Fauth et al. (2014b, S. 8) unter *Classroom management* veröffentlicht und hier übersetzt zu:

Tabelle 4.4: Skala 4 - Effizienz der Klassenführung

	Text
22	Im Mathematikunterricht stört keiner den Unterricht.
23	Im Mathematikunterricht sind die Schüler leise, wenn die Lehrerin/ der Lehrer spricht.
24	Im Mathematikunterricht hören alle zu und sind leise.
25	Im Mathematikunterricht stört niemand mit Reden.
26	Im Mathematikunterricht folgen alle der Lehrerin/ dem Lehrer.

Die Selbstkonzeptskala

Die hier verwendete Skala zur Messung des mathematikbezogenen Selbstkonzeptes ist bereits in Bos et al. (2010, S. 62) erwähnt und dort in ihrer psychometrischen Güte beschrieben. Ursprünglich wurde sie für die BIJU⁶ - Erhebung Anfang der 1990er Jahre entwickelt.

Tabelle 4.5: Skala - Mathematikbezogenes Selbstkonzept

	Text
27	Für Mathematik habe ich einfach keine Begabung.
28	Bei manchen Sachen im Fach Mathematik weiß ich von vornherein: „Das verstehe ich nie.“
29	Obwohl ich mir bestimmt Mühe gebe, fällt mir das Fach Mathematik schwerer als vielen meiner Mitschüler und Mitschülerinnen.
30	Was wir in Mathematik durchnehmen, kann ich mir schlecht merken.

Diese Selbstkonzeptskala wurde nur zum ersten und zum letzten Messzeitpunkt an die Schülerinnen und Schüler ausgegeben, das mathematikbezogene Selbstkonzept ist über die Zeit relativ stabil.

⁶Abkürzung: Bildungsverläufe und psychosoziale Entwicklung im Jugendalter

Allgemeines

Die zuvor aufgeführten Skalen sind 4-stufig mit den Antwortkategorien *1= trifft gar nicht zu*, *2=trifft eher nicht zu*, *3=trifft eher zu* und *4=trifft völlig zu* versehen. Um mit aufsteigendem Zahlenwert eine positivere Besetzung zu verbinden, wurden die Items 9, 10, 13 sowie die Items 27 bis 30 in der Auswertung umkodiert.

Die Daten zum Geschlecht und zum Geburtsdatum wurden im Fragebogen nicht erfasst, sie entstammen den Klassenlisten.

4.3.2 Reliabilität

Übersicht über die interne Konsistenz der Skalen

In der Tabelle 4.6 ist die interne Konsistenz, Cronbachs-Alpha, zu jedem Messzeitpunkt, für jede Qualitätsdimension für die Gesamtstichprobe zu sehen. Da α , von wenigen Ausnahmen abgesehen, Werte $\alpha > .70$ und zum Teil sogar Werte $\alpha > .90$ erreicht, ist die Skalenreliabilität durchweg als mindestens akzeptabel einzustufen, (Schnell, Hill & Esser, 2005, S. 153ff.). Die Skalenreliabilitäten unterscheiden sich in ihrer Güte erwartungsgemäß. Wie in Abschnitt 2.4.3 bereits erwähnt, können Schülerinnen und Schüler die Kognitive Aktivierung schlechter beurteilen als die Klassenführung (Kunter et al., 2011).

Tabelle 4.6: Darstellung der internen Konsistenz zu jedem Messzeitpunkt für jede Qualitätsdimension der Gesamtstichprobe

		1	2	3	4	5	6	7	8
KA	α	.69	.52	.63	.77	.78	.78	.74	.77
	N _G	79	77	80	77	78	80	80	78
IU	α	.61	.64	.75	.76	.66	.74	.76	.72
	N _G	76	75	77	73	79	78	79	78
EU	α	.52	.80	.82	.83	.81	.84	.80	.84
	N _G	79	73	76	78	80	75	81	78
KF	α	.86	.91	.90	.95	.93	.93	.91	.94
	N _G	79	79	80	78	82	82	85	80

Eine grafische Darstellung der Skalenreliabilität der Skalen zur Erfassung der Dimensionen der Unterrichtsqualität zu den 8 Messzeitpunkten liefert die Abbildung 4.2. Darin

ist besonders gut zu sehen, dass Schwankungen der Reliabilitäten mit zunehmender Messung geringer werden und sich die Reliabilität stabilisiert.

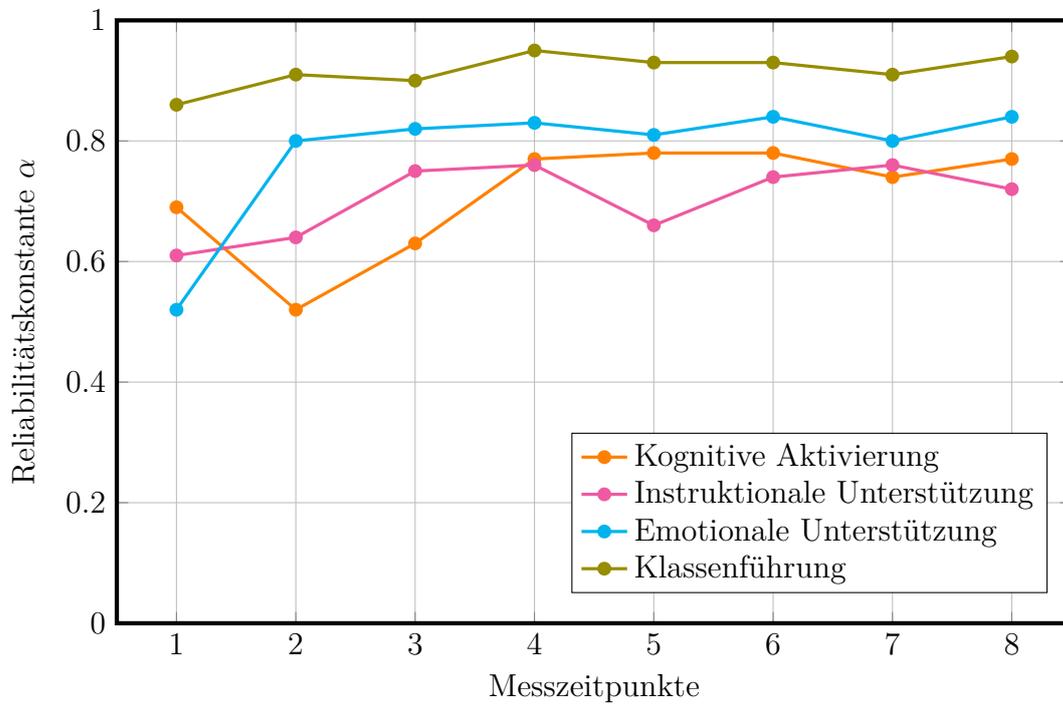


Abbildung 4.2: Darstellung der Reliabilitätskonstante α zu den 8 Messzeitpunkten

Die Reliabilitäten der Selbstkonzeptskala sind entsprechend in der nachfolgenden Tabelle 4.7 erfasst. Da alle berechneten Cronbachs-Alpha Werte von $\alpha > .80$ erreichen, ist die Skala nach Bortz und Döring (2006) als reliabel zu betrachten. Damit bestätigt sich hier die bereits in Bos et al. (2010) eingeschätzte Reliabilität dieser Skala.

Tabelle 4.7: Reliabilität der Selbstkonzeptskala zu den zwei Messzeitpunkten

		1	8
SK	α	.85	.86
	N_G	80	82

4.3.3 Leistungsmessung

Die Leistungsmessung durch Leistungstests

KERMIT 5

Das Akronym KERMIT steht für *Kompetenzen ermitteln*. Diese standardisierten Tests werden seit dem Schuljahr 2012/13 an allen hamburger Schulen durchgeführt. KERMIT ist damit das KESS Nachfolgeinstrument des IfBQ, Hamburgs.

Das Instrument dient der Feststellung, inwieweit die Schülerinnen und Schüler den nationalen und Bildungsanforderungen und den hamburger Bildungsplänen entsprechen. Stärken und Schwächen der Testpersonen und Testklassen werden aufgezeigt und werden den Lehrpersonen zur Verfügung gestellt (für Bildungsmonitoring und Qualitätsentwicklung, o.D. b). Sie sollen zur pädagogischen Unterstützung dienen. Die Testungen finden mit unterschiedlichen Schwerpunkten in den Jahrgangsstufen 2, 3, 5, 7, 8 (ersetzt VERA 8), 9 und 10 statt.

KERMIT 5 ist nach einem Rasch-Modell (Moosbrugger, 2012) skaliert (Thonke & Lücken, 2014). Der Mittelwert liegt dank entsprechender Skalierung hamburgweit genau bei $M = 500$ Punkten und die Standardabweichung bei $SD = 100$ Punkten. Bedeutsame Unterschiede machen bei dieser Skalierung nach Thonke und Lücken (2014) 30 Punkte aus.

LAU

Die Abkürzung LAU steht für *Aspekte der Lernausgangslage und der Lernentwicklung* und ist eine Längsschnittstudie, die Lernstände, Lernentwicklungen und schulbezogene Einstellungen vom Ende der Grundschulzeit bis zur neunten Klasse erfasst (Lehmann, Peek & Gänsfuß, 2011). In Hamburg nahmen dabei im Schuljahr 1996/97 alle Schülerinnen und Schüler staatlicher Schulen zu Beginn der fünften Klassenstufe, an der Vollerfassung LAU 5 teil. Zur Bestimmung der Leistungsstände wurden zwei standardisierte Leistungstests herangezogen. Der für diese Untersuchung relevante Teil zur Erfassung der mathematischen Leistungsfähigkeit geht in dem *Hamburger kombinierten Schulleistungstest für vierte und fünfte Klassen*, kurz KS HAM 4/5 auf. Näheres dazu findet sich unter anderem in Mietzel und Willenberg (1996). Die Fragebögen zur Erfassung des mathematischen Leistungsstandes sind bisher nicht veröffentlicht und wurden für diese Erhebung durch das IPN Kiel zur Verfügung gestellt. Laut Lehmann, Peek

und Gänsfuß (1997) ist jedoch klar, dass 30 Fragen in den Leistungstest aufgenommen wurden und eine interne Konsistenz von $\alpha = .85$, bei $N = 12250$ Schülerinnen und Schülern hatten.

Näheres zu LAU 5 kann man unter anderem in Lehmann et al. (1997) nachlesen.

Welche konkreten 30 Fragen in LAU verwendet wurden, ist nicht veröffentlicht. Es wurden in der Leistungsmessung für diese Arbeit daher alle zur Verfügung gestellten 51 Testfragen verwendet. Anders als bei LAU fand hier bei der Auswertung keine Rasch-Skalierung Anwendung. Für die Auswertung dieser Erhebung wurden die Ergebnisse so skaliert, dass 1000 Punkte bei vollständiger und richtiger Lösung zu erreichen waren. Ein richtiger Vergleich zu den mittels KERMIT 5 eingangs erhobenen Leistungsdaten ist dennoch nicht möglich, denn diese waren ja nach dem Rasch-Modell skaliert (Koller, Alexandrowicz & Hatzinger, 2012).

Es bleibt also zu betonen, dass der Leistungstest, der in dieser Arbeit genutzt wird, zwar auf der Grundlage von LAU-Items entstanden ist, es sich aber ausdrücklich nicht um das LAU Instrument handelt.

Die Leistungsmessung durch Noten

Als weiteres Maß der Leistungsmessung in den drei Testklassen werden die Noten der vier im Schuljahr geschriebenen Klassenarbeiten herangezogen. Es wurden je zwei in jedem Halbjahr in den drei Klassen inhaltlich mit gleicher Aufgabenstellung und zeitlich parallel geschrieben und nach gleichem Bewertungsmaßstab benotet. Des Weiteren konnten die Halbjahresnoten, die sich aus der laufenden Kursarbeit während des Halbjahres und den je zwei schriftlichen Klassenarbeiten zusammensetzten, genutzt werden. Die beiden Halbjahresnoten sind unabhängig von einander entstanden, das heißt, dass die Note des zweiten Halbjahres nicht etwa auf Noten des ersten Halbjahres zurückgreift. Alle Noten liegen als Tendenznoten vor, die Bewertungsskala umfasst also 15 Teilschritte von 1+ bis zur 6.

Die Validität, Reliabilität und Objektivität von Schulnoten sind weithin diskutiert. Brügelmann (2018) kommt beispielsweise zu dem Schluss, dass Noten informationsarm und zwischen verschiedenen Klassen nicht vergleichbar sind, da sich Bewertungen in der Regel am Durchschnitt einer Klasse orientieren und unterschiedliche Leistungsprofile mit der gleichen Note belegt werden. In Weber und Petermann (2016) werden sogar die Zusammenhänge zwischen den Schulnoten und Schulangst, Schulunlust und Anstrengungsvermeidung herausgearbeitet. Diese Diskussionen sollen hier nicht geführt werden,

zumal es sich in in dieser Studie um drei verschiedene bewertende Lehrer und Lehrerinnen handelte. Allerdings geben die erteilten Noten den unterrichteten Schülern und Schülerinnen unstrittig eine Rückmeldung, die bei diesen eine Wirkung erzielte, die vielleicht eine Ursache für bestimmte Einschätzungen und Wahrnehmungen darstellte. Ihre Bedeutung für die Testpersonen dürfte, Validität hin oder her, daher eine größere Bedeutung haben als die geprüften Testinstrumente KERMIT oder LAU.

4.3.4 Ergebnisse der Leistungsmessung deskriptiv

Die Tabelle 4.8 führt die deskriptive Beschreibung der Leistungsmessungen auf.

Tabelle 4.8: Deskriptive Beschreibung der Leistung

Test	M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	N	N _{5b}	N _{5c}	N _{5d}
KER	521,78	532,44	517,57	515,71	76,43	72,87	76,25	81,50	83	27	28	28
LAU	753,43	762,50	756,84	741,72	147,57	157,03	117,90	164,53	82	28	25	29
K 1	10,64	10,74	9,64	11,54	2,38	1,91	1,95	2,83	83	27	28	28
K 2	9,10	9,42	8,96	8,93	3,33	3,91	2,99	3,15	82	26	28	28
K 3	9,05	8,89	9,46	8,79	2,95	3,59	1,99	3,11	84	28	28	28
K 4	7,79	7,29	8,21	7,24	3,32	3,15	1,85	4,43	85	28	28	29
H 1	9,78	10,04	9,64	9,68	2,25	2,39	1,68	2,64	83	27	28	28
H 2	9,04	9,29	9,43	8,41	2,57	2,71	1,69	3,07	85	28	28	29

Die nachfolgende Tabelle 4.9 zeigt die Korrelationen r der jeweiligen Leistungsmessungen. Es zeigt sich, dass alle signifikant sind und ausschließlich starke Effekte haben, da $r > .50$, gemessen werden konnten (Cohen, 1988).

Tabelle 4.9: Korrelationen r der jeweiligen Leistungsmessungen

	KERMIT 5	LAU	K 1	K 2	K 3	K 4	H 1	H 2
KERMIT 5	1	.68*	.56*	.67*	.54*	.70*	.67*	.69*
N	83	80	83	82	83	83	83	83
LAU		1	.60*	.60*	.60*	.67*	.62*	.66*
N		82	80	79	81	82	80	82
K 1			1	.46*	.52*	.57*	.70*	.58*
N			83	82	83	83	83	83
K 2				1	.58*	.62*	.75*	.66*
N				82	82	82	82	82
K 3					1	.64*	.60*	.75*
N					84	84	83	84
K 4						1	.76*	.90*
N						85	83	85
H 1							1	.83*
N							83	83
H 2								1
N								85

4.3.5 Modellfits allgemein

Die berechneten Modelle, wie zum Beispiel die in Abschnitt 5.1.1 berechneten linearen Strukturgleichungsmodelle, sind nur dann verwertbar, wenn die Modellfitparameter geeignete Werte erzielen. Leider weichen die Modellfitparameter häufig von den in der Literatur empfohlenen Idealwerten ab. Solche idealen Referenzwerte sind zum Beispiel in Geiser (2010) vorgestellt. Geiser (2010) seinerseits stützt die Zusammenstellung der Fitparameter auf Hu und Bentler (1999) und Schermelleh-Engel, Moosbrugger und Müller (2003). Hiernach werden folgende Bedingungen gefordert: Die Werte für χ^2 entscheiden abhängig von den Freiheitsgraden über die Signifikanz des Tests und damit über eine Ablehnung der Nullhypothese. Für einen Wert des CFI mit $.80 < CFI < .89$ gilt ein Zielmodell als akzeptabel (Segars & Grover, 1998) und es sollte dann dem Unabhängigkeitsmodell vorgezogen werden. Für Werte $CFI > .95$ gilt ein Modell als gut angepasst (Geiser, 2010). Gleiche Cutoff-Grenzen gelten für den TLI , dieser ist allerdings stichprobenabhängig und tendiert bei großen Stichproben zu höheren Werten (Fan, Thompson & Wang, 1999). Für den approximativen Datenfit sollte der $RMSEA$ -Wert kleiner aus-

fallen als .05. Gleiches gilt für den *SRMR*-Koeffizienten für die Gesamtmodellbewertung (Geiser, 2010).

Die zum Beispiel in Abschnitt 5.1.1 dargestellten Berechnungsergebnisse zeigen auf der einen Seite gute Modellanpassungen nach χ^2 und akzeptable Werte für den *TLI* und den *CFI* von größer als .88, auf der anderen Seite jedoch auch zu hohe Werte für den *RMSEA* und den *SRMR*. Widersprüchliche Anpassungsergebnisse werden in der Literatur unterschiedlich diskutiert. Es sind Für und Wider der einzelnen Kriterien im Kontext der Modellanpassung abzuwägen. Die Diskrepanz zwischen den für das Modell günstigen Werten, wie sie der *CFI* aufweist und den, das Modell ablehnenden Werten des *RMSEA*, kann verschiedene Ursachen haben. Wann, welcher Wert wie genutzt werden sollte, ist zum Beispiel in Lai und Green (2016) komplex diskutiert. In dem hier vorliegendem Fall empfiehlt es sich auf Grund der vergleichsweise kleinen Stichprobe ($N = 85$) den *RMSEA* nicht zu verwenden, da dieser bei kleinem Stichprobenumfang zu groß ausfällt (Kenny, Kanistan & McCoach, 2014). Bei Stichprobenumfängen von $N < 250$ schlagen schon Hu und Bentler (1999) einen Fokus vor allem auf den *CFI* zur Gütebeurteilung vor, so besteht hier vor allem gestützt auf den *CFI* ein akzeptabler Modellfit. Dieser Argumentation folgend, werden die linearen Strukturgleichungsmodelle für die vier Qualitätsdimensionen nachfolgend als den Daten angepasst bewertet.

5 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Datenanalysen zur Beantwortung der Fragen aus Kapitel 3 dargestellt. Demnach werden zunächst Analyseergebnisse, die Stabilität der Unterrichtswahrnehmung betreffend, aufgliedert in die Stabilität des Niveaus der Unterrichtswahrnehmung und in die Stabilität der Rangfolge, vorgestellt. Die Frage nach der Stabilität zwischen und innerhalb von Unterrichtsthemen wird in Abschnitt 5.1.3 zusammengetragen. Die Beantwortung der Frage stützt sich auf Analyseergebnisse, die in verschiedenen Unterabschnitten unter anderen Fragestellungen verwendet werden. Den Ergebnissen der Analysen zur Frage nach dem Zusammenhang zwischen Unterrichtswahrnehmung und Leistung widmet sich der letzte Abschnitt 5.2 dieses Kapitels.

Obwohl die Skalen nicht direkt miteinander vergleichbar sind, sie entstammen ja verschiedenen Quellen und wurden unabhängig voneinander entwickelt, so zeigen sich doch interessante Korrelationen zwischen den Qualitätsdimensionen. Diese Berechnungen seien den dezidierten Beantwortungen der Forschungsfragen vorangestellt.

Korrelation zwischen den Qualitätsdimensionen im Mittel über die Messzeiten

Die Tabelle 5.1 zeigt die Korrelationen zwischen den Mittelwerten der Qualitätseinschätzungen über die acht Messzeitpunkte zwischen den vier Qualitätsdimensionen an. Anders könnte man formulieren, dass in 5.1 die Korrelation der eingeschätzten längsschnittlichen Unterrichtsqualität zwischen den vier Qualitätsdimensionen gezeigt wird. Werte mit der Kennzeichnung * sind auf dem Niveau $p < .05$ signifikant.

Tabelle 5.1: Korrelationen zwischen den Qualitätsdimensionen im Mittel über die Messzeiten; Mit * gekennzeichnete Werte sind signifikant auf dem Niveau $p < .05$.

	\bar{M}_{KA}	\bar{M}_{IU}	\bar{M}_{EU}	\bar{M}_{KF}
\bar{M}_{KA}	1	.583*	.711*	.487*
N	68	68	68	68
\bar{M}_{IU}		1	.819*	.532*
N		68	68	68
\bar{M}_{EU}			1	.568*
N			68	68
\bar{M}_{KF}				1
N				68

Die Tabelle 5.1 macht deutlich, dass die Skalen zur Unterrichtswahrnehmung miteinander korrelieren. Alle Korrelationen sind signifikant. Besonders hohe Werte der Korrelation erzielen die Skalen der Unterstützungsdimensionen untereinander. Es lässt sich daraus deuten, dass hohe Werte in der einen Dimension zu hohen Werten in den anderen Dimensionen führen. Das leitet zu Fragen der Einschätzungsstabilität über.

5.1 Stabilität der Unterrichtswahrnehmung

In diesem Abschnitt sollen die zur Beantwortung der Fragen zur Stabilität der Unterrichtswahrnehmungen der Schülerinnen und Schülern notwendigen Analyseergebnisse dargestellt werden.

Zunächst werden die Ergebnisse der Qualitätsmessungen in den Qualitätsdimensionen deskriptiv aufgeführt. Die Stabilität des Qualitätsniveaus auf der Individualebene wurde mit Hilfe linearer Strukturgleichungsmodelle berechnet, deren Ergebnisse aufgeführt werden. Zur Bestimmung der Rangstabilität werden die Korrelationen zwischen den Messzeitpunkten vorgestellt und im Anschluss die Abhängigkeiten zwischen den Wirkungen der Ausprägungen von Messzeitpunkt und Klassenzugehörigkeiten beschrieben (Interaktionseffekte). Zum Ende des Abschnittes werden verschiedene Analysen herangezogen, um die Stabilität zwischen und innerhalb von Unterrichtsthemen zu untersuchen.

Tabelle 5.4: Deskriptive Beschreibung der Emotionalen Unterstützung

	M	M_{5b}	M_{5c}	M_{5d}	SD	SD_{5b}	SD_{5c}	SD_{5d}	N	N_{5b}	N_{5c}	N_{5d}
M 1	3,49	3,57	3,33	3,58	0,21	0,25	0,28	0,17	79	27	27	25
M 2	3,48	3,65	3,26	3,55	0,20	0,18	0,32	0,17	73	25	27	21
M 3	3,32	3,45	3,26	3,24	0,17	0,27	0,23	0,19	76	26	25	25
M 4	3,36	3,31	3,19	3,59	0,10	0,19	0,27	0,16	78	25	26	27
M 5	3,26	3,42	3,11	3,27	0,15	0,27	0,17	0,22	80	25	27	28
M 6	3,26	3,36	3,20	3,23	0,16	0,21	0,26	0,18	75	22	26	27
M 7	3,25	3,32	3,11	3,32	0,13	0,23	0,24	0,21	81	27	28	26
M 8	3,29	3,33	3,23	3,30	0,11	0,15	0,15	0,14	78	27	24	27
\bar{M}	3,35	3,41	3,23	3,41								

Tabelle 5.5: Deskriptive Beschreibung der Klassenführung

	M	M_{5b}	M_{5c}	M_{5d}	SD	SD_{5b}	SD_{5c}	SD_{5d}	N	N_{5b}	N_{5c}	N_{5d}
M 1	2,39	2,70	1,96	2,50	0,24	0,18	0,31	0,26	79	26	26	27
M 2	2,29	2,75	1,68	2,50	0,23	0,15	0,29	0,27	79	26	28	25
M 3	2,17	2,39	1,68	2,43	0,24	0,22	0,26	0,26	80	26	26	28
M 4	2,15	2,38	1,65	2,45	0,18	0,19	0,20	0,19	78	24	27	27
M 5	2,31	2,56	1,97	2,41	0,13	0,12	0,12	0,18	82	27	26	29
M 6	2,10	2,21	1,87	2,21	0,13	0,07	0,15	0,19	82	27	27	28
M 7	2,10	2,27	1,88	2,17	0,16	0,09	0,22	0,25	85	28	28	29
M 8	2,10	2,33	1,75	2,18	0,12	0,12	0,15	0,13	80	28	24	28
\bar{M}	2,21	2,39	1,83	2,40								

Um den Blick auf die Individualebene lenken zu können, wurden lineare Strukturgleichungsmodelle (LGM) zur Messung der Veränderungen der eingeschätzten Unterrichtsdimensionen über die acht Messzeitpunkte mit der Software *Mplus* (L. K. Muthén & Muthén, 2014) berechnet. Die errechneten Modellfits der jeweiligen Modelle für die jeweilige Qualitätsdimension und die daraus folgenden Berechnungen der Modelle für die vier Qualitätsdimensionen werden hier zusammenfassend aufgeführt:

Modellfits

Kognitive Aktivierung

$\chi^2(28) = 457.40$ mit $p < 0.0001$, $CFI = .87$, $TLI = .89$, $RMSEA = .14$, $SRMR = .24$

Instruktionale Unterstützung

$\chi^2(28) = 514.03$ mit $p < 0.0001$, $CFI = .92$, $TLI = .93$, $RMSEA = .12$, $SRMR = .37$

Emotionale Unterstützung

$\chi^2(28) = 517.85$ mit $p < 0.0001$, $CFI = .86$, $TLI = .88$, $RMSEA = .16$, $SRMR = .40$

Klassenführung

$\chi^2(28) = 389.08$ mit $p < 0.0001$, $CFI = .91$, $TLI = .92$, $RMSEA = .12$, $SRMR = .17$

Wie bereits in Abschnitt 4.3.5 auf Seite 49 ausgeführt, werden die hier aufgeführten berechneten Modelle als den Daten angepasst bewertet.

Modell

Die Schülerinnen und Schüler schätzen die Qualitätsdimensionen im Durchschnitt zum ersten Messzeitpunkt folgendermaßen ein (Mittelwert des Intercept-Faktors). Alle Berechnungen sind signifikant:

KA $M_{interc} = 3.119$, $z = 77.377$, $p < .001$ (Zweiseitiger Test)

IU $M_{interc} = 3.217$, $z = 78.473$, $p < .001$ (Zweiseitiger Test)

EU $M_{interc} = 3.454$, $z = 77.498$, $p < .001$ (Zweiseitiger Test)

KF $M_{interc} = 2.330$, $z = 36.782$, $p < .001$ (Zweiseitiger Test)

Die interindividuellen Unterschiede im Ausgangsniveau sind alle signifikant:

KA $VAR_{Intercept} = 0.096$, $z = 4.585$, $p < .001$

IU $VAR_{Intercept} = 0.111$, $z = 5.005$, $p < .001$

EU $VAR_{Intercept} = 0.117$, $z = 4.483$, $p < .001$

KF $VAR_{Intercept} = 0.252$, $z = 4.676$, $p < .001$

Für die Höhe der linearen Veränderung im Durchschnitt (Mittelwert des Slope-Faktors) sind die Veränderungen der Unterstützungsdimensionen und der Qualitätsdimension der Klassenführung signifikant. Nicht so die lineare Veränderung für die Kognitive Aktivierung. Im Durchschnitt kommt es demnach zu keiner linearen Veränderung in dieser Qualitätsdimension:

KA	$M_{Linear} = -0.010,$	$z = -1.436,$	$p = .151$ (Zweiseitiger Test)
IU	$M_{Linear} = -0.029,$	$z = -4.632,$	$p < .001$ (Zweiseitiger Test)
EU	$M_{Linear} = -0.036,$	$z = -4.619,$	$p < .001$ (Zweiseitiger Test)
KF	$M_{Linear} = -0.037,$	$z = -3.633,$	$p < .001$ (Zweiseitiger Test)

Die interindividuellen Unterschiede in der linearen Veränderung sind hingegen alle auf $p < .05$ signifikant:

KA	$VAR_{Linear} = 0.002,$	$z = 2.938,$	$p = .003$ (Zweiseitiger Test)
IU	$VAR_{Linear} = 0.002,$	$z = 2.943,$	$p = .003$ (Zweiseitiger Test)
EU	$VAR_{Linear} = 0.002,$	$z = 2.537,$	$p = .011$ (Zweiseitiger Test)
KF	$VAR_{Linear} = 0.004,$	$z = 2.919,$	$p = .004$ (Zweiseitiger Test)

Die Korrelation des Ausgangsniveaus mit der Veränderung (Korrelation von Intercept- und Slope-Faktor) wird im Folgenden aufgeführt. Zu sehen ist, dass nur die Qualitätsdimension der Emotionalen Unterstützung die Signifikanzanforderungen von $p < .05$ erfüllt. Alle anderen Qualitätsdimensionen liegen außerhalb des Signifikanzbereiches. Nur für die Emotionale Unterstützung gibt es einen Zusammenhang zwischen dem Ausgangsniveau und der Veränderung:

KA	$r = 0.089,$	$z = 0.407,$	$p = .684$ (Zweiseitiger Test)
IU	$r = -0.064,$	$z = -0.343,$	$p = .732$ (Zweiseitiger Test)
EU	$r = 0.526,$	$z = 2.036,$	$p = .042$ (Zweiseitiger Test)
KF	$r = -0.174,$	$z = -0.958,$	$p = .338$ (Zweiseitiger Test)

Der Determinationskoeffizient R^2 zeigt nur signifikante Werte für die 8 Messzeitpunkte in der Reihenfolge der Messungen:

Kognitive Aktivierung

0.502, 0.640, 0.493, 0.549, 0.665, 0.669, 0.689, 0.738 mit jeweils $p < .001$

Instruktionale Unterstützung

0.617, 0.622, 0.694, 0.681, 0.713, 0.700, 0.745, 0.681 mit jeweils $p < .001$

Emotionale Unterstützung

0.559, 0.505, 0.609, 0.613, 0.747, 0.721, 0.741, 0.697 mit jeweils $p < .001$

Klassenführung

0.548, 0.577, 0.648, 0.508, 0.520, 0.625, 0.590, 0.735 mit jeweils $p < .001$

Kognitive Aktivierung

Es ergibt sich das Modell $y = -0.010x + 3.119$. Eine grafische Veranschaulichung bietet die Abbildung 5.1.

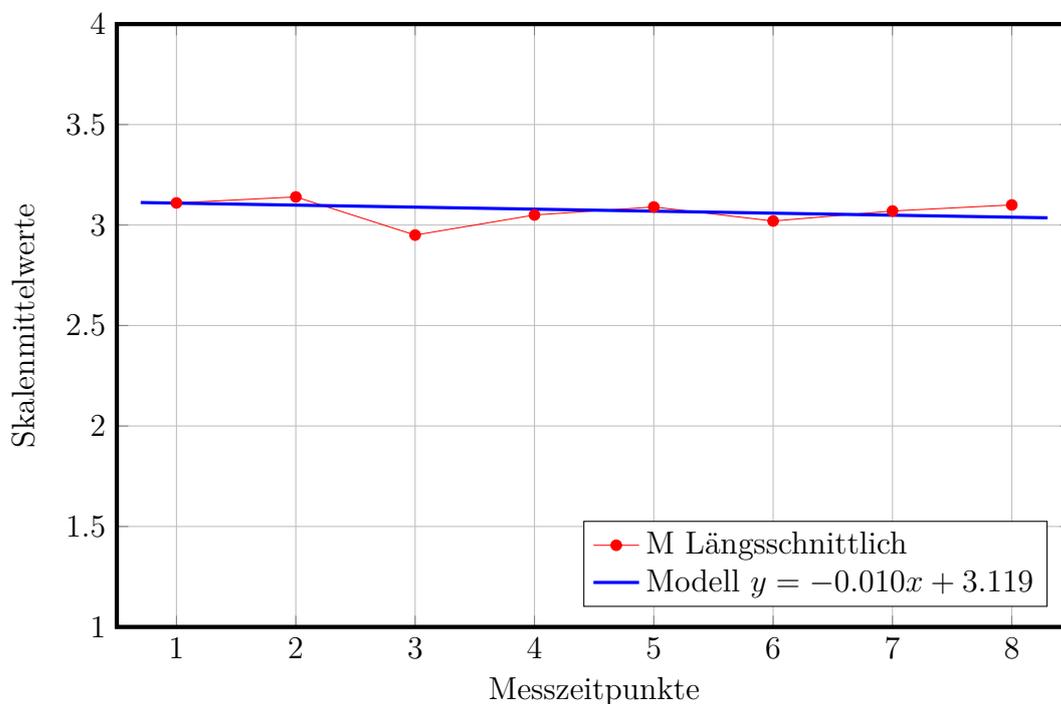


Abbildung 5.1: Lineares Strukturgleichungsmodell der Kognitiven Aktivierung

Die Schülerinnen und Schüler schätzen die Kognitive Aktivierung im Durchschnitt zum ersten Messzeitpunkt mit $M_{Intercept} = 3.119$ ein. Dieser Mittelwert des Intercepts hat signifikante Variabilität, das heißt, es sind interindividuelle Unterschiede im Ausgangsniveau vorhanden. Bei einem Signifikanzniveau von $p < 0.05$ zeigt das Modell keine signifikante lineare Veränderung der Kognitiven Aktivierung über die Zeit der 8 Messungen. Allerdings gibt es signifikante interindividuelle Unterschiede in der linearen Veränderung. Es

liegt deutlich keine signifikante Korrelation zwischen dem Ausgangsniveau (*Intercept*) und der Veränderung (*slope*) vor. Zwischen 49.3% und 73.8% der Varianz der beobachteten Variablen zu den 8 Messzeitpunkten¹ der Kognitiven Aktivierung werden durch die latenten Wachstumsvariablen erklärt.

Instruktionale Unterstützung

Es ergibt sich das Modell $y = -0.029x + 3.217$. Eine grafische Veranschaulichung bietet die Abbildung 5.2.

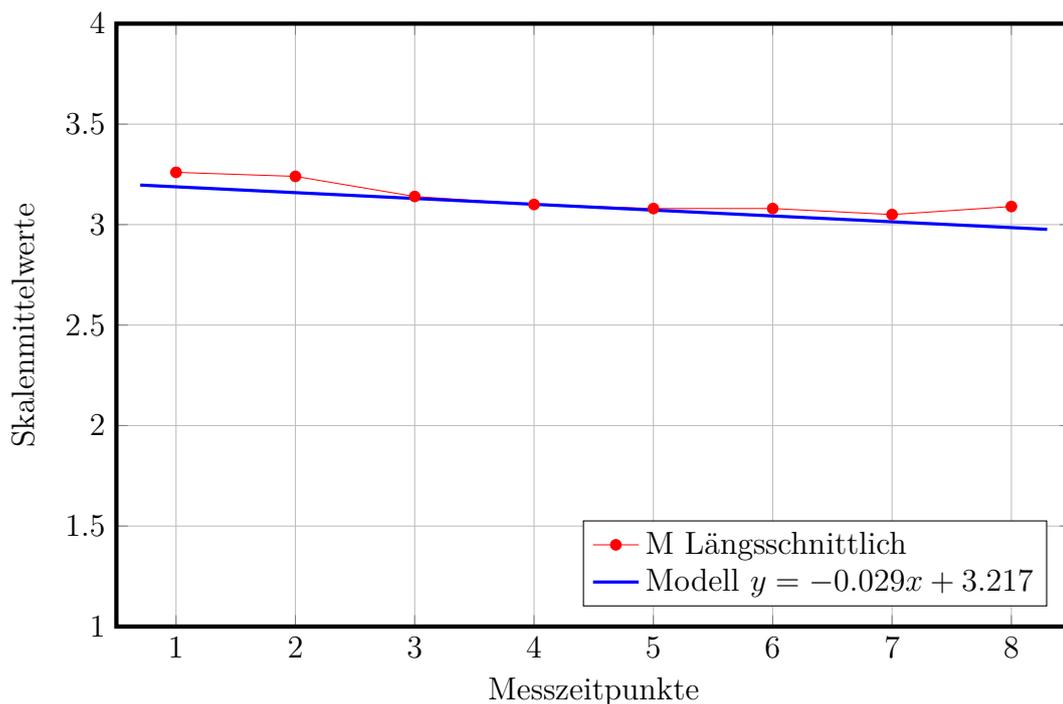


Abbildung 5.2: Lineares Strukturgleichungsmodell $y = -0.029x + 3.217$ der Instruktionalen Unterstützung mit den 8 Messzeitpunkten

Die Schülerinnen und Schüler schätzen die Instruktionale Unterstützung im Durchschnitt zum ersten Messzeitpunkt mit $M_{Intercept} = 3.217$ ein. Dieser Mittelwert des Intercepts hat signifikante Variabilität, das heißt, es sind interindividuelle Unterschiede im Ausgangsniveau vorhanden. Bei einem Signifikanzniveau von $p < 0.05$ zeigt das Modell eine signifikante Abnahme der linearen Veränderung in Höhe von im Mittel $M_{Linear} = -0,029$ der Instruktionalen Unterstützung über die Zeit der 8 Messungen. Diese Veränderung

¹In der Reihenfolge der Messungen: 50,2%, 64,0%, 49,3%, 54,9%, 66,5%, 66,9%, 68,9%, 73,8%

unterliegt signifikanten interindividuellen Unterschieden. Es liegt deutlich keine signifikante Korrelation zwischen dem Ausgangsniveau (*Intercept*) und der Veränderung (*slope*) vor. Zwischen 61.7% und 74.5% der Varianz der beobachteten Variablen zu den 8 Messzeitpunkten² der Instrukionalen Unterstützung werden durch die latenten Wachstumsvariablen erklärt.

Emotionale Unterstützung

Es ergibt sich das Modell $y = -0.036x + 3.454$. Eine grafische Veranschaulichung bietet die Abbildung 5.3.

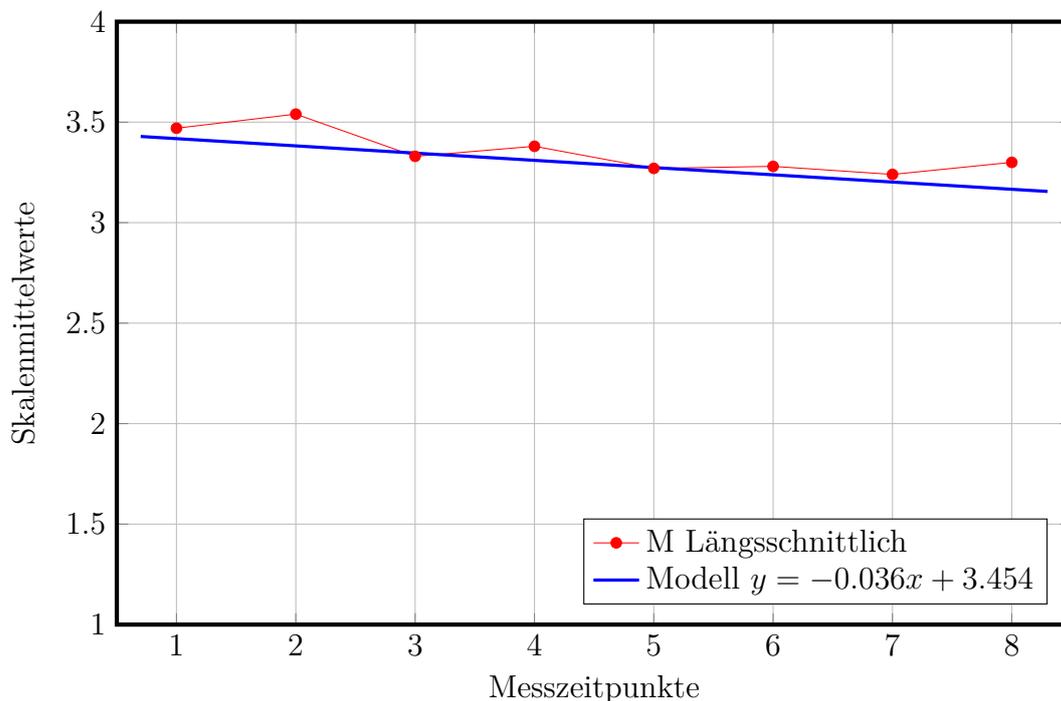


Abbildung 5.3: Lineares Strukturgleichungsmodell $y = -0.036x + 3.454$ der Emotionalen Unterstützung mit den 8 Messzeitpunkten

Die Schülerinnen und Schüler schätzen die Emotionale Unterstützung im Durchschnitt zum ersten Messzeitpunkt mit $M_{Intercept} = 3.454$ ein. Dieser Mittelwert des Intercepts hat signifikante Variabilität, das heißt, es sind interindividuelle Unterschiede im Ausgangsniveau vorhanden. Das Modell zeigt eine signifikante Abnahme der linearen Veränderung in Höhe von im Mittel $M_{Linear} = -0,036$ der Emotionalen Unterstützung über die Zeit

²In der Reihenfolge der Messungen: 61,7%, 62,2%, 69,4%, 68,1%, 71,3%, 70,0%, 74,5%, 68,1%

der 8 Messungen. Diese Veränderung unterliegt signifikanten interindividuellen Unterschieden. Bei einem Signifikanzniveau von $p = 0.042$ (zweiseitiger Test) liegt eine Korrelation von $r = 0.526$ zwischen dem Ausgangsniveau und der linearen Veränderung vor. Das heißt, je höher eine Schülerin oder ein Schüler zu Beginn der Messung die Emotionale Unterstützung durch die Lehrkraft einschätzt, desto größer ist die lineare Abnahme der Bewertung durch die Schülerin oder den Schüler mit der Zeit. Schülerinnen und Schüler, die von vornherein die Emotionale Unterstützung als gering bewerten, verringern ihre Bewertung mit der Zeit auch geringer. Zwischen 50.5% und 74.1% der Varianz der beobachteten Variablen zu den 8 Messzeitpunkten³ der Emotionalen Unterstützung werden durch die latenten Wachstumsvariablen erklärt.

Klassenführung

Es ergibt sich das Modell $y = -0.037x + 2.330$. Eine grafische Veranschaulichung bietet die Abbildung 5.4.

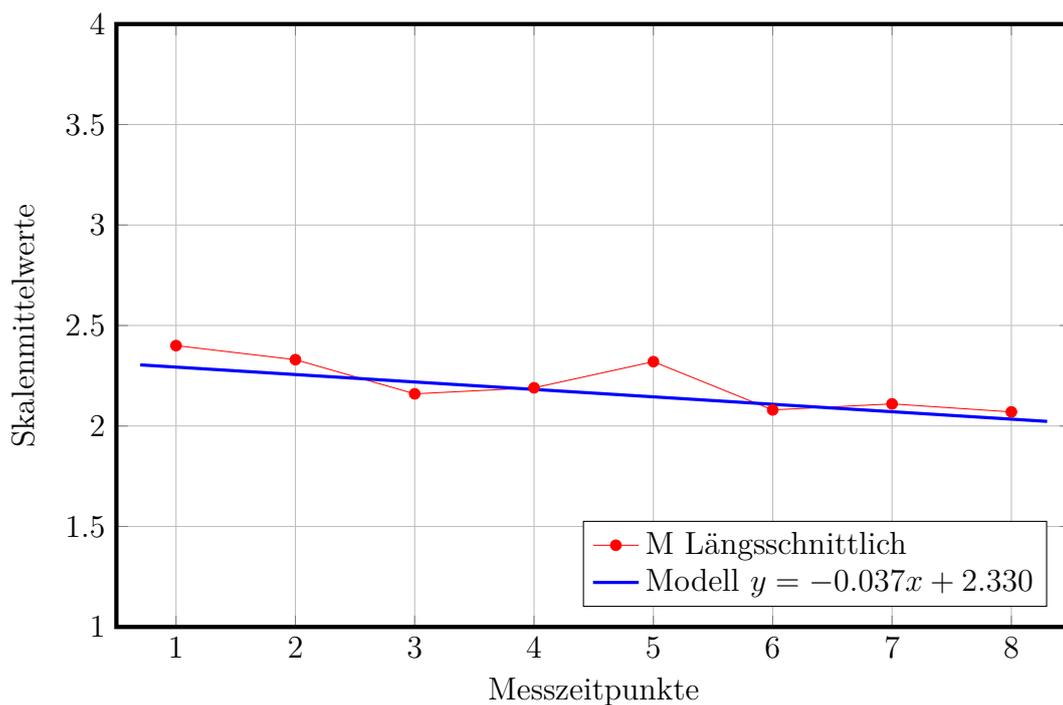


Abbildung 5.4: Lineares Strukturgleichungsmodell $y = -0.037x + 2.330$ der Klassenführung mit den 8 Messzeitpunkten

³In der Reihenfolge der Messungen: 55,9%, 50,5%, 60,9%, 61,3%, 74,7%, 72,1%, 74,1%, 69,7%

Die Schülerinnen und Schüler schätzen die Klassenführung im Durchschnitt zum ersten Messzeitpunkt mit $M_{Intercept} = 2.330$ ein. Dieser Mittelwert des Intercepts hat signifikante Variabilität, das heißt, es sind interindividuelle Unterschiede im Ausgangsniveau vorhanden. Das Modell zeigt eine signifikante Abnahme der linearen Veränderung in Höhe von im Mittel $M_{Linear} = -0,037$ der Klassenführung über die Zeit der 8 Messungen. Diese Veränderung unterliegt signifikanten interindividuellen Unterschieden. Es liegt deutlich keine signifikante Korrelation zwischen dem Ausgangsniveau (*Intercept*) und der Veränderung (*slope*) vor. Zwischen 50.8% und 73.5% der Varianz der beobachteten Variablen zu den 8 Messzeitpunkten⁴ der Klassenführung werden durch die latenten Wachstumsvariablen erklärt.

⁴In der Reihenfolge der Messungen: 54,8%, 57,7%, 64,8%, 50,8%, 52,0%, 62,5%, 59,0%, 73,5%

Tabelle 5.9: Zentrierte Korrelationen der Qualitätsdimension Klassenführung zu den 8 Messzeitpunkten; alle mit ** gekennzeichneten Korrelationen sind signifikant auf dem Signifikanzniveau $p < .01$, die durch * gekennzeichneten Werte sind signifikant auf dem Niveau von $p < .05$ (zweiseitiger Test).

	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8
M 1	1	.23*	.28*	.24*	.22*	.25*	.36**	.27*
N	81	78	78	77	80	79	81	78
M 2		1	.53**	.60**	.38**	.43**	.43**	.45**
N		79	77	76	78	77	79	76
M 3			1	.58**	.65**	.61**	.52**	.52**
N			80	77	79	78	80	78
M 4				1	.53**	.55**	.41**	.58**
N				80	79	78	80	77
M 5					1	.51**	.56**	.56**
N					84	82	84	81
M 6						1	.69**	.66**
N						83	83	80
M 7							1	.55**
N							85	82
M 8								1
N								82

Die in Tabelle 5.6 auf Seite 62 dargestellten signifikanten Korrelationen r erreichen Werte zwischen $.37 < r < .77$ für die Qualitätsdimension Kognitive Aktivierung. Sie sind daher überwiegend als gute bis sehr gute positive Zusammenhänge zu bewerten (Cohen, 1988). Das lässt auf eine hohe interindividuelle Stabilität der Qualitätsdimension Kognitive Aktivierung schließen. Ein Schüler, eine Schülerin, die zu einem Messzeitpunkt einen Wert angekreuzt hat, wird mit einer hohen Wahrscheinlichkeit zu einem anderen Messzeitpunkt einen sehr ähnlichen Wert ankreuzen. Zu sehr ähnlichen Ergebnissen kommen die Berechnungen für die anderen Qualitätsdimensionen: Für die Instruktionale Unterstützung zeigt die Tabelle 5.7 signifikante Korrelationen r zwischen $.31 < r < .82$, überwiegend sogar von über $r > .70$. Für die in Tabelle 5.8 signifikanten Korrelationen r der Emotionalen Unterstützung sind überwiegend Werte von über $.60$ und vielfach Werte von über $.80$ zu sehen. Allerdings besteht zwischen den Messergebnissen des dritten Messzeitpunktes und den Messergebnissen der Messzeitpunkte 1, 2, 4, 6, 7 und 8 keine

signifikante Korrelation für diese Qualitätsdimension. Für die Klassenführung sind in Tabelle 5.9 die für die signifikanten Korrelationen r erreichten Werte zwischen $.36 < r < .69$ dargestellt. Die Messergebnisse von Messung 1 sind überwiegend nicht signifikant auf dem Niveau $p > .01$ zu den anderen Messergebnissen der Qualitätsdimension Klassenführung, sondern nur auf dem Niveau $p < .05$. Es hat für alle vier Qualitätsdimensionen den Anschein, dass der Korrelationskoeffizient in den ersten Messungen geringer ausfällt als zu späteren. Des Weiteren sind die Werte der Korrelationskoeffizienten zwischen frühen und späten Messzeitpunkten im Vergleich zu anderen, dichter bei einander liegenden Messzeitpunkten bei allen Qualitätsdimensionen geringer.

Abhängigkeiten zwischen den Wirkungen der Ausprägungen von Messzeitpunkt und Klassenzugehörigkeit

Im Folgenden werden die Ergebnisse der Interaktionseffekte zwischen den beiden unabhängigen Variablen Messzeitpunkt und Klassenzugehörigkeit für die einzelnen 4 Qualitätsdimensionen dargestellt.

Für alle vier Qualitätsdimensionen, Kognitive Aktivierung, Instruktionale Unterstützung, Emotionale Unterstützung und Klassenführung, ergab der Mauchly-Test auf Sphärizität eine Verletzung der Sphärizitätsannahme, so dass die Anwendung eines Korrekturverfahrens notwendig wurde. Nach den Empfehlungen von Girten (1992) wurde für alle vier Qualitätsdimensionen das Verfahren von Huynh-Feldt verwendet. Danach errechneten sich $\epsilon = 0,848$ und mit dieser Korrektur ein $F(11.87, 385, 86) = 2,24$ mit einem signifikanten ($p < 0,05$) Interaktionseffekt, dem partiellem $\eta^2 = 0,065$ für die Kognitive Aktivierung. Für die Instruktionale Unterstützung führte das Verfahren nach Huynh-Feldt zu $\epsilon = 0,818$. Es gab für die Instruktionale Unterstützung keine signifikante Interaktion zwischen den Messzeitpunkten und den drei verschiedenen Klassen mit der Korrektur nach Huynh-Feldt $F(5, 73; 372, 40) = 0,73$, $p = 0,715$ und partiellem $\eta^2 = 0,022$. Die Berechnungen nach gleichem Verfahren für die Emotionale Unterstützung ergaben $\epsilon = 0,840$ und ebenfalls keine signifikante Interaktion zwischen den Messzeitpunkten und den drei verschiedenen Klassen mit der Korrektur nach Huynh-Feldt $F(11, 76; 382, 31) = 1,40$, $p = 0,164$ und partiellem $\eta^2 = 0,041$. Anders als bei den bisher betrachteten Qualitätsdimensionen kam nach den Empfehlungen von Girten (1992) das Verfahren von Greenhouse-Geisser zur Anwendung. Es berechnete ein $\epsilon = 0,838$. Mit dieser Korrektur ergibt sich ein $F(11.73, 381, 24) = 2,24$ mit einem signifikanten ($p < 0,05$) Interaktionseffekt von $\eta^2 = 0,064$ für die Klassenführung.

Entsprechende Profildiagramme der Mittelwerte der drei Testklassen zu den 8 Messzeiten zur Messung der vier Qualitätsdimensionen zeigen die Abbildungen 5.5, 5.6, 5.7 und 5.8.

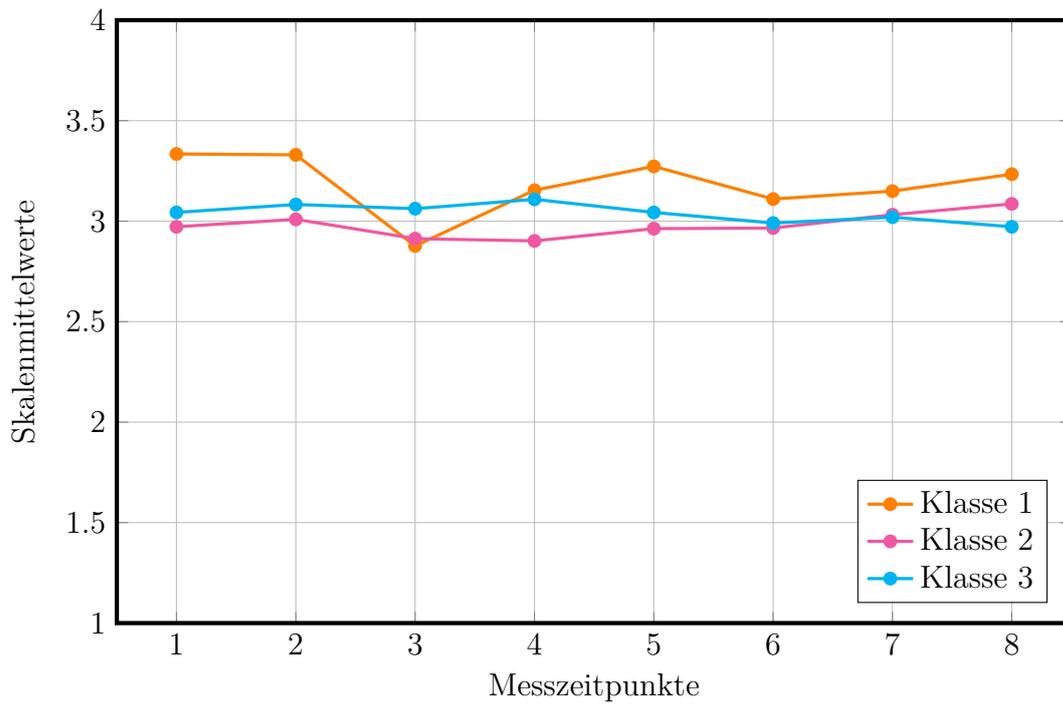


Abbildung 5.5: Mittelwerte der Kognitiven Aktivierung der 3 Klassen

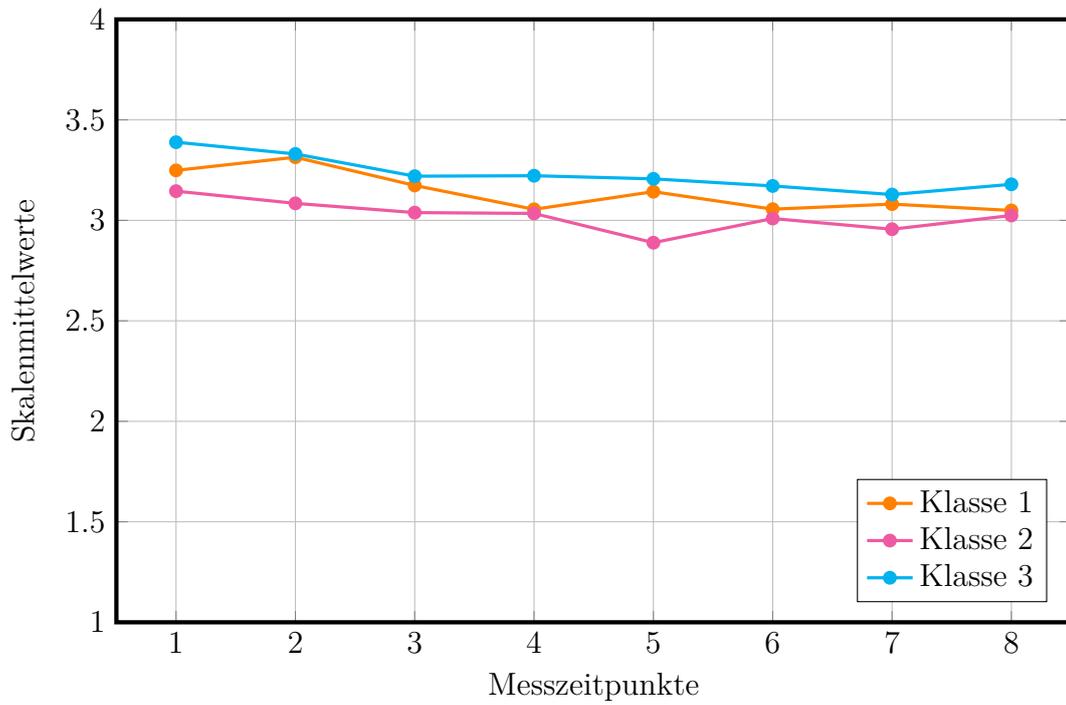


Abbildung 5.6: Mittelwerte der Instrukionalen Unterstützung der 3 Klassen

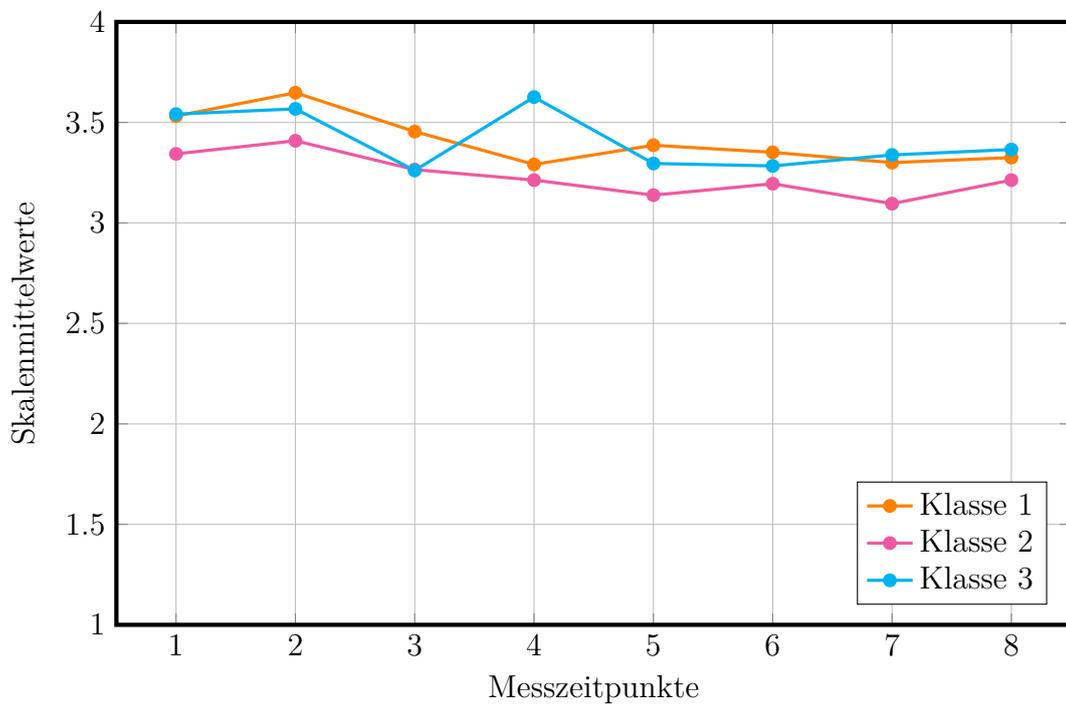


Abbildung 5.7: Mittelwerte der Emotionalen Unterstützung der 3 Klassen

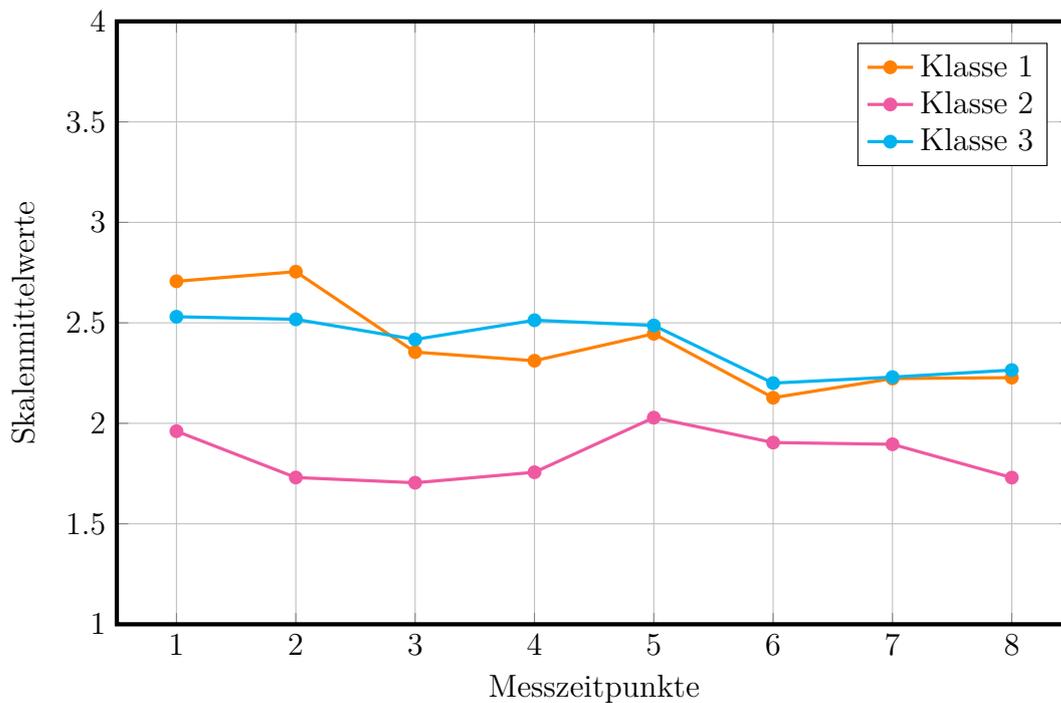


Abbildung 5.8: Mittelwerte der Klassenführung der 3 Klassen

5.1.3 Stabilität zwischen und innerhalb der Unterrichtsthemen

Um die Stabilität der beobachteten Unterrichtswahrnehmungen zwischen und innerhalb der Unterrichtsthemen empirisch zu prüfen, werden die Korrelationen der Messungen zwischen den Themenblöcken mit denen innerhalb von Themenblöcken verglichen, dargestellt in den Tabellen 5.6, 5.7, 5.8 und 5.9. Die erste Messung fand zu einem Zeitpunkt statt, als die Klassen zum Thema „Natürliche Zahlen“ harmonisiert unterrichtet wurden. Die Messungen 2, 3 und 4 wurden während der Unterrichtseinheit „Geometrie: Figuren und Körper“, die Messungen 5 und 6 während der Einheit „Flächen und Rauminhalt“ und die Messungen 7 und 8 während der Einheit „Brüche und Dezimalzahlen“ durchgeführt. Für die einzelnen Qualitätsdimensionen ergaben sich folgende Messergebnisse:

Für die Kognitive Aktivierung zeigt die Tabelle 5.6, Seite 62, dass die höchste Korrelation zwischen zwei Messungen überhaupt bei den Messungen 5 und 6 mit $r = .77$ besteht. Beide Messungen liegen in der Unterrichtseinheit „Flächen und Rauminhalt“. Die niedrigste Korrelation mit $r = .37$ zwischen zwei Messungen besteht zwischen Messung 1 und 7, die nicht in einer Unterrichtseinheit stattfanden. Von diesen beiden Ausnahmen

abgesehen, findet sich keine Auffälligkeit die auf eine Abhängigkeit zwischen den Unterrichtsthemen und der Beurteilung der Kognitiven Aktivierung hinweist. So korreliert beispielsweise Messung 2 in gleichem Maße mit Messung 1 und 3 und zum Beispiel noch höher mit Messung 5, obwohl diese nicht im gleichen Themenbereich erhoben wurden. Das in Abbildung 5.9, auf Seite 78 abgebildete Pfaddiagramm zeigt an, dass das individuelle Messergebnis in Messung 1 einen besseren Prädiktor für das Messergebnis 2 darstellt als das Ergebnis von Messung 2 für Messung 3. Genauso verhält es sich bei den anderen Messungen: Die vorherige Messung stellt in aller Regel den besseren Prädiktor für die folgende Messung dar, unabhängig davon welches Thema unterrichtet wurde.

Tabelle 5.7, Seite 63, führt die Korrelationen zwischen den Messungen zur Instruktionalen Unterstützung auf. Auch darin ist die höchste gemessene Korrelation zwischen zwei Messungen mit $r = .82$ innerhalb einer Unterrichtseinheit, nämlich zwischen Messung 3 und 4 entstanden. Zu sehen ist aber auch, dass Messung 3 ansonsten viel stärker mit den Messungen 5, 6 und 7 korreliert, als mit Messung 2, die ebenfalls in der Unterrichtseinheit liegt. Das Pfaddiagramm, dargestellt in Abbildung 5.10, Seite 80, zeigt auch keine besondere Vorhersagekraft für die Messungen innerhalb einer Unterrichtseinheit im Vergleich zu den anderen Messungen an den Themengrenzen an.

Aus Tabelle 5.8, Seite 64 wird für die Emotionale Unterstützung beispielsweise deutlich, dass Messung 3 themenintern mit Messung 2 und 4 nicht signifikant korreliert, allerdings mit der Messung 5 ($r = .81$), des Folgethemas eine sehr hohe und signifikante Korrelation erreicht. Auch Messung 4 korreliert höher mit Messung 5 ($r = .80$) und 6 ($r = .80$) als mit Messung 2 ($r = .52$). Das Pfaddiagramm, dargestellt in Abbildung 5.11, Seite 82, zeigt, dass Messungen innerhalb eines Themas keine deutlich stärkeren Prädiktoren für Folgemessungen darstellen als Messungen zwischen den Unterrichtseinheiten. Insgesamt kann durch die Messungen keine Themenabhängigkeit nachgewiesen werden.

Nach der gleichen Argumentationskette wie für die Emotionale und Instruktionale Unterstützung und die Kognitive Aktivierung zuvor, finden sich für die Klassenführung keine Auffälligkeiten der Korrelationen, in dem Sinne, dass die Korrelationen innerhalb eines Themas höhere Werte erzielen als zwischen den Themen. Das lässt sich gut an der Messung 3 nachvollziehen, die zwar hohe Korrelationen mit Messung 4 ($r = .58$) hat, aber auch hohe Werte zu Messung 5 ($r = .65$) und 6 ($r = .61$) erhält. Auch die Messungen 7 und 8 erhalten hohe Werte für den Korrelationskoeffizienten, wie die Tabelle 5.9, Seite 65, zeigt. Lediglich die Ergebnisse der ersten Messung erzielen nur mit den Messergebnissen der Messung 7 signifikante Werte für den Korrelationskoeffizienten. Anders als bei der Instruktionalen Unterstützung zeigt das Pfaddiagramm zur Klassenführung, dargestellt in Abbildung 5.12, Seite 84, zwar auch eine geringere Prädiktivität

der Messung 1 zu 2 und 4 zu 5, als beispielsweise zwischen Messung 2 zu 3 und 3 zu 4 an, allerdings ist die Prädiktivität der Messungen 4 zu 5 und 6 zu 7 stärker, als die der Messung 5 zu 6.

5.2 Unterrichtswahrnehmung und Leistung

Der Zusammenhang zwischen Leistungsdaten und Unterrichtsbeurteilungen kann durch verschiedene Modelle modelliert werden. In diesem Abschnitt wurden drei Schritte angewandt, um sich der Frage zu nähern. Zunächst wurden im ersten Schritt in 5.2.1 lineare Strukturgleichungsmodelle mit Kovariaten untersucht. Dieser Abschnitt liefert Hinweise darauf, inwieweit Leistung und Selbstkonzept zu Beginn des Schuljahres einen Effekt auf das Level und die Veränderung der Unterrichtswahrnehmungen im weiteren Schuljahr haben. Im zweiten Schritt, in Abschnitt 5.2.2, werden lineare Strukturgleichungsmodelle und Korrelationen zur Leistung berechnet. Hier werden vor allem Variablen eingeschlossen, die im Laufe des Schuljahres erhoben wurden. Anders als im Abschnitt 5.2.1 soll hier die andere Kausalrichtung erschlossen werden: Welchen Einfluss hat das Level der Veränderung der Unterrichtswahrnehmung auf die dann folgende Lernleistung?

Zur besseren Veranschaulichung der folgenden Regressionsanalyse wurden schließlich Pfaddiagramme in 5.2.3 erstellt.

5.2.1 Lineare Strukturgleichungsmodelle mit Kovariaten

In diesem Abschnitt sollen die Kovariaten Leistung und mathematikbezogenes Selbstkonzept als Prädiktorvariablen, also deren Einfluss auf das lineare Strukturgleichungsmodell für die Einschätzung der vier Qualitätsdimensionen guten Unterrichts durch die Schülerinnen und Schüler, untersucht werden. Dazu wurden die in Abschnitt 5.1.1 vorgestellten Modelle entsprechend der Syntax aus (L. K. Muthén & Muthén, 2017) mit dem Befehl *on* erweitert. Die Auswahl der konkreten Prädiktorvariablen als Kovariate wurde auf die zu Schuljahresbeginn erhobenen Merkmale KERMIT5, H1 und MWSKM1 eingeschränkt. Welchen Einfluss sie im weiteren Schuljahresverlauf auf die Unterrichtswahrnehmung haben, wird im Folgenden für die einzelnen Qualitätsdimensionen aufgeführt.

Kognitive Aktivierung

Die Variablen hatten keinen signifikanten Einfluss auf das Ausgangsniveau (Intercept).

Die Variablen hatten folgenden Einfluss auf die lineare Veränderung (Slope) der Qualitätseinschätzungen der Schülerinnen und Schüler:

KERMIT5: $\hat{\beta}_1 = -0.361$ mit $p = .081$ (bei einem zweiseitigen Test)

MWSKM1: $\hat{\beta}_1 = 0.231$ mit $p = .203$ (bei einem zweiseitigen Test)

Auf einem Signifikanzniveau $p < 0.1$ für einen zweiseitigen Test zeigt sich der Einfluss der durch KERMIT 5 gemessenen Ausgangsleistung in Mathematik mit dem Regressionskoeffizienten $\hat{\beta}_1 = -0.361$. Schülerinnen und Schüler mit einer höheren Mathematikleistung schätzen demnach die Kognitive Aktivierung des Unterrichts mit zunehmender Messung (also zunehmender Unterrichtszeit) geringer ein. Je höher die mathematische Leistungsfähigkeit, desto stärker ist der lineare Abfall der Einschätzung der Kognitiven Aktivierung über die Zeit.

Der Einfluss des mathematikbezogenen Selbstkonzepts als Prädiktor auf die Einschätzung der Kognitiven Aktivierung ist bei den Berechnungen nicht eindeutig, mit $p = 0.203$. Bei einem einseitigen Test kann nicht von klaren Signifikanz der Variablen *MWSKM1* gesprochen werden.

Immerhin stellen die Werte die Möglichkeit in den Raum, dass ein höheres mathematikbezogenes Selbstkonzept durch mehr Zutrauen in die eigenen Fähigkeiten eher zu einer sich über die Zeit steigernden Bewertung der Kognitiven Aktivierung des Unterrichts führt.

Instruktionale Unterstützung

Die Variablen hatten folgenden Einfluss auf das Ausgangsniveau (Intercept):

MWSKM1: $\hat{\beta}_1 = 0.422$ mit $p = .001$ (bei einem zweiseitigen Test)

Die Variablen hatten folgenden Einfluss auf die lineare Veränderung (Slope) der Qualitätseinschätzungen der Schülerinnen und Schüler:

KERMIT5: $\hat{\beta}_1 = -0.333$ mit $p = .119$ (bei einem zweiseitigen Test)

H1: $\hat{\beta}_1 = 0.521$ mit $p = .015$ (bei einem zweiseitigen Test)

MWSKM1: $\hat{\beta}_1 = -0.254$ mit $p = .164$ (bei einem zweiseitigen Test)

Die Berechnungen zeigen einen klaren Einfluss des mathematikbezogenen Selbstkonzeptes auf die Ausgangseinschätzung der Instrukionalen Unterstützung. Mit einem Signifikanzniveau von $p < 0.005$ kann man vorhersagen, dass jemand mit einem höheren mathematikbezogenem Selbstkonzept die Instruktionale Unterstützung bei der ersten Messung höher einschätzt, der Regressionskoeffizient beträgt $\hat{\beta}_1 = 0.422$. Bei zweifelhafter Signifikanz ($p = .164$, bei einem zweiseitigen Test) kann man durch die Berechnungen sagen, dass diese Einschätzung dieser Schülerinnen und Schüler im Verlauf der Zeit jedoch stärker sinken wird als bei anderen (mit dem Regressionskoeffizienten $\hat{\beta}_1 = -0.254$).

Ein wesentlicher Einflussfaktor für die Beurteilung der Instrukionalen Unterstützung stellt den Berechnungen zufolge die erste Halbjahresnote dar. Bei einer Signifikanz von $p = .015$ lässt sich sagen, dass die Instruktionale Unterstützung von den Schülerinnen und Schülern zukünftig umso besser eingeschätzt wird ($\hat{\beta}_1 = 0.521$), je höher die Ausgangsnote ist.

Einen auf Grund der zweifelhaften Signifikanz von $p = .119$ weniger klaren Einfluss auf die zukünftige Einschätzung der Instrukionalen Unterstützung hat die durch KERMIT 5 ermittelte anfängliche Leistungsfähigkeit in Mathematik. Hiernach würde jemand mit einer höheren Leistungsfähigkeit tendenziell die Qualitätsdimension Instruktionale Unterstützung mit dem Zeitverlauf stärker abfallend bewerten.

Emotionale Unterstützung

Die Variablen hatten folgenden Einfluss auf das Ausgangsniveau (Intercept):

KERMIT5: $\hat{\beta}_1 = -0.249$ mit $p = .152$ (bei einem zweiseitigen Test)

MWSKM1: $\hat{\beta}_1 = 0.277$ mit $p = .061$ (bei einem zweiseitigen Test)

Die Variablen hatten folgenden Einfluss auf die lineare Veränderung (Slope) der Qualitätseinschätzungen der Schülerinnen und Schüler:

KERMIT5: $\hat{\beta}_1 = -0.457$ mit $p = .036$ (bei einem zweiseitigen Test)

H1: $\hat{\beta}_1 = 0.615$ mit $p = .006$ (bei einem zweiseitigen Test)

Die eingangs als höher ermittelte Leistungsfähigkeit führt mit zweifelhafter Signifikanz von $p = .152$ zu einer niedrigeren Einschätzung der Emotionalen Unterstützung bei der ersten Messung. Prädiktiv lässt sich signifikant ($p = 0.036$) einschätzen, dass Schülerinnen und Schüler mit höheren Ausgangsleistungen einen stärkeren Abfall ($\hat{\beta}_1 = -0.457$) der Unterrichtsbewertung in der Qualitätsdimension Emotionale Unterstützung haben werden als andere.

Es zeigt sich, dass das mathematikbezogene Selbstkonzept der Schülerinnen und Schüler einen signifikanten ($p = 0.061$) Einfluss auf die Bewertung der Emotionalen Unterstützung hat. Schülerinnen und Schüler mit höherem Selbstkonzept bewerten diese Qualitätsdimension in der ersten Messung höher. Für den weiteren Verlauf zeigt sich kein signifikanter Einfluss.

Anders verhält es sich bei der Note des ersten Halbjahres. Wer eine bessere Note erhält wird zukünftig die Emotionale Unterstützung deutlich höher einschätzen ($p = .006$ und $\hat{\beta}_1 = 0.615$).

Klassenführung

Die Variablen hatten folgenden Einfluss auf das Ausgangsniveau (Intercept):

KERMIT5: $\hat{\beta}_1 = -0.231$ mit $p = .172$ (bei einem zweiseitigen Test)

MWSKM1: $\hat{\beta}_1 = 0.390$ mit $p = .005$ (bei einem zweiseitigen Test)

Die Variablen hatten folgenden Einfluss auf die lineare Veränderung (Slope) der Qualitätseinschätzungen der Schülerinnen und Schüler:

H1: $\hat{\beta}_1 = 0.355$ mit $p = 0.118$ (bei einem zweiseitigen Test)

MWSKM1: $\hat{\beta}_1 = -0.425$ mit $p = 0.018$ (bei einem zweiseitigen Test)

Hier ist der Einfluss des Selbstkonzeptes erwähnenswert. Schülerinnen und Schüler mit einem höheren Selbstkonzept bewerten die Qualität der Klassenführung in der ersten Messung höher ($\hat{\beta}_1 = 0.390$ bei $p = .005$) als andere Schülerinnen und Schüler einerseits und andererseits unterliegt deren Einschätzung einem stärkeren ($\hat{\beta}_1 = -0.425$ und $p = .018$) linearen Abfall als die Einschätzung von Schülerinnen und Schülern mit weniger starkem Selbstkonzept im Laufe der 8 Messungen.

Der Einfluss der Leistung auf das Ausgangsniveau der Bewertung der Qualität der Klassenführung unterliegt zweifelhafter Signifikanz $p = .172$, deutet aber darauf hin, dass leistungsstärkere Schülerinnen und Schüler diese niedriger bewerten als andere.

Ebenso unsicher ($p = .118$) ist der Einfluss der ersten Halbjahresnote auf den linearen Verlauf der Einschätzungen. Es deutet sich aber an, dass bei besserer Halbjahresnote der Verlauf weniger stark ($\hat{\beta}_1 = 0.355$) linear abfällt.

5.2.2 Lineare Strukturgleichungsmodelle und Korrelation zur Leistung

Die in Abschnitt 5.1.1 entwickelten linearen Modelle wurden dahingehend durch die in L. K. Muthén und Muthén (2017) beschriebene Syntax mit *with* erweitert, sodass untersucht werden konnte ob die Modellausgangsniveaus (*Intercepts*) und die linearen Veränderungen (*slope*) zu den Leistungsmerkmalen der Schülerinnen und Schüler korrelieren. Es wurden nur diejenigen Modellabwandlungen berechnet, auf die die Unterrichtswahrnehmung im Laufe des Schuljahres einen Einfluss gehabt haben könnte. Das sind folgende Leistungsmerkmale: die Klassenarbeiten (K2, K3 und K4) und H2. Die Modellergebnisse werden im Einzelnen für die Qualitätsdimensionen aufgeführt, wenn sie einen signifikanten Einfluss haben sollten:

Kognitive Aktivierung

Korrelationen des Ausgangsniveaus (Intercept) mit

K4: $r = -0.364$ mit $p = .090$ (bei einem zweiseitigen Test) - eingeschränkt signifikant

Korrelation der linearen Veränderung (Slope) mit

K3: $r = -0.334$ mit $p = .030$ (bei einem zweiseitigen Test)

H2: $r = 0.477$ mit $p = .021$ (bei einem zweiseitigen Test)

Instruktionale Unterstützung

Es liegt keine signifikante Korrelation zwischen den Leistungsmerkmalen und dem Ausgangsniveau sowie mit der linearen Veränderung für die Leistungsmerkmale K2, K3, K4 und H2 vor.

Emotionale Unterstützung

Es liegt keine signifikante Korrelation zwischen den Leistungsmerkmalen und dem Ausgangsniveau vor.

Korrelation der linearen Veränderung (Slope) mit

K4: $r = -0.089$ mit $p = .085$ (bei einem zweiseitigen Test) - eingeschränkt signifikant

H2: $r = 0.127$ mit $p = .015$ (bei einem zweiseitigen Test)

Klassenführung

Korrelationen des Ausgangsniveaus (Intercept) mit

K2: $r = 0.183$ mit $p = .074$ (bei einem zweiseitigen Test) - eingeschränkt signifikant

K3: $r = -0.264$ mit $p = .084$ (bei einem zweiseitigen Test) - eingeschränkt signifikant

K4: $r = -0.508$ mit $p = .043$ (bei einem zweiseitigen Test)

Es liegt keine signifikante Korrelation zwischen den Leistungsmerkmalen und den linearen Veränderungen vor.

5.2.3 Regressionsanalysen zur Leistung und Unterrichtsqualität

In diesem Abschnitt sind vereinfachte graphische Darstellungen berechneter Strukturgleichungsmodelle als Pfaddiagramme dargestellt. Autoregressionen höherer Ordnung sind dabei zugelassen. Es werden die Stabilität und der Einfluss verschiedener Prädiktoren der Leistung deutlich. Die behandelten Unterrichtsinhalte werden in den Grafiken mit dargestellt. Gerichtete Zusammenhänge werden durch gerichtete Pfeile indiziert. Manifeste Variablen werden durch Rechtecke umgrenzt. Residualvariablen werden aus Gründen der Übersichtlichkeit weggelassen. Die Darstellungen folgen den Konventionen, wie sie unter anderem in Bollen (1989), Kaplan (2009) und Reinecke (2005) verwendet werden.

Die hier abgebildeten Modelle stellen eine Auswahl dar. Nicht aufgeführte Modelle, mit weiteren Pfaden, führten in den Rechnungen zu deutlichen Verschlechterungen der Modellfits. Die Modelle waren ausnahmslos zu verwerfen, die hier abgebildeten Modelle hatten die einzig annehmbaren Passungen.

Kognitive Aktivierung

Das Pfaddiagramm in Abbildung 5.9 zeigt die Visualisierung eines Regressionsmodells mit Autoregressionen höherer Ordnung. Die Messungen und Leistungen sind dabei zentriert.

Modellfits

Das berechnete Modell genügt folgender Passung:

$$\chi^2(81) = 119.71 \text{ mit } p = 0.01, CFI = .96, TLI = .95, RMSEA = .07, SRMR = .08$$

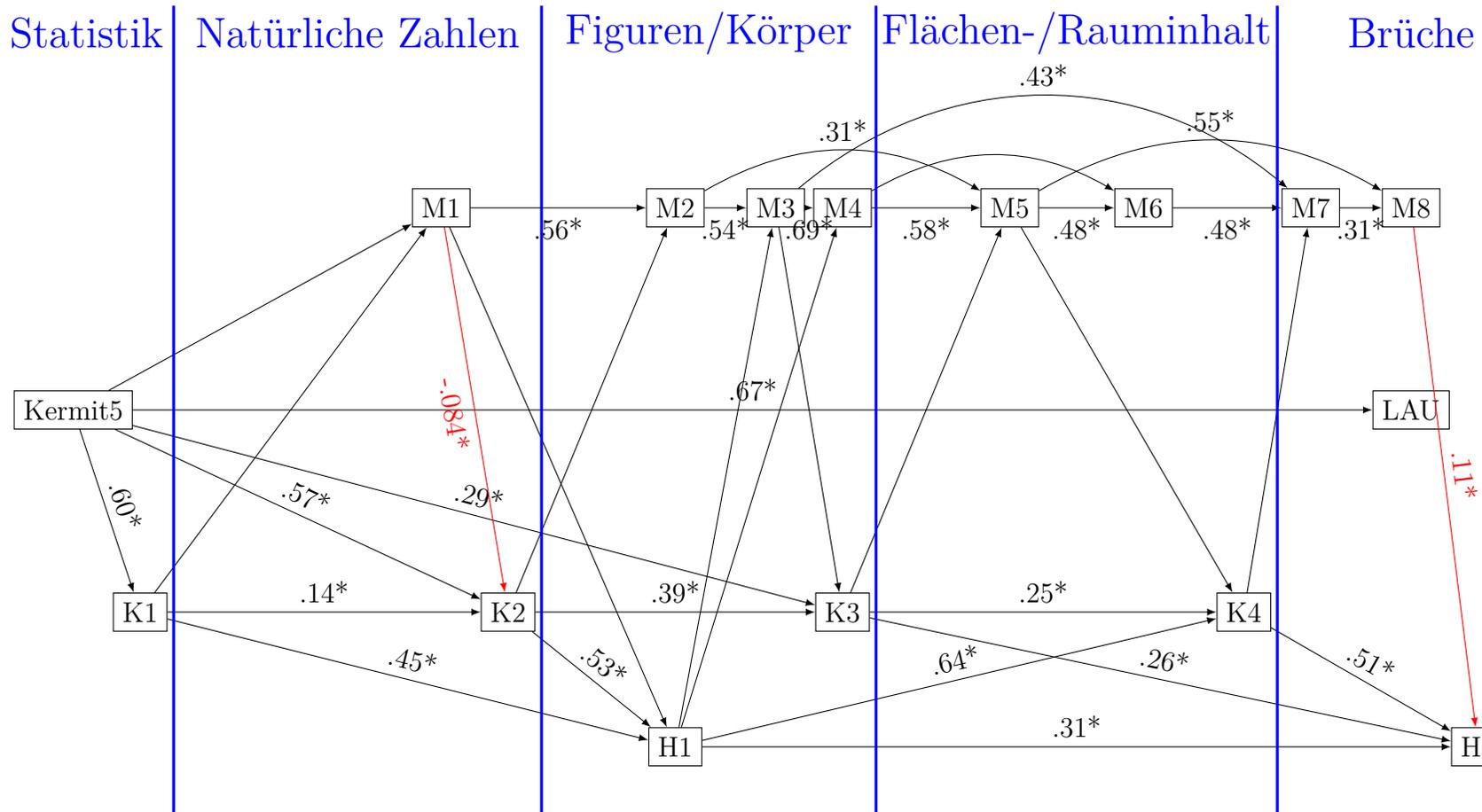


Abbildung 5.9: Pfaddiagramm zur Kognitiven Aktivierung mit Autoregressionen höherer Ordnung; Messungen und Leistungen sind zentriert; signifikante ($p < .05$) Regressionen sind mit * gekennzeichnet; nicht beschriftete Pfade sind nicht signifikant.

Gemäß den in Abschnitt 4.3.5 beschriebenen Cutt-Off Grenzen gilt das hier berechnete Modell als gut angepasst. Der *CFI* erreicht einen Wert $> .95$. *RMSEA* und *SRMR* sind trotz kleiner Stichprobengröße nahe an der idealen Grenze von $< .05$ dran und überschreiten diese nur leicht.

Auffallend ist, dass keine Messung der Kognitiven Aktivierung durch einen Leistungsparameter beeinflusst wird. Weder KERMIT5 noch eine Unterrichtsnote zeigen einen signifikanten Effekt auf eine nachfolgende Unterrichtsbeobachtung. Andersherum gilt das gleiche: Die eingeschätzte Unterrichtsqualität hat auch keinen Effekt auf die nachfolgende Leistungsmessung, beziehungsweise die Leistungsbeurteilung. Lediglich die Messungen 1 und 8 scheinen die zweite Klassenarbeit und die zweite Halbjahresnote signifikant zu erklären, allerdings nimmt der Regressionskoeffizient β jeweils nur geringe Wert an.

Als stabil über die Zeit und unabhängig vom Wechsel unterrichtlicher Inhalte ist die Dimension der Kognitiven Aktivierung einzuschätzen. Vorherige Messungen erklären spätere. Der Regressionskoeffizient β nimmt signifikante Werte zwischen $.31$ und $.69$ an. Das Qualitätsmerkmal der Kognitiven Aktivierung weist damit nicht nur über die Zeit, sondern auch über die Unterrichtsinhalte hinweg eine hohe interindividuelle Stabilität auf.

Das trifft auf die Leistungsmessung so nicht zu. Beispielsweise ist ein Ergebnis in der ersten Klassenarbeit ist nicht gut geeignet, ein Ergebnis in der zweiten Klassenarbeit vorherzusagen. Die Prognosekraft ist zum Teil nur von schwacher Signifikanz $p > .05$ und damit sehr unsicher.

Instruktionale Unterstützung

Das Pfaddiagramm in Abbildung 5.10 zeigt die Visualisierung eines Regressionsmodells mit Autoregressionen höherer Ordnung. Die Messungen und Leistungen sind dabei zentriert.

Modellfits

Das berechnete Modell erreicht folgende Passung:

$$\chi^2(82) = 158.24 \text{ mit } p < 0.0001, CFI = .93, TLI = .90, RMSEA = .11, SRMR = .08$$

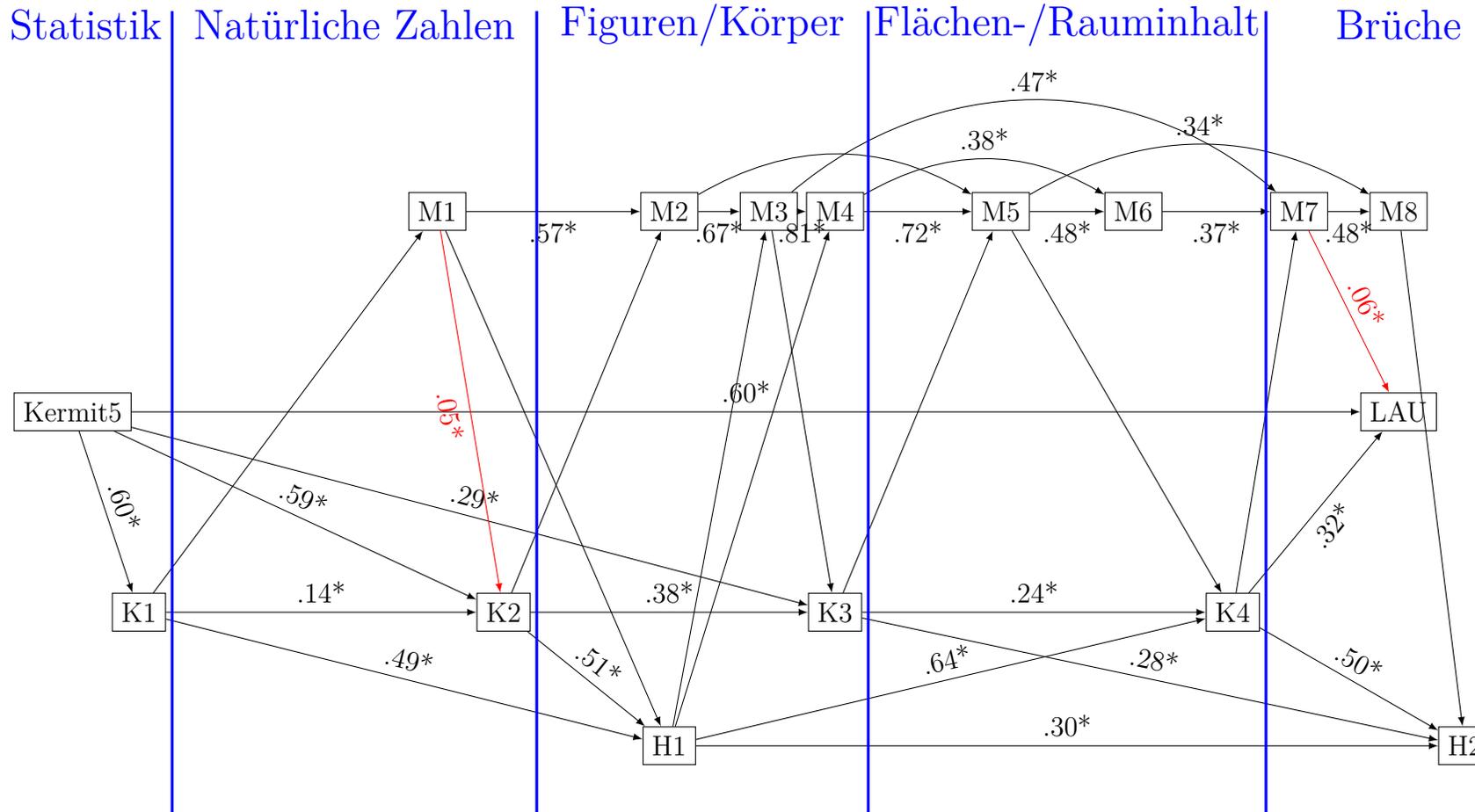


Abbildung 5.10: Pfaddiagramm zur Instrukionalen Unterstützung mit Autoregressionen höherer Ordnung; Messungen und Leistungen sind zentriert; signifikante ($p < .05$) Regressionen sind mit * gekennzeichnet; nicht beschriftete Pfade sind nicht signifikant.

Wie im vorherigen Abschnitt auch, ist, der Argumentation aus Abschnitt 4.3.5 folgend, die Modellpassung als geeignet zu bewerten.

Auch hier ist kein signifikanter Einfluss der Leistungsmerkmale auf die Qualitätseinschätzung der Schülerinnen und Schüler und umgekehrt zu erkennen. Die zweite Klassenarbeit hat mit einem nur wenig signifikanten Regressionskoeffizienten $\beta = .18$ einen Einfluss auf die zweite Messung der Unterrichtsqualität. Wer hier ein besseres Ergebnis hatte, schätzt diese Unterstützungsdimension in der zweiten Messung etwas höher ein. Andere Einflüsse konnten nicht nachgewiesen werden.

Die im Pfaddiagramm in Abbildung 5.10 dargestellten Werte der Regressionskoeffizienten zeigen wieder ein hohes Maß an interindividueller Stabilität des Qualitätsmerkmals auf. Der Regressionskoeffizient β nimmt Werte zwischen .37 und .81 themenunabhängig an. Die interindividuelle Stabilität wird durch die Regressionskoeffizienten der Autoregressionen höherer Ordnung gestützt. Wie das in Abbildung 5.9 abgebildete Modell auch, zeigen die Berechnungen dieses Modells, dass der Einfluss einer Klassenarbeitsnote auf die folgende, wenn überhaupt signifikant, nur gering ist, $\beta < .40$.

Emotionale Unterstützung

Das Pfaddiagramm in Abbildung 5.11 zeigt die Visualisierung eines Regressionsmodells mit Autoregressionen höherer Ordnung. Die Messungen und Leistungen sind dabei zentriert.

Modellfits

Das berechnete Modell unterliegt der folgenden Passung:

$$\chi^2(82) = 127.98 \text{ mit } p < 0.0001, CFI = .95, TLI = .93, RMSEA = .08, SRMR = .08$$

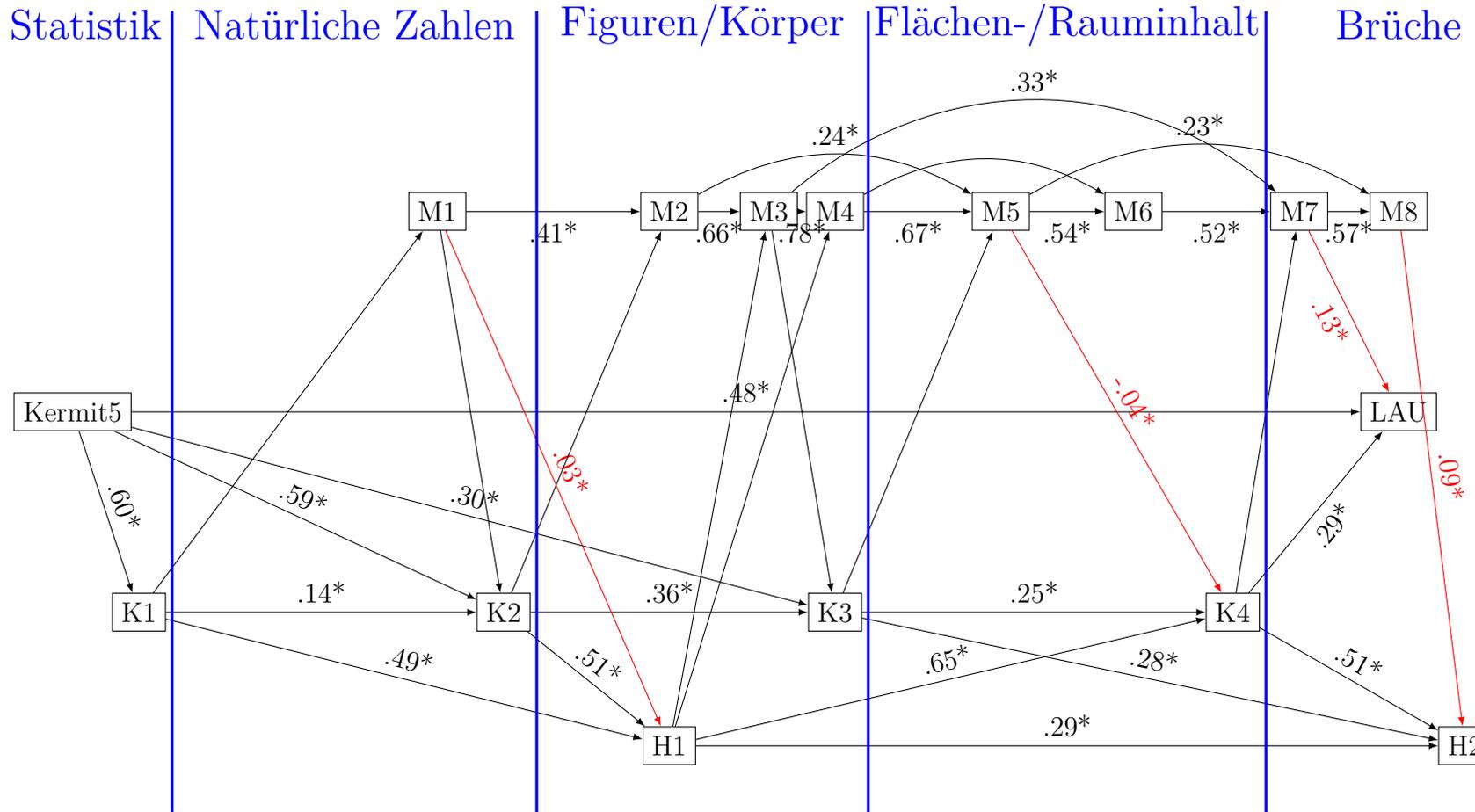


Abbildung 5.11: Pfaddiagramm zur Emotionalen Unterstützung mit Autoregressionen höherer Ordnung; Messungen und Leistungen sind zentriert; signifikante ($p < .05$) Regressionen sind mit * gekennzeichnet; nicht beschriftete Pfade sind nicht signifikant.

Auch dieses Modell ist in seiner Passung, wie in Abschnitt 4.3.5 begründet, als gut angepasst einzustufen.

Auch das in Abbildung 5.11 dargestellte Modell zeigt keine entscheidenden anderen Ergebnisse als die beiden, zuvor diskutierten Modelle für die Kognitive Aktivierung und die Instruktionale Unterstützung. Wieder zeigt sich eine hohe interindividuelle Stabilität der Qualitätsdimension, die unabhängig von beurteilter Leistung und unterrichteten Themen besteht.

Klassenführung

Das Pfaddiagramm in Abbildung 5.12 zeigt die Visualisierung eines Regressionsmodells mit Autoregressionen höherer Ordnung. Die Messungen und Leistungen sind dabei zentriert.

Modellfits

Das berechnete Modell hat folgende Passung:

$$\chi^2(81) = 180.24 \text{ mit } p < .0001, CFI = .88, TLI = .82, RMSEA = .12, SRMR = .10$$

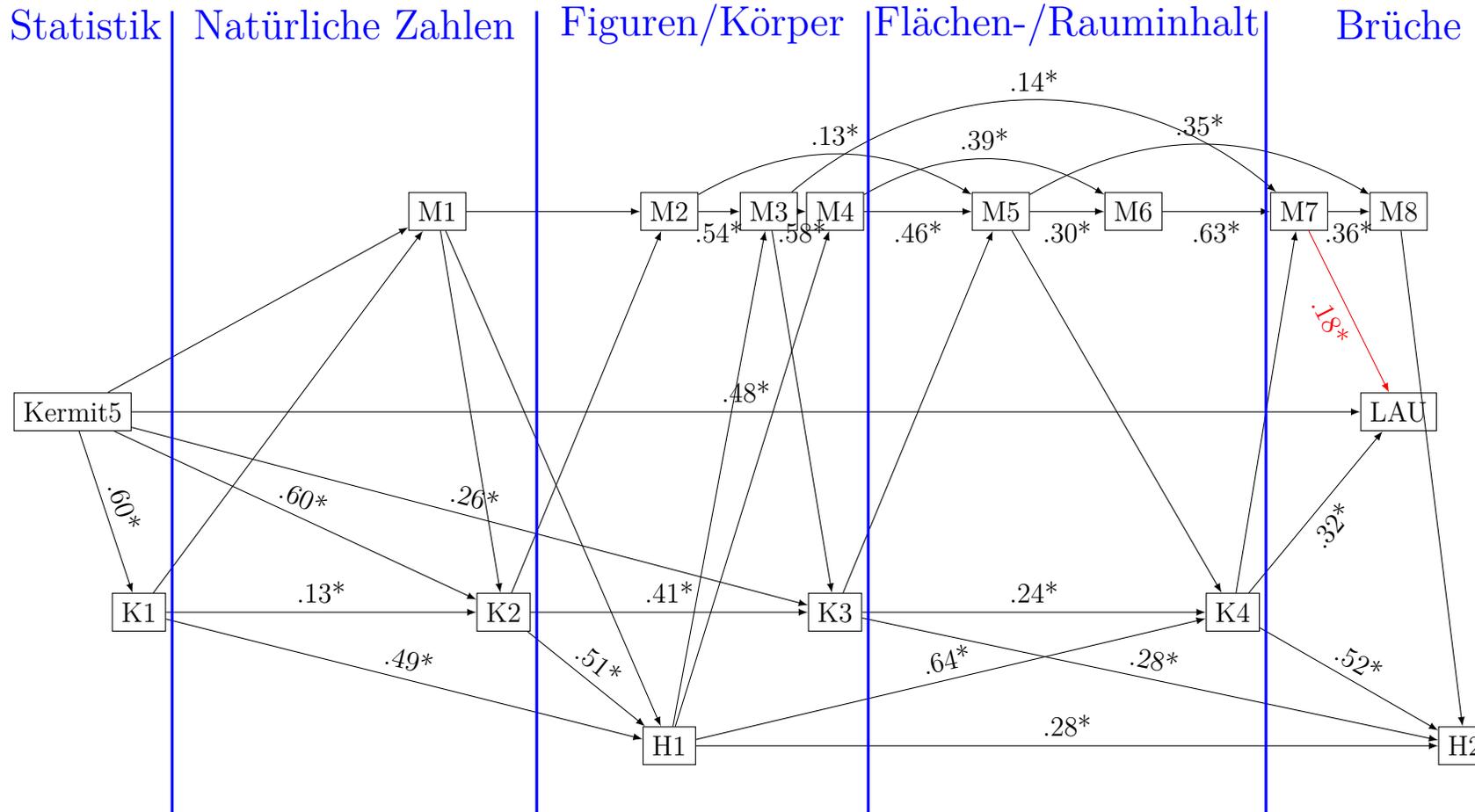


Abbildung 5.12: Pfaddiagramm zur Klassenführung mit Autoregressionen höherer Ordnung; Messungen und Leistungen sind zentriert; signifikante ($p < .05$) Regressionen sind mit * gekennzeichnet; nicht beschriftete Pfade sind nicht signifikant.

Das berechnete Regressionsmodell passt, der Argumentation aus Abschnitt 4.3.5 folgend, gut zu den Daten.

Auch die Regressionsanalyse der Klassenführung zeigt keine anderen Ergebnisse als die der anderen drei Qualitätsdimensionen zuvor.

Alle vier Qualitätsdimensionen zeigen eine hohe interindividuelle Stabilität und sind weitgehend unabhängig von Unterrichtsinhalten und Benotungen.

6 Diskussion der Ergebnisse

Die in Kapitel 5 vorgestellten Ergebnisse der Datenanalysen sollen im Folgenden interpretiert werden, um die Fragen aus Abschnitt 3 zu beantworten. Dabei sollen bei der Beurteilung der Validität der Aussagen die in Abschnitt 4.3.5 beschriebenen Kriterien zur Modellgüte Grundlage sein. Abschließend wird im Abschnitt 6.4 eine Bilanz gezogen, inwieweit die gestellten Forschungsfragen beantwortet wurden. Zum Abschluss wird auf die Stärken und die Schwächen der vorliegenden Untersuchung eingegangen.

6.1 Diskussion der Stabilität der Unterrichtswahrnehmung

Hier werden die Ergebnisse aus Abschnitt 5 in Hinblick auf die Stabilität der Unterrichtswahrnehmung der Schülerinnen und Schüler ausgewertet. Maßgeblich für die Diskussionsabfolge sollen die in Abschnitt 3 formulierten Fragestellungen sein.

6.1.1 Stabilität des Qualitätsniveaus

In Abschnitt 5.1.1 wurde für jede Qualitätsdimension ein lineares Strukturgleichungsmodell (LGM) berechnet. Auf diese Weise werden Trends der Veränderung auf der Individualebene deutlich: Es lässt sich anhand der vorliegenden Daten, vgl. Abschnitt 5.1.1, ab Seite 53ff. sagen, dass die Qualitätsdimension der Kognitiven Aktivierung stabil in ihrem Ausgangsniveau über die Zeit der 8 Messungen von den Schülerinnen und Schülern eingeschätzt wird. Es erfolgt weder ein Anstieg noch ein Abfallen des Qualitätsniveaus über die Zeit, unabhängig davon, wie die Schülerinnen und Schüler das Ausgangsniveau bei der ersten Messung einschätzten. Die Berechnungen zeigen, dass die Einschätzung des Ausgangsniveaus der Instrukionalen Unterstützung im Laufe der 8 Messungen linear abnahm. Dabei spielt es keine Rolle, in welcher Höhe das Ausgangsniveau durch einen Schüler, eine Schülerin eingeschätzt wurde, es erfolgte im Mittel die gleiche lineare Abnahme des Ausgangsniveaus über die Zeit für diese Qualitätsdimension. Für die Qualitätsdimension Emotionale Unterstützung zeigte das lineare Strukturgleichungsmodell

auf, dass die Ausgangseinschätzungen über die 8 Messzeiten im Mittel linear abnahmen. Dabei hing die Stärke der linearen Abnahme davon ab, wie hoch der jeweilige Schüler, die jeweilige Schülerin, das Qualitätsniveau in der ersten Messung einschätzte. Je höher die Bewertung erfolgte, desto größer ist die individuelle, lineare Abnahme der Einschätzung im Laufe der Zeit. Das erklärt warum die Varianz der beobachteten Variablen mit steigender Wahrscheinlichkeit (von zu Beginn 55.9% auf später, bei Messung 7 auf 74,1%) durch die latenten Wachstumsvariablen unterstützt werden. Des Weiteren zeigen die Berechnungen, dass die Einschätzungen des Ausgangsniveaus der Klassenführung im Laufe der 8 Messungen linear abnahmen. Dabei spielt es keine Rolle, in welcher Höhe das Ausgangsniveau durch einen Schüler, eine Schülerin eingeschätzt wurde, es erfolgte im Mittel die gleiche lineare Abnahme des Ausgangsniveaus über die Zeit.

Insgesamt lassen sich zu allen Qualitätsdimensionen lineare Trends mit Hilfe der berechneten LGM nachweisen. Die Qualitätsdimension der Kognitiven Aktivierung verzeichnet im Gegensatz zu den anderen Qualitätsdimensionen im Mittel keine lineare Abnahme des Ausgangsniveaus. Es ist auch keine Zunahme der Ausgangseinschätzung nachweisbar. Das Ausgangsniveau der Einschätzung bleibt im Mittel über die Zeit erhalten, es ist stabil. Im Gegensatz zu der in Abschnitt 3 formulierten These liegt hier die im Vergleich höchste Stabilität der Qualität vor.

Bei allen anderen Qualitätsdimensionen nehmen im Mittel die Ausgangseinschätzungen linear ab. Die Unterrichtsqualität sinkt demnach im Mittel über die Zeit leicht. Nur in der Qualitätsdimension der Emotionalen Unterstützung hängt die Stärke der Abnahme individuell von der Ausgangseinschätzung ab. Hier lässt sich sagen, je höher das Ausgangseinschätzungsniveau, desto größer die lineare Abnahme mit der Zeit. Es erfolgt über die Zeit ein Angleichen der Einschätzungen. Bei allen anderen Qualitätsdimensionen gibt es diese Abhängigkeit von der Ausgangseinschätzung nicht.

Die gefundenen Untersuchungsergebnisse ordnen sich dabei in die bereits zitierte Literatur ein und stellen zumindest keinen Widerspruch zu ihr dar. So fassen beispielsweise Wagner et al. (2015) und Praetorius et al. (2014) den vorliegenden Wissenstand damit zusammen, dass die überwiegende Mehrheit von Studien davon ausgeht, dass die Lehrqualität von Lehrerinnen und Lehrern stabil ist. Das leichte, aber kontinuierliche Absinken der Lehrqualität der zugrundeliegenden Beobachtungen dieser Studie könnte an der Neuzusammenstellung der Testklassen liegen. Alle Schülerinnen und Schüler waren neu an der Testschule und begannen ihr fünftes Schuljahr euphorisch und wurden behutsam an das neue Leistungsniveau herangeführt. Es könnte also sein, dass die Schülerinnen und Schüler erst nach der Eingewöhnungsphase im Laufe des Schuljahres authentischen Unterricht auf einem schultypischen Niveau erlebt haben. Eine folge

Untersuchung in späteren Schuljahren könnte hier mehr Klarheit bringen.

6.1.2 Stabilität der Rangfolge

Insgesamt haben die Berechnungen ein hohes Maß an Stabilität der Einschätzungen der vier Qualitätsdimensionen gezeigt. Die in den Abbildungen 5.9, 5.10, 5.11 und 5.12 aus Abschnitt 5.2.3, ab Seite 78 dargestellten Pfaddiagramme zeigen Autoregressionen höherer Ordnung. Diese Autoregressionen höherer Ordnung erreichen überwiegend signifikante Werte des Regressionskoeffizienten von $r > .50$. Auch die in Abschnitt 3 erwartete geringe Stabilität der Einschätzungen der Qualitätsdimension Kognitiven Aktivierung hat sich nicht bestätigt. Allerdings zeigen die anderen drei Qualitätsdimensionen, trotz nicht signifikanter Korrelationen zwischen vereinzelt Messergebnissen, eine höhere Einschätzungsstabilität an, vor allem die Qualitätsdimension der Instrukionalen Unterstützung. Damit bestätigen die Ergebnisse letztlich den Kern der Aussagen von Praetorius et al. (2014), wonach es mehr Beobachtungszeit für die Einschätzung der Kognitiven Aktivierung bedarf als zum Beispiel für die Qualitätsdimension Klassenführung. Darin könnte eine Begründung der leichten graduellen Unterschiede in der Rangstabilität der einzelnen Qualitätsdimensionen liegen.

6.1.3 Stabilität zwischen und innerhalb der Unterrichtsthemen

Für die Kognitive Aktivierung wurde in Abschnitt 2.4.3 auf die Studie von Fauth et al. (2014b) hingewiesen, die zu der Hypothese führte, dass für die Qualitätsdimension Kognitive Aktivierung eine Themenabhängigkeit bestehen könnte. Diese Hypothese hat sich hier nicht bestätigt. Die Korrelationen zwischen den Messungen lassen diesen Schluss nicht zu und legen eher eine Themenunabhängigkeit bei der Bewertung der Kognitiven Aktivierung durch die Schülerinnen und Schüler nahe. Zum gleichen Ergebnis kommt man bei der Auswertung der Abbildung 5.9, wie bereits in Abschnitt 5.1.3 beschrieben. Ähnlich wie bei der Beurteilung der Kognitiven Aktivierung ist in der Dimension der Instrukionalen Unterstützung keine Themenabhängigkeit auffällig. Darin widersprechen die erhaltenen Messergebnisse der in Abschnitt 3 formulierten Hypothese, dass die Instruktionale Unterstützung themenabhängig sein würde. Die aufgebrachte Hypothese, die Beurteilung der Emotionalen Unterstützung sei themenunabhängig, kann anhand der Tabelle 5.8, wie in Abschnitt 5.1.3 beschrieben, gestützt werden. Insgesamt kann durch die Messungen keine Themenabhängigkeit der Emotionalen Unterstützung nachgewiesen werden. Für die Qualitätsdimension Klassenführung kann in der Summe keine

Themenabhängigkeit bei der Beurteilung, wie in Abschnitt 5.1.3 beschrieben, nachgewiesen werden.

Die Berechnungen haben für alle 4 Unterrichtsdimensionen gezeigt, dass es keine Abhängigkeit der Unterrichtswahrnehmung von den unterrichteten Themen gab. Damit kommt diese Untersuchung zu einem anderen Ergebnis als die Darlegungen in Göllner et al. (2016), auch gestützt auf die Untersuchung von Wagner et al. (2013), haben vermuten lassen. Hier sah man bei der Beurteilung der Unterrichtsqualität durch Schülerinnen und Schüler nur teilweise eine Themenunabhängigkeit gegeben.

6.2 Diskussion Unterrichtswahrnehmung und Leistung

Um die Zusammenhänge zwischen Unterrichtswahrnehmung und Lernleistungen von Schülerinnen und Schülern zu untersuchen, werden verschiedene Analysen herangezogen, die zusammengeführt ein Bild ergeben sollen.

Die im Abschnitt 5.2.2 dargestellten Ergebnisse können als Anhaltspunkte für spätere Forschungen dienen, immerhin deuten sich für alle 4 Qualitätsdimensionen Hinweise auf Korrelationen zwischen den linearen Modellen und der Leistung an:

Beispielsweise zeigt die Höhe der Ausgangseinschätzung der Schülerinnen und Schüler sehr unterschiedliche Korrelationen zur Leistung bei den Qualitätsdimensionen Kognitive Aktivierung und der Klassenführung auf. Korrelationen zum Anstieg konnten ebenfalls bei allen vier Qualitätsdimensionen gefunden werden.

Es werden Regressionsanalysen zur Unterrichtswahrnehmung und Leistung herangezogen, die auch als Pfaddiagramme in Abschnitt 5.2.3 veranschaulicht wurden: Für alle berechneten Regressionsanalysen und deren Pfaddiagramme gilt zu sagen, dass die Pfade höherer Autoregressionen notwenig sind, um annähernd passende Modellfitparameter für die Modellation zu erhalten. Bei geringerem Grad der Autoregressionen verschlechtern sich die Passungen erheblich. Die Kriterien der Modellpassung, wie sie in 4.3.5, auf Seite 49 beschrieben sind, gelten auch hier und konnten durch die Zulassung der Autoregressionen eingehalten werden.

Damit lässt sich insgesamt für den Zusammenhang zwischen Unterrichtswahrnehmung und Leistung sagen, dass die vorgestellten Berechnungen einen Zusammenhang zwischen Unterrichtswahrnehmung und Leistung aufzeigen. Es zeigen sich Einflüsse der Leistungen auf den Einschätzungstrend (das lineare Strukturgleichungsmodell), beispielsweise fällt auf, dass eingangs leistungsstärkere Schülerinnen und Schüler die Unterrichtsqualität insgesamt negativer einschätzen und auch zukünftig stärker negativ bewerten als

Schülerinnen und Schüler mit weniger, in KERMIT5 ermittelter Eingangsleistungsfähigkeit. Es zeigt sich weiter, dass das mathematikbezogene Selbstkonzept ebenfalls einen Einfluss auf die Bewertung der Unterrichtsqualität in allen ihren Facetten hat.

Die Regressionsanalyse der Unterrichtswahrnehmung und Leistung zeigt hingegen keine deutlichen Prädiktoren für zukünftige Ergebnisse an. Weder ist aus einer Leistungseinschätzung eindeutig und sicher eine Unterrichtswahrnehmung schlussfolgerbar noch andersherum.

Im Abschnitt 2.6 wurde der unzureichende und unscharfe Wissenstand der Forschung zu diesem Themenfeld aufgezeigt. Die vorliegende Untersuchung ordnet sich hier ein. Fauth et al. (2014b) kommt zu dem Schluss, dass es auf der Individualebene betrachtet keinen prädiktiven Zusammenhang zwischen Unterrichtswahrnehmung und Leistung gibt. Diese Aussage wird durch diese Studie bestätigt, darüber hinaus wurden aber auch Ansätze für weitere Untersuchungen aufgezeigt.

6.3 Diskussion des methodischen Vorgehens

Diese Arbeit stellt erstmals einen deutschsprachigen Fragebogen, der die vier Qualitätsdimensionen guten Unterrichts *Kognitive Aktivierung*, *Instruktionale Unterstützung*, *Emotionale Unterstützung* und *Klassenführung* erfasst, bereit und bestätigt seine Verwendbarkeit durch eine hohe nachgewiesene Güte. Das methodische Vorgehen, das Instrument zu 8 Messzeiten einzusetzen, stellt im Vergleich zu anderen Studien eine hohe Anzahl dar. Diese 8 Messzeitpunkte sind auch mit Blick auf die Stabilität innerhalb und zwischen Unterrichtsthemen bewusst gewählt und innerhalb des Schuljahres abgestimmt.

Der harmonisierte Unterricht in den drei Testklassen, auf denen die Untersuchung fußt, ist in anderen Studien nicht zu finden und bietet daher weiterführende Erkenntnisse zum Forschungsstand. Durch dieses kontrolliert abgestimmte Arrangement erlaubt die Studie eine themenspezifische Betrachtung und die Herstellung themenbezogener Analysen. Die Erkenntnisse zur Themenabhängigkeit des Forschungsfeldes sind wegen des schwer einzurichtenden harmonisierten Unterrichtsfeldes bisher als gering zu bewerten. Hier schafft die vorliegende Arbeit einen Erkenntniszuwachs.

Andere Untersuchungen des Forschungsfeldes führen in der Mehrheit Analysen an, die auf Klassenmittelwerten basieren. Eine Stärke der vorliegenden Arbeit besteht nun dar-

in, dass die individuelle Schülerebene längsschnittlich mit Hilfe von linearen Strukturgleichungsmodellen betrachtet wurden. Darüber hinaus wurde der klassenspezifische Einfluss bei Bewertungen der Unterrichtsqualität und der Unterrichtswahrnehmungen durch eine Zentrierung der klassenspezifischen Werte herausgerechnet, auch das ist als Neuerung zu bezeichnen.

Grenzen der vorliegenden Arbeit betreffen vor allem nachfolgende Aspekte:

Die geringe Anzahl von Testklassen und damit verbunden die geringe Anzahl von teilnehmenden Schülerinnen und Schülern stellt die größte Einschränkung der Aussagekraft der Untersuchungen dar. Die Verallgemeinerung der Erkenntnisse ist dadurch schwierig und erforderte eine Replikation in größerem Umfang. Damit begründet ist die Validität der berechneten Modelle nicht in jedem Fall optimal. Die Modellfits bei kleinen Datenumfängen sind wie in Abschnitt 4.3.5 oftmals schlechter an die Daten angepasst. Auch hier würde eine größere Testkohorte Genauigkeit und eine höhere Präzision schaffen. Eine Ausweitung der Untersuchung auf andere Jahrgangsstufen und Fächer könnte die Ergebnisse absichern.

Die im Testinstrument verwendeten Items hatten nicht alle die Schülerinnen und Schüler als Referenten. Das wäre bei einer Folgeuntersuchung sicher vorteilhaft zu verändern.

6.4 Bilanz und Ausblick

Ziel der vorliegenden Arbeit war es, die in Abschnitt 3 formulierten Forschungsfragen zu beantworten. Es galt zum einen die Stabilität der Unterrichtswahrnehmung in Hinblick auf die Stabilität ihres Qualitätsniveaus, ihre Rangstabilität und die Stabilität zwischen und innerhalb verschiedener Unterrichtsthemen zu untersuchen. Konkret waren folgende Forschungsfragen formuliert:

- (a) Wie stabil ist das beobachtete Qualitätsniveau in den einzelnen Dimensionen?
- (b) Wie stabil sind die Einschätzungen der Unterrichtsqualität in den vier Dimensionen Kognitive Aktivierung, Instruktionale Unterstützung, Emotionale Unterstützung und Klassenführung über die Zeit?
- (c) Sind die Beurteilungen der Unterrichtsqualität themenunabhängig?

Zum anderen sollten Zusammenhänge zwischen Unterrichtswahrnehmung und Lernleistung aufgezeigt oder negiert werden.

Welchen Zusammenhang zeigen die Mathematikleistungen und das mathematikbezogene Selbstkonzept mit den Unterrichtsbeurteilungen der Schülerinnen und Schüler?

Die nachgewiesenen Ergebnisse sollten in den in Abschnitt 2 beschriebenen Forschungsstand eingeordnet werden.

Dabei konnten die formulierten Forschungsfragen mit Einschränkungen beantwortet werden. Die vorliegende Arbeit liefert in der Bilanz Erkenntniszuwachs und leistet dadurch einen Mehrwert für das Forschungsfeld.

Um die dieser Arbeit zugrundeliegenden Forschungsfragen umfangreicher und valider zu beantworten, müssten Folgestudien größeren Umfangs erfolgen. Nur so könnten die Limitationen dieser Studie überwunden werden.

Letztlich liefert diese Arbeit der Praxis des Schulalltags einen deutschsprachigen Fragebogen zur Erfassung der vier Qualitätsdimensionen, dessen psychometrische Güte bestätigt werden konnte. Sie stützt die Argumentation derjenigen, die Unterrichtsfeedback von Schülerinnen und Schülern regelmäßig einholen und zur Qualitätssicherung ihres Unterrichts nutzen. Damit stellt diese Arbeit nicht nur einen Mehrwert für das Forschungsfeld, sondern auch für die Schule, die gelebte Praxis, dar. Auch das war ein Anspruch dieser Arbeit.

Symbolverzeichnis

Symbol	Beschreibung
α	Chronbachs Alphakoeffizient
β	(geschätzter) Regressionskoeffizient
<i>CFI</i>	Comparative Fit Index
<i>d</i>	Effektstärke nach Cohen
η^2	Empirisches η^2 ; Effektstärkemaß
<i>F</i>	F-Wert
<i>M</i>	Mittelwert; arithmetisches Mittel
\bar{M}	Mittelwert; je Schülerin/Schüler über alle 8 Messzeitpunkte einer Skala
<i>N</i>	Mächtigkeit einer Stichprobe; Anzahl der Teilnehmer
<i>p</i>	<i>p</i> -Wert; Signifikanzwert; Überschreitungswahrscheinlichkeit
<i>r</i>	Korrelationskoeffizient nach Bravais-Pearson
<i>r_{it}</i>	Trennschärfekoeffizient
<i>RMSEA</i>	Root Mean Square Error of Approximation
R^2	Bestimmtheitsmaß
<i>SD</i>	Standardabweichung
<i>SRMR</i>	Standardized Root Mean Square Residual
<i>TLI</i>	Tucker-Lewis Index
χ^2	Hypothesentest mit Chi-Quadrat-verteilter Testprüfgrößen, hier als Maß einer Anpassungsgüte verwendet

Symbole 1: Spezielle Symbole - Teil 1

Symbol	Beschreibung
BIJU	Bildungsverläufe und psychosoziale Entwicklung im Jugendalter
CLASS	Classroom Assessment Scoring System
COACTIV	Cognitive Activation in the Classroom: The Orchestration of Learning Opportunities for the Enhancement of Insightful Learning in Mathematics
DESI	Deutsch-Englisch-Schülerleistungen-International
IGLU	Deutsch für Internationale Grundschul-Lese-Untersuchung; international mit PIRLS bezeichnet
KERMIT	Kompetenzen ermitteln
KESS	Kompetenzen und Einstellung von Schülerinnen und Schülern
KS HAM 4/5	Hamburger kombinierten Schulleistungstest für vierte und fünfte Klassen
LAU	Aspekte der Lernausgangslage und Lernentwicklung
MINT	Gemeint sind die Fächer: Mathematik, Informatik, Naturwissenschaften und Technik
OECD	Organisation for Economic Cooperation und Development
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PLUS	Professionswissen von Lehrkräften, naturwissenschaftlicher Unterricht und Zielerreichung im Übergang von Primar- zur Sekundarstufe
TIMSS	Third International Mathematics and Science Study 1999
VERA	VERgleichsArbeiten

Symbole 2: Abkürzungen - Studien und Diagnoseinstrumente

Symbol	Beschreibung
ANOVA	Varianzanalyse
DISK	DISK-Gitter; Differentielle schulische Selbst-Gitter
G	Steht als Fußnote für die gesamte Stichprobe
IfBQ	Institut für Bildungsmonitoring und Qualitätsentwicklung Hamburg
PASW	Predictive Analysis SoftWare; Vermarktung von SPSS zwischen 2009 und 2010
SPSS	Statistical Package; Software von IBM
LGM	Lineares Strukturgleichungsmodell
<i>Mplus</i>	Statistik Software von L. K. Muthén & Muthén

Symbole 3: Abkürzungen

Symbol	Beschreibung
BMBF	Bundesministerium für Bildung und Forschung
BPH	Bright- Person-Hypothese
EU	Qualitätsdimension Emotionale Unterstützung
H1 /H2	Halbjahresnote 1. Halbjahr bzw. Halbjahresnote 2. Halbjahr
IU	Qualitätsdimension Instruktionale Unterstützung
J	Steht als Fußnote für Jungen einer Stichprobe
K1 - 4	Note Klassenarbeit 1 bis 4
KA	Qualitätsdimension Kognitive Aktivierung
KF	Qualitätsdimension Klassenführung
KTH	Knowledgeable-Teacher-Hypothese
M	Steht als Fußnote für Mädchen einer Stichprobe
M1 - 8	Messung zu den Messzeitpunkten 1 bis 8
<i>MWSKMi</i>	Item-ID i- Item Selbstkonzeptskala

Symbole 4: Abkürzungen dieser Arbeit

A Literaturverzeichnis

- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaften*, (9), 469–520.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W. & Neubrand, M. (2004). PISA 2003. Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs. In P.-K. Deutschland (Hrsg.), (Kap. Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. S. 314–354). Münster: Waxmann.
- Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. & Neubrand, J. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Wiesbaden: Springer.
- Blömeke, S. (2004). Empirische Befunde zur Wirksamkeit der Lehrerbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & E. Wild (Hrsg.), (S. 59–91). Bad Heilbrunn: Klinkhardt.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler*. 4. Heidelberg.
- Bos, W., Dudas, D.-F., Gröhlich, C., Guill, K. & Scharenberg, K. (2010). *KESS 8. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Bromme, R. & Haag, L. (2004). Forschung zur Lehrerpersönlichkeit. In W. Helsper & J. Böhme (Hrsg.), (S. 777–793). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brophy, J. (2000). Teaching. *Educational Practices Series, 1*. Zugriff unter <http://www.ibe.unesco.org>
- Brophy, J. (2006). Handbook of educational psychology. In P. A. Alexander & P. Winne (Hrsg.), (Kap. Observational research on generic aspects of classroom teaching, S. 755–780). Mahwah, NJ: Erlbaum.
- Brown, J. L., Jones, S. M., LaRusso, M. D. & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4Rs program. *Journal of Educational Psychology*, (102), 153–167.

- Brügelmann, H. (2018). Sind Noten nützlich - und nötig? Ziffernzensuren und ihre Alternativen im empirischen Vergleich. *Lehren und Lernen*, (44), 37–39.
- Brunnhuber, P. (1977). *Prinzipien effektiver Unterrichtsgestaltung*. Donauwörth: Auer.
- Camron, C. E., Connor, C. M. & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology*, (43), 61–85.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: LEA Lawrence Erlbaum Associates.
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, (77), 113–143.
- Corp., I. (Hrsg.). (2015). *IBM Corp. Released 2015. IBM SPSS Statistics for Macintosh, Version 23.0.02*. Armonk, NY.
- Den Brok, P., Brekelmans, M. & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement*, (15), 407–442.
- Derry, S., Pea, R., Barron, B., A. Engle, R., Erickson, F., Goldman, R., ... Sherin, B. (2010). Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology and Ethics. *Journal of the Learning Sciences*, (19), 3–53.
- DESI. (2018). Zugriff unter <https://www.kmk.org/themen/qualitaets-sicherung-in-schulen/bildungsmonitoring/internationale-schulleistungsvergleiche/desi.html>
- Desimone, L. M., Smith, T. & Frisvold, D. E. (2010). Survey Measures of Classroom Instruction: Comparing Student and Teacher Reports. *Educational Policy*, (24), 267–329.
- Ditton, H. (2006). *Unterrichtsqualität*. Handbuch Unterricht. Bad Heilbrunn: Klinkhardt.
- Douven, I. (2017). Abduction. Zugriff unter <https://plato.stanford.edu/entries/abduction>
- Drucksache 20/7094*. (2013). Institut für Bildungsmonitoring und Qualitätssicherung. Senat der Freien und Hansestadt Hamburg. Zugriff unter <http://www.hamburg.de/contentblob/4025290/9cfd3e28b23342acb6874838d21a3961/data/pdf-hh-sozialindex-drucksache-20-7094.pdf>
- 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.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.

- Epstein, R. M. & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, (287), 226–235.
- Fan, X., Thompson, B. & Wang, L. (1999). Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56–83.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014a). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, (28), 127–137.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014b). Student ratings quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, (29), 1–9.
- Feldon, D. F. (2007). Cognitive load and classroom teaching: The doubleedged sword of automaticity. *Educational Psychologist*, 123–137.
- for Economic Co-operation, O. .-. O. & Development. (2005). *Teachers matter: Attracting, developing, and recruiting effective teachers*. Organisation for Economic Co-operation und Development. Paris.
- for Mathematics for Teaching Project, L. (2010). Measuring the mathematical quality for instruction. *Journal of Mathematics Teacher Education*, 1(14), 25–47.
- Frischer Wind für die Qualitätsentwicklung - Dokumentation der Veranstaltung zur Qualitätsentwicklung an allgemeinbildenden Schulen am 25. Oktober 2018*. (2018). Behörde für Schule und Berufsbildung. Zugriff unter <https://www.hamburg.de/contentblob/11946916/a3300c6c8ea7416f1a3832579f3c9050/data/frischer-wind.pdf>
- für Bildungsmonitoring und Qualitätsentwicklung, I. (o.D. a). Zugriff unter <http://www.hamburg.de/bsb/hamburger-sozialindex/>
- für Bildungsmonitoring und Qualitätsentwicklung, I. (o.D. b). Zugriff unter <https://www.kermit-hamburg.de/>
- Geiser, C. (2010). *Datenanalyse mit Mplus - Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gigliotti, R. J. & Buchtel, F. (1990). Attributional bias and course evaluations. *Journal of Educational Psychology*, (82), 341–351.
- Girden, E. R. (1992). ANOVA: Repeated measures. *Sage university papers. Quantitative applications in the social sciences*, (07-084).
- 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.
- Göllner, R., Wagner, W., Klieme, E., Lüdke, O., Nagengast, B. & Trautwein, U. (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments*. 63–82.
- Greenwald, A. G. & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, (52), 1209–1217.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Haag, L. & Lohrmann, K. (2006). *Lehrerhandeln: Lehrerkognitionen und Lehrerexpertise*. Handbuch Unterricht. Bad Heilbrunn: Klinkhardt.
- Hamre, B. K. & Pianta, R. (2010). *Classroom environments and developmental processes: Conceptualization, measurement, & improvement*. Handbook of research on schools, schooling and human development. New York, NY: Routledge.
- Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Hattie, J. A. (2012). *Visible Learning for Teachers. Maximizing impact on Learning*. London: Routledge.
- Helmke, A. (1991). *Schule und Persönlichkeitsentwicklung. Ein Resümee der Längsschnittforschung* (R. Pekrun & H. Fend, Hrsg.). Enke.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität* (A. Helmke, Hrsg.). Seelze-Velber: Klett-Kallmeyer.
- Helmke, A., Piskol, K., Pikowsky, B. & Wagner, W. (2009). Schüler als Experten von Unterricht. Unterrichtsqualität aus Schülerperspektive. *Lernende Schule*, 46–47, 98–105.
- Hochweber, J., Hosenfeld, I. & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology*, (104), 289–300.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6:1, 1–55.
- Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions* (2nd, Hrsg.). Newbury Park, CA: Sage Publications.
- Kauertz, A., Kleickmann, T., Ewerhardy, A., Fricke, K., Lange, K., Ohle, A., ... Möller, K. (2011). Dokumentation der Erhebungsinstrumente im Projekt PLUS. Zugriff unter <https://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet/>

- Derivate- 36697 / Dokumentation _ der _ Erhebungsinstrumente _ im _ Projekt _ PLUS_ 2013_ final2.pdf
- Kennedy, M. M., Ahn, S. & Choi, J. (2008). The value added by teacher education. In M. Cochran-Smith, S. Feimann-Nemser, D. J. McIntyre & K. E. Demers (Hrsg.), (Bd. 3, S. 1249–1273). New York, NY: Routledge.
- Kenny, D., Kanistan, B. & McCoach, D. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507.
- Klieme, E. (2018). Zusammenfassung zentraler Ergebnisse der DESI-Studie. Zugriff unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2006/2006_03_01-DESI-Ausgewaehlte-Ergebnisse.pdf
- Klieme, E., Hartig, J. & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & Leutner (Hrsg.), (S. 3–22). Göttingen: Hogrefe & Huber.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projektes "Pythagoras". In M. Prenzel & L. Allolio-Näcke (Hrsg.), (S. 127–146). 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.), (S. 137–160). Waxmann.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung. In BMBF (Hrsg.), (Kap. TIMSS - Impulse für Schule und Unterricht, Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente, S. 43–58). Bonn: BMBF.
- Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: Eine Einführung in eRm*. Wien: Uni-Taschenbücher (UTB), Facultas Verlags- und Buchhandels AG.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Kuger, S. (2016). *Curriculum and learning time in international school achievement studies*. Assessing contexts of learning. Berlin: Springer.
- Künsting, J., Neuber, V. & Lipowsky, F. (2016). Teacher self-efficacy as a long-term predictor of instructional quality in the classroom. *European Journal Psychology of Education*, (31), 299–322.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.

- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, (9), 231–251.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U. & Krauss, S. (2011). *Professionelle Kompetenz von Lehrkräften - Ergebnisse des Forschungsprogramms COACTIV* (M. Neubrand, Hrsg.). Münster: Waxmann.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W. & Neubrand, M. (o.D.). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler, Schulformunterschiede in der Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaften*, (8), 502–520.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and students development. *Journal of Educational Psychology*, (105), 805–820.
- Kunter, M. & Voss, T. (2013). *The model of instructional quality in COACTIV: A multicriteria analysis*. Cognitive activation in mathematics classroom and professional competence of teachers. Results from COACTIV project. New York: Springer.
- Lai, K. & Green, S. B. (2016). The Problem with Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree. *Multivariate Behavioral Research*, 51:2-3, 220–239.
- Lange, B. (2005). Bildungsstandards un Unterrichtsplanung - Konsequenzen für didaktisches Denken un Planen. *Lehren und Lernen*, (5), 3–10.
- Lehmann, R. H., Peek, R. & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage und der Lernentwicklung- Bericht über die Erhebung im September 1996 (LAU 5)*. BSB Hamburg. Zugriff unter <http://bildungsserver.hamburg.de/contentblob/2815702/3b66049d4257501a0d44dce9b7ca449c/data/pdf-schulleistungstest-lau-5.pdf>
- Lehmann, R. H., Peek, R. & Gänsfuß, R. (2011). *LAU - Aspekte der Lernausgangslage und der Lernentwicklung Klassenstufen 5,7 und 9* (B. für Schule und Berufsbildung, Hrsg.). Waxmann.
- Lipowsky, F. (2002). *Zur Qualität offener Lernsituationen im Spiegel empirischer Forschung - Auf die Mikroebene kommt es an*. (U. Drews & U. Wallrabenstein, Hrsg.). Frankfurt a. M.: Grundschulverband.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), (Kap. 4, Bd. 2, 69ff). Heidelberg: Springer.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E. & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of Pythagorean Theorem. *Learning and Instruction*, (19), 527–537.

- Lüdke, O., Trautwein, U., Kunter, M. & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment – A re-analysis of TIMSS data. *Learning Environments Research*, (9), 215–230.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, (58), 103–117.
- Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, (52), 1187–1197.
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen.
- Meyer, H. (2007). Zehn Merkmale guten Unterrichts. Anleitung zur Überarbeitung um zum Einsatz eines Beobachtungsbogens. In W. Endres (Hrsg.), (S. 166–187). Weinheim: Beltz.
- Mietzel, G. & Willenberg, H. (1996). *Hamburger Kombiniertes Schulleistungstest für vierte und fünfte Klassen (KS HAM 4/5)*. Unveröffentlichtes Testverfahren. Göttingen: Hogrefe.
- Moosbrugger, H. (2012). Testtheorie und Fragebogenkonstruktion. In H. Moosbrugger & A. Kelava (Hrsg.), (Kap. Item-Response-Theorie (IRT)). Berlin: Springer.
- Muthén, L. K. & Muthén, B. O. (2014). *MPlus [Version 8.1, Computer Software]*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K. & Muthén, B. O. (2017). *Mplus User's Guide*. (8. Aufl.). Los Angeles: Muthén & Muthén.
- Orgin, S., Silber, S., Friedrich, A., Trautwein, U. & Schmitz, B. (2017). Entwicklung und empirische Prüfung einer Lehrkräftefortbildung zur Förderung von Selbstregulationskompetenz und mathematischer Kompetenz bei Schülerinnen und Schülern der Haupt- und Werkrealschule („Lernen mit Plan“). In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals - Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (1. Aufl., S. 194–214). Berlin Heidelberg New York: Springer-Verlag.
- Oser, F. K. & Baeriswyl, F. J. (2001). *Choreographies of teaching: Bridging instruction to learning*. Handbook of research on teaching. Washington, DC: American Educational Research Association.
- Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, (13), 423–451.
- Pianta, R. C., La Paro, K. M. & Hamre, B. K. (2004). *Classroom assessment scoring system [CLASS]*. Charlottesville, VA.

- Pianta, R., Hamre, B. & Allen, J. (2012). *Handbook of Research on Student Engagement* (S. Christenson, A. Reschly & C. Wylie, Hrsg.). New York: Springer.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings* (D. H. Rost, Hrsg.). Münster: Waxmann.
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of 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.
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht*. Münster: Waxmann.
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg Verlag.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 3(13), 273–304.
- 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.
- Scheltwort, P. (2006). *Lehrerhandeln wahrnehmen, erfassen, bewerten. Theoretische und empirische Analysen zum systemanalytischen Beobachtungsverfahren „Lehrerhandeln im Unterricht“*. Unveröffentlichte Dissertation, Pädagogische Hochschule Heidelberg.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, 8(2), 23–74.
- Schneider, M. & Stern, E. (2010). The cognitive perspective on learning: Ten cornerstone findings. *The nature of learning: Using research to inspire practice*, 69–90.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. München: Oldenbourg Verlag.
- Schulte, K., Hartig, J. & Pietsch, M. (2014). Grundlagen für eine daten- und theoriegestützte Schulentwicklung. Konzeption und Anspruch des Hamburger Instituts für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) Hanse - Hamburger Schriften zur Qualität im Bildungswesen. In D. Fickermann & N. Maritzen (Hrsg.), (Kap. Der Sozialindex für Hamburger Schulen, 13, S. 67–80). Münster: Waxmann.

- Schumacher, R., Barth, A., Lipscher, J. & Hänger-Surer, B. (2017). Kognitive Aktivierung - Wie lassen sich mathematisch-naturwissenschaftliche Inhalte lernwirksam vermitteln? *Schulmanagement Handbuch*, (163).
- Segars, A. H. & Grover, V. (1998). Strategic Information Systems Planning Success: An Investigation of the Construct and Its Measurement. *MIS Quarterly*, 22(2), 139–163.
- Seidel, T. (2003). *Lehr-Lernskripts im Unterricht*. Münster: Waxmann.
- Seidel, T. (2015). *Pädagogische Psychologie* (E. Wild & J. Möller, Hrsg.). Heidelberg: Springer.
- 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*, 4(77), 454–499.
- Shulman, L. S. (1986). *Paradigms and research programs in the study of teaching: a contemporary perspective*. Handbook of research on teaching. London: Macmillan.
- Stern, E., Schalk, L. & Schumacher, R. (2016). Basiswissen Lehrerbildung: Schule und Unterricht. Lehren und Lernen. In J. Möller, M. Köller & T. Riecke-Baulecke (Hrsg.), (Kap. Lernen, S. 106–120). Seelze: Klett-Verlag.
- Stigler, J. & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.
- Taut, S. & Rakoczy, K. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction*, (46), 45–60.
- Thonke, F. & Lücken, M. (2014). KERMIT - Kompetenzen ermitteln. 19. EMSE-Tagung, IFBQ. Zugriff unter https://www.emse-netzwerk.de/uploads/Main/19E_Luecken_Thonke_EMSE_19_KERMIT_HH.pdf
- Trautwein, U., Lüdke, O., Klieme, E., Nagengast, B. & Wagner, W. (2011). *Erfassung der Unterrichtsqualität in Large- Scale- Studien: Optimierung der Modellierung und Itemauswahl*. Tübingen.
- Vazire, S. & Solomon, B. C. (2015). Self- and other- knowledge of personality. In M. Mikulincer, P. Shaver, M. L. Cooper & R. Larsen (Hrsg.), (S. 261–281). Washington, DC: American Psychological Association.
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U. & Lüdtke, O. (2013). Construction 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.
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B. & Trautwein, U. (2015). Student and Teacher Ratings of Instructional Quality: Consistency of Ratings Over Time, Agreement, and Predictive Power. *Journal of Educational Psychology*.

- Walberg, H. J. & Paik, S. J. (2000). *Effective educational practices*. Brussels: International Academy of Education & International Bureau of Education. Zugriff unter <http://www.ibe.unesco.org>
- Weber, H. M. & Petermann, F. (2016). Der Zusammenhang zwischen Schulangst, Schullust, Anstrengungsvermeidung und den Schulnoten in den Fächern Mathematik und Deutsch. *Zeitschrift für Pädagogik*, (62), 551–570.
- Weinert, F. E. (2001). Qualifikation und Unterricht zwischen gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In W. Melzer & U. Sandfuchs (Hrsg.), (S. 65–85). Weinheim: Juventa.
- Weinert, F. E., Schrader, F. .-.-. W. & Helmke, A. (1989). Quality of Instruction and achievement outcomes. *International Journal of Educational Research*, (13), 895–914.
- Zaunbauer, A., Gebauer, K., Retelsdorf, J. & Möller, J. (2013). Motivationale Veränderung von Grundschulkindern in Englisch, Deutsch und Mathematik im Immersions- und Regelunterricht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, (45), 91–102.

B Itemanalyse

Analyse der einzelnen Messzeitpunkte

Hier finden sich die Ergebnisse der Itemanalysen zu den einzelnen Messzeitpunkten für die Skalen zur Erfassung der Unterrichtsdimensionen und des Selbstkonzeptes für die drei Klassen im Gesamten sowie in der Einzelaufschlüsselung 5b, 5c und 5d. In der ersten Zeile pro Skala sind die Mittelwerte und Standardabweichungen der betreffenden Skala auf Schülerebene (S) und auf Klassenebene (K) angegeben. Die Itemnummerierung entspricht der des Fragebogens.

Messzeitpunkt 1

Die Messung fand in der Zeit vom 05.12.16 bis zum 09.12.16 statt. Das Selbstkonzept wurde in dieser Messung auch mit abgefragt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Es haben in der Klasse 5c insgesamt $N_G = 25$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,52$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,46$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 79$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet

haben, ergibt ein Cronbachs α mit $\alpha = 0,69$. Die interne Konsistenz der Skala für die Gesamtstichprobe ist damit als gerade noch akzeptabel zu bewerten.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,70$.

Es haben in der Klasse 5c insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,51$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,56$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 76$ Schülerinnen und Schülern, die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,61$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,41$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,59$.

Es haben in der Klasse 5d insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,46$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 79$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,52$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,73$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 79$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$.

Mathematikbezogenes Selbstkonzept

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 4 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,89$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 4 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,75$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 4 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der 4 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,85$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,16	3,42	2,98	3,08	0,63	0,48	0,67	0,75				
	K	3,16	3,42	2,97	3,08	0,78	0,68	0,75	0,82				
	1	3,00	3,30	2,84	2,85	0,78	0,67	0,75	0,86	0,45	0,43	0,33	0,42
	2	3,49	3,72	3,44	3,30	0,75	0,53	0,77	0,87	0,40	0,64	0,05	0,49
	3	2,89	3,30	2,72	2,63	0,72	0,62	0,61	0,74	0,63	0,59	0,57	0,59
	4	2,95	3,22	2,80	2,82	0,77	0,75	0,82	0,68	0,40	0,51	0,14	0,34
	5	3,40	3,43	3,12	3,63	0,75	0,76	0,78	0,63	0,34	0,62	0,39	0,04
	6	3,00	3,22	2,76	3,00	0,78	0,74	0,83	0,73	0,30	0,56	0,33	-0,18
7	3,39	3,78	3,00	3,37	0,69	0,42	0,57	0,79	0,29	0,62	0,06	-0,08	
Instruktionale Unterstützung	S	3,27	3,28	3,14	3,38	0,74	0,80	0,75	0,68				
	K	3,27	3,28	3,14	3,38	0,82	0,88	0,80	0,76				
	8	3,59	3,72	3,52	3,54	0,61	0,46	0,65	0,71	0,41	0,62	0,62	0,21
	9	2,56	2,26	2,48	2,92	0,87	0,88	0,82	0,80	0,24	0,34	0,16	0,18
	10	3,18	3,36	3,00	3,19	0,76	0,74	0,65	0,85	0,24	0,34	0,02	0,27
	11	3,19	3,38	2,84	3,35	0,72	0,78	0,69	0,56	0,41	0,55	0,35	0,18
	12	3,63	3,68	3,32	3,89	0,56	0,54	0,56	0,43	0,28	0,22	0,29	0,11
	13	2,83	2,44	2,84	3,19	0,91	0,87	0,94	0,80	0,08	-0,02	-0,16	0,49
	14	3,32	3,18	3,20	3,58	0,72	0,78	0,76	0,58	0,43	0,42	0,42	0,40
15	3,61	3,70	3,64	3,50	0,65	0,61	0,57	0,76	0,35	0,59	0,25	0,37	
16	3,41	3,68	3,28	3,27	0,80	0,63	0,79	0,92	0,31	0,63	0,37	0,17	
Emo Unterstützung	S	3,49	3,57	3,33	3,58	0,53	0,55	0,55	0,50				
	K	3,49	3,55	3,33	3,58	0,67	0,76	0,67	0,62				
	17	3,83	3,85	3,78	3,86	0,41	0,36	0,42	0,45	0,29	0,39	0,26	0,20
	18	3,27	3,17	3,13	3,52	0,71	0,94	0,55	0,51	0,19	0,16	0,25	0,22
	19	3,41	3,61	3,19	3,42	0,74	0,68	0,74	0,76	0,37	0,31	0,53	0,11
	20	3,53	3,67	3,44	3,48	0,64	0,62	0,58	0,71	0,35	0,27	0,34	0,44
21	3,39	3,57	3,11	3,50	0,70	0,58	0,80	0,61	0,29	0,02	0,40	0,30	
Klassenführung	S	2,39	2,70	1,96	2,50	0,47	0,41	0,48	0,51				
	K	2,36	2,67	1,93	2,49	0,75	0,68	0,70	0,67				
	22	2,01	2,50	1,58	2,11	0,72	0,71	0,58	0,58	0,68	0,61	0,71	0,42
	23	2,58	2,81	2,12	2,82	0,71	0,63	0,65	0,62	0,62	0,45	0,72	0,38
	24	2,42	2,67	2,04	2,56	0,63	0,73	0,45	0,51	0,72	0,86	0,52	0,55
	25	2,23	2,52	1,78	2,41	0,77	0,70	0,71	0,69	0,80	0,81	0,74	0,70
26	2,62	2,90	2,39	2,63	0,72	0,53	0,80	0,74	0,55	0,50	0,61	0,44	

Im ersten Messzeitpunkt wurde noch das Selbstkonzept erhoben. Hier die dazugehörige Itemanalyse.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Selbstkonzept	S	2,99	3,09	2,78	3,10	0,48	0,39	0,52	0,52				
	K	2,99	3,08	2,78	3,10	0,92	0,87	0,86	1,00				
	27	2,99	3,04	2,93	3,00	0,91	0,92	0,73	1,07	0,69	0,74	0,61	0,73
	28	2,87	3,02	2,67	2,93	0,99	0,94	0,92	1,11	0,68	0,68	0,57	0,73
	29	3,01	3,14	2,79	3,11	0,94	0,84	0,89	1,05	0,74	0,82	0,62	0,74
	30	3,08	3,14	2,74	3,37	0,85	0,82	0,90	0,74	0,65	0,79	0,40	0,79

Messzeitpunkt 2

Die Messung fand in der Zeit vom 23.01.17 bis zum 27.01.17 statt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 24$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,56$.

Es haben in der Klasse 5c insgesamt $N_G = 28$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,55$.

Es haben in der Klasse 5d insgesamt $N_G = 25$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,22$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 77$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,52$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 24$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,48$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,54$.

Es haben in der Klasse 5d insgesamt $N_G = 24$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,74$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 75$ Schülerinnen und Schülern, die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,64$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,63$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$.

Es haben in der Klasse 5d insgesamt $N_G = 21$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,67$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 73$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,80$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$.

Es haben in der Klasse 5d insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,78$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 79$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,91$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,15	3,36	3,01	3,08	0,68	0,56	0,67	0,82				
	K	3,15	3,36	3,01	3,08	0,76	0,62	0,76	0,85				
	1	2,98	3,25	2,79	2,94	0,67	0,44	0,69	0,77	0,20	0,29	0,1	0,06
	2	3,40	3,83	3,04	3,40	0,78	0,38	0,84	0,82	0,27	0,22	0,30	-0,01
	3	2,96	3,29	2,86	2,76	0,70	0,46	0,59	0,88	0,23	0,00	0,17	0,20
	4	2,91	3,00	2,86	2,88	0,74	0,71	0,76	0,78	0,19	0,31	0,26	-0,02
	5	3,46	3,54	3,36	3,48	0,74	0,66	0,73	0,82	0,33	0,25	0,50	0,18
	6	2,84	2,96	2,86	2,70	0,76	0,62	0,85	0,79	0,17	0,39	0,22	-0,07
7	3,44	3,63	3,30	3,42	0,67	0,49	0,69	0,79	0,45	0,62	0,42	0,31	
Instruktionale Unterstützung	S	3,22	3,34	3,02	3,33	0,74	0,75	0,80	0,65				
	K	3,22	3,34	3,02	3,32	0,84	0,81	0,88	0,77				
	8	3,53	3,79	3,33	3,48	0,67	0,39	0,68	0,80	0,27	0,61	0,21	0,10
	9	2,73	2,79	2,37	3,08	0,79	0,72	0,84	0,65	0,23	0,24	0,02	0,27
	10	3,13	3,33	2,85	3,25	0,86	0,76	0,86	0,90	0,41	0,39	0,19	0,48
	11	3,15	3,27	3,00	3,21	0,68	0,77	0,62	0,66	0,24	-0,09	0,26	0,38
	12	3,47	3,65	3,22	3,58	0,74	0,56	0,89	0,65	0,58	0,48	0,63	0,52
	13	2,73	2,29	2,79	3,13	1,00	1,08	0,97	0,80	-0,01	0,06	-0,27	0,53
	14	3,43	3,50	0,68	3,46	0,68	0,66	0,68	0,72	0,40	-0,03	0,40	0,69
15	3,54	3,79	3,35	3,50	0,61	0,44	0,65	0,66	0,52	0,34	0,61	0,46	
16	3,19	3,67	2,94	3,00	0,90	0,56	0,94	0,98	0,44	0,37	0,50	0,40	
Emo Unterstützung	S	3,48	3,65	3,26	3,55	0,44	0,38	0,52	0,40				
	K	3,48	3,65	3,26	3,55	0,68	0,56	0,82	0,59				
	17	3,80	3,88	3,74	3,76	0,44	0,33	0,53	0,44	0,52	0,48	0,58	0,41
	18	3,28	3,46	3,07	3,33	0,79	0,68	0,87	0,80	0,51	0,35	0,65	0,20
	19	3,32	3,52	2,93	3,57	0,74	0,59	0,83	0,60	0,64	0,31	0,77	0,41
	20	3,52	3,80	3,33	3,43	0,60	0,41	0,73	0,51	0,72	0,57	0,75	0,73
21	3,43	3,56	3,11	3,67	0,71	0,51	0,89	0,48	0,63	0,34	0,68	0,57	
Klassenführung	S	2,29	2,75	1,68	2,50	0,43	0,37	0,39	0,54				
	K	2,27	2,73	1,67	2,46	0,81	0,61	0,64	0,75				
	22	1,99	2,56	1,27	2,2	0,94	0,85	0,48	0,91	0,74	0,72	0,50	0,55
	23	2,42	2,90	1,70	2,74	0,78	0,58	0,63	0,44	0,79	0,70	0,81	0,37
	24	2,34	2,77	1,75	2,56	0,75	0,53	0,59	0,71	0,85	0,77	0,81	0,70
	25	2,15	2,64	1,61	2,24	0,74	0,52	0,63	0,66	0,79	0,73	0,73	0,59
26	2,56	2,90	2,07	2,76	0,78	0,57	0,66	0,83	0,71	0,60	0,60	0,63	

Messzeitpunkt 3

Die Messung fand in der Zeit vom 13.02.17 bis zum 17.02.17 statt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,64$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,71$.

Es haben in der Klasse 5d insgesamt $N_G = 28$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,49$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,63$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 23$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,76$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,58$.

Es haben in der Klasse 5d insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,83$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 77$ Schülerinnen und Schülern, die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,75$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,83$.

Es haben in der Klasse 5c insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,89$.

Es haben in der Klasse 5d insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,69$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 76$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,82$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,88$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,90$.

Es haben in der Klasse 5d insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,80$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,90$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	2,98	2,95	2,90	3,09	0,74	0,72	0,70	0,78				
	K	2,98	2,95	2,90	3,09	0,85	0,82	0,84	0,87				
	1	2,70	2,69	2,65	2,79	0,75	0,74	0,69	0,83	0,46	0,54	0,46	0,39
	2	3,13	3,12	2,96	3,29	0,85	0,77	0,82	0,94	0,46	0,47	0,40	0,50
	3	2,80	2,77	2,73	2,89	0,72	0,65	0,60	0,88	0,44	0,47	0,50	0,38
	4	2,81	2,81	2,85	2,79	0,78	0,69	0,88	0,79	0,26	0,18	0,40	0,22
	5	3,24	2,96	3,19	3,54	0,90	0,99	0,90	0,74	0,29	0,20	0,53	0,11
	6	2,76	2,92	2,62	2,75	0,90	0,84	1,10	0,75	0,28	0,39	0,49	-0,07
7	3,41	3,39	3,27	3,57	0,76	0,85	0,67	0,74	0,23	0,31	0,18	0,14	
Instruktionale Unterstützung	S	3,14	3,22	3,04	3,17	0,72	0,78	0,75	0,64				
	K	3,14	3,22	3,04	3,17	0,85	0,90	0,83	0,83				
	8	3,49	3,65	3,35	3,50	0,79	0,71	0,85	0,79	0,56	0,54	0,56	0,57
	9	2,68	2,61	2,54	2,86	0,90	0,78	1,03	0,85	0,20	0,17	0,06	0,37
	10	3,10	3,26	3,04	3,04	0,85	0,86	0,87	0,84	0,55	0,49	0,38	0,72
	11	2,92	3,04	2,85	2,88	0,72	0,82	0,67	0,69	0,47	0,60	0,18	0,56
	12	3,53	3,78	3,31	3,54	0,74	0,67	0,68	0,79	0,59	0,65	0,44	0,60
	13	2,78	2,52	2,92	2,86	0,90	1,04	0,80	0,85	-0,03	0,11	-0,36	0,19
	14	3,43	3,44	3,35	3,50	0,66	0,73	0,63	0,64	0,49	0,58	0,39	0,51
	15	3,36	3,52	3,23	3,36	0,78	0,73	0,76	0,83	0,59	0,41	0,64	0,67
16	3,12	3,61	2,77	3,04	0,83	0,66	0,76	0,84	0,60	0,59	0,46	0,72	
Emo Unterstützung	S	3,32	3,45	3,26	3,24	0,51	0,50	0,47	0,55				
	K	3,32	3,44	3,26	3,25	0,77	0,73	0,81	0,73				
	17	3,58	3,65	3,56	3,52	0,70	0,75	0,65	0,71	0,67	0,74	0,73	0,58
	18	3,20	3,00	3,28	3,32	0,71	0,75	0,79	0,56	0,48	0,52	0,67	0,39
	19	3,15	3,46	2,92	3,04	0,81	0,71	0,81	0,84	0,67	0,59	0,85	0,57
	20	3,34	3,66	3,24	3,12	0,78	0,56	0,93	0,73	0,67	0,60	0,81	0,58
	21	3,37	3,50	3,24	3,36	0,76	0,86	0,83	0,57	0,56	0,75	0,63	0,14
Klassenführung	S	2,17	2,39	1,68	2,43	0,43	0,46	0,36	0,46				
	K	2,14	2,36	1,65	2,38	0,78	0,74	0,71	0,66				
	22	1,94	2,15	1,39	2,25	0,79	0,73	0,57	0,75	0,77	0,72	0,80	0,61
	23	2,29	2,46	1,77	2,61	0,83	0,81	0,82	0,63	0,82	0,69	0,90	0,72
	24	2,18	2,46	1,62	2,43	0,74	0,71	0,64	0,57	0,84	0,81	0,87	0,64
	25	1,95	2,19	1,54	2,11	0,69	0,69	0,65	0,57	0,75	0,74	0,75	0,55
	26	2,51	2,69	2,08	2,75	0,75	0,68	0,74	0,65	0,62	0,59	0,56	0,42

Messzeitpunkt 4

Die Messung fand in der Zeit vom 27.02.17 bis zum 03.03.17 statt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 24$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,87$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,70$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,72$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 77$ Schülerinnen und Schüler die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,77$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 23$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,77$.

Es haben in der Klasse 5c insgesamt $N_G = 24$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,64$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,82$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 73$ Schülerinnen und Schülern, die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,76$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,88$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 78$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,83$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 24$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,96$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,94$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 78$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,95$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,05	3,14	2,93	3,08	0,65	0,58	0,69	0,67				
	K	3,05	3,15	2,93	3,08	0,84	0,82	0,85	0,83				
	1	2,77	2,96	2,48	2,89	0,78	0,81	0,75	0,71	0,53	0,75	0,61	0,15
	2	3,29	3,50	3,11	3,27	0,84	0,66	1,01	0,78	0,53	0,55	0,53	0,54
	3	2,78	3,00	2,56	2,81	0,74	0,66	0,70	0,80	0,64	0,74	0,55	0,60
	4	2,92	2,96	3,04	2,77	0,81	1,04	0,59	0,76	0,36	0,59	0,16	0,36
	5	3,34	3,19	3,33	3,50	0,72	0,70	0,68	0,76	0,44	0,70	0,22	0,55
	6	2,82	3,04	2,59	2,85	0,91	0,81	1,01	0,88	0,53	0,74	0,55	0,28
7	3,44	3,63	3,33	3,39	0,75	0,65	0,73	0,85	0,46	0,59	0,24	0,53	
Instruktionale Unterstützung	S	3,11	3,08	3,05	3,20	0,69	0,79	0,67	0,63				
	K	3,11	3,09	3,05	3,19	0,85	0,92	0,79	0,85				
	8	3,43	3,35	3,38	3,54	0,69	0,65	0,65	0,76	0,48	0,70	0,17	0,51
	9	2,70	2,52	2,07	2,85	0,91	0,90	0,86	0,97	0,22	0,33	0,21	0,14
	10	3,00	3,04	3,00	2,96	0,85	0,93	0,66	0,96	0,53	0,35	0,53	0,70
	11	3,06	3,09	3,00	3,08	0,69	0,79	0,59	0,69	0,48	0,53	0,52	0,42
	12	3,36	3,44	3,13	3,50	0,79	0,84	0,74	0,76	0,60	0,54	0,45	0,75
	13	2,58	2,17	2,67	2,85	0,97	0,94	0,82	1,05	0,08	-0,08	-0,26	0,44
	14	3,36	3,44	3,21	3,42	0,73	0,73	0,78	0,70	0,54	0,71	0,34	0,56
	15	3,22	3,30	3,13	3,23	0,80	0,76	0,90	0,76	0,59	0,60	0,58	0,65
16	3,23	3,52	3,00	3,19	0,83	0,85	0,83	0,75	0,58	0,65	0,53	0,67	
Emo Unterstützung	S	3,36	3,31	3,19	3,59	0,45	0,48	0,51	0,35				
	K	3,37	3,30	3,20	3,59	0,77	0,87	0,77	0,60				
	17	3,56	3,36	3,73	3,59	0,69	0,86	0,45	0,69	0,64	0,84	0,42	0,80
	18	3,35	3,00	3,19	3,82	0,83	0,96	0,85	0,40	0,58	0,61	0,69	0,46
	19	3,30	3,28	3,04	3,56	0,82	0,89	0,77	0,75	0,69	0,67	0,72	0,66
	20	3,37	3,52	3,23	3,37	0,67	0,71	0,71	0,56	0,67	0,78	0,69	0,65
21	3,37	3,40	3,12	3,59	0,67	0,87	0,52	0,50	0,59	0,67	0,50	0,46	
Klassenführung	S	2,15	2,38	1,65	2,45	0,32	0,36	0,29	0,32				
	K	2,14	2,36	1,65	2,38	0,86	0,67	0,79	0,88				
	22	1,95	2,21	1,37	2,30	0,82	0,69	0,59	0,82	0,85	0,75	0,83	0,82
	23	2,23	2,58	1,63	2,52	0,90	0,79	0,65	0,89	0,93	0,75	0,94	0,95
	24	2,13	2,38	1,63	2,41	0,87	0,79	0,71	0,89	0,88	0,73	0,92	0,85
	25	2,06	2,29	1,63	2,30	0,83	0,79	0,55	0,91	0,85	0,57	0,94	0,84
26	2,44	2,67	1,93	2,74	0,88	0,83	0,70	0,86	0,79	0,64	0,80	0,74	

Messzeitpunkt 5

Die Messung fand in der Zeit vom 03.04.17 bis zum 07.04.17 statt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,75$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,83$.

Es haben in der Klasse 5d insgesamt $N_G = 29$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,71$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 83$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,78$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,78$.

Es haben in der Klasse 5c insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,29$.

Es haben in der Klasse 5d insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,70$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 79$ Schülerinnen und Schülern die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,66$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,83$.

Es haben in der Klasse 5d insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,93$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,93$.

Es haben in der Klasse 5d insgesamt $N_G = 29$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,90$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 82$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,93$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,08	3,31	2,95	3,01	0,58	0,56	0,52	0,65				
	K	3,08	3,31	2,94	3,01	0,78	0,71	0,77	0,81				
	1	2,95	3,26	2,74	2,86	0,71	0,66	0,71	0,69	0,52	0,50	0,61	0,32
	2	3,27	3,48	3,00	3,31	0,80	0,64	0,88	0,81	0,65	0,70	0,67	0,56
	3	2,98	3,26	2,89	2,79	0,71	0,71	0,68	0,68	0,62	0,51	0,75	0,53
	4	2,95	3,15	2,89	2,83	0,71	0,66	0,64	0,80	0,41	0,43	0,57	0,21
	5	3,20	3,11	3,15	3,33	0,84	0,80	0,82	0,91	0,37	0,40	0,48	0,42
	6	2,87	3,19	2,78	2,67	0,80	0,83	0,80	0,71	0,51	0,62	0,44	0,41
7	3,34	3,70	3,04	3,28	0,75	0,47	0,76	0,84	0,50	0,16	0,57	0,51	
Instruktionale Unterstützung	S	3,06	3,15	2,88	3,15	0,76	0,77	0,82	0,69				
	K	3,06	3,14	2,88	3,15	0,87	0,88	0,87	0,82				
	8	3,33	3,39	3,21	3,40	0,80	0,70	0,83	0,87	0,39	0,64	0,18	0,33
	9	2,75	2,62	2,36	3,32	0,94	0,94	0,91	0,69	0,16	0,60	-0,12	-0,26
	10	2,98	3,19	2,54	3,24	0,89	0,90	0,88	0,72	0,24	0,35	-0,30	0,44
	11	2,98	3,12	2,82	3,00	0,70	0,71	0,72	0,65	0,52	0,55	0,31	0,67
	12	3,29	3,54	3,01	3,28	0,79	0,65	0,77	0,89	0,66	0,73	0,58	0,72
	13	2,61	2,23	2,54	3,08	0,99	0,95	1,00	0,86	-0,09	0,07	-0,34	-0,06
	14	3,34	3,58	3,18	3,28	0,73	0,58	0,77	0,79	0,43	0,26	0,39	0,65
	15	3,29	3,39	3,25	3,22	0,72	0,75	0,65	0,79	0,53	0,53	0,44	0,69
16	3,01	3,23	2,96	2,84	0,88	0,86	0,84	0,94	0,45	0,64	0,42	0,43	
Emo Unterstützung	S	3,26	3,42	3,11	3,27	0,46	0,47	0,46	0,43				
	K	3,26	3,39	3,11	3,27	0,77	0,78	0,82	0,72				
	17	3,53	3,67	3,41	3,52	0,68	0,56	0,80	0,66	0,62	0,67	0,60	0,60
	18	3,19	3,00	3,07	3,48	0,80	0,82	0,87	0,63	0,51	0,62	0,59	0,51
	19	3,13	3,40	3,00	3,05	0,83	0,82	0,85	0,77	0,62	0,74	0,59	0,51
	20	3,24	3,28	3,15	3,29	0,70	0,79	0,66	0,66	0,69	0,56	0,81	0,73
21	3,23	3,60	3,07	3,05	0,78	0,71	0,78	0,75	0,60	0,66	0,58	0,64	
Klassenführung	S	2,31	2,56	1,97	2,41	0,32	0,28	0,32	0,36				
	K	2,29	2,54	1,96	2,39	0,83	0,72	0,88	0,76				
	22	2,23	2,48	2,00	2,19	0,80	0,70	0,93	0,71	0,81	0,84	0,84	0,74
	23	2,39	2,60	2,16	2,45	0,86	0,75	0,99	0,78	0,85	0,86	0,85	0,82
	24	2,30	2,52	1,92	2,43	0,80	0,64	0,89	0,75	0,86	0,80	0,91	0,79
	25	2,20	2,44	1,85	2,29	0,78	0,75	0,78	0,70	0,78	0,81	0,72	0,77
26	2,52	2,74	2,12	2,67	0,88	0,76	0,91	0,85	0,77	0,75	0,82	0,67	

Messzeitpunkt 6

Die Messung fand in der Zeit vom 01.05.17 bis zum 05.05.17 statt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,59$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,78$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,80$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,44$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,82$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 78$ Schülerinnen und Schülern, die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,74$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 22$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,82$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,85$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,87$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 75$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,93$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,92$.

Es haben in der Klasse 5d insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,94$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 82$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,93$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,02	3,15	2,91	3,00	0,64	0,63	0,57	0,73				
	K	3,02	3,15	2,91	3,00	0,84	0,87	0,76	0,86				
	1	2,89	2,96	2,82	2,90	0,66	0,65	0,68	0,66	0,41	0,69	0,53	-0,07
	2	3,23	3,32	3,00	3,39	0,81	0,87	0,78	0,75	0,63	0,85	0,53	0,44
	3	2,85	3,00	2,67	2,89	0,71	0,73	0,73	0,65	0,52	0,81	0,53	0,04
	4	3,00	3,37	2,93	2,69	0,81	0,63	0,78	0,88	0,42	0,41	0,58	0,35
	5	3,11	3,00	3,15	3,19	0,95	0,96	0,86	1,06	0,50	0,55	0,57	0,53
	6	2,74	2,74	2,82	2,65	0,94	1,10	0,79	0,94	0,46	0,58	0,55	0,27
7	3,36	3,67	3,11	3,31	0,80	0,78	0,70	0,84	0,60	0,67	0,51	0,59	
Instruktionale Unterstützung	S	3,06	3,05	3,02	3,12	0,73	0,76	0,77	0,66				
	K	3,06	3,05	3,02	3,12	0,86	0,92	0,84	0,83				
	8	3,28	3,00	3,44	3,37	0,84	1,04	0,60	0,79	0,53	0,55	0,33	0,66
	9	2,61	2,56	2,42	2,83	0,89	0,87	0,95	0,82	0,10	0,19	-0,27	0,28
	10	3,18	3,36	3,00	3,19	0,82	0,76	0,94	0,74	0,31	0,55	0,02	0,42
	11	3,00	3,04	3,00	2,96	0,70	0,59	0,75	0,76	0,43	0,51	0,44	0,42
	12	3,37	3,48	3,27	3,37	0,74	0,71	0,67	0,84	0,66	0,73	0,60	0,67
	13	2,64	2,20	2,85	2,85	0,99	1,05	0,92	0,91	0,03	0,07	-0,25	0,20
	14	3,21	3,28	3,15	3,20	0,74	0,84	0,61	0,76	0,60	0,69	0,51	0,59
15	3,24	3,16	3,27	3,30	0,84	0,90	0,78	0,87	0,66	0,71	0,39	0,77	
16	3,01	3,04	2,96	3,04	0,89	0,93	0,87	0,90	0,64	0,74	0,43	0,71	
Emo Unterstützung	S	3,26	3,36	3,20	3,23	0,44	0,45	0,46	0,42				
	K	3,27	3,39	3,22	3,23	0,81	0,83	0,76	0,86				
	17	3,56	3,59	3,69	3,41	0,76	0,59	0,68	0,93	0,51	0,65	0,50	0,49
	18	3,20	3,05	3,15	3,37	0,74	0,72	0,61	0,84	0,55	0,46	0,65	0,64
	19	3,13	3,46	3,04	2,96	0,84	0,74	0,77	0,94	0,81	0,70	0,85	0,86
	20	3,25	3,36	3,19	3,20	0,74	0,79	0,63	0,81	0,70	0,78	0,58	0,75
21	3,28	3,50	3,15	3,22	0,80	0,67	0,83	0,85	0,69	0,53	0,74	0,75	
Klassenführung	S	2,10	2,21	1,87	2,21	0,32	0,25	0,36	0,33				
	K	2,08	2,19	1,85	2,18	0,80	0,68	0,83	0,83				
	22	1,97	2,17	1,78	1,96	0,81	0,72	0,85	0,84	0,80	0,76	0,80	0,82
	23	2,18	2,30	1,91	2,32	0,81	0,67	0,86	0,86	0,89	0,86	0,89	0,90
	24	2,09	2,22	1,82	2,23	0,77	0,64	0,79	0,83	0,88	0,84	0,91	0,87
	25	1,99	2,11	1,78	2,07	0,78	0,70	0,85	0,77	0,86	0,87	0,85	0,85
26	2,29	2,26	2,15	2,45	0,81	0,71	0,82	0,88	0,67	0,71	0,53	0,78	

Messzeitpunkt 7

Die Messung fand in der Zeit vom 05.06.17 bis zum 09.06.17 statt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,83$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,74$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,63$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,74$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,68$.

Es haben in der Klasse 5c insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,79$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 79$ Schülerinnen und Schüler die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,76$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,85$.

Es haben in der Klasse 5c insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,77$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,79$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 81$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,80$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,92$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,92$.

Es haben in der Klasse 5d insgesamt $N_G = 29$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,91$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 85$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,91$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,09	3,16	3,05	3,05	0,64	0,63	0,58	0,70				
	K	3,09	3,16	3,05	3,06	0,79	0,81	0,73	0,82				
	1	2,93	2,89	2,89	3,00	0,71	0,75	0,65	0,73	0,46	0,63	0,35	0,39
	2	3,26	3,22	3,23	3,33	0,76	0,80	0,82	0,68	0,53	0,69	0,57	0,31
	3	3,06	3,19	2,89	3,11	0,66	0,62	0,59	0,75	0,50	0,64	0,36	0,49
	4	2,98	3,22	2,98	2,74	0,82	0,80	0,64	0,94	0,40	0,54	0,33	0,37
	5	3,23	3,19	3,35	3,17	0,87	0,79	0,80	1,02	0,49	0,48	0,66	0,46
	6	2,81	2,78	2,77	2,87	0,81	0,97	0,76	0,67	0,29	0,50	0,32	-0,04
7	3,48	3,67	3,35	3,43	0,63	0,68	0,63	0,57	0,59	0,68	0,58	0,48	
Instruktionale Unterstützung	S	3,05	3,09	2,96	3,09	0,73	0,79	0,71	0,69				
	K	3,05	3,09	2,96	3,09	0,87	0,91	0,84	0,87				
	8	3,38	3,19	3,46	3,50	0,81	1,00	0,51	0,81	0,37	0,40	0,50	0,40
	9	2,77	2,85	2,54	2,92	0,86	0,77	0,95	0,84	0,35	0,02	0,45	0,49
	10	3,01	3,07	2,92	3,04	0,76	0,78	0,80	0,72	0,39	0,15	0,58	0,41
	11	2,93	3,02	2,92	2,85	0,71	0,66	0,74	0,73	0,41	0,48	0,54	0,26
	12	3,33	3,52	3,23	3,23	0,71	0,80	0,59	0,71	0,57	0,59	0,43	0,70
	13	2,38	2,07	2,42	2,65	0,95	0,87	0,95	0,98	0,09	-0,05	0,14	0,23
	14	3,28	3,44	3,08	3,31	0,80	0,75	0,85	0,79	0,69	0,60	0,61	0,83
15	3,32	3,52	3,15	3,27	0,74	0,70	0,67	0,83	0,70	0,71	0,67	0,71	
16	3,07	3,33	2,92	2,94	0,90	0,83	0,89	0,96	0,59	0,54	0,57	0,67	
Emo Unterstützung	S	3,25	3,32	3,11	3,32	0,52	0,53	0,52	0,53				
	K	3,25	3,35	3,11	3,32	0,79	0,94	0,74	0,74				
	17	3,48	3,44	3,54	3,44	0,76	0,85	0,64	0,80	0,75	0,92	0,54	0,83
	18	3,18	2,93	3,03	3,60	0,79	0,92	0,70	0,57	0,37	0,51	0,45	0,28
	19	3,14	3,41	2,93	3,08	0,86	0,93	0,72	0,89	0,74	0,85	0,62	0,75
	20	3,25	3,48	3,07	3,19	0,78	0,80	0,81	0,69	0,49	0,44	0,53	0,56
21	3,21	3,41	3,00	3,23	0,74	0,84	0,72	0,59	0,57	0,59	0,61	0,47	
Klassenführung	S	2,10	2,27	1,88	2,17	0,35	0,29	0,38	0,38				
	K	2,09	2,26	1,88	2,13	0,86	0,74	0,95	0,83				
	22	2,01	2,36	1,75	1,93	0,88	0,73	1,00	0,80	0,84	0,79	0,89	0,83
	23	2,15	2,32	1,86	2,28	0,91	0,77	0,93	0,96	0,86	0,77	0,93	0,85
	24	1,98	2,14	1,68	2,12	0,80	0,71	0,82	0,80	0,87	0,92	0,85	0,86
	25	2,01	2,21	1,86	1,97	0,87	0,79	1,00	0,78	0,68	0,70	0,66	0,68
26	2,37	2,30	2,25	2,53	0,80	0,71	0,89	0,78	0,66	0,78	0,64	0,66	

Messzeitpunkt 8

Die Messung fand in der Zeit vom 03.07.17 bis zum 07.07.17 statt. Das Selbstkonzept wurde in dieser Messung auch mit abgefragt.

Kognitive Aktivierung

Es haben in der Klasse 5b insgesamt $N_G = 28$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,75$.

Es haben in der Klasse 5c insgesamt $N_G = 21$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Es haben in der Klasse 5d insgesamt $N_G = 29$ Schülerinnen und Schüler alle der 7 Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,73$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 78$ Schülerinnen und Schülern, die alle der Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 7 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,77$.

Instruktionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,79$.

Es haben in der Klasse 5c insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,57$.

Es haben in der Klasse 5d insgesamt $N_G = 26$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 9 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,73$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 78$ Schülerinnen und Schülern die alle der 9 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,72$.

Emotionale Unterstützung

Es haben in der Klasse 5b insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,85$.

Es haben in der Klasse 5c insgesamt $N_G = 24$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,82$.

Es haben in der Klasse 5d insgesamt $N_G = 27$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 78$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Klassenführung

Es haben in der Klasse 5b insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,94$. Die interne Konsistenz ist damit als gut zu bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 24$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,94$.

Es haben in der Klasse 5d insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 5 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,91$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 80$ Schülerinnen und Schülern, die alle der 5 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,94$.

Mathematikbezogenes Selbstkonzept

Es haben in der Klasse 5b insgesamt $N_G = 28$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 4 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,92$. Die interne Konsistenz ist damit als gut zu

bewerten.

Es haben in der Klasse 5c insgesamt $N_G = 25$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 4 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,84$.

Es haben in der Klasse 5d insgesamt $N_G = 29$ Schülerinnen und Schüler alle der Fragen beantwortet. Die interne Konsistenz dieser Skala für alle diejenigen, die alle 4 Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,81$.

Damit ergibt sich ein Gesamtstichprobenumfang aus den drei untersuchten Testklassen von insgesamt $N_G = 82$ Schülerinnen und Schülern, die alle der 4 Fragen beantwortet haben. Die interne Konsistenz dieser Skala für alle diejenigen, die alle Fragen beantwortet haben, ergibt ein Cronbachs α mit $\alpha = 0,86$.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Kognitive Aktivierung	S	3,09	3,23	3,07	2,97	0,61	0,58	0,58	0,67				
	K	3,09	3,23	3,06	2,97	0,80	0,75	0,79	0,84				
	1	2,93	3,07	2,86	2,85	0,70	0,60	0,73	0,77	0,60	0,69	0,71	0,42
	2	3,26	3,43	3,00	3,29	0,76	0,63	0,95	0,70	0,60	0,63	0,60	0,60
	3	2,96	3,18	3,00	2,72	0,76	0,67	0,77	0,80	0,62	0,60	0,70	0,55
	4	2,90	3,14	3,19	2,47	0,79	0,71	0,68	0,78	0,39	0,37	0,55	0,36
	5	3,21	3,14	3,05	3,38	0,93	0,97	0,80	0,98	0,44	0,35	0,58	0,57
	6	2,90	3,00	3,76	2,91	0,79	0,86	0,83	0,71	0,42	0,69	0,52	0,10
Instruktionale Unterstützung	S	3,07	3,02	3,06	3,13	0,72	0,77	0,71	0,67				
	K	3,07	3,02	3,06	3,13	0,85	0,91	0,81	0,82				
	8	3,46	3,37	3,56	3,44	0,71	0,69	0,51	0,90	0,47	0,65	0,37	0,40
	9	2,67	2,52	2,52	2,96	0,88	0,89	0,92	0,77	0,24	0,33	0,15	0,19
	10	2,99	2,93	2,92	3,12	0,76	0,87	0,81	0,59	0,32	0,34	0,39	0,24
	11	3,03	2,96	3,08	3,06	0,74	0,85	0,64	0,73	0,43	0,34	0,51	0,49
	12	3,45	3,59	3,24	3,50	0,68	0,57	0,66	0,76	0,53	0,65	0,45	0,58
	13	2,51	2,11	2,52	2,92	1,00	0,93	1,00	0,94	0,10	0,52	-0,35	0,06
	14	3,32	3,44	3,28	3,23	0,76	0,58	0,68	0,99	0,41	0,40	0,48	0,48
	15	3,19	3,19	3,24	3,15	0,70	0,79	0,60	0,73	0,65	0,69	0,39	0,79
Emo Unterstützung	S	3,29	3,33	3,23	3,30	0,46	0,49	0,42	0,47				
	K	3,29	3,35	3,23	3,30	0,76	0,92	0,69	0,78				
	17	3,46	3,48	3,50	3,41	0,73	0,75	0,59	0,84	0,68	0,82	0,58	0,63
	18	3,21	3,07	3,13	3,41	0,74	0,78	0,68	0,75	0,55	0,60	0,51	0,58
	19	3,17	3,26	3,17	3,07	0,83	0,90	0,76	0,83	0,70	0,78	0,68	0,66
	20	3,30	3,37	3,29	3,22	0,71	0,79	0,55	0,75	0,54	0,41	0,56	0,70
	21	3,31	3,37	3,21	3,33	0,73	0,74	0,66	0,78	0,72	0,75	0,73	0,68
Klassenführung	S	2,10	2,33	1,75	2,18	0,28	0,28	0,23	0,33				
	K	2,10	2,33	1,72	2,19	0,78	0,68	0,74	0,80				
	22	1,98	2,29	1,58	2,00	0,73	0,66	0,58	0,77	0,74	0,81	0,64	0,66
	23	2,09	2,29	1,79	2,14	0,73	0,60	0,72	0,80	0,87	0,85	0,89	0,84
	24	2,01	2,14	1,58	2,18	0,79	0,69	0,65	0,86	0,87	0,89	0,90	0,82
	25	2,08	2,32	1,63	2,21	0,78	0,67	0,71	0,79	0,85	0,91	0,94	0,70
	26	2,29	2,54	1,92	2,38	0,86	0,79	0,88	0,82	0,82	0,76	0,84	0,82

Wie zum ersten Messzeitpunkt wurde auch im 8 Messzeitpunkt das Selbstkonzept getestet. Entsprechend hier die Itemanalyse.

Item		M	M _{5b}	M _{5c}	M _{5d}	SD	SD _{5b}	SD _{5c}	SD _{5d}	r _{it}	r _{it5b}	r _{it5c}	r _{it5d}
Selbstkonzept	S	2,88	2,89	2,86	2,90	0,46	0,36	0,46	0,55				
	K	2,88	2,89	2,86	2,90	0,91	0,98	0,86	0,90				
	27	2,87	2,93	2,72	2,93	0,97	0,94	1,02	0,96	0,70	0,89	0,65	0,55
	28	2,90	2,86	2,80	3,02	0,88	0,93	0,87	0,87	0,70	0,83	0,61	0,63
	29	2,72	2,71	2,88	2,59	0,96	1,08	0,83	0,95	0,71	0,75	0,67	0,74
	30	3,05	3,05	3,04	3,05	0,83	0,97	0,73	0,78	0,73	0,81	0,78	0,59

Übersicht interne Konsistenz der Skalen

		1	2	3	4	5	6	7	8
Kognitive Aktivierung	α	0,69	0,52	0,63	0,77	0,78	0,78	0,74	0,77
	α_{5b}	0,81	0,56	0,64	0,87	0,75	0,86	0,83	0,75
	α_{5c}	0,52	0,55	0,71	0,70	0,83	0,81	0,74	0,84
	α_{5d}	0,46	0,22	0,49	0,72	0,71	0,59	0,63	0,N
	N_G	79	77	80	77	78	80	80	78
	N_{5b}	27	24	26	24	27	27	27	28
	N_{5c}	25	28	26	27	27	27	26	21
	N_{5d}	27	25	28	26	29	26	27	29
Instrukt Unterstützung	α	0,61	0,64	0,75	0,76	0,66	0,74	0,76	0,72
	α_{5b}	0,70	0,48	0,76	0,77	0,78	0,80	0,68	0,79
	α_{5c}	0,51	0,54	0,58	0,64	0,29	0,44	0,79	0,57
	α_{5d}	0,56	0,74	0,83	0,82	0,70	0,82	0,81	0,73
	N_G	76	75	77	73	79	78	79	78
	N_{5b}	25	24	23	23	26	25	27	27
	N_{5c}	25	27	26	24	28	26	26	25
	N_{5d}	26	24	28	26	25	27	26	26
Emotionale Unterstützung	α	0,52	0,80	0,82	0,83	0,81	0,84	0,80	0,84
	α_{5b}	0,41	0,63	0,83	0,88	0,84	0,82	0,85	0,85
	α_{5c}	0,59	0,86	0,89	0,81	0,83	0,85	0,77	0,82
	α_{5d}	0,46	0,67	0,69	0,81	0,81	0,87	0,79	0,84
	N_G	79	73	76	78	80	75	81	78
	N_{5b}	27	25	26	25	25	22	27	24
	N_{5c}	27	27	25	26	27	26	28	27
	N_{5d}	25	21	25	27	28	27	26	27
Klassenführung	α	0,86	0,91	0,90	0,95	0,93	0,93	0,91	0,94
	α_{5b}	0,84	0,86	0,88	0,86	0,93	0,93	0,92	0,94
	α_{5c}	0,84	0,86	0,90	0,96	0,93	0,92	0,92	0,94
	α_{5d}	0,73	0,78	0,80	0,94	0,90	0,94	0,91	0,91
	N_G	79	79	80	78	82	82	85	80
	N_{5b}	26	26	26	24	27	27	28	28
	N_{5c}	26	28	26	27	26	27	28	24
	N_{5d}	27	25	28	27	29	28	29	28

Fragebogennummer:

«Kodierung»

Deckblatt

Bitte das Deckblatt vor der Abgabe abreißen!

«Vorname» «Nachname»

In diesem Fragebogen geht es um deine Einschätzung zum Mathematikunterricht seit Beginn des Schuljahres. Bei den Antworten gibt es kein „richtig“ und kein „falsch“.

Bitte kreuze die zutreffende Antwort an!

Fragebogennummer:

«Kodierung»

Fragebogennummer:

«Kodierung»

		trifft gar nicht zu	trifft eher nicht zu	trifft eher zu	trifft völlig zu
1	Wir bearbeiten im Mathematikunterricht Aufgaben, über die ich sehr gründlich nachdenken muss.				
2	Unsere Mathematiklehrerin/ unser Mathematiklehrer fragt mich, was ich verstanden habe und was nicht.				
3	Unsere Mathematiklehrerin/ unser Mathematiklehrer stellt Fragen, über die ich sehr gründlich nachdenken muss.				
4	Unsere Mathematiklehrerin/ unser Mathematiklehrer stellt uns Aufgaben, die auf den ersten Blick schwierig wirken.				
5	Unsere Mathematiklehrerin/ unser Mathematiklehrer fragt uns bei einem neuen Thema, was wir schon wissen.				
6	Unsere Mathematiklehrerin/ unser Mathematiklehrer stellt uns Aufgaben, über die ich gerne nachdenke.				
7	Unsere Mathematiklehrerin/ unser Mathematiklehrer möchte, dass ich meine Antworten erkläre.				
8	Unsere Mathematiklehrerin/ unser Mathematiklehrer weist darauf hin, was wir uns merken sollen.				
9	Im Mathematikunterricht geht es oft um zu viele Fragen gleichzeitig.				
10	Im Mathematikunterricht weiß ich oft nicht, worüber wir gerade sprechen.				
11	Beim Arbeiten im Mathematikunterricht weiß ich genau, was ich tun soll.				
12	Unsere Mathematiklehrerin/ unser Mathematiklehrer interessiert sich für unsere Erklärungen.				
13	Unsere Mathematiklehrerin/ unser Mathematiklehrer erklärt oft mit Fremdwörtern, die wir nicht verstehen.				
14	Unsere Mathematiklehrerin/ unser Mathematiklehrer hilft mir nur so viel, dass ich auch alleine eine Erklärung finden kann.				
15	Wenn ich einen Fehler mache, sagt unsere Mathematiklehrerin/ unser Mathematiklehrer mir, wie ich es besser machen kann.				
16	Unsere Mathematiklehrerin/ unser Mathematiklehrer sagt mir, worin ich schon gut bin und was ich noch lernen muss.				

Fragebogennummer:

«Kodierung»

Fragebogennummer:

«Kodierung»

		trifft gar nicht zu	trifft eher nicht zu	trifft eher zu	trifft völlig zu
17	Unsere Mathematiklehrerin/ unser Mathematiklehrer ist auch dann nett zu mir, wenn ich einen Fehler mache.				
18	Ich bin unserer Mathematiklehrerin/ unserem Mathematiklehrer wichtig.				
19	Unsere Mathematiklehrerin/ unser Mathematiklehrer ermutigt mich, wenn ich eine Aufgabe schwierig finde.				
20	Unsere Mathematiklehrerin/ unser Mathematiklehrer lobt mich, wenn ich etwas gut gemacht habe.				
21	Unsere Mathematiklehrerin/ unser Mathematiklehrer glaubt, dass ich schwierige Aufgaben lösen kann.				
22	Im Mathematikunterricht stört keiner den Unterricht.				
23	Im Mathematikunterricht sind die Schüler leise, wenn die Lehrerin/ der Lehrer spricht.				
24	Im Mathematikunterricht hören alle zu und sind leise.				
25	Im Mathematikunterricht stört niemand mit Reden.				
26	Im Mathematikunterricht folgen alle der Lehrerin/ dem Lehrer.				
27	Für Mathematik habe ich einfach keine Begabung.				
28	Bei manchen Sachen im Fach Mathematik weiß ich von vornherein: "Das verstehe ich nie."				
29	Obwohl ich mir bestimmt Mühe gebe, fällt mir das Fach Mathematik schwerer als vielen meiner Mitschüler und Mitschülerinnen.				
30	Was wir in Mathematik durchnehmen kann ich mir schlecht merken.				

Fragebogennummer:

«Kodierung»