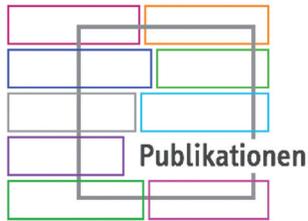


IfE



Erziehungswissenschaftliche
Studien
Band 12

Katharina Dreiling

Feedback aus der Sicht von Schülerinnen
und Schülern im Deutschunterricht der
gymnasialen Oberstufe

Längsschnittliche Analysen zur differenziellen
Wahrnehmung und motivationalen Wirkung von Feedback

Feedback
Motivationsentwicklung
Schülerwahrnehmung
Latent-Change-Analysen
Messinvarianz
Deutschunterricht
Sekundarstufe II

Universitätsverlag Göttingen

Katharina Dreiling
Feedback aus der Sicht von Schülerinnen und Schülern
im Deutschunterricht der gymnasialen Oberstufe

Dieses Werk ist lizenziert unter einer
[Creative Commons
Namensnennung - Weitergabe unter gleichen Bedingungen
4.0 International Lizenz.](https://creativecommons.org/licenses/by-sa/4.0/)



erschieden als Band 12 der Reihe „Erziehungswissenschaftliche Studien“
im Universitätsverlag Göttingen 2023

Katharina Dreiling

Feedback aus der Sicht von
Schülerinnen und Schülern
im Deutschunterricht der
gymnasialen Oberstufe

Längsschnittliche Analysen
zur differenziellen Wahrnehmung
und motivationalen Wirkung
von Feedback

Erziehungswissenschaftliche
Studien

Band 12

Universitätsverlag Göttingen
2023

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://dnb.dnb.de> abrufbar.

Herausgeber:innen der Reihe „Erziehungswissenschaftliche Studien“:

Prof. Dr. Ariane Willems (Schriftleitung), Prof. Dr. Klaus-Peter Horn,
Prof. Dr. Katharina Kunze, Prof. Dr. Jessica Löser, Prof. Dr. Kerstin Rabenstein,
Prof. Dr. Hermann Veith

Georg-August-Universität Göttingen
Institut für Erziehungswissenschaft
Waldweg 26
37073 Göttingen
(<https://www.uni-goettingen.de/ife>)

Die vorliegende Arbeit wurde von der Sozialwissenschaftlichen Fakultät der Georg-August-Universität Göttingen unter demselben Titel als Dissertation angenommen.

Dieses Werk ist auch als freie Onlineversion über die Verlagswebsite sowie über den Göttinger Universitätskatalog (GUK) bei der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (<https://www.sub.uni-goettingen.de>) zugänglich. Es gelten die Lizenzbestimmungen der Onlineversion.

Satz und Layout: Katharina Dreiling
Umschlaggestaltung: Margo Bargheer



© 2023 Universitätsverlag Göttingen, Göttingen
<https://univerlag.uni-goettingen.de>
ISBN: 978-3-86395-615-8
DOI: <https://doi.org/10.17875/gup2023-2481>
ISSN: 2199-5133
eISSN: 2512-6024

Zusammenfassung

Das lern- und leistungsbezogene Feedback, das Schülerinnen und Schüler im Unterricht von ihrer Lehrkraft erhalten, stellt ein wichtiges Merkmal eines guten Unterrichts und eines qualitätvollen Unterrichtsgesprächs dar. Empirische Forschungsbefunde zeigen, dass sich das Feedback der Lehrkraft nicht zwangsläufig – oder bei allen gleichermaßen – förderlich auf den Lernprozess auswirkt, sondern erst dadurch, dass Lernende in Abhängigkeit ihrer individuellen Lernvoraussetzungen das Feedback subjektiv wahrnehmen und zur Verbesserung ihres Lernprozesses nutzen. Angesichts dieser differenziellen Wirkung von Feedback rücken neuere Untersuchungen vermehrt die Schülerinnen und Schüler und ihre subjektive Wahrnehmung des im Unterricht erteilten Feedbacks ins Zentrum des Interesses. An diese Forschungslage knüpft die vorliegende Studie an, deren übergreifendes Ziel es ist, die Wahrnehmung von Feedback aus der Sicht von Schülerinnen und Schülern im Deutschunterricht der gymnasialen Oberstufe differenziert zu analysieren. Ausgehend von den Annahmen theoretischer Angebots-Nutzungs-Modelle steht die Frage im Mittelpunkt, inwiefern Feedback als Qualitätsmerkmal von Unterricht in Abhängigkeit der individuellen Merkmale von Schülerinnen und Schülern differenziell wahrgenommen wird und wie die wahrgenommene Feedbackqualität mit der Motivationsentwicklung zusammenhängt.

Zur Beantwortung der Fragestellungen wurden im Rahmen der Längsschnittstudie FeeHe 807 Schülerinnen und Schüler in 49 Deutschkursen der gymnasialen

Oberstufe standardisiert zu ihrer Wahrnehmung verschiedener Dimensionen der Feedbackqualität sowie zu ihrer Lernmotivation befragt. Die Befunde der Latent-True-Change-Analysen zeigen, dass sich die Schülerinnen und Schüler im Deutschunterricht der gymnasialen Oberstufe nicht nur bedeutsam in der Wahrnehmung von Feedback zu Beginn des Schuljahres, sondern auch in ihrer Wahrnehmungsveränderung im Verlauf des Schulhalbjahres unterscheiden. Dabei erweist sich die Vorleistung als prädiktiv für interindividuelle Unterschiede in der Wahrnehmung der Feedbackqualität. Die Befunde aus bivariaten Latent-True-Change-Modellen belegen positive Korrelationen zwischen der wahrgenommenen Feedbackqualität und der Ausgangsmotivation der Schülerinnen und Schüler zu Beginn des Schuljahres. Ebenso stehen die Veränderungen der Feedbackwahrnehmung und der intrinsischen Motivation im Laufe des Schulhalbjahres in einem positiven Zusammenhang. Mediationsanalysen weisen nach, dass sowohl die wahrgenommene Nützlichkeit von Feedback als auch das subjektive Kompetenzerleben, die als Merkmale der Nutzung von Feedback konzeptualisiert werden, den Zusammenhang zwischen Dimensionen der wahrgenommenen Feedbackqualität und der Motivationsveränderung der Schülerinnen und Schüler erklären können.

Insgesamt erweitert die vorliegende Untersuchung den bestehenden Forschungsstand zu interindividuellen Unterschieden in der Wahrnehmung von Feedback und der Entstehung differenzieller Wirkungen auf die Motivation, womit auch ein Beitrag zur Überprüfung zentraler Annahmen von Angebots-Nutzungs-Modellen geleistet werden soll.

Abstract

Feedback that students receive from their teacher during class is considered a key factor of effective teaching and classroom discourse. However, empirical findings show that the impact of teacher feedback on the learning process is not necessarily positive, indicating that learners perceive and use feedback differently depending on their individual learning characteristics. Against this background, students' individual perception of feedback in classroom has been the focus of much recent research. Taking their findings into account, the present study aims to investigate the perception of feedback from the perspective of students in German language classes in upper secondary schools. Based on theoretical assumptions of supply-use models, the focus of this study is on the extent to which feedback quality is perceived differently depending on the individual characteristics of the students and in how far the perceived feedback quality is related to the development of motivation.

The analyses are based on data from the longitudinal study "FeeHe", in which 807 students from 49 German language classes in upper secondary school were surveyed about their perception of different dimensions of feedback quality as well as their learning motivation. Results of latent change models show that students in German lessons in upper secondary school not only differ significantly in their perception of feedback at the beginning of the school year, but also in their changes of perception over the course of a school semester. Students' individual performance predicts interindividual changes in the perception of feedback quality. Findings

from bivariate latent change models show that the level and change in perceived feedback quality and in students' motivation are positively related. Further analyses indicate that both the perceived usefulness of feedback and the perceived competence support mediate the link between dimensions of perceived feedback quality and change in students' motivation.

Overall, the present study expands the current state of research on interindividual differences in the perception of feedback and on differences in the effects of feedback on motivation.

Inhaltsverzeichnis

1	Einleitung.....	13
1.1	Ausgangslage.....	13
1.2	Ziele der Arbeit.....	15
1.3	Aufbau der Arbeit.....	16
2	Feedback als Unterrichtsmerkmal.....	19
2.1	Definition von Feedback.....	20
2.2	Dimensionen von Feedback.....	21
2.3	Empirische Befunde zu Wirkungen von Feedback.....	27
2.4	Empirische Befunde zur Häufigkeit von Feedback im Unterricht.....	30
2.5	Die Rolle der Wahrnehmung von Feedback.....	32
2.5.1	Theoretische Modelle zur Wahrnehmung von Feedback.....	32

2.5.2	Zur standardisierten Erfassung der Wahrnehmung von Feedback	36
2.5.3	Empirische Befunde zur differenziellen Feedbackwahrnehmung	41
2.5.4	Veränderung der Wahrnehmung – theoretische Annahmen	44
2.5.5	Empirische Befunde zur (differenziellen) Veränderung der Wahrnehmung	46
3	Feedback und Motivation	49
3.1	Konzeption von Lernmotivation und Bedeutung für das Lernen	50
3.2	Die Selbstbestimmungstheorie	51
3.3	Empirische Befunde zur Entwicklung der Lernmotivation	54
3.4	Forschungsüberblick zum Zusammenhang zwischen Feedback und Motivation	57
3.4.1	Die Bedeutung von Feedback für die Entwicklung der Lernmotivation	57
3.4.2	Reziproke Effekte von Feedback und Lernmotivation	59
4	Zusammenfassung und Rahmenmodell	63
5	Forschungsfragen und Hypothesen	69
5.1	Differenzielle Wahrnehmung von Feedback	71
5.2	Zusammenhang zwischen Feedback und Lernmotivation	74
5.3	Indirekte Wirkungen von Feedback auf die Lernmotivation	75
6	Methodisches Vorgehen	77
6.1	Die Messwiederholungsstudie FeeHe	77
6.2	Beschreibung der Stichprobe	79
6.3	Erhebungsinstrumente	81
6.3.1	Qualität von Feedback	81
6.3.2	Wahrgenommene Nützlichkeit von Feedback	83

6.3.3	Wahrgenommenes Kompetenzerleben.....	83
6.3.4	Intrinsische Lernmotivation.....	84
6.3.5	Hintergrundvariablen.....	85
6.4	Hierarchische Struktur der Daten.....	86
6.5	Umgang mit fehlenden Werten.....	86
6.6	Statistische Analysemethoden.....	90
6.6.1	Konfirmatorische Faktorenanalyse.....	90
6.6.2	Messinvarianzanalysen.....	91
6.6.3	Modellierung und Analyse von Veränderung über die Zeit.....	93
6.6.4	Mediationsmodelle.....	97
6.6.5	Parameterschätzung.....	99
6.6.6	Modellgütekriterien.....	100
6.6.7	Modellvergleiche.....	103
7	Ergebnisse.....	105
7.1	Wahrnehmung von Feedback im Deutschunterricht.....	106
7.1.1	Faktorenstruktur der Skalen zur Feedbackwahrnehmung.....	106
7.1.2	Deskriptive Ergebnisse.....	109
7.1.3	Modellierung der Feedbackqualität im Längsschnitt.....	110
7.2	Messinvarianz der Feedbackqualität.....	116
7.2.1	Messinvarianz über die Zeit.....	116
7.2.2	Messinvarianz über die Geschlechtergruppen.....	118
7.3	Differenzielle Veränderung der Feedbackwahrnehmung.....	120
7.3.1	Deskriptive Entwicklungen.....	120
7.3.2	Veränderung im Latent-True-Change-Modell.....	123
7.3.3	Effekte auf die Veränderung der Feedbackwahrnehmung.....	127
7.4	Zusammenhänge zwischen Feedback und Lernmotivation.....	129
7.4.1	Faktorenstruktur und deskriptive Statistik der Skala Lernmotivation.....	129
7.4.2	Modellierung der Lernmotivation im Längsschnitt.....	131

7.4.3	Messinvarianz der Lernmotivation über die Zeit.....	134
7.4.4	Messinvarianz der Lernmotivation über die Geschlechtergruppen.....	135
7.4.5	Differenzielle Veränderung der Lernmotivation	137
7.4.6	Bivariate Latent-True-Change-Modelle.....	141
7.5	Indirekte Wirkungen von Feedback auf die Lernmotivation	145
7.5.1	Faktorenstruktur und deskriptive Statistik der Skala Nützlichkeit.....	145
7.5.2	Faktorenstruktur und deskriptive Statistik der Skala Kompetenzerleben.....	146
7.5.3	Mediationsanalysen.....	148
8	Zusammenfassung und Diskussion der Ergebnisse	155
8.1	Zusammenfassung der zentralen Befunde.....	155
8.1.1	Differenzielle Wahrnehmung von Feedback.....	155
8.1.2	Zusammenhänge zwischen Feedback und Lernmotivation.....	161
8.1.3	Indirekte Effekte von Feedback auf die Lernmotivation.....	165
8.2	Grenzen der Studie.....	168
8.3	Implikationen für die Forschung.....	170
8.4	Implikationen für die Praxis.....	174
9	Verzeichnisse	177
9.1	Literaturverzeichnis	177
9.2	Abbildungsverzeichnis	207
9.3	Tabellenverzeichnis	209

1 Einleitung

1.1 Ausgangslage

Die Qualität schulischer Lernprozesse hängt in hohem Maße von individuellen Merkmalen der einzelnen Schülerinnen und Schüler ab (Buff et al., 2010; Helmke, 2015; Seidel et al., 2016). In Wechselwirkung mit Merkmalen der Lernsituation beeinflussen diese die Wahrnehmung des Unterrichtsgeschehens und vermittelt darüber den Lernerfolg (Buff et al., 2010). Ausgehend von diesen theoretischen Annahmen werden in empirischen Studien der Unterrichtsforschung zunehmend *differenzielle* Analyseansätze genutzt, indem individuelle Wahrnehmungsprozesse und Lernvoraussetzungen von Schülerinnen und Schülern bei der Beschreibung der Unterrichtsqualität und -wirkungen stärker als bisher einbezogen werden (Iglar et al., 2019; Jurik et al., 2015; Stang & McElvany, 2020; Willems, 2022a; Willems & Dreiling, 2022). Dabei wird angenommen, dass aufgrund der unterschiedlichen Voraussetzungen von Schülerinnen und Schülern sogenannte „Mikro-Umwelten“ in Klassenzimmern entstehen (Seidel et al. 2016, S. 65; vgl. auch Jurik et al., 2015; Seidel, 2006; Willems, 2022a; Willems & Dreiling, 2022), die zu differenziellen Wahrnehmungsprozessen zwischen den Schülerinnen und Schülern innerhalb des gleichen Unterrichts führen. Das Zusammenspiel des Unterrichts und den Merkmalen der Schülerinnen und Schüler wird auch in theoretischen Angebots-Nutzungs-Modellen veranschaulicht (z.B. Helmke, 2015; Seidel, 2014; Vieluf et al., 2020). Der Kern-

gedanke des Modells ist, dass der schulische Lernerfolg nicht nur von der Gestaltung des Lernangebots, sondern im Wesentlichen auch von dessen Wahrnehmung und Nutzung durch die Schülerinnen und Schüler abhängt. Im Rahmen dieses Forschungsansatzes werden Unterrichtsmerkmale daher häufig aus der Perspektive von Schülerinnen und Schülern erfasst. Deren Einschätzungen haben sich als prädiktiv für die Vorhersage kognitiver und motivational-affektiver Lernergebnisse erwiesen (Clausen, 2002; Fauth et al., 2014; Kunter et al., 2007; Stahns et al., 2020; Wagner et al., 2016).

Ein Merkmal des Unterrichtsangebots, das aufgrund seiner lernförderlichen Wirkungen immer wieder im Fokus der Forschung steht, ist das Feedback der Lehrkraft (Hattie, 2009; Hattie & Timperley, 2007; Lipnevich & Smith, 2018). Im Kontext des Schulunterrichts wird Feedback häufig im Anschluss an die Aufgabenbearbeitung oder eine Antwort bereitgestellt. Es informiert Lernende über die Qualität ihrer Aufgabenbearbeitung bzw. ihrer Antwort (Lipowsky, 2020). Die Bedeutsamkeit von Feedback für den allgemeinen Lernerfolg konnte anhand von Meta-Analysen und Reviews demonstriert werden (Hattie, 2009; Hattie & Zierer, 2019; Kluger & DeNisi, 1996; Wisniewski et al., 2020). In ihrem Review berichten Hattie und Timperley (2007), dass Feedback zu den lernwirksamsten Faktoren im Unterricht gehört. Die Ergebnisse einer kürzlich durchgeführten Meta-Analyse (Wisniewski et al., 2020) deuten auf einen mittleren Effekt von Feedback auf das Lernen der Schülerinnen und Schüler hin. Allerdings wird die Wirkung von Feedback wesentlich davon beeinflusst, welche Form von Feedback vermittelt wird (Wisniewski et al., 2020). Im Hinblick auf die Gestaltung gilt Feedback dann als besonders förderlich, wenn es Informationen über das *Ziel* oder das *Ergebnis* einer Aufgabenbearbeitung, Hinweise zum *Prozess* der Aufgabenbearbeitung sowie Impulse zur *Selbstregulation* enthält (Hattie & Timperley, 2007). Im Sinne der Förderung einer ko-konstruktiven Wissenskonstruktion (Lipowsky, 2020; Pauli & Reusser, 2018) sollte Feedback im Unterricht auch zum gegenseitigen Austausch bzw. *Dialog* zwischen den Schülerinnen und Schülern anregen (Willems & Dreiling, 2022).

Vor dem Hintergrund der differenziellen Effekte von Feedback richtet die Forschung zunehmend ihren Blick auf Merkmale von Schülerinnen und Schülern, die die Wirksamkeit von Feedback beeinflussen. Auf Basis von Angebots-Nutzungs-Modellen wird argumentiert, dass – neben der Qualität der Gestaltung von Feedback – die individuelle Wahrnehmung des Feedbacks durch die Schülerinnen und Schüler beeinflusst, ob sich das Feedback positiv auf den individuellen Lernprozess auswirkt (Willems & Dreiling, 2022). Beispielsweise konnte Hoya (2019) zeigen, dass das durch Grundschülerinnen und -schüler wahrgenommene Feedback der Lehrkraft zur Erklärung des Leseverständnisses beitragen kann. Allerdings wirkt sich Feedback nicht auf alle Schülerinnen und Schüler gleichermaßen positiv aus, da diese das Feedback aufgrund ihres Geschlechts und ihrer Leistung (Hoya, 2021; Sortkær, 2019) sowie ihren motivationalen Lernvoraussetzungen (Rakoczy et al., 2013; Willems & Dreiling, 2022) unterschiedlich wahrnehmen und verarbeiten. Die meisten der genannten Befunde beziehen sich auf die Primar- oder Sekundarstufe I.

Untersuchungen in der gymnasialen Oberstufe stellen damit eine weiterhin bestehende Forschungslücke dar (vgl. Dreiling et al., 2018; Willems & Dreiling, 2022). Ein zentrales Anliegen der vorliegenden Arbeit besteht daher in der Betrachtung interindividueller Unterschiede in der Wahrnehmung von Dimensionen der Feedbackqualität im Deutschunterricht der gymnasialen Oberstufe sowie in der Untersuchung der prädiktiven Bedeutsamkeit der Feedbackwahrnehmung für die Entwicklung der intrinsischen Lernmotivation als relevantes Zielkriterium von Unterricht.

Ein weiterer Forschungsschwerpunkt dieser Studie leitet sich aus dem Desiderat ab, dass sich – mit wenigen Ausnahmen (u.a. Lazarides et al., 2019; Lazarides, Fauth et al., 2021) – bislang kaum längsschnittlich ausgerichtete Untersuchungen zu der Frage finden, inwiefern sich Wahrnehmungen von Unterrichtsqualitätsmerkmalen wie Feedback über die Schulzeit hinweg differenziell verändern können und in welchem Zusammenhang diese Veränderungen mit der Motivationsentwicklung von Schülerinnen und Schülern stehen. Insbesondere für den Bereich der Sekundarstufe II mangelt es an längsschnittlichen Untersuchungen dieses Zusammenhangsmusters. Mit Blick auf die heterogene Schülerschaft in den leistungsdifferenzierten Kursen der Sekundarstufe II (Baumert & Köller, 2000; Nagy et al., 2006; Dreiling & Willems, 2020) und die damit einhergehenden Herausforderungen hinsichtlich der Anpassungen des unterrichtlichen Handelns an die interindividuellen Schülerinnen- und Schülerbedürfnisse, scheint es, dass Veränderungen der (wahrgenommenen) Feedbackqualität und der Motivation auch noch in der gymnasialen Oberstufe möglich sind.

1.2 Ziele der Arbeit

Die vorliegende Forschungsarbeit stellt die Perspektive der Lernenden in den Mittelpunkt der Untersuchung. In Ergänzung zu Videostudien, die die Gestaltung von Feedback im Unterricht aus einer eher „objektiven“ Sicht heraus beschreiben (Lotz, 2016), erlauben Schülerbefragungen einen Einblick in das subjektive Erleben des von der Lehrkraft erteilten Feedbacks (Waldis et al., 2010). Das übergeordnete Ziel ist die systematische und differenzierte Analyse der wahrgenommenen Qualität von Feedback aus der Sicht von Schülerinnen und Schülern und die Untersuchung differenzieller Wahrnehmungen von Feedback zu Beginn und zur Mitte eines Schuljahres in der Sekundarstufe II. Im Fokus der Untersuchung steht außerdem auch die längsschnittlich ausgerichtete Frage nach der Veränderung der wahrgenommenen Qualität von Feedback im Laufe eines Schulhalbjahres und der prädiktiven Bedeutsamkeit der Feedbackwahrnehmung für die Entwicklung der intrinsischen Lernmotivation. Die zentralen Fragestellungen lauten:

- (1) Wie nehmen Schülerinnen und Schüler das Feedback der Lehrkraft im Deutschunterricht der gymnasialen Oberstufe zu Beginn und zur Mitte eines Schulhalbjahres wahr?
- (2) In welchem Zusammenhang steht die wahrgenommene Feedbackqualität mit der intrinsischen Lernmotivation?
- (3) Wird der Zusammenhang zwischen der wahrgenommenen Feedbackqualität und der Veränderung der intrinsischen Lernmotivation durch die wahrgenommene Nützlichkeit von Feedback und das Kompetenzerleben mediiert?

Die vorliegende Arbeit ist in das längsschnittlich angelegte Projekt „Feedback im Kontext von Heterogenität“ (Willems et al., 2020) eingebettet, in der in 49 Deutschkursen der gymnasialen Oberstufe 807 Schülerinnen und Schüler zu zwei Messzeitpunkten (Beginn und Mitte eines Schuljahres) standardisiert zu ihrer selbstbestimmten Lernmotivation sowie zu ihrer Wahrnehmung verschiedener Dimensionen der Feedbackqualität im Deutschunterricht befragt wurden. Aus methodischer Sicht wird zur Beantwortung der Fragestellungen mit dem Verfahren des *Latent-True-Change-Modeling* (Steyer et al., 2000) ein Strukturgleichungsansatz der Veränderungsmessung gewählt, bei dem es möglich ist, sowohl *intraindividuelle* Veränderungen der Feedbackwahrnehmung und der intrinsischen Motivation (Veränderung der Konstrukte bei einer Person über die Zeit) als auch *interindividuelle* Unterschiede (Unterschiede zwischen den Schülerinnen und Schülern) in den Ausgangswerten sowie Veränderungen zu betrachten und diese wechselseitig aufeinander zu beziehen (Dietrich, 2019; Geiser, 2011; Reuter et al., 2010).

1.3 Aufbau der Arbeit

Theoretischer Teil

Der theoretische Teil der Arbeit umfasst zwei übergreifende Kapitel: In *Kapitel 2* werden ausgehend von unterschiedlichen Definitionen von Feedback zentrale theoretische Modelle und empirische Befunde zur Qualität und Wirksamkeit von Feedback im schulischen Unterricht vorgestellt. Hierbei wird insbesondere auf Modelle der Feedbackforschung sowie der pädagogisch-psychologisch orientierten Unterrichtsforschung eingegangen, die Prozesse der Wahrnehmung und Nutzung von Feedback durch die Schülerinnen und Schüler als wichtige Prädiktoren der Wirksamkeit von Feedback herausstellen. Ausgehend von empirischen Befunden werden sodann geschlechtsspezifische Unterschiede sowie der Einfluss der Leistung auf die Feedbackwahrnehmung thematisiert. Die Beschreibung theoretischer Ansätze und aktueller Befunde zu Veränderungen von Unterrichts- und Feedbackwahrnehmungen bildet den Abschluss von *Kapitel 2*.

Vor dem Hintergrund des übergreifenden Ziels der Arbeit, die Wirkungen des wahrgenommenen Feedbacks auf die Veränderung der intrinsischen Lernmotivation zu untersuchen, steht in *Kapitel 3* die Aufarbeitung des theoretischen und empirischen Hintergrunds zur Lernmotivation und deren Zusammenhang zu Feedback im Unterricht im Vordergrund. Nach einer Darstellung von motivationstheoretischen Grundlagen sowie aktuellen empirischen Befunden zur Motivationsentwicklung werden theoretische Ansätze und Forschungsbefunde zum Zusammenhang zwischen Feedback und der intrinsischen Lernmotivation näher vorgestellt. In *Kapitel 4* wird ein Rahmenmodell zur Analyse der Wahrnehmung und Wirkung von Feedback im schulischen Unterricht hergeleitet, welches auf einer Zusammenführung von theoretischen Ansätzen der Feedback- und Unterrichtsforschung beruht.

Empirischer Teil

Die im theoretischen Teil erarbeiteten Ausführungen und empirischen Erkenntnisse bilden die Grundlage für die Untersuchung der Feedbackwahrnehmung der in der FeeHe-Studie befragten Schülerinnen und Schüler im Deutschunterricht der gymnasialen Oberstufe. Hierfür werden in *Kapitel 5* die Fragestellungen und Hypothesen zu den drei Hauptforschungsfragen der Untersuchung vorgestellt. Im Kapitel zum methodischen Vorgehen wird zunächst die FeeHe-Studie beschrieben, bevor auf die verwendeten Skalen zur Erfassung der Feedbackqualität im Unterrichtsgespräch und der motivationalen Konstrukte eingegangen wird (*Kapitel 6*). Das Kapitel beschäftigt sich abschließend mit der Darstellung der zentralen statistischen Analysen, die dieser Arbeit zugrunde liegen.

Die Darstellung der Ergebnisse gliedert sich entlang der aufgestellten Fragestellungen (*Kapitel 7*). Im Anschluss an die Analysen zur Dimensionalität und Messinvarianz der wahrgenommenen Feedbackqualität werden die Ergebnisse zur differenziellen Wahrnehmung und Veränderung von Feedback im Verlauf eines Schulhalbjahres beschrieben (*Hauptforschungsfrage I*). Daran anschließend erfolgt die Darstellung der Befunde zu den Zusammenhangs- und Mediationsanalysen der wahrgenommenen Feedbackqualität mit der intrinsischen Lernmotivation der Lernenden (*Hauptforschungsfragen II und III*).

In der Diskussion (*Kapitel 8*) werden die zentralen Ergebnisse der Arbeit zusammengefasst und vor dem theoretischen und empirischen Hintergrund diskutiert. Gemäß der Reihenfolge der aufgestellten Fragestellungen wird in Abschnitt 8.1.1 zunächst die Dimensionalität von Feedback – als ein relevantes Qualitätsmerkmal von Unterricht – aus der Sicht von Schülerinnen und Schülern beschrieben. Es wird diskutiert, inwieweit sich die Hypothesen zum Einfluss des Geschlechts und der Leistung auf differenzielle Wahrnehmungen und Veränderungen von Feedback im Laufe eines Schulhalbjahres bestätigen lassen (*Hauptforschungsfrage I*). Der Abschnitt 8.1.2 fokussiert auf die Kernfrage der vorliegenden Arbeit nach dem Zusammenhang zwischen der Feedbackwahrnehmung und der intrinsischen Lernmotivation sowie der Veränderung der beiden Konstrukte über die Zeit (*Hauptforschungsfrage II*).

Hierbei soll insbesondere die theoretische Annahme von Angebots-Nutzungs-Modellen diskutiert werden, die von einer reziproken Beziehung zwischen dem Angebot und Merkmalen der Lernenden ausgehen. Im Abschnitt 8.1.3 werden schließlich die Ergebnisse der Mediationsanalysen berichtet und in den bestehenden empirischen Forschungsstand eingebettet (*Hauptforschungsfrage III*). Das Kapitel 8.1 schließt mit einer Übersicht über Hypothesen und empirische Befunde der eigenen Studie ab (vgl. Tabelle 50). Es folgt eine Diskussion der Grenzen der Arbeit (Kapitel 8.2) und daran anknüpfend ein Ausblick auf weiterführende Forschungsfragen für die Feedback- und Unterrichtsforschung (Kapitel 8.3). Abschließend werden mögliche Implikationen im Hinblick auf eine förderliche und adaptive Gestaltung von Feedback im Deutschunterricht der Sekundarstufe II und die damit einhergehenden Anforderungen an die Professionalisierung von Lehrkräften im Rahmen der Aus- und Weiterbildung aufgezeigt (Kapitel 8.4).

2 Feedback als Unterrichtsmerkmal

Im folgenden Kapitel werden die für diese Forschungsarbeit relevanten theoretischen und empirischen Grundlagen zu Feedback im schulischen Unterricht dargestellt. Hierfür werden zunächst in Abschnitt 2.1 verschiedene Definitionsansätze von Feedback kurz beleuchtet. Eine zentrale Bedingung der Wirksamkeit von Feedback ist, wie das Feedback gestaltet ist; daher werden in Abschnitt 2.2 unterschiedliche Dimensionen der inhaltlichen Gestaltung von Feedback aufgeführt. Hierbei erfolgt auch eine Definition und theoretische Abgrenzung der in dieser Arbeit erfassten Feedbackdimensionen (Ergebnis, Prozess, Selbstregulation und Dialog). Daran anschließend werden zentrale empirische Befunde zur Wirksamkeit von Feedback auf leistungs- und motivationsbezogene Zielkriterien von Unterricht (Abschnitt 2.3) und zur Ausprägung von Feedback im Unterricht aus einer „objektiven“ Beobachterperspektive (Abschnitt 2.4) und zusammenfassend vorgestellt. Anhand von unterschiedlichen theoretischen Modellen, wie dem Interaktionalen Rahmenmodell zur Gestaltung und Verarbeitung von Feedback nach Strijbos und Müller (2014) sowie den allgemeinen Angebots-Nutzungs-Modellen unterrichtlicher Wirkungen (Helmke, 2015; Vieluf et al., 2020), wird sodann aufgezeigt, dass neben der Gestaltung von Feedback auch die Wahrnehmung dieses Feedbacks durch die Schülerinnen und Schüler ein wichtiger Prädiktor im Prozess der Wirksamkeit von Feedback sein kann (Abschnitt 2.5). Hierzu werden zunächst zentrale theoretische Modelle

aus der Feedback- und Unterrichtsforschung skizziert (Abschnitt 2.5.1) und anschließend werden die Herausforderungen bei der Erfassung der Wahrnehmung von Feedback thematisiert (Abschnitt 2.5.2). Dabei steht die methodische Fragestellung im Vordergrund, wie die Schülersicht auf Feedback im Unterricht operationalisiert wird. Abschnitt 2.5.3 widmet sich der Beschreibung und Diskussion aktueller empirischer Befunde zu Prädiktoren von differenziellen Feedbackwahrnehmungen. Ein Anliegen der vorliegenden Arbeit ist die längsschnittliche Untersuchung der Feedbackwahrnehmung im Laufe eines Schulhalbjahres. Theoretische Perspektiven auf und empirische Belege zu Veränderungen von Unterrichts- und Feedbackwahrnehmung sind daher Gegenstand des Abschnitts 2.5.4.

2.1 Definition von Feedback

In der vorliegenden Arbeit liegt der Fokus auf dem Feedback durch die Lehrkraft im Sinne der Bereitstellung eines Unterrichts- bzw. Lernangebots. Auf den Kontext des Schulunterrichts bezogen kann Feedback „alle verbalen und nonverbalen Reaktionen der Lehrkraft auf Schüleräußerungen oder Schülerverhalten“ (Lotz, 2016, S. 196) umfassen. Dies ist allerdings eine weite Definition von Feedback und würde auch Formen von Feedback einschließen, die nicht im Zusammenhang mit der Leistung von Schülerinnen und Schülern stehen (z.B. Lob und Tadel). Daher wird für die vorliegende Arbeit eine engere Definition von Feedback gewählt: „Feedback wird als jede Art von Rückmeldung verstanden, die sich auf die Leistung oder das Verständnis des Lernenden bezieht, diesen über die Richtigkeit seiner Antwort bzw. seiner Aufgabenlösung informiert [...] oder ihm inhaltliche und/oder strategische Hilfen und Informationen zu seinem Bearbeitungsprozess zur Verfügung stellt [...]“ (Lipowsky, 2020, S. 83). Nach Hattie und Timperley (2007) kann Feedback von einer Lehrkraft, einem Mitschüler bzw. einer Mitschülerin oder vom Lernenden selbst erteilt werden. Davon abgrenzend wird das Erteilen von Lob und Tadel auch in dieser Arbeit nicht als Feedback klassifiziert, da Lob und Tadel häufig eher Bekräftigungen oder Sanktionierungen infolge eines Schülerverhaltens enthalten und sich in der Regel nicht auf die erbrachte Leistung von Lernenden beziehen (Hattie, 2009; Hattie & Timperley, 2007; Jacobs, 2002; Lipowsky, 2020; Lotz & Lipowsky, 2015).

In der Forschungsliteratur wird Feedback als ein Merkmal eines guten Unterrichts beschrieben (Brophy & Good, 1986; Hattie, 2009; Helmke, 2015; Lipowsky, 2020; Pianta et al., 2008), da es die Lernprozesse der Schülerinnen und Schüler unterstützt. Aus kognitionspsychologischer Sicht hat Feedback eine informierende Funktion, die darin besteht, die Diskrepanz zwischen der aktuellen Leistung und dem angestrebten Lernziel aufzuzeigen und zu verringern (Hattie & Timperley, 2007; Krause et al., 2004; Ramaprasad, 1983). Einerseits kann es dabei die Richtigkeit einer Antwort oder Aufgabenbearbeitung bestätigen und so bestehende Wissensstrukturen oder Lösungswege festigen; andererseits kann es den Lernenden bei

einer falschen Antwort bzw. Aufgabenbearbeitung korrigieren und somit bereits bestehendes Wissen verändern (Kulhavy & Stock, 1989; Lotz, 2016; Huth, 2004; Mory, 2004). Auf diese Weise kann Feedback kognitive Prozesse bei Schülerinnen und Schülern aktivieren, also zu einer vertieften Auseinandersetzung mit dem Lerngegenstand führen (Lotz, 2016). Neben der Relevanz für kognitive Lernprozesse ist auch die Bedeutsamkeit von Feedback für die Unterstützung motivationaler (Vollmeyer & Rheinberg, 2005) und selbstregulativer (Butler & Winne, 1995) Prozesse der Lernenden unumstritten (zfsd. Huth, 2004; Narciss, 2006).

In der Forschung zur Unterrichtskommunikation wird Feedback außerdem als ein Element eines qualitativvollen Unterrichtsgesprächs beschrieben (Denn, 2021; Pauli, 2006, 2010; Pehmer et al., 2015; Seidel et al., 2003). Der Stellenwert von Feedback im Unterrichtsgespräch zeigt sich am Beispiel der IRE-Sequenz (*Initiation-Reply-Evaluation*-Sequenz; vgl. Lüders, 2003; Mehan, 1981; Mercer & Dawes, 2014; Richert, 2005), die als charakteristisches Muster der Interaktion zwischen Lehrkräften und Lernenden im Unterricht gilt (Howe & Abedin, 2013; Lüders, 2003; Richert, 2005). Demnach besteht das Unterrichtsgespräch aus einer Abfolge aus (1) Frage oder Impuls der Lehrkraft (*Initiation*), (2) Antwort der Schülerinnen und Schüler (*Response*) und (3) Feedback der Lehrkraft auf die Schülerantwort (*Evaluation*). Mit der Qualität von Feedback in diesem Dreischritt der Abfolge im Unterrichtsgespräch haben sich viele Studien beschäftigt (u.a. Richert, 2005). Damit die Lernenden im Unterrichtsgespräch nicht nur als „Stichwortgeber“ (Kobarg & Seidel, 2003, S. 174) fungieren, muss das Feedback so gestaltet sein, dass es die Schülerinnen und Schüler zu einer aktiven und vertieften Auseinandersetzung mit dem Lerngegenstand anregt und ihr selbstbestimmtes Lernen unterstützt (Lotz & Lipowsky, 2016; Pauli, 2010). Daran anknüpfend wird im folgenden Abschnitt auf Dimensionen einer qualitativvollen Gestaltung von Feedback eingegangen.

2.2 Dimensionen von Feedback

Die Qualität der Gestaltung von Feedback im Unterricht nimmt einen wesentlichen Einfluss auf die Wahrnehmung und Verarbeitung durch die Lernenden (Kopp & Mandl, 2014a; Narciss, 2014) und auf eine förderliche Entwicklung motivational-affektiver sowie kognitiver Lernergebnisse der Schülerinnen und Schüler (Hattie & Timperley, 2007). In diesem Zusammenhang sind zahlreiche theoretische Modelle entstanden, die unterschiedliche Dimensionen der Feedbackqualität zusammenfassend darstellen (u.a. Hattie & Timperley 2007; Kulhavy & Stock, 1989; Narciss, 2006; Shute, 2008). In der Feedbackforschung hat sich die Unterscheidung einer einfachen und elaborierten Komponente von Feedback etabliert (u.a. Jacobs, 2002; Kulhavy & Stock, 1989; Narciss, 2006). Eine ausführliche Klassifizierung einfacher und elaborierter Feedbackformen findet sich bei Narciss (2006) (vgl. Tabelle 1).

Tabelle 1: Klassifikation unterschiedlicher Feedbackkomponenten (Narciss, 2006, S. 23)

	Bezeichnung	Beispiele für Feedbackinhalte
Einfache Komponenten	Knowledge of performance (KP)	<ul style="list-style-type: none"> • 15 von 20 Aufgaben richtig • 85% der Aufgaben korrekt gelöst
	Knowledge of result/response (KR)	<ul style="list-style-type: none"> • Falsch/ richtig • Nicht richtig/ stimmt
	Knowledge of correct result (KCR)	<ul style="list-style-type: none"> • Angabe der korrekten Lösung • Markierung der korrekten Antwort
Elaborierte Komponenten	Knowledge on task constraints (KTC)	<ul style="list-style-type: none"> • Hinweise auf Art der Aufgabe • Hinweise auf Bearbeitungsregeln • Hinweise auf Teilaufgaben • Hinweise auf Aufgabenanforderungen
	Knowledge about concepts (KC)	<ul style="list-style-type: none"> • Hinweise auf Fachbegriffe • Beispiele für Begriffe • Hinweise auf Begriffskontext • Erklärungen zu Begriffen
	Knowledge about mistakes (KM)	<ul style="list-style-type: none"> • Anzahl der Fehler • Ort der Fehler/des Fehlers • Art der Fehler/des Fehlers • Ursache/n des/r Fehler(s)
	Knowledge on how to proceed (KH)	<ul style="list-style-type: none"> • Fehlerspezifische Korrekturhinweise • Aufgabenspezifische Lösungshinweise • Hinweise auf Lösungsstrategien • Leitfragen • Lösungsbeispiele
	Knowledge on meta-cognition (KMC)	<ul style="list-style-type: none"> • Hinweise auf meta-kognitive Strategien • Meta-kognitive Leitfragen

Das Feedback bezieht sich dabei in der Regel auf Aufgaben mit einer eindeutigen Lösung bzw. Antwort, weshalb die Klassifizierung häufig im Rahmen von computerbasierten Lernumgebungen Anwendung findet. Als einfache Komponente werden solche Feedbackarten gewertet, die den Lernenden im Anschluss an eine Antwort oder eine Aufgabenbearbeitung darüber informieren, wie der aktuelle Leistungsstand ist (*knowledge of performance*), ob eine Antwort richtig oder falsch ist (*knowledge of result*) oder wie die richtige Lösung ist (*knowledge of correct result*) (Narciss, 2006). Eine weitere Form einfachen Feedbacks ist *answer until correct* (u.a. Jacobs, 2002; Kulhavy & Stock, 1989; Richert, 2005). Hierbei fordert die Lehrkraft den Schüler oder die Schülerin nach einer falschen Antwort zu weiteren Antwortversuchen auf, bis

die richtige Lösung genannt wird (Lotz, 2016; Richert, 2005). Während diese einfachen Feedbackarten dem Lernenden nur wenige Informationen bereitstellen und nicht über die Mitteilung der richtigen Antwort hinausgehen, enthält elaboriertes Feedback ergänzende Hinweise, die zur Verbesserung der Aufgabenbearbeitung beitragen können. Hierzu zählen nach Narciss (2006) u.a. Hinweise zur Aufgabenstellung (*knowledge on task constraints*), zur Korrektur oder Lösung einer Aufgabe (*knowledge on how to proceed*), zu Fehlern bei der Aufgabenbearbeitung (*knowledge about mistakes*) oder zu metakognitiven Strategien (*knowledge on meta-cognition*).

Auf Basis von Ergebnissen aus Metaanalysen entwickeln Hattie und Timperley (2007) in ihrem Review ein konzeptuelles Modell von Feedback im unterrichtlichen Kontext. Im Wesentlichen werden in dem Modell zwei Bereiche der Gestaltung von effektivem Feedback beschrieben (vgl. Abbildung 1): Der erste Bereich beinhaltet drei unterschiedliche Fragen von effektivem Feedback. Demnach beantwortet Feedback die Frage nach dem Lernziel einer Aufgabenstellung (*Feed Up*), nach dem aktuellen Leistungsstand und möglichen Diskrepanzen zu dem Lernziel (*Feed Back*) und nach Hilfestellungen sowie Strategien zum Erreichen des Lernziels (*Feed Forward*). Im zweiten Bereich des Modells werden inhaltlich vier unterschiedliche Ebenen von Feedback unterschieden (i) Aufgabe (*task-level*), (ii) Prozess (*process-level*), (iii) Selbstregulation (*self-regulation level*) und (iv) Person (*self-level*).

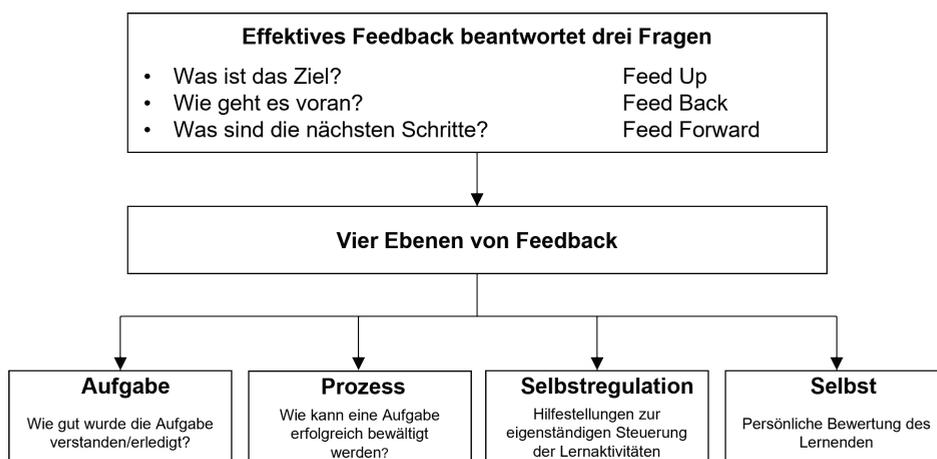


Abbildung 1: Ebenen des Feedbacks (nach Hattie und Timperley, 2007; vereinfachte Darstellung)

- (i) Feedback auf der Aufgabenebene enthält Informationen, wie gut eine Aufgabe bearbeitet wurde und inwieweit die angestrebten Lernziele erreicht wurden. Diese Feedbackebene, vor allem die aufgabenbezogene Ergebnisrückmeldung, findet nach Hattie und Timperley die häufigste Verwendung im Unterricht.

- (ii) Ein Feedback auf der Prozessebene fokussiert auf das *Wie* des Lernens, also die Art und Weise, wie der Lernprozess gestaltet wird und wie die Lernziele erreicht werden können. Hierbei erhält der Lernende konkrete Hinweise, ob die einzelnen Schritte und angewandten Strategien angemessen sind oder ein alternativer Lernweg eingeschlagen werden muss, um das Lernziel zu erreichen. Während Feedback auf der Aufgabenebene kein Tiefenverständnis befördern kann, liefert Feedback auf der Prozessebene den nötigen Anstoß, Informationen zu verarbeiten und Lernstrategien effektiver anzuwenden, um zur Lösung der Aufgabe zu gelangen (vgl. Berger et al., 2013).
- (iii) Feedback auf der Ebene der Selbstregulation informiert Lernende darüber, wie gut sie ihre Lernaktivitäten eigenständig überwachen, steuern und regulieren können, um das jeweilige Lernziel zu erreichen. Die Herausforderung bei dieser Feedbackebene besteht darin, dass Lernende Fähigkeiten der Selbsteinschätzung mitbringen, mit denen sie eigenständig Diskrepanzen in ihrer Aufgabenbearbeitung erkennen und Strategien entwickeln, um die Diskrepanzen zu beheben und ihrem Lernziel näherzukommen (Hattie & Timperley, 2007).
- (iv) Das Feedback auf der Ebene der Person gilt als am wenigsten förderlich, sofern es nur Lob an den Lernenden enthält und keine Informationen in Bezug auf die Leistung oder den Lernprozess bereithält (Hattie & Timperley, 2007; Hattie & Clarke, 2019). Die Autoren unterscheiden in dieser Hinsicht zwischen einem lernunwirksamen Lob bezogen auf das Selbst (z.B. „Gut gemacht!“) und einem Lob, das mit der Bewältigung der Aufgabe verbunden ist (z.B. Lob zur Anstrengungsbereitschaft oder zum Engagement).

Anknüpfend an die Forschungsarbeit von Hattie und Timperley (2007) wurde im Rahmen des Projekts FeeHe (Willems & Dreiling, 2022; Willems et al., 2020; vgl. auch Dreiling et al., 2018) ein Modell der Feedbackqualität entwickelt, das die vier Dimensionen *Ergebnisorientierung*, *Prozessorientierung*, *Selbstregulationsorientierung* und *Dialogorientierung* umfasst (vgl. Abbildung 2). In den Untersuchungen von Willems und Dreiling (2022) steht die Wahrnehmung der Dimensionen der Feedbackqualität aus der Perspektive der Schülerinnen und Schüler im Kontext des Unterrichtsgesprächs im Mittelpunkt. Daher orientieren sich die Autorinnen bei der Operationalisierung der Dimensionen von Feedback nicht nur an dem Feedback-Modell nach (Hattie & Timperley, 2007), sondern auch an theoretischen Konzepten einer qualitätsvollen und produktiven Gestaltung von Unterrichtsgesprächen mit der gesamten Klasse (Lipowsky & Lotz, 2015; Pauli & Reusser, 2018; Pehmer et al., 2015).

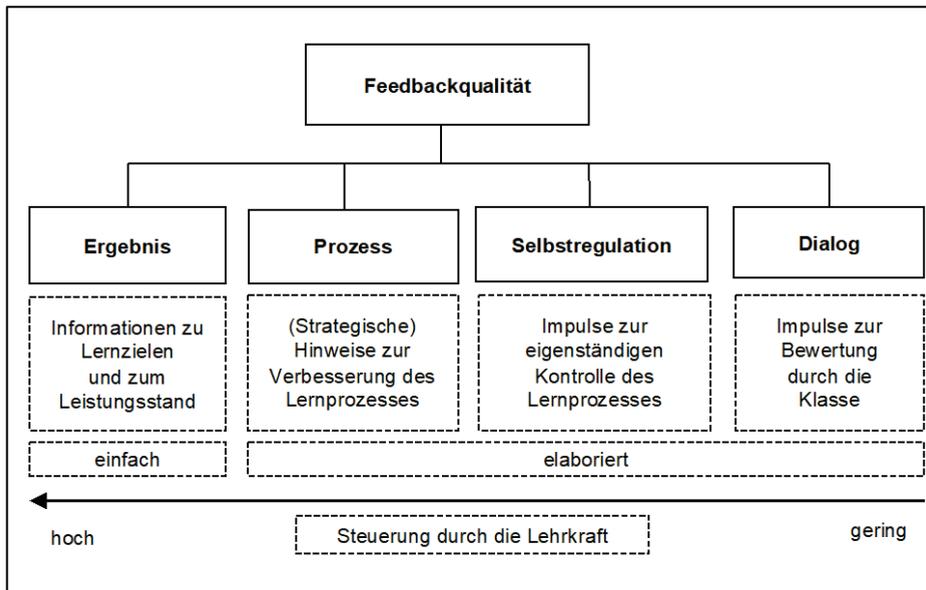


Abbildung 2: Dimensionen der Feedbackqualität (eigene Darstellung).

Die *ergebnisorientierte* Dimension von Feedback beinhaltet einfache Informationen über die Richtigkeit einer Antwort und darüber, ob die Antwort den Erwartungen der Lehrkraft entspricht oder was die angestrebten Ziele der Aufgabenstellung sind. Im Vordergrund steht dabei die Annahme, dass Feedback effektiver ist, wenn es sich auf konkrete Lernziele bezieht (Hattie & Timperley, 2007; zfsd. Hoya, 2019). Während die ergebnisorientierte Dimension dem Lernenden eher einfache Informationen bereitstellt, enthält die *prozessorientierte* Dimension ergänzende, meist strategische Hinweise (z.B. zur Aufgabenstellung, zum konzeptuellen Verständnis), damit Lernende ihre Antwort verbessern können. Die prozessorientierte Dimension zählt zu den elaborierten Dimensionen von Feedback, da sie weiterführende Informationen enthält. Neben der aufgaben- und prozessorientierten Feedbackdimension werden in dem Modell von Dreiling et al. (2018) und Willems und Dreiling (2022) noch zwei weitere elaborierte Dimensionen unterschieden: eine *selbstregulationsorientierte* und *dialogorientierte* Dimension. Bei der selbstregulationsorientierten Dimension von Feedback gibt die Lehrkraft durch gezielte Rückfragen und Denkipulse dem einzelnen Lernenden die Möglichkeit zur eigenständigen Bewertung und Verbesserung der Antwort, sodass die Fähigkeiten zur Regulation des Lernprozesses angeregt werden. Die selbstregulationsorientierte Dimension ist daher ein Feedback auf der metakognitiven Ebene. Die dialogorientierte Dimension von Feedback umfasst, dass die Beiträge der Schülerinnen und Schüler nicht mehr bloß von der Lehrkraft evaluiert, sondern zum Nachdenken und kritischen Bewerten an die gesamte Klasse weitergegeben werden.

Mit der Einführung einer dialogorientierten Feedbackdimension knüpfen Dreiling et al. (2018) und Willems und Dreiling (2022) – so argumentieren sie – ursprünglich an die Überlegungen von Pauli (2006, 2010; vgl. auch Pauli & Reusser, 2018) zum produktiven Unterrichtsgespräch an, welches eine Abweichung vom stark lehrkraftgeleiteten Unterrichtsgespräch darstellt und stattdessen die aktive Beteiligung der Lernenden am Unterrichtsgespräch anregt. Hiernach besteht die Funktion von Feedback auch darin, dass Prozesse des ko-konstruktiven Ideen- und Wissensaufbaus (Lipowsky, 2020; Pauli, 2010; Pauli & Reusser, 2018) ausgelöst werden, bei denen die Lernenden durch den gegenseitigen Austausch neue Ideen, Argumente und Lösungen entwickeln (Lipowsky, 2020; Orsmond et al., 2013).

Ergänzend kann an dieser Stelle argumentiert werden, dass auf der Ebene der aufgaben- und prozessorientierten Dimension der Feedbackgestaltung noch eine stärkere Steuerung und Anleitung des Lernprozesses durch die Lehrkraft stattfinden (vgl. Abbildung 2). Indem konkrete Informationen zum Leistungsstand und zur Verbesserung des Lernprozesses erteilt werden, werden dem Lernenden vergleichsweise wenig Möglichkeiten zur aktiven Beteiligung an der Evaluation des Beitrags gegeben. Demgegenüber sind die selbstregulations- und dialogorientierte Dimension dadurch gekennzeichnet, dass das Feedback der Lehrkraft weniger Anleitungen zur Verbesserung des Lernprozesses enthält. Stattdessen gibt die Lehrkraft die Verantwortung für die Bewertung und Verbesserung des Beitrags an den Beitragenden *selbst* (selbstregulationsorientierte Dimension) oder die *gesamte Klasse* (dialogorientierte Dimension) weiter. So gesehen haben beide Dimensionen einen stärkeren Dialogcharakter und fördern damit die Selbstevaluation sowie -regulation – nur liegt der Fokus bei der Feedbackgestaltung *entweder* auf dem Feedbackdialog zwischen der Lehrkraft mit dem einzelnen Schüler bzw. der einzelnen Schülerin *oder* auf dem Feedbackdialog zwischen der Lehrkraft mit der Klasse und zwischen den *peers* untereinander (Orsmond et al., 2013). Theoretisch gestützt wird diese Argumentation durch die neuere Feedbackforschung, die fordert, dass Feedback als Dialog gestaltet werden sollte (Nicol, 2010; Nicol & Macfarlane-Dick, 2006; Orsmond et al., 2013).

Vor diesem Hintergrund ist auch die Kategorisierung von Feedback bei Sortkaer (2018) entstanden. Der Autor unterscheidet zwischen zwei Dimensionen, die im Grad der Anleitung und Steuerung durch die Lehrkraft variieren: Das *directive feedback* und das *facilitative feedback*. Das *directive feedback* kann sowohl einfache als auch elaborierte Informationen enthalten, die den Lernenden dazu anleiten, wie die Leistung oder der Lernprozess verbessert werden kann. Hier findet kein Dialog zwischen der Lehrkraft und dem Lernenden statt, sodass auch die selbstregulativen Prozesse eingeschränkt werden (Sortkaer, 2018). Dieses Feedback ist außerdem spezifischer und detaillierter als das *facilitative feedback*, das durch elaborierte Hinweise den Lernenden zur eigenständigen Regulation des Lernprozesses anregt. Der Autor argumentiert, dass beim *facilitative feedback* mehr Raum für den Feedbackdialog zwischen der Lehrkraft und dem Schüler bzw. der Schülerin geschaffen wird, was als förderlicher für die Selbstregulation und folglich den Lernerfolg angesehen wird (Sortkaer, 2018 in Anlehnung an Nicol, 2010 und Orsmond et al., 2013).

2.3 Empirische Befunde zu Wirkungen von Feedback

Im vorangegangenen Kapitel wurde aufgezeigt, welche Dimensionen ein lernförderliches Feedback auszeichnen. Gegenstand des folgenden Kapitels ist die Frage, wie das Feedback – als Angebot – mit der Leistung sowie Motivation von Lernenden – als Zielkriterien des Unterrichts – zusammenhängt. Aus methodischer Sicht soll dabei sowohl auf Studien eingegangen werden, in denen diese Zusammenhänge durch Videoanalysen geprüft wurden, als auch auf Studien, die die Wirksamkeit des Erteilens von Feedback im Rahmen von experimentellen Lernsettings untersucht haben.

Im Allgemeinen ist die Bedeutsamkeit von Feedback für den Lernerfolg von Schülerinnen und Schülern anhand zahlreicher Metaanalysen dargelegt worden (Bangert-Drowns et al., 1991; Hattie, 2009; Hattie & Zierer, 2019; Kluger & DeNisi, 1996; Kulhavy & Stock, 1989; Seidel & Shavelsson, 2007; Wisniewski et al., 2020). In ihrem Review berichten Hattie und Timperley (2007), dass Feedback mit einer durchschnittlichen Effektstärke von $d = .79$ einer der lernwirksamsten Faktoren im Unterricht darstellt. Demzufolge ist das Erteilen von Feedback bei einer Aufgabebearbeitung förderlicher für den Lernerfolg als eine Aufgabebearbeitung, bei der Lernende kein Feedback für ihre Leistung erhalten (Bangert-Drowns et al., 1991; Labuhn et al., 2010; Wollenschläger et al., 2011). Einschränkend ist festzuhalten, dass sich über alle Studien hinweg große Streuungen der Effektstärken zeigen; zum Teil kann Feedback keine oder sogar negative Auswirkungen auf den Lernprozess haben. Auch die Überlegenheit elaborierter Feedbackdimensionen gegenüber einfachen Feedbackdimensionen konnte nicht eindeutig belegt werden (Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996; Kulhavy & Stock 1989; Hattie, 2009). Neben einer hohen methodischen Variabilität der Studien (u.a. Fach, Lernkontext, Alter der Lernenden) liegt eine weitere Ursache für die uneinheitlichen Ergebnisse auch darin begründet, dass unterschiedliche Dimensionen von Feedback auch unterschiedlich wirksam sind. Demzufolge ist die Effektivität von Feedback maßgeblich davon abhängig, wie qualitativ die Information ist, die dem Lernenden bereitgestellt wird (Hattie & Timperley, 2007; Wisniewski et al., 2020). Insofern ergibt sich ein einheitlicheres Bild der Befundlage, wenn bei der Betrachtung und Interpretation der Effekte nach den Qualitätsdimensionen von Feedback differenziert wird (Lipowsky, 2020; Wisniewski et al., 2020).

Entsprechend wird argumentiert, dass einfaches, ergebnisorientiertes Feedback am wenigsten förderlich für den Lernprozess ist, während elaboriertes Feedback zum Prozess der Aufgabebearbeitung (prozessorientierte Dimension) und zur Selbstregulation (selbstregulationsorientierte Dimension) vergleichsweise effektiver ist (Hattie & Timperley, 2007). Aus kognitionspsychologischer Sicht kann die Lernwirksamkeit von solchen elaborierten Dimensionen von Feedback darin begründet liegen, dass sie zu einer intensiveren Auseinandersetzung mit dem Lerngegenstand und damit zu einem vertieften Verständnis der Lerninhalte beitragen (Lipowsky &

Lotz, 2015; Pauli & Reusser, 2018). Aus motivationspsychologischer Sicht kann davon ausgegangen werden, dass sich elaboriertes, positiv unterstützendes Feedback – vermittelt über das Kompetenzerleben – positiv auf die selbstbestimmte Motivation von Lernenden auswirken sollte (vgl. Deci & Ryan, 2000; Kiemer et al., 2015; Kobarg & Seidel, 2007; Pauli, 2010; Pauli & Reusser, 2018; Rakoczy et al., 2008; Ryan & Deci, 2000).

Diese Annahmen zur Überlegenheit elaborierter Feedbackdimensionen werden auch durch verschiedene experimentell angelegte Studien empirisch untermauert, in denen Feedback – häufig unter Laborbedingungen – nach einer Antwort oder einer Aufgabenbearbeitung (computerbasiert) erteilt wird. So berichten Berner et al. (2022) in einer quasi-experimentellen Untersuchung, dass das Erteilen von elaboriertem Feedback im Vergleich zu einfachem Feedback zu positiveren Effekten auf den Erwerb mathematischer Kompetenzen bei Lernenden im Vorschulalter führt. Huth (2004) und Narciss und Huth (2006) stellen in ihren experimentellen Untersuchungen heraus, dass sich elaboriertes Feedback, das neben der Information zur Richtigkeit der Aufgabenbearbeitung auch weiterführende Hinweise zur Aufgabenbearbeitung enthält, positiv auf die Leistung und die Motivation von Grundschulkindern bei einem Rechentest auswirkt. Auch im Rahmen von experimentellen Studien mit Studierenden konnten positive Effekte auf die Leistung durch ein elaboriertes Feedback, welches neben Informationen zum Leistungsstand auch Erklärungen oder detaillierte Hinweise zur Verbesserung des Lernprozesses beinhaltet, herausgestellt werden (Krause et al., 2004; Lipnevich & Smith, 2009).

Anhand der Analysen im Rahmen des Co²Ca-Projekts (Klieme et al., 2010) können die leistungs- und motivationsförderlichen Effekte elaborierten Feedbacks im schulischen Unterricht aufgezeigt werden (Harks et al., 2014; Rakoczy et al., 2013; Rakoczy et al., 2019). Schülerinnen und Schüler der neunten Klassenstufe erhielten im Mathematikunterricht entweder ein einfaches (notenorientiertes) oder ein elaboriertes (prozessorientiertes) Feedback nach einem Test. Beim prozessorientierten Feedback wurden die Schülerinnen und Schüler schriftlich über den individuellen Lernstand im Vergleich zu den Lernzielen und über Strategien zum Erreichen des Lernziels informiert. Die Befunde verdeutlichen, dass sich prozessorientiertes Feedback positiver auf die Änderung der Mathematikleistung und des Interesses auswirkt als einfaches Feedback in Form von Noten (Harks et al., 2014; Rakoczy et al., 2013). Dieser Effekt des prozessorientierten Feedbacks wurde zum Teil über die von den Lernenden wahrgenommene Nützlichkeit (Harks et al., 2014; Rakoczy et al., 2013; Rakoczy et al., 2019) und die wahrgenommene Kompetenzunterstützung (Rakoczy et al., 2013) mediiert.

Darüber hinaus liegen auch Befunde von Videostudien vor, die Zusammenhänge zwischen Feedback im Unterrichtsgespräch und der Leistungs- sowie Motivationsentwicklung von Schülerinnen und Schülern beschreiben (Jurik et al., 2014; Kiemer et al., 2015; Rakoczy et al., 2008). So untersuchen Rakoczy et al. (2008) Feedback im Mathematikunterricht der Sekundarstufe I aus der Beobachterper-

spektive. Die Autoren berichten von positiven Wirkungen eines einfachen (positiven) Feedbacks zur Richtigkeit einer Antwort und eines elaborierten (prozessorientierten) Feedbacks, das Hinweise zum weiteren Vorgehen enthält, auf die von Schülerinnen und Schülern eingeschätzte intrinsische Lernmotivation im Fach Mathematik. Allerdings konnten weder für einfaches Feedback noch für elaboriertes Feedback bedeutsame Zusammenhänge mit der Leistungsentwicklung festgestellt werden, was die Autoren darauf zurückführen, dass sie nicht zwischen unterschiedlichen Dimensionen von elaboriertem Feedback differenziert haben.

Kiemer et al. (2015) untersuchen im Rahmen einer längsschnittlichen Interventionsstudie über ein Schuljahr, inwiefern eine videobasierte Intervention zur Entwicklung von professionellen Kompetenzen im Bereich des produktiven Unterrichtsgesprächs (u.a. Feedback) zu einer positiven Entwicklung des Interesses und der selbstbestimmten Motivation der Schülerinnen und Schüler für das Lernen im mathematisch-naturwissenschaftlichen Unterricht führt. Dabei wurde auch das von Lehrkräften erteilte Feedback im Unterrichtsgespräch videobasiert analysiert, während die motivationalen Merkmale über die Einschätzungen von Schülerinnen und Schülern erfasst wurden. Die Ergebnisse zeigen, dass Schülerinnen und Schüler der Interventionsgruppe, in der die Lehrkräfte mehr elaboriertes Feedback (prozess- und selbstregulationsorientiert) und weniger einfaches (ergebnisorientiertes) Feedback im Unterrichtsgespräch erteilten, über eine höhere wahrgenommene Autonomie- und Kompetenzunterstützung sowie eine höhere selbstbestimmte Motivation berichteten. Außerdem war das Feedback in der Interventionsgruppe prädiktiv für eine positive Entwicklung des fachspezifischen Interesses der Schülerinnen und Schüler.

In der Videostudie von Jurik et al. (2014) im Physikunterricht des neunten Jahrgangs wurden unterschiedliche Dimensionen von Feedback (Aufgabe, Prozess, Selbstregulation, Selbst), das von der Lehrkraft im Unterrichtsgespräch als Reaktion auf Schülerantworten erteilt wurde, durch externe Beobachterinnen und Beobachter kodiert und in Beziehung zu den kognitiven und motivationalen Merkmalen der Schülerinnen und Schüler gesetzt. Die Befunde verweisen darauf, dass das „objektiv“ erfasste Feedback die von Schülerinnen und Schülern eingeschätzte kognitive Lernaktivität und die intrinsische Lernmotivation positiv vorhersagt. Allerdings können keine Rückschlüsse gezogen werden, welche Dimension von Feedback am förderlichsten sind, da in den Analysen nicht mehr zwischen den unterschiedlichen Feedbackdimensionen differenziert wurde.

Zusammenfassend konnte anhand der aufgeführten Befunde belegt werden, dass das Erteilen von elaboriertem Feedback sich bedeutsam auf die Motivation und Leistung von Lernenden auswirkt. Dabei wird deutlich, dass in den dargestellten Studien häufig das erteilte Feedback im Rahmen quasi-experimenteller Designs im Fokus steht, wobei lediglich zwischen einem einfachen und einem elaborierten Feedback differenziert wird. Unklar bleibt, inwiefern sich die Effekte zwischen den in der Theorie abgeleiteten elaborierten Dimensionen der Feedbackqualität (prozessorientiert, selbstregulationsorientiert) unterscheiden (vgl. Jurik et al., 2014).

2.4 Empirische Befunde zur Häufigkeit von Feedback im Unterricht

Trotz der nachgewiesenen Bedeutung von Feedback für kognitive und motivationale Lernprozesse verweisen Studien darauf, dass Feedback von Lehrkräften in der Unterrichtspraxis nur selten den Qualitätsanforderungen entspricht und das Vorkommen von einfachem Feedback, welches nur wenige Informationen bereitstellt, im Unterricht dominiert (Denn et al., 2015; Kobarg & Seidel, 2007; Krammer et al., 2010; Lotz, 2016; Pohlmann-Rother et al., 2020; Voermann et al., 2012). Um die Gestaltung von Feedback im Unterricht zu beschreiben und zu bewerten, greifen solche Untersuchungen methodisch häufig auf Videoanalysen zurück (Lotz, 2016). Anders als Studien, deren Datengrundlage Befragungen von Schülerinnen und Schülern sind, ermöglichen Untersuchungen unter Verwendung von Beobachtungsdaten die Erfassung einer objektiveren Perspektive auf den Unterricht (Lotz, 2016; Pianta & Hamre, 2009; Seidel & Thiel, 2017) und bieten vor allem „punktuelle Einsichten“ (Waldis et al., 2010, S. 173) in die unterrichtliche Feedbackpraxis. Auf einzelne Forschungsbefunde zur videobasierten Erfassung der Ausprägung von Feedback im Schulunterricht soll im folgenden Abschnitt näher eingegangen werden.

Seidel et al. (2003) untersuchen anhand von Videoanalysen im Physikunterricht die Ausprägung von Feedback durch die Lehrperson im Unterrichtsgespräch. Die deskriptiven Auswertungen der einbezogenen Physikklassen verweisen auf relativ geringe Anteile von „sachlich-konstruktiven“ und „positiven“ Rückmeldungen (minimal 3.1% und maximal 22.2%) während des Unterrichtsgesprächs. Darüber hinaus verweisen vertiefende Analysen auf negative Effekte einer hohen Engführung des Klassengesprächs auf selbstbestimmte Formen der Lernmotivation und auf das physikbezogene Interesse der Schülerinnen und Schüler. Die Engführung des Klassengesprächs zeichnet sich u.a. durch ein geringes Vorkommen sachlich-konstruktiver Rückmeldungen aus.

Ähnlich verweisen auch die Auswertungen der IPN-Videostudie von Kobarg und Seidel (2007), dass einfaches Feedback in Form von Bestätigungen oder Falsifizierungen der Antworten von Schülerinnen und Schüler den größten Teil aller Rückmeldungen im Physikunterricht darstellt ($MW = 87.54\%$, $SD = 7.89$). Demgegenüber kommen sachlich-konstruktive Rückmeldungen ($MW = 5.01\%$, $SD = 3.57$), die Informationen zur Verbesserung der Lernprozesse und -strategien beinhalten, sowie positiv-unterstützende Rückmeldungen ($MW = 7.35\%$, $SD = 7.24$), welche die Motivation zur aktiven Beteiligung im Unterricht unterstützen, nur selten vor. Einschränkend ist festzuhalten, dass die Standardabweichungen auf Unterschiede in der Häufigkeit der Arten von Feedback zwischen den untersuchten Klassen verweisen.

Lotz (2016) untersucht im Rahmen der PERLE-Videostudie das Feedback von Lehrkräften im Leseunterricht der Grundschule. Die Ergebnisse zeigen, dass Lehrkräfte zwar insgesamt häufig Feedback erteilen ($MW_{\text{pro Minute}} = 4.75$). Allerdings fällt in den detaillierten Kodierungen der einzelnen Reaktionen auf, dass einfaches Feedback am häufigsten erteilt wird, während der durchschnittliche Anteil elaborierten Feedbacks, das über die Richtigkeit der Antwort hinausgehende Hinweise enthält, bei lediglich 18.03% liegt. Im Einzelnen zeigen die Analysen, dass Lehrkräfte am häufigsten Feedback zur Lösung (40.20%) erteilen. In 39.09% der Fälle kommt eine Kombination aus Feedback zur Lösung und zum Prozess der Aufgabenbearbeitung vor. Seltener wird in den untersuchten Lerngruppen ausschließlich prozessbezogenes Feedback beobachtet (12.32%). Lediglich 4.92% der Feedbacks der Lehrkräfte hatten keinen inhaltlichen Bezug zur Leseübung.

Die Befunde von Denn et al. (2015), die das Feedback von Lehrkräften im Mathematikunterricht der Grundschule untersuchen, weisen die Dominanz von einfachem Feedback zur Lösung nach (80.60%). Feedback zum Prozess der Aufgabenbearbeitung wird hingegen in nur 6.18% der Fälle erteilt (gefolgt von Feedback zum Arbeits- und Sozialverhalten und Feedback zu persönlichen Merkmalen der Schülerinnen und Schüler). Bemerkenswert ist auch der Befund, dass im Mittel nur die Hälfte der Feedbacks an einen Schüler oder an eine Schülerin überhaupt als mathematikbezogen kategorisiert wurde ($MW = 47.32\%$).

Mittels Videodaten untersuchen Pohlmann-Rother et al. (2020), wie häufig und welche Form von Feedback Grundschullehrkräfte während einer Schülerarbeitsphase beim Schreiben erteilen. Die Ergebnisse demonstrieren deutliche Unterschiede im Feedbackverhalten der Lehrkräfte: Die Mehrheit aller in der Stichprobe erfassten Rückmeldungen (66.88%) enthält keine elaborierten Informationen in Form von Begründungen, Erläuterungen und Hinweisen zur richtigen bzw. falschen Antwort. Demgegenüber werden einfache Formen von Feedback (*knowledge of result* oder *knowledge of correct result*) häufiger erteilt.

Auch für den internationalen Schulkontext konnte die Studie von Voermann et al. (2012) belegen, dass Lehrkräfte nur selten Feedback zum individuellen Lernfortschritt (*progress feedback*) oder Feedback zur Leistung in Bezug zu den angestrebten Lernzielen (*discrepancy feedback*) bereitstellen. Weitaus häufiger verwenden die Lehrkräfte unspezifisches Feedback ohne Bezug zu den Leistungen oder zum Lernprozess der Schülerinnen und Schüler (z.B. Lob).

Zusammenfassend zeigt sich in den videobasierten Untersuchungen zum einen die Dominanz von Feedback in Form von einfachen Evaluationen. Zum anderen konnten teilweise aber auch beträchtliche Unterschiede im Feedbackverhalten zwischen den einzelnen Klassen bzw. Lehrpersonen nachgewiesen werden, die darauf hindeuten, dass es auch durchaus Klassen gibt, in denen ein höherer Anteil elaborierter Feedbackdimensionen erteilt wird. In diesem Zusammenhang gehen einzelne Untersuchungen auch der Frage nach, ob es von bestimmten Merkmalen wie dem Geschlecht der Schülerinnen und Schüler abhängt, welche Dimension von Feed-

back Lehrkräfte erteilen. Die empirischen Befunde verweisen insbesondere auf geschlechtsspezifische Unterschiede im Erteilen von *nicht-leistungsbezogenem* Feedback wie Lob und Tadel (Denn et al., 2015; Irvine, 1986). Die Befunde in Bezug auf die Zusammenhänge zwischen dem Erteilen von *leistungsbezogenem* Feedback und dem Geschlecht der Schülerinnen und Schüler sind nicht konsistent. Während Denn et al. (2015) für den Leseunterricht der Grundschule keine geschlechtsspezifischen Unterschiede im inhaltlichen Bezug von Feedback feststellen konnten, weisen andere Studien nach, dass Jungen tendenziell mehr Feedback erhalten (Pohlmann-Rother et al., 2020).

2.5 Die Rolle der Wahrnehmung von Feedback

2.5.1 Theoretische Modelle zur Wahrnehmung von Feedback

Die Befundlage zu Wirkungen von Feedback verdeutlicht, dass nicht alle Schülerinnen und Schüler gleichermaßen von Feedback als Lernangebot profitieren (Hattie & Timperley, 2007; Wisniewski et al., 2020). Zusätzlich wurde in Videostudien, die die Ausprägung von Feedback aus einer Beobachterperspektive untersuchen, aufgezeigt, dass Lehrkräfte vermehrt einfache Feedbackdimensionen im schulischen Unterricht erteilen (u.a. Denn et al., 2015; Krammer et al., 2010). In diesem Zusammenhang muss allerdings berücksichtigt werden, dass das (aus „objektiver“ Sicht) erteilte Feedback der Lehrkraft nicht zwangsläufig mit der subjektiven Wahrnehmung der Schülerinnen und Schüler übereinstimmt (Havnes et al., 2012; Hoya, 2019). So muss beispielsweise ein von Beobachterinnen und Beobachtern identifizierte laborierte und für den Lernprozess förderliche Form von Feedback im Unterricht nicht gleichermaßen auch von Schülerinnen und Schülern als solche wahrgenommen und für die Verbesserung der Leistungen genutzt werden.

Theoretisch wird angenommen, dass nicht nur Dimensionen der Feedbackgestaltung, sondern auch die Wahrnehmung und aktive Nutzung von Feedback dessen Wirksamkeit beeinflussen (Kopp & Mandl, 2014a). Entsprechend zeigen schon ältere Forschungsergebnisse einer Metaanalyse von Bangert-Drowns et al. (1991), dass Feedback nur dann förderlich für den Lernprozess ist, wenn es *bewusst* wahrgenommen und bei der Aufgabenbearbeitung genutzt wird. Die Autoren sprechen in diesem Zusammenhang von „*mindful reception*“ (Bangert-Drowns et al., 1991, S. 233), also einer bewussten Rezeption von Feedback (zsf. Narciss, 2014). In Übereinstimmung dazu heben auch Rakoczy et al. (2013) die zentrale Rolle der Wahrnehmung von Feedback hervor und unterscheiden in Anlehnung an Ilgen et al. (1979) zwischen vier Phasen der Wirksamkeit von Feedback auf die Verhaltensänderung des Lernenden: „*On a theoretical level they differentiated four stages in understanding how feedback results in behavior change: Feedback needs to (a) be perceived, (b) be accepted as accurate, (c) be perceived as useful, and (d) lead to actual behavior change*“ (Rakoczy et al., 2013, S. 64).

Diese Perspektive auf den Feedbackprozess lässt sich auch lerntheoretisch begründen. Vor dem Hintergrund einer konstruktivistischen Lerntheorie (Gerstenmaier & Mandl, 1995; zsf. Reinmann-Rothmeier & Mandl, 2006) wird Feedback als ein Lernangebot verstanden, mit dem sich Lernende aktiv auseinandersetzen und das sie in unterschiedlicher Weise für die Verbesserung des eigenen Lernprozesses nutzen (Knickenberg, 2018; Kopp & Mandl, 2014b; Krause, 2007; Krause et al., 2004; Thurlings et al., 2013). Ob Lernende das angebotene Feedback wahrnehmen und nutzen, hängt im Wesentlichen von dem individuellen Lernpotenzial der Lernenden ab.

Aus diesem Grund liegt der Fokus der neueren Feedbackforschung nicht mehr nur auf der Gestaltung von Feedback durch die Lehrkraft aus einer „objektiven“ Perspektive, sondern vermehrt auf der subjektiven Sicht der Lernenden als Adressantinnen und Adressaten des Feedbacks (Brooks et al., 2019; Hoya, 2019; Knickenberg, 2018; Lipnevich et al., 2016; Lipnevich & Panadero, 2021; Panadero & Lipnevich, 2022; Van der Kleij & Lipnevich, 2021; Willems & Dreiling, 2022; Winstone et al., 2017).

Interaktionales Rahmenmodell zur Gestaltung und Verarbeitung von Feedback (Strijbos & Müller, 2014)

Eine differenzierte, theoretische Sichtweise auf das Zusammenspiel von individuellen Merkmalen und dem Prozess der Wahrnehmung und Nutzung von Feedback bietet das interaktionale Rahmenmodell von Strijbos und Müller (2014). Die Autoren greifen auf bestehende Modelle zurück, in denen die individuelle Wahrnehmung und die individuellen Merkmale von Lernenden bei der Wirkungsweise von Feedback berücksichtigt werden (u.a. Bangert-Drowns et al., 1991; Butler & Winne, 1995; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Narciss, 2006, 2008; Shute, 2008). Diese Modelle erweitern sie um eine stärker interaktionsanalytische Perspektive. Das Rahmenmodell (vgl. Abbildung 3) nimmt dabei sowohl die Gestaltung von Feedback als auch dessen Verarbeitung und Nutzung durch die Lernenden in den Blick. Beide Prozesse werden durch komplexe Interaktionen zwischen Gestaltungsmerkmalen des Feedbacks selbst, der Lernsituation sowie kognitiven und motivationalen Merkmalen des Lernenden (u.a. Vorwissen, Selbstwirksamkeit, Zielorientierungen) und des Lehrenden (u.a. Kompetenz) beeinflusst (Strijbos & Müller, 2014). So ist es zum einen vorstellbar, dass Lernende ein Feedback aufgrund ihres Leistungsniveaus oder ihrer motivationalen Ausprägungen unterschiedlich wahrnehmen. Das von der Lehrkraft erteilte Feedback wird dadurch sozusagen „ausgeseibt“ (Knickenberg, 2018, S. 19) und kann anders als von der Lehrkraft ursprünglich intendiert wahrgenommen und interpretiert werden (Strijbos & Müller, 2014). Dadurch können die Wahrnehmungen ein und desselben Feedbacks zwischen den Schülerinnen und Schülern einer Schulklasse je nach individuellen motivationalen und kognitiven Ausprägungen variieren und zu differenziellen motivationalen und kognitiven Effekten führen (Hoya, 2019; Knickenberg, 2018). Auf der anderen Seite

ist in Anlehnung an das interaktionale Modell von Strijbos und Müller (2014) auch vorstellbar, dass Lehrkräfte ihr Feedback in Abhängigkeit von den Bedürfnissen der Lernenden und ihren heterogenen Lernausgangslagen anpassen, d.h. dass auch die Gestaltung von Feedback durch die Lehrkraft von der Wahrnehmung und den Merkmalen der Lernenden beeinflusst werden kann. Folglich liefert das theoretische Rahmenmodell einen wichtigen Beitrag zur Erklärung der differenziellen Wirksamkeit von Feedback und verdeutlicht die Komplexität von Feedbackprozessen im schulischen Unterricht.

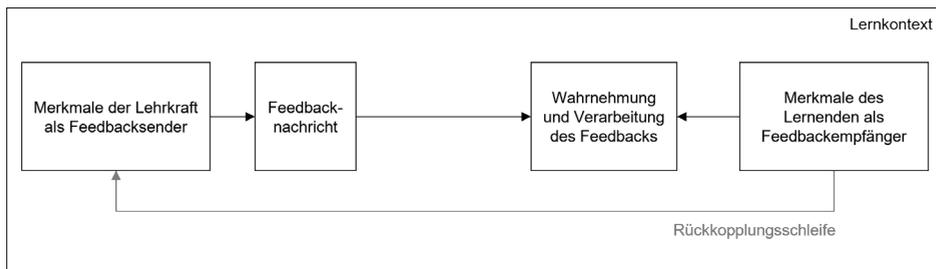


Abbildung 3: Interaktionales Rahmenmodell der Gestaltung und Verarbeitung von Feedback (nach Strijbos & Müller, 2014; modifiziert durch die Autorin)

Modell zur Interaktion zwischen Feedback und Lernenden (Lipnevich et al., 2016)

Ein aktuelles theoretisches Modell, das sich explizit auf die Annahme stützt, dass Lernende eine aktive Rolle im Feedbackprozess einnehmen, stammt von Lipnevich et al. (2016; vgl. auch Lipnevich & Panadero, 2021). Das Modell (vgl. Abbildung 4) wurde in den Arbeiten von Van der Kleij & Lipnevich (2021) sowie Lipnevich & Smith (2022) weiterentwickelt. Einzelne Komponenten aus diesen erweiterten Modellen wurden in Abbildung 4 aufgenommen. Im Modell steht die individuelle Wahrnehmung von und die Reaktion („*Response*“) auf das Feedback durch die Schülerinnen und Schüler im Mittelpunkt des gesamten Feedbackprozesses. Die zentrale Annahme ist, dass neben Merkmalen des spezifischen Lernkontextes, in dem das Feedback erteilt wird, insbesondere die individuellen Merkmale der Lernenden, wie das kognitive Fähigkeitsniveau und die Motivation, zu differenziellen Reaktionen auf das Feedback führen (Lipnevich et al., 2016; Lipnevich & Panadero, 2021). Dabei können die Reaktionen auf das Feedback kognitive, affektive und verhaltensbezogene Prozesse des Lernenden umfassen. Zur kognitiven Reaktion gehört beispielsweise die Einschätzung, ob das Feedback verstanden wird. Die affektive Reaktion kann Emotionen wie Freude oder Ärger über das erteilte Feedback umfassen. In den erweiterten Modellen von Van der Kleij & Lipnevich (2021) sowie Lipnevich & Smith (2022) wird außerdem eine verhaltensbezogene Reaktion aufgeführt, die die weitere Nutzung des Feedbacks im Lernprozess betrifft. Kognitive,

affektive und verhaltensbezogene Reaktionen auf das Feedback können sich wechselseitig beeinflussen. So wird der Lernende umso positiver auf ein Feedback reagieren, je besser er das Feedback versteht (Lipnevich et al., 2016). Diese Verarbeitung wiederum beeinflusst die konkreten Lernhandlungen bzw. die Nutzung von Feedback („Action“). Im erweiterten Modell von Lipnevich & Smith (2022) wird der Faktor Wirkungen („Outcome“) auf die konkrete Leistung oder den Lernprozess ergänzt.

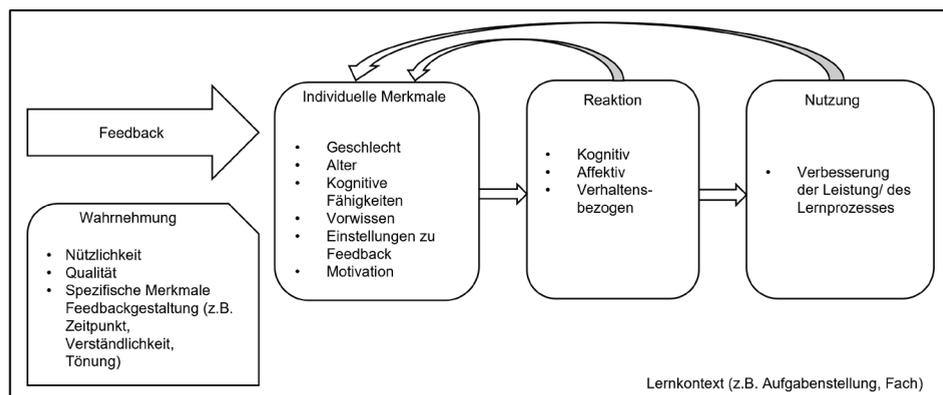


Abbildung 4: Modell der Interaktion zwischen Feedback und Lernenden (nach Lipnevich et al., 2016 und Lipnevich & Smith, 2022; übersetzt und modifiziert durch die Autorin)

Angebots-Nutzungs-Modell (Helmke, 2015; Vieluf et al., 2020)

Neben theoretischen Modellen der Feedbackforschung dienen auch Angebots-Nutzungs-Modelle der pädagogisch-psychologisch orientierten Unterrichtsforschung (Helmke, 2015; Lipowsky, 2006; Reusser & Pauli, 2010; Seidel, 2014; Vieluf et al., 2020) als geeigneter theoretischer Rahmen zur Untersuchung der Wahrnehmung von Feedback aus Sicht der Schülerinnen und Schüler (vgl. Abbildung 5). In Angebots-Nutzungs-Modellen wird Unterricht als Angebot von Lerngelegenheiten verstanden. Ob und inwiefern der Unterricht Veränderungen der kognitiven oder motivationalen Lernergebnisse von Schülerinnen und Schülern bewirkt, hängt nicht nur von der Qualität des Unterrichtsangebots selbst ab, sondern auch davon, wie Schülerinnen und Schüler das Angebot der Lehrkraft überhaupt wahrnehmen und interpretieren (Helmke, 2015). Vom Ausgang dieser Wahrnehmungsprozesse hängt ab, welche Lernaktivitäten (Nutzung) auf Seiten der Schülerinnen und Schüler resultieren. Die Wahrnehmung und Nutzung des Angebots stehen wiederum in Abhängigkeit von den individuellen Merkmalen und Dispositionen der Schülerinnen und Schüler. Unterricht wirkt sich demnach nicht automatisch oder bei allen Schülerinnen und Schülern in gleicher Weise auf den Lernerfolg aus. Vielmehr findet eine Interaktion zwischen dem Unterrichtsangebot (z.B. Feedback) und den Merkmalen von Schülerinnen und Schülern statt (Lipowsky, 2020; Vieluf et al., 2020).

Diese reziproke Beziehung wird in verschiedenen Angebots-Nutzungs-Modellen (Kleickmann et al., 2019; Lipwosky, 2020; Reusser & Pauli, 2010; Seidel, 2014; Vieluf et al., 2020) durch in beide Richtungen weisende Pfeile zwischen dem Feld „Unterrichtsangebot“ und dem Feld „Wahrnehmung und Nutzung“ des Unterrichtsangebots durch die Lernenden abgebildet (vgl. Abbildung 5). Die heterogenen Lernvoraussetzungen der Schülerinnen und Schüler bedingen, dass das gleiche Lernangebot von den Lernenden innerhalb einer Klasse unterschiedlich wahrgenommen wird und dies wiederum zu differenziellen Nutzungen und Lernentwicklungen führt (Jurik et al., 2015; Seidel, 2006). Dabei stellt die Nutzung ebenfalls einen komplexen Prozess dar, in dem sich kognitive Prozesse (z.B. Verarbeitungstiefe) sowie emotionales und motivationales Erleben (z.B. Autonomie und Kompetenz) gegenseitig beeinflussen (Jansen et al., 2022; Vieluf et al., 2020).

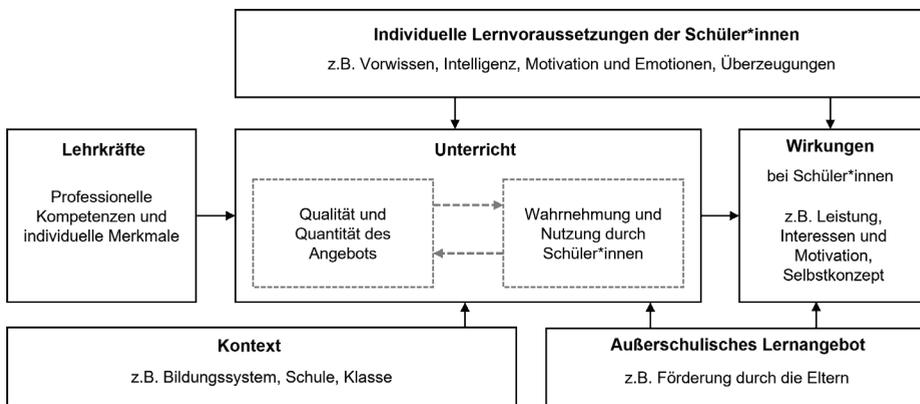


Abbildung 5: Vereinfachtes Angebots-Nutzungs-Modell (in Anlehnung an Kleickmann et al., 2019, S. 209)

Angelehnt an die Logik klassischer Angebots-Nutzungs-Modelle wird Feedback in dieser Arbeit als Unterrichtsmerkmal aufgefasst, das im Sinne eines Lernangebots von Schülerinnen und Schülern individuell wahrgenommen und genutzt wird und auf Basis dieser subjektiven Wahrnehmung und Nutzung differenzielle leistungs- und motivationsbezogene Wirkungen erzielen kann.

2.5.2 Zur standardisierten Erfassung der Wahrnehmung von Feedback

Im vorangegangenen Kapitel wurde die Bedeutung der individuellen Wahrnehmung von Feedback hervorgehoben, die beeinflusst, ob das Feedback genutzt wird und welche Wirkungen das Feedback entfaltet. Hierfür ist es zunächst einmal entscheidend, zu klären, wie die Perspektive der Schülerinnen und Schüler auf das Feedback, das im Unterricht von der Lehrkraft erteilt wird, empirisch erfasst wird. Einschränk-

kend ist festzuhalten, dass sich die Studien in der Operationalisierung der Feedbackwahrnehmung und im methodischen Zugang sehr unterscheiden (zsf. Van der Kleij & Lipnevich, 2021). Bei der folgenden Darstellung bestehender Instrumente zur Erfassung der Feedbackwahrnehmung wird daher zwischen unterschiedlichen Forschungssträngen unterschieden, die beeinflussen, wie die Wahrnehmung von Feedback methodisch operationalisiert wird.

Instrumente in der Feedbackforschung

In den letzten Jahren sind infolge des Perspektivwechsels der Feedbackforschung hin zum Feedback aus der Sicht der Lernenden (vgl. Lipnevich et al., 2016; Van der Kleij & Lipnevich, 2021; Winstone et al., 2017) zunehmend empirische Studien entstanden, die die Wahrnehmung von Feedback fragebogenbasiert untersuchen. Dabei liegt der Fokus der aktuellen Forschung auf unterschiedlichen Bereichen der Wahrnehmung, dazu gehören u.a.:

- die Art und Weise, wie das Feedback von der Lehrkraft erteilt wird (*feedback delivery*; vgl. Kyaruzi et al., 2019),
- die *Nützlichkeit* bzw. der Nutzung von Feedback (Brooks et al., 2019; Harks et al., 2014; Kyaruzi et al., 2019; Rakoczy et al., 2013; Rakoczy et al., 2019; Strijbos et al., 2010; Strijbos et al., 2021) oder
- die *Akzeptanz* von Feedback (Lipnevich & Lopera-Oquendo, 2022; Strijbos et al., 2010; Strijbos et al., 2021).

So entwickeln Brooks et al. (2019) in Anlehnung an das Feedbackmodell von Hattie und Timperley (2007) den SFPQ-Fragebogen („*Student Feedback Perception Questionnaire*“) mit 31 Items zur Erfassung der wahrgenommenen Nützlichkeit von unterschiedlichen Dimensionen von Feedback (u.a. die drei Ebenen Aufgabe, Prozess, Selbstregulation) aus der Sicht von Schülerinnen und Schülern. Zur Einschätzung sollen die Schülerinnen und Schüler auf einer fünfstufigen Skala angeben, wie hilfreich sie die spezifische Ebene des Feedbacks finden. Allerdings wird nicht erfasst, ob und wie diese Dimensionen von Feedback im Unterricht aus der Sicht der Schülerinnen und Schülern ausgeprägt sind.

Rakoczy et al. (2013), Rakoczy et al. (2019) und Harks et al. (2014) erfassen anhand einer Skala mit fünf Items die wahrgenommene Nützlichkeit von Feedback, das von der Lehrkraft im Mathematikunterricht erteilt wird. Konkret wurden die Schülerinnen und Schüler retrospektiv gefragt, inwiefern das Feedback ihnen geholfen hat, sich zu verbessern. Die Schülerinnen und Schüler wurden gebeten, ihre Zustimmung zu den unterschiedlichen Aussagen auf einer dreistufigen Skala anzugeben. In den Analysen wurde die wahrgenommene Nützlichkeit von Feedback als Mediatorvariable zur Erklärung der Wirkung von erteiltem Feedback auf die Leistungs- und Interessensentwicklung eingesetzt. Aufgrund des experimentellen Designs bleibt unklar, wie hoch die Dimensionen von Feedback (notenorientiert und

prozessorientiert) aus der Sicht der Schülerinnen und Schüler im Unterricht ausgeprägt sind.

Lipnevich et al. (2020) entwickeln mit den RIF-Skalen („*Receptivity to Instructional Feedback*“) einen Fragebogen zur differenzierten Messung der wahrgenommenen Akzeptanz von Feedback aus der Sicht von Schülerinnen und Schülern. Im Einzelnen werden mit den eingesetzten Skalen vier Merkmale der wahrgenommenen Akzeptanz operationalisiert: 1) eine affektive Komponente (z.B. Freude über das Feedback), 2) eine wertbezogene Komponente (z.B. Bewertung der Nützlichkeit von Feedback), 3) eine kognitive Komponente (z.B. Verständnis des Feedbacks) und 4) eine verhaltensbezogene Komponente (z.B. Verhalten in Bezug auf die Nutzung von Feedback). Die Schülerinnen und Schüler gaben ihre Zustimmung zur Ausprägung der unterschiedlichen Aussagen auf einer fünfstufigen Antwortskala an. Die latente, vierfaktorielle Struktur der Feedbackwahrnehmung konnte in unterschiedlichen Stichproben (USA, Neuseeland, Singapur) im schulischen und universitären Kontext bestätigt werden (Lipnevich et al., 2020; Lipnevich & Loper-Oquendo, 2022). Außerdem konnte die Messäquivalenz des Strukturmodells auch für unterschiedliche Gruppen (Mädchen und Jungen) im Rahmen von Messinvarianzanalysen nachgewiesen werden.

Ähnlich erfassen auch Strijbos et al. (2021) mit dem FPQ-Fragebogen („*Feedback Perceptions Questionnaire*“; vgl. auch Strijbos et al., 2010) zwei übergreifende Komponenten der Feedbackwahrnehmung, die wiederum unterschiedliche Subkomponenten beinhalten: eine kognitive Komponente sowie eine motivationale Komponente. Die kognitive Komponente bezieht sich auf die kognitive Funktion von Feedback und beinhaltet die drei Subkomponenten wahrgenommene Fairness, Nützlichkeit und Akzeptanz der Feedbackinformation. Die motivationale Komponente bezieht sich auf die motivierende Funktion von Feedback und umfasst die zwei Subkomponenten wahrgenommene Bereitschaft, die Leistung infolge des Feedbacks zu verbessern, sowie affektive Reaktionen, die das Feedback beim Feedbackempfänger auslösen. Im Rahmen der psychometrischen Auswertungen weisen die Autoren für das Instrument sowohl Konstruktvalidität als auch Messinvarianz in Bezug auf unterschiedliche Gruppierungsvariablen (z.B. Geschlecht) nach.

Zusammenfassend ist dem oben skizzierten Forschungsstand zu bestehenden Messinstrumenten zur Erfassung der Feedbackwahrnehmung zu entnehmen, dass auf der Ebene der Operationalisierung erfasst wird, wie Lernende auf Feedback reagieren, wobei diese Reaktion als multidimensionales Konstrukt (u.a. kognitive, affektiv-motivationale Komponenten usw.) konzeptualisiert wird. Einen Indikator der kognitiven Komponente bildet dabei u.a. die wahrgenommene Nützlichkeit von Feedback (Strijbos et al., 2021). In diesem Zusammenhang fällt auf, dass in den dargestellten Studien nicht erfasst wird, wie Schülerinnen und Schüler die Ausprägung von Feedback wahrnehmen. Dabei ist theoretisch davon auszugehen, dass das Feedback bei den Schülerinnen und Schülern zunächst „ankommen“ muss, bevor es als nützlich oder hilfreich wahrgenommen und für konkrete Lernhandlungen genutzt werden kann (Rakoczy et al., 2013). Insgesamt fehlt es folglich an empirischen

Untersuchungen, die systematisch erfassen, wie Schülerinnen und Schüler die Ausprägung von Feedback der Lehrkraft im Unterricht retrospektiv einschätzen. Damit einher geht auch die Frage, inwiefern unterschiedliche Dimensionen der Qualität der Feedbackgestaltung aus der Sicht von Schülerinnen und Schülern beschrieben werden können und wie diese Dimensionen der Feedbackqualität mit den Lernergebnissen zusammenhängen.

Instrumente in der Unterrichtsforschung

In der Unterrichtsforschung werden Einschätzungen von Schülerinnen und Schülern zum Unterricht oft dazu genutzt, die Dimensionen der Unterrichtsqualität differenziert zu beschreiben oder Zusammenhänge zwischen Unterrichtsmerkmalen und Lernergebnissen herzustellen (Stahns et al., 2020). Mit der Schülersicht auf das Unterrichtsgeschehen erhält man „authentische Berichte des individuellen Erlebens“ (Waldis et al., 2010, S. 172) der Schülerinnen und Schüler. Schüleraussagen gelten als eine zuverlässige Datenquelle zur Beschreibung der Qualität des Unterrichts (Clausen, 2002; Göllner et al., 2021; Lüdtke et al., 2009; Kunter et al., 2007; Seidel et al., 2007; Wagner et al., 2013, Wagner et al., 2016; Waldis et al., 2010) und stehen nachweislich im Zusammenhang mit kognitiven und motivational-affektiven Zielkriterien des Unterrichts (u.a. Fauth et al., 2014; Göllner et al., 2018; Iglar et al., 2019; Kunter et al., 2013; Schiepe-Tiska et al., 2016; Stahns et al., 2020; Wagner et al., 2016; Willems, 2011, 2022b), wobei Unterschiede zwischen den Qualitätsdimensionen berücksichtigt werden müssen (Göllner et al., 2021). Ein Vorteil von Schülerbefragungen gegenüber Beobachtungen durch externe Raterinnen und Rater ist, dass Schülerinnen und Schüler den Unterricht einer Lehrkraft über einen längeren Zeitraum und in verschiedenen Situationen beobachtet haben und somit als Expertinnen und Experten des Unterrichts angesehen werden können (Baumert et al., 2004; Clausen, 2002; Ditton, 2002; Göllner et al., 2016; Göllner et al., 2021; Wagner et al., 2016). Darüber hinaus können Schülereinschätzungen dazu genutzt werden, das Unterrichtsgeschehen differenziert zu beschreiben und die Trennung von unterschiedlichen Dimensionen der Unterrichtsqualität empirisch nachzuweisen (Fauth et al., 2014; Göllner et al., 2016; Stahns et al., 2020; Wagner et al., 2013). So analysierten Wagner et al. (2013) anhand der Daten der Studie Deutsch-Englisch-Schulleistungen International (DESI) die Schülerbeurteilungen des Unterrichts für die Fächer Deutsch und Englisch in 280 Schulklassen. Die Ergebnisse eines Mehrebenen-Strukturgleichungsmodells bestätigen u.a., dass Schülerinnen und Schüler unterschiedliche Dimensionen der Qualität des Deutsch- und Englischunterrichts (Schülermotivierung, Verständlichkeit, Schülerorientierung, Strukturiertheit und Klassenführung) differenziert beurteilen können, wobei die Trennung der fünf Dimensionen nicht nur auf der Ebene „innerhalb der Klassen“ (nicht geteilte Wahrnehmung), sondern auch „zwischen den Klassen“ (geteilte Wahrnehmung) empirisch nachgewiesen wurde. Vor diesem Hintergrund sollen im Folgenden einzelne Studien vorgestellt werden, die an diese Forschungstradition angelehnt sind und

unterschiedliche Dimensionen der Feedbackqualität auf Basis von Schülerwahrnehmungen untersuchen.

Die meisten Studien erfassen die wahrgenommene Feedbackqualität häufig als eindimensionales Konstrukt. So wurde in der internationalen Schulleistungstudie PISA 2015 die Häufigkeit von Feedback der Lehrkraft im naturwissenschaftlichen Unterricht aus der Sicht der Schülerinnen und Schülern erhoben (Bayer et al., 2016). Das Feedback wurde mittels fünf Items erfasst (Beispielitem: „Die Lehrkraft sagt mir, wie ich meine Leistungen verbessern kann“, Mang et al., 2019, S. 83). In der PISA-Studie 2018, in der das Feedback der Lehrkraft im Lese- bzw. Deutschunterricht im Mittelpunkt stand, wurde die Skala sogar auf drei Items reduziert (Mang et al., 2021). Die Schülerinnen und Schüler wurden gebeten, die Aussagen auf einer 4-stufigen Likert-Skala von „nie oder fast nie“ bis „jede Stunde oder fast jede Stunde“ zu beantworten, wobei höhere Werte häufigere Rückmeldungen anzeigten. Im Rahmen konfirmatorischer Analysen der in PISA erhobenen Feedback-Skala diskutieren Vries et al. (2022) auch, dass Feedback ein komplexes Konstrukt ist und unterschiedliche Dimensionen beinhaltet, die nicht mit nur fünf bzw. drei Items abgebildet werden können. Stattdessen plädieren die Autoren dafür, unterschiedliche Dimensionen von Feedback zu erfassen, die mit dem Lernerfolg der Schülerinnen und Schüler in Zusammenhang stehen.

Auch die Studie von Sortkær (2019) widmet sich der Untersuchung der Feedbackskalen, die in PISA 2012 aus der Perspektive der Schülerinnen und Schüler in naturwissenschaftlichen Fächern der Sekundarstufe I erhoben wurden. Sortkær (2019) geht theoretisch von zwei Dimensionen von Feedback aus (vgl. Abschnitt 2.2): 1) Ein direktives bzw. anleitendes Feedback (*directive feedback*), welches Informationen von der Lehrkraft enthält, die den Lernenden dazu anleiten, wie die Leistung oder der Lernprozess verbessert werden kann (Beispielitem: „The teacher tells me what I need to do to become better in mathematics“ (Sortkær, 2019, S. 735)) und 2) ein stärker selbstregulatives Feedback (*facilitative feedback*), welches durch Hinweise zur eigenständigen Überwachung und Regulation des Lernprozesses anregt (Beispielitem: „The teacher asks us to explain how we have solved a problem“ (Sortkær, 2019, S. 735)). Die Ergebnisse der konfirmatorischen Faktorenanalysen für unterschiedliche Stichproben (nordische Staaten) weisen auf eine zweifaktorielle Struktur des wahrgenommenen Feedbacks auf der Grundlage von Schülereinschätzungen hin.

Bislang liegen nur wenige Untersuchungen vor, die die wahrgenommene Feedbackqualität multidimensional erfassen. Hoya (2019) untersucht die subjektiv wahrgenommene Häufigkeit von unterschiedlichen Dimensionen der Feedbackqualität aus der Sicht von Grundschulkindern im Leseunterricht (vgl. auch Knickenberg, 2018). In seiner Studie setzt der Autor einen Fragebogen bestehend aus drei Subskalen zum wahrgenommenen einfachen („Lob“ und „Tadel“) sowie elaborierten („instruktionalen“) Feedback ein. Die Kinder wurden gebeten die Ausprägungen der verschiedenen Formen von Feedback anhand von insgesamt 18 Items auf einer

fünfstufigen Likertskala von „nie“ bis „sehr oft“ anzugeben. Im Rahmen von Faktorenanalysen kann die Dimensionalität der drei Faktoren der Feedbackwahrnehmung nachgewiesen werden (Hoya, 2019; Knickenberg, 2018). Allerdings erfassen die Autoren lediglich eine allgemeine Form von elaboriertem Feedback, obschon argumentiert wird, dass elaboriertes Feedback unterschiedliche Formen beinhalten kann. Die Ergebnisse aus Mehrebenen-Analysen zeigen, dass das Leseverständnis der Kinder auf der Individual- und Klassenebene signifikant durch das wahrgenommene negative und elaborierte Feedback erklärt werden kann. Da sich die Ergebnisse auf die Wahrnehmung von Grundschulkindern beziehen, können sie nicht ohne Weiteres auf den Kontext der vorliegenden Arbeit übertragen werden.

Zum gegenwärtigen Zeitpunkt liegen nur die Untersuchungen von Dreiling et al. (2018) und Willems und Dreiling (2022) vor, die die wahrgenommene Ausprägung von unterschiedlichen Dimensionen der Feedbackqualität aus der Sicht von Schülerinnen und Schülern differenziert erfassen. In Anlehnung an die Forschungsarbeiten von Hattie und Timperley (2007) zu effektivem Feedback im Schulunterricht sowie Pauli (2006, 2010) zu Feedback als Merkmal eines produktiven Unterrichtsgesprächs entwickeln die Autorinnen vier Subskalen, die die theoretisch angenommenen Dimensionen Aufgaben-, Prozess-, Selbstregulations- und Dialogorientierung von Feedback abbilden (vgl. Abschnitt 2.2). Die Schülerinnen und Schüler werden konkret gefragt, wie sie die Ausprägung dieser Dimensionen von Feedback, das sie im Unterrichtsgespräch nach einem Beitrag oder einer Antwort erhalten, retrospektiv einschätzen. Im Rahmen von zwei Studien zur Qualität von Feedback im Deutschunterricht der Oberstufe weisen die Autorinnen das Modell mit vier Faktoren auf Basis der Einschätzungen der Schülerinnen und Schüler empirisch nach. Ungeklärt ist bislang, ob sich die Struktur des Modells für unterschiedliche Gruppen von Schülerinnen und Schülern (z.B. Geschlechtergruppen) und zu mehreren Zeitpunkten (z.B. im Laufe eines Schuljahres) in gleicher Weise abbilden lässt. Außerdem bleibt offen, inwiefern die aus Schülersicht erfassten Dimensionen der Feedbackqualität mit der Leistungs- und Motivationsentwicklung der Schülerinnen und Schüler zusammenhängen. Insofern ist die Bedeutung des Modells von Feedbackqualität für den Deutschunterricht der Oberstufe noch nicht hinreichend nachgewiesen.

2.5.3 Empirische Befunde zur differenziellen Feedbackwahrnehmung

Unterricht – und damit auch das Feedback – wird nicht von allen Schülerinnen und Schülern gleichermaßen wahrgenommen. Durch die individuellen Merkmale, die Schülerinnen und Schüler mitbringen, entstehen interindividuelle Unterschiede in ihren Unterrichtswahrnehmungen (Igeler et al., 2019; Jurik et al., 2015; Schenke et al., 2018; Tsai et al., 2008; Willems, 2022a; Willems & Dreiling, 2022; Willems et al., im Druck), die sich auch in unterschiedlichen Lernentwicklungen niederschlagen können. Nichtsdestotrotz ist bislang wenig darüber bekannt, inwiefern Schülerinnen und Schüler aufgrund ihrer individuellen Merkmale das Feedback im Unterricht

unterschiedlich wahrnehmen (Gamlem & Smith, 2013; Havnes et al., 2012; Hoya, 2021; Sortkær, 2018, 2019) und noch weniger Studien widmen sich der Frage der differenziellen Feedbackwahrnehmungen im Deutschunterricht (Hoya, 2021).

Geschlecht als Prädiktor

Ein Merkmal, das sich als prädiktiv für differenzielle Unterrichts- und Feedbackwahrnehmungen im Deutschunterricht herausgestellt hat, ist das *Geschlecht* der Schülerinnen und Schüler (Ditton, 2002; Hoya, 2021; Jurik et al., 2015; Stang & McElvany, 2020; Wagner, 2008). Dabei finden sich insbesondere Belege für den Deutschunterricht der Grundschule und der Sekundarstufe I. So untersuchen Stang und McElvany (2020), ob zwischen Grundschülerinnen und Grundschulern systematische Unterschiede in der Wahrnehmung von allgemeinen Basisdimensionen von Unterrichtsqualität (Klassenführung, kognitive Aktivierung und Unterrichtsklima) im Deutschunterricht bestehen. Die Autorinnen belegen anhand von Daten der Internationalen Grundschul-Lese-Untersuchung 2016 (IGLU), dass Mädchen das unterstützende Unterrichtsklima – auch unter Kontrolle der individuellen Schülerleistung (Lesekompetenz) – positiver wahrnehmen als Jungen. Für den Deutschunterricht der Sekundarstufe I konnte nachgewiesen werden, dass die Gesamteinschätzung des Unterrichts bei Mädchen positiver ausfällt als bei Jungen (Jurik et al., 2015; Wagner, 2008).

In Bezug auf Unterschiede in der Wahrnehmung von Feedback im Deutschunterricht liegt die Studie von Hoya (2021) vor. Die Befragung von 665 Schülerinnen und Schülern der dritten und vierten Jahrgangsstufe zu dem aus ihrer Sicht wahrgenommenen positiven und negativen Feedback ergab, dass Mädchen im Vergleich zu Jungen seltener negatives Feedback und häufiger positives Feedback ihrer Lehrkräfte wahrnehmen. Der Geschlechtereffekt der Wahrnehmung von negativem Feedback bleibt auch unter Kontrolle der Leseleistung erhalten, nicht aber, wenn zusätzlich die intrinsische Lesemotivation und die Selbstkonzepte der Kinder kontrolliert werden. Für die Wahrnehmung von positivem Feedback konnten unter Kontrolle der Leseleistung, der intrinsischen Lesemotivation und der lesebezogenen Selbstkonzepte der Kinder keine geschlechtsspezifischen Unterschiede mehr nachgewiesen werden. Der Autor argumentiert, dass Wahrnehmungsunterschiede stärker über die Leistung und motivationale Merkmale erklärt werden können als über das Geschlecht (Hoya, 2021). Einschränkend ist festzuhalten, dass sich die Studie auf den Primarbereich bezieht, sodass nicht zwangsläufig auf die Feedbackwahrnehmung an weiterführenden Schulen geschlossen werden kann. Außerdem werden keine Ergebnisse in Bezug auf geschlechtsspezifische Unterschiede in der Wahrnehmung von leistungsbezogenem Feedback (z.B. zur Aufgabe oder zum Prozess der Aufgabenbearbeitung) berichtet.

Daneben wurden auch einige Fragebogenstudien zur Wahrnehmung von Feedback aus der Sicht von Mädchen und Jungen im internationalen Raum durchgeführt. Eine norwegische Fragebogenstudie zeigt, dass Schüler der Sekundarstufe II

über eine höhere Qualität von Feedback im Unterricht unterschiedlicher Fächer berichten als Schülerinnen (Havnes et al., 2012). Die Qualität von Feedback wurde dabei über verschiedene Merkmale erfasst, wie beispielsweise Länge des Feedbacks, Informationen zum Leistungsstand, summativ versus formativ (Havnes et al., 2012). Auf der Grundlage von Daten der PISA-Studie belegt Sortkær (2018, 2019), dass die Wahrnehmung von Feedback über alle nordischen Staaten hinweg in Abhängigkeit des Geschlechts variiert, wobei Jungen berichten, häufiger Feedback zum Leistungsstand und zum Lernprozess im mathematisch-naturwissenschaftlichen Unterricht zu erhalten als Mädchen. Auch nicht-domänenspezifische Untersuchungen im internationalen Kontext belegen, dass Jungen häufiger negatives Feedback zu ihren Leistungen wahrnehmen als Mädchen (Chen et al., 2011), Mädchen hingegen berichten allgemein mehr leistungsbezogenes Feedback zu erhalten als Jungen (Carvalho et al., 2014). Im Rahmen der Validitätsprüfungen ihres Fragebogens zu unterschiedlichen Dimensionen der Akzeptanz von Feedback aus Sicht von Schülerinnen und Schülern der Sekundarstufe I an Schulen in Singapur weisen Lipnevich und Lopera-Oquendo (2022) auch geschlechtsspezifische Unterschiede in der Feedbackwahrnehmung nach. Demzufolge nehmen Mädchen im Vergleich zu Jungen das Feedback der Lehrkraft insgesamt positiver wahr und berichten über eine höhere Bereitschaft, das Feedback zu nutzen.

Leistung als Prädiktor

Unterschiede in der Unterrichts- und Feedbackwahrnehmung können auch aufgrund des individuellen *Leistungsniveaus* der Schülerinnen und Schüler entstehen (Igeler et al., 2019; Hoya, 2021; Jurik et al., 2015; Seidel, 2006). Konkret wird davon ausgegangen, dass leistungsschwächere Schülerinnen und Schüler den Unterricht anders einschätzen als leistungsstärkere Schülerinnen und Schüler (Stang & McElvany, 2021). Entsprechend zeigen die empirischen Befunde, dass Unterschiede in der Wahrnehmung der Unterrichts- und Feedbackqualität im Deutschunterricht je nach Leistungsniveau bestehen. Ein (negativer) Zusammenhang zwischen der Vorleistung (Deutschnote) und der Einschätzung des Unterrichtsmerkmals „Kooperation und Diskussion“ konnte für den Deutschunterricht der Sekundarstufe I aufgezeigt werden, wonach leistungsschwächere Schülerinnen und Schüler die Kooperation und Diskussion im Unterricht positiver wahrnehmen als leistungsschwächere Schülerinnen und Schüler (Jurik et al., 2015). Dieser Befund steht allerdings im Widerspruch zu den Ergebnissen von Wagner (2008), der von einer positiveren individuellen Unterrichtswahrnehmung von Schülerinnen und Schülern mit höheren Leistungszuwächsen beim Deutschttest berichtet. Auf Klassenebene hängen außerdem bessere Deutschnoten am Ende des Schuljahres mit einer positiveren Unterrichtswahrnehmung zusammen (Wagner, 2008). Im Grundschulunterricht hängen das Vorwissen und die kognitiven Fähigkeiten der Schülerinnen und Schüler mit ihrer Einschätzung der kognitiven Aktivierung zusammen (Igeler et al., 2019). Hoya (2021) konnte belegen, dass die Leseleistung von Grundschulinnen und -schülern

eine stärkere Erklärungskraft für Unterschiede in der Wahrnehmung von positivem und negativem Feedback besitzt als ihr Geschlecht. Teilweise kann unter Hinzunahme der Leistung im Regressionsmodell kein signifikanter Effekt des Geschlechts mehr belegt werden (Hoya, 2021). Ähnlich wie Jurik et al. (2015), die für den Deutschunterricht nachweist, dass eine geringere Leistung mit einer positiveren Unterrichtseinschätzung einhergeht, zeigt auch Sortkær (2018), dass Schülerinnen und Schüler mit einer geringeren Leistung im Naturwissenschaftstest über mehr Feedback berichten als Schülerinnen und Schüler mit einer besseren Leistung im Naturwissenschaftstest. Bei der Erfassung von zwei Dimensionen von Feedback in PISA 2012 zeigt sich außerdem, dass eine höhere Mathematikleistung mit einer geringeren Wahrnehmung von direktivem Feedback (*directive feedback*) und einer höheren Wahrnehmung von selbstregulativem Feedback (*facilitative feedback*) einhergeht (Sortkær, 2019). Dieses Ergebnis könnte darauf hindeuten, dass Lehrkräfte ihr Feedbackverhalten im Sinne einer adaptiven Unterrichtsgestaltung anpassen und leistungsschwächeren Schülerinnen und Schülern mehr Unterstützung in Form von (direktivem) Feedback geben, welches konkrete Informationen zur Verbesserung des Lernprozesses liefert, während leistungsstarke Schülerinnen und Schüler stärker darin gefördert werden, ihren Lernprozess selbst zu kontrollieren (Sortkær, 2018, 2019).

2.5.4 Veränderung der Wahrnehmung – theoretische Annahmen

Neben interindividuellen Unterschieden in der Feedbackwahrnehmung in Abhängigkeit von Schülermerkmalen steht auch die Frage nach der zeitlichen Variabilität der wahrgenommenen Feedbackqualität im Fokus dieser Untersuchung. Damit einher geht zunächst die Frage, inwiefern die Unterrichtsqualität einer Lehrkraft nicht nur Veränderungen in der Lernentwicklung beeinflussen kann, sondern auch selbst einer Veränderung unterliegt (Göllner et al., 2016). In der Unterrichtsforschung wird die Annahme vertreten, dass Unterrichtshandeln einer Lehrkraft eine stabile Größe ist (Göllner et al., 2016). Ebenso wird angenommen, dass Schülereinschätzungen der Unterrichtsqualität über die Zeit stabil sind, weshalb die Erfassung von Unterrichtsmerkmalen in vielen Studien entsprechend nur einmal erfolgt, um Aussagen über die „allgemeine“ Unterrichtsqualität einer Lehrkraft zu machen (Wagner et al., 2016). Dabei wird für bestimmte Unterrichtsmerkmale – wie Klassenführung und konstruktive Unterstützung – eine höhere zeitliche Stabilität angenommen als für andere Unterrichtsmerkmale – wie kognitive Aktivierung –, die stärker von den Unterrichtsinhalten abhängig sind und sich somit von Unterrichtsstunde zu Unterrichtsstunde in ihrer Qualität unterscheiden können (Praetorius et al., 2014; Thomen et al., 2021).

Trotz der angenommenen Stabilität des unterrichtlichen Handelns einer Lehrkraft über die Zeit, können Veränderungen der Unterrichtswahrnehmung nicht ausgeschlossen werden (Curby et al., 2011; Göllner et al., 2016; Karlen, 2016; Wagner et al., 2016) und sind aus unterschiedlichen Gründen anzunehmen:

Erstens sind auf der individuellen Ebene der Schülerinnen und Schüler Veränderungen der Wahrnehmung des Unterrichts über die Zeit denkbar (Ruzek & Schenke, 2019). Entsprechend kommen auch Studien wie die von Wagner et al. (2016) zu dem Ergebnis, dass Varianzen in den Schülereinschätzungen der Unterrichtsqualität über die Zeit eher auf intraindividuelle Unterschiede in den Wahrnehmungen einzelner Schülerinnen und Schüler (92% der Varianz) zurückzuführen sind als auf tatsächliche Unterschiede in der beobachteten Unterrichtsqualität in den unterschiedlichen Klassen (8%). Diese intraindividuellen Unterschiede können dadurch entstehen, dass Schülerinnen und Schüler die Lerngelegenheiten des Unterrichts in Abhängigkeit ihrer individuellen motivationalen, affektiven und kognitiven Merkmale unterschiedlich interpretieren (Ruzek & Schenke, 2019) – eine Annahme, die auf theoretischen Angebots-Nutzungs-Modellen (vgl. Abschnitt 2.5.1) gründet.

Zweitens können zeitliche Variabilitäten in den Wahrnehmungen durch den Zeitraum der Befragung eintreten (Göllner et al., 2016; Wagner et al., 2016). Auf den Kontext der FeeHe-Studie angewendet kann eine Veränderung der Unterrichtswahrnehmung dadurch hervorgerufen werden, dass der zweite Befragungszeitpunkt kurz vor Bekanntgabe der Halbjahresnoten stattfand. Dementsprechend ist es plausibel anzunehmen, dass es aufgrund der bevorstehenden Zeugnisnoten Schwerpunktverschiebungen des Unterrichts gab (Göllner et al., 2016), die sich auch im Feedbackverhalten der Lehrkraft äußern können. Gleichzeitig können sich aufgrund der bevorstehenden Zeugnisnoten auch die Bedürfnisse der Schülerinnen und Schüler an das Feedback und die Unterstützung der Lehrkraft ändern. Beides – das veränderte Lehrkrafthandeln und die veränderten Bedürfnisse der Schülerinnen und Schüler – kann sich in einer Veränderung der Feedbackwahrnehmung niederschlagen. Außerdem spielt der Zeitraum, der zwischen den Messzeitpunkten der Untersuchung der Unterrichtswahrnehmung liegt, eine Rolle für die Veränderbarkeit: Ein kurzer Zeitraum der Untersuchung von einigen Tagen oder wenigen Wochen macht eine „wirkliche Veränderung der Unterrichtsqualität eher unwahrscheinlich“ (Göllner et al., 2016, S. 72); während ein größerer Zeitraum von einem halben oder einem Jahr für die Abbildung von Veränderung geeigneter zu sein scheint (Göllner et al., 2016).

Drittens stellt Unterricht ein komplexes Geschehen dar, das durch eine Vielzahl an Interaktionen zwischen der Lehrkraft und den Schülerinnen und Schülern gekennzeichnet ist. Entsprechend ist anzunehmen, dass unterschiedliche Faktoren der Lernumgebung und der individuellen Schülerinnen und Schüler dazu beitragen können, dass Lehrkräfte die Qualität ihres Unterrichtshandelns im Laufe eines Schuljahres verändern können. So wäre denkbar, dass sie bestimmte Merkmale der Unterstützung und Schülerorientierung – und dazu gehört auch das Erteilen von Feedback – stärker in den Mittelpunkt ihres Unterrichtshandelns stellen, um auf die spezifischen Bedürfnisse der Schülerinnen und Schüler zu reagieren (Curby et al., 2011; Göllner et al., 2016). Insbesondere in der gymnasialen Oberstufe, in der die Fach-

kurse infolge des Kurswahlsystems auf unterschiedlichem Anforderungsniveau unterrichtet werden, sind gewisse Schwerpunktsetzungen im Unterricht als Reaktion auf den jeweiligen Leistungsstand der Schülerinnen und Schüler zu erwarten. Diese Veränderung des unterrichtlichen Handelns einer Lehrkraft könnte sich auch in den Schülerwahrnehmungen widerspiegeln.

Viertens können auch Kontextmerkmale zu Veränderungen der Wahrnehmung von Unterrichtsqualität führen. Mit den Kurswahlen beim Übergang in die gymnasiale Oberstufe können veränderte Bedürfnisse der Schülerinnen und Schüler einhergehen. Sie können das Anspruchsniveau selbst wählen und damit bis zu einem gewissen Grad über die Gestaltung des Lernangebots mitbestimmen. In empirischen Studien zum Kurswahlverhalten konnte nachgewiesen werden, dass Schülerinnen und Schüler vornehmlich Kurse mit erhöhtem Anforderungsniveau wählen, die ihren fachspezifischen Interessen und Fähigkeiten entsprechen (Köller et al., 2000; Nagy et al., 2008; Eckert et al., im Druck). Mit diesen interessensgesteuerten Kurswahlen dürften auch veränderte Erwartungen und Bedürfnisse an das Lernangebot einhergehen, die sich auch in einer Veränderung der Wahrnehmung der Unterrichtsqualität im Laufe der Sekundarstufe II manifestieren könnten.

Bei der Betrachtung von Veränderungen können zwei Dimensionen unterschieden werden (Dietrich, 2019): (1) Auf der Ebene der *intraindividuellen* Dimension wird die Wahrnehmung ein und desselben Schülers bzw. ein und derselben Schülerin zu verschiedenen Zeitpunkten untersucht. (2) Auf der Ebene der *interindividuellen* Dimension werden die Unterschiede in den Wahrnehmungen zwischen den einzelnen Schülerinnen und Schülern zu verschiedenen Zeitpunkten betrachtet. Diese differenzielle Sichtweise berücksichtigt die Variabilität in den Veränderungen und untersucht folglich, inwiefern die individuelle Veränderung der Wahrnehmung von der durchschnittlichen Veränderung aller Schülerinnen und Schüler abweichen und welche Faktoren ggf. interindividuelle Unterschieden im Veränderungsprozess erklären können (Gruber et al., 2014).

Im nachfolgenden Abschnitt werden empirische Befunde zu Veränderungen der Unterrichtswahrnehmung auf diesen beiden Ebenen vorgestellt.

2.5.5 Empirische Befunde zur (differenziellen) Veränderung der Wahrnehmung

Bisher liegen nur wenige Studien vor, die die zeitliche Stabilität und Veränderung der Wahrnehmung von Unterrichtsmerkmalen wie Feedback im Rahmen längsschnittlicher Studien untersuchen. Diese Studien legen den Fokus meistens auf Veränderungen beim Übergang von der Grundschule in die Sekundarstufe I (Lazarides et al., 2019; Lazarides, Fauth et al., 2021; Maulana et al., 2013; Maulana et al., 2016; Praetorius et al., 2017) oder auf Veränderungen im Laufe der Sekundarstufe I (De Wit et al., 2010; Praetorius et al., 2017; Way et al., 2007). Dabei wird auf der Basis von Motivationstheorien (*Stage-Environment-Fit*-Theorie, Eccles et al., 1993; Eccles & Roeser, 2009; *Self-Determination*-Theorie, Ryan & Deci, 2000, vgl. Abschnitt 3.2)

häufig argumentiert, dass die im Laufe der Sekundarstufe I beobachtete Abnahme in der Motivation von Schülerinnen und Schülern mit einer Abnahme in der Qualität des Unterrichts zusammenhängt, was insgesamt auf eine fehlende Passung zwischen den individuellen Bedürfnissen der Schülerinnen und Schüler (z.B. mehr soziale Eingebundenheit und Autonomie) und der Gestaltung der Lernumgebung hindeutet (De Wit et al., 2010; Lazarides et al., 2019; Lazarides, Fauth et al., 2021; Maulana et al., 2016; Way et al., 2007).

Beispielsweise konnten im Rahmen der TRAIN-Studie Veränderungen in der aus Schülersicht wahrgenommenen Unterstützung durch die (Mathematik-)Lehrkraft (Lazarides et al., 2019; Lazarides, Fauth et al., 2021) sowie der Klarheit und des Klassenmanagements (Lazarides, Fauth et al., 2021) nach dem Übergang von der Grundschule in die Sekundarstufe I nachgewiesen werden. Konkret weisen die Ergebnisse der *Latent-Change*-Analysen auf eine signifikante Abnahme der wahrgenommenen Unterrichtsqualitätsmerkmale im Laufe des Schuljahres hin. In Übereinstimmung zeigen auch weitere (internationale) Studien eine Abnahme in der (wahrgenommenen) Unterrichtsqualität in der Sekundarstufe I (Maulana et al., 2013; Maulana et al., 2016; Way et al., 2007; Reddy et al., 2003). Im Einzelnen konnten Maulana et al. (2016) in einer längsschnittlichen Studie mit Schülerinnen und Schülern im ersten Jahr der Sekundarstufe I an niederländischen Schulen nachweisen, dass die wahrgenommene Klarheit, das Klassenmanagement sowie die Steuerung durch die Mathematik- und Englischlehrkräfte aus der Sicht der Schülerinnen und Schüler im Laufe eines Schuljahres sinken. Anhand von latenten Wachstumskurvenmodellen weisen sie außerdem interindividuelle Unterschiede in der Veränderung der Unterrichtswahrnehmung nach. Demnach sind das Geschlecht und das Leistungsniveau der Klassen bedeutsame Prädiktoren von differenziellen Wahrnehmungen und Veränderungen der Unterrichtsqualität. Way et al. (2007) beschreiben in einer längsschnittlichen, domänenunspezifischen Studie an Schulen in den USA, dass die wahrgenommene Unterstützung durch die Lehrkraft in der Schule – auch unter Kontrolle des Geschlechts und des sozioökonomischen Status – zu Beginn der sechsten Jahrgangsstufe zwar hoch ausfällt, im Laufe der Sekundarstufe I aber sinkt. Interindividuelle Unterschiede in der Wahrnehmung der Unterstützung zum ersten Messzeitpunkt, nicht aber Unterschiede in der Veränderung der Wahrnehmung, konnten durch das Geschlecht erklärt werden (Way et al., 2007). De Wit et al. (2010) weisen auch für den Übergang in die *High School* (von Klasse 9 in Klasse 10) mittels latenter Wachstumskurvenmodelle nach, dass die wahrgenommene Unterstützung aus der Sicht der Schülerinnen und Schüler sinkt, wobei diese Veränderung für Jungen stärker ausfällt als für Mädchen.

Insgesamt verdeutlichen die Ergebnisse der längsschnittlichen Untersuchungen, dass sich das Lehrkraftverhalten in Bezug auf unterschiedliche Unterrichtsmerkmale aus der Sicht von Schülerinnen und Schülern im Laufe der Schulzeit verändern kann. Obwohl Befunde in Bezug auf die Veränderung der Feedbackwahrnehmung fehlen, können die Ergebnisse zu längsschnittlichen Entwicklungen von ähnlichen Konstrukten (z.B. Unterstützung) ein Hinweis darauf sein, dass eine Untersuchung

der Veränderung des Feedbackverhaltens aus der Sicht der Schülerinnen und Schüler inhaltlich aufschlussreich sein kann. Einschränkend ist festzuhalten, dass die meisten längsschnittlichen Untersuchungen zu Veränderungen in der wahrgenommenen Unterrichtsqualität fachunspezifisch sind (De Wit et al., 2010; Lazarides, Fauth et al., 2021; Way et al., 2007) und somit keine Rückschlüsse auf bestimmte Fächer zulassen. Zudem beziehen sich die bisherigen Ergebnisse häufig auf Veränderungen der Unterrichtsqualität in der Sekundarstufe I, sodass eine Untersuchung im Deutschunterricht der Sekundarstufe II neue Erkenntnisse in diesem Forschungsbereich liefern kann. Veränderungen der Qualität des Unterrichts – im Speziellen des Feedbacks – sind auch für diesen Lernkontext zu erwarten, da sich durch die Leistungsdifferenzierung und interessensgeleiteten Kurswahlen sowohl die Gestaltung des Feedbacks durch die Lehrkraft als auch die Bedürfnisse der Schülerinnen und Schüler an das Feedback der Lehrkraft ändern können. Beides kann sich in einer veränderten Wahrnehmung von Feedback aus der Sicht der Schülerinnen und Schüler widerspiegeln.

Neben der Analyse der intraindividuellen Veränderung der Feedbackwahrnehmung scheint es auch relevant, Prädiktoren (wie das Geschlecht und das Leistungsniveau) einzubeziehen, welche interindividuelle Unterschiede in der Veränderung der Wahrnehmung erklären können (De Wit et al., 2010; Maulana et al., 2016).

3 Feedback und Motivation

Im Rahmen der vorliegenden Arbeit wird das übergreifende Ziel verfolgt, die Zusammenhänge der aus Schülersicht wahrgenommenen Feedbackqualität mit der Veränderung der intrinsischen Lernmotivation von Schülerinnen und Schülern im Deutschunterricht der gymnasialen Oberstufe zu untersuchen. Gegenstand des vorliegenden Kapitels ist daher der theoretische und empirische Hintergrund zur Lernmotivation sowie deren Zusammenhang zu Feedback im Unterricht. Dazu gliedert sich das Kapitel wie folgt: Zunächst wird der motivationstheoretische Hintergrund dargelegt. Hierzu wird der Begriff der Lernmotivation definiert und deren Bedeutung für den Lernerfolg von Schülerinnen und Schülern herausgestellt (Abschnitt 3.1). Ausgehend hiervon widmet sich Abschnitt 3.2 der Selbstbestimmungstheorie von Deci und Ryan (2000, 2002), wobei auch die postulierten Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit thematisiert werden. Sodann werden in Abschnitt 3.3 aktuelle empirische Befunde zur Entwicklung der selbstbestimmten Motivation näher erörtert. Vor dem Hintergrund des Ziels der vorliegenden Arbeit werden in Abschnitt 3.4 theoretische Bezüge und empirische Befunde zum Zusammenhang zwischen Feedback und der intrinsischen Lernmotivation diskutiert. Im Einzelnen widmet sich Abschnitt 3.4.1 der Bedeutung von Feedback im Unterricht für die Entwicklung der intrinsischen Motivation im Kontext der Selbstbestimmungstheorie. In Abschnitt 3.4.2 wird im Rahmen von theoretischen Angebots-Nutzungs-Modellen und aktuellen empirischen Befunden der

Unterrichtsforschung auf das reziproke Verhältnis zwischen der Feedbackwahrnehmung und der intrinsischen Lernmotivation sowie deren Veränderungen im Laufe der Schulzeit eingegangen.

3.1 Konzeption von Lernmotivation und Bedeutung für das Lernen

Die Motivation von Schülerinnen und Schülern gilt als eine wichtige Voraussetzung für positive Lernentwicklungen (Cerasoli et al., 2014; Spinath & Spinath, 2005; Wigfield & Cambria, 2010) und wird daher neben fachlichen und überfachlichen Kompetenzen als eigenständiges Ziel der schulischen Bildung deklariert (Kunter, 2005; Rakoczy et al., 2022; Willems, 2011). Ein effektiver Unterricht schlägt sich demnach nicht nur in den kognitiven Lernergebnissen wieder, sondern ist auch durch eine erhöhte Lernmotivation gekennzeichnet (Kunter, 2005). Im schulischen Kontext wird die Lernmotivation als die individuelle Bereitschaft verstanden, zielgerichtet Lernhandlungen aufzunehmen und aufrechtzuerhalten (Krapp et al., 2014; Schiefele et al., 2018; Schiefele & Schaffner, 2020; Willems, 2018). Ausgehend von dieser Konzeption und je nach Art der intendierten Ziele kann zwischen der intrinsischen und der extrinsischen Lernmotivation differenziert werden (Ryan & Deci, 2000, 2017, 2020). Die intrinsische Motivation beschreibt den Wunsch oder die Absicht, eine bestimmte Lernhandlung auszuführen, weil damit positive Erlebensaspekte (u.a. Freude und Spaß) verbunden sind. Weiterhin kann zwischen einer *tätigkeitszentrierten* und *gegenstandszentrierten* Form der intrinsischen Motivation unterschieden werden (Schiefele & Streblov, 2005; Schiefele et al., 2018; Schiefele & Schaffner, 2020). Die *tätigkeitszentrierte* Form ist dadurch gekennzeichnet, dass Lernende eine Lernhandlung ausführen, weil sie Freude an einer bestimmten Aktivität (z.B. Lesen) zeigen. Die *gegenstandszentrierte* Form der intrinsischen Motivation charakterisiert Lernende, die Interesse an dem Lerngegenstand (z.B. Thema eines Textes) zeigen. Allerdings konnte die Trennung zwischen diesen theoretisch angenommenen Komponenten der intrinsischen Motivation empirisch nicht belegt werden, was dafür spricht die intrinsische Motivation als einheitliches Konstrukt zu betrachten (vgl. Schiefele & Schaffner, 2020).

Im Gegensatz zur intrinsischen Motivation wird extrinsisch motiviertes Lernen als der Wunsch oder die Absicht aufgefasst, eine Lernhandlung auszuführen, weil damit positive Folgen herbeigeführt oder negative Konsequenzen (z.B. schlechte Noten) vermieden werden können (Ryan & Deci, 2000). Für den schulischen Kontext wird angenommen, dass die intrinsische Lernmotivation aufgrund von curricularen Anforderungen, die den Lernenden von außen vorgegeben werden, nicht die dominante Motivationsqualität darstellt (Kunter, 2005). Umso bedeutsamer erscheint es, dass der Unterricht so gestaltet ist, dass er es den Lernenden ermöglicht, sich mit den Inhalten und Zielen des Unterrichts zu identifizieren und schließlich intrinsische Motivation zu entwickeln (Rakoczy et al., 2008; Rakoczy et al., 2022).

Im schulischen Kontext zeigt sich insbesondere die intrinsische bzw. selbstbestimmte Motivation als förderlich für erfolgreiche Lernprozesse. Entsprechend finden sich auch zahlreiche empirische Belege für den positiven Zusammenhang zwischen der intrinsischen Motivation und der schulischen Leistung (u.a. Burton et al., 2006; Hayenga & Corpus, 2010; Lepper et al., 2005). Demgegenüber wurden für die extrinsische Motivation vereinzelt sogar negative Leistungseffekte nachgewiesen (Corpus et al., 2009; Hayenga & Corpus, 2010; Lepper et al., 2005; Taylor et al., 2014). Neben dem Einfluss auf kognitive Merkmale geht die intrinsische Motivation auch mit einer höheren Ausdauer und Intensität bei der Bearbeitung von Lernaufgaben einher (Vallerand et al., 1997; Vollmeyer & Rheinberg, 2000) und fördert den Einsatz von Lernstrategien, die mit einer tieferen Verarbeitung der Lerninhalte einhergehen (Pintrich & Groot, 1990; Schiefele & Schreyer, 1994; Walker et al., 2006). Die intrinsische Motivation ist auch für den Deutschunterricht besonders relevant. So wird beispielsweise davon ausgegangen, dass eine hohe Motivation für das Lesen im Deutschunterricht das Leseverhalten (im Sinne der Lesemenge bzw. Lesehäufigkeit) auch außerhalb des Deutschunterrichts begünstigt, was sich positiv auf die Lesekompetenz auswirkt (Guthrie & Wigfield, 1999; McElvany et al., 2008; Möller & Schiefele, 2004). Die Lesekompetenz wiederum beeinflusst nicht nur den schulischen Lernerfolg, sondern ist auch eine über den schulischen Kontext hinausgehende wichtige Voraussetzung für den beruflichen Erfolg sowie die Bewältigung des alltäglichen Lebens (McElvany et al., 2008). Mit Blick auf die Ergebnisse internationaler Schulleistungsstudien wie PISA des Jahres 2018 sowie IGLU des Jahres 2016 zeigt sich allerdings, dass die intrinsische Motivation im Deutschunterricht vergleichsweise gering ausgeprägt ist (Diedrich et al., 2019; Goy et al., 2017).

3.2 Die Selbstbestimmungstheorie

Ein theoretischer Ansatz zur Erklärung der Entwicklung der intrinsischen Motivation bei Lernenden ist die Selbstbestimmungstheorie (*Self-Determination-Theory*, SDT; Deci & Ryan, 1993, 2000; Ryan & Deci, 2000, 2017, 2020). In der SDT werden verschiedene Formen oder Qualitäten der Motivation unterschieden, die hinsichtlich des subjektiv wahrgenommenen Orts der Handlungssuche (internal vs. external) und somit im Grad der wahrgenommenen Selbst- oder Fremdbestimmtheit variieren (Deci & Ryan, 2000, 2020). Je nach Grad der erlebten Selbstbestimmung können die Qualitäten der Motivation auf einem Kontinuum abgebildet werden, wobei die intrinsische und die extrinsische Motivation die äußeren Pole des Motivationskontinuums darstellen (vgl. Abbildung 6).

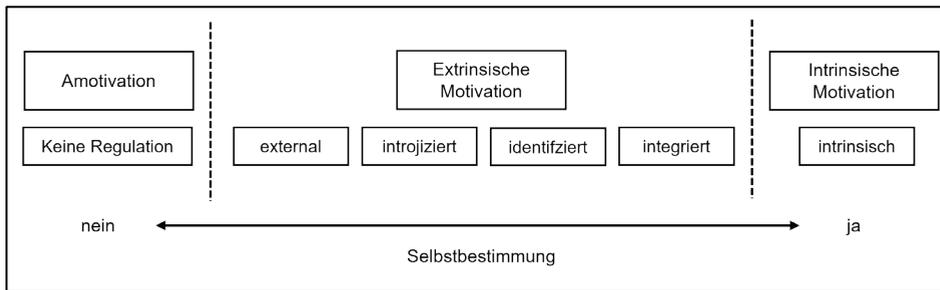


Abbildung 6: Motivationsformen, Regulationsstile und erlebte Selbstbestimmung (Buff et al., 2010, S. 257 in Anlehnung an Ryan & Deci, 2000, 2002)

Während sich intrinsisch motiviertes Handeln durch eine hohe Ausprägung der Selbstbestimmung auszeichnet, ist extrinsisch motiviertes Handeln durch die geringste Ausprägung der Selbstbestimmung gekennzeichnet. Bei der extrinsischen Motivation gibt es aber auch unterschiedliche Abstufungen – von gänzlich fremdbestimmter, von außen kontrollierter externaler Regulation bis hin zur stärker selbstbestimmten, integrierten Regulation (Schiefele & Schaffner, 2020). Diese vier Stufen, die als *externale*, *introjierte*, *identifizierte* und *integrierte* Verhaltensregulation bezeichnet werden, unterscheiden sich nach ihrem jeweiligen Grad der wahrgenommenen Fremd- bzw. Selbstbestimmung (Deci & Ryan, 2000, 2002; Ryan & Deci, 2020):

- Die *externale Regulation* ist durch das höchste Ausmaß der Fremdbestimmung gekennzeichnet. Das Verhalten wird ausschließlich durch externe Handlungsziele reguliert. Ein Schüler bzw. eine Schülerin erfüllt eine Lernhandlung nur, um eine Belohnung (z.B. gute Note) zu erhalten oder negative Konsequenzen (z.B. schlechte Note) zu vermeiden.
- Die *introjierte Regulation* ist eine eher fremdbestimmte Form der extrinsischen Motivation und liegt vor, wenn eine Person ein Handlungsziel verfolgt, ohne sich damit persönlich zu identifizieren bzw. diese als persönlich wichtig zu erachten. Ein Schüler bzw. eine Schülerin erfüllt Lernhandlungen nur, um beispielsweise den Erwartungen der Eltern oder der Lehrkraft zu entsprechen oder von den Mitschülerinnen und -schülern anerkannt zu werden.
- Die *identifizierte Regulation* ist eine stärker selbstbestimmte Form der extrinsischen Motivation und liegt vor, wenn Lernende die ursprünglich von außen vorgegeben Lernziele als persönlich wichtig erachten und sich damit identifizieren. Beispielsweise übt ein Schüler bzw. eine Schülerin eine Lernhandlung aus, um später eine bestimmte Ausbildung machen zu können. Allerdings kann die Lernhandlung weiterhin unverbunden im Selbstkonzept der Lernenden repräsentiert sein.

- Die *integrierte Regulation* stellt die höchste Ausprägungsform der auf Selbstbestimmung beruhenden extrinsischen Motivation dar. Sie ist dadurch gekennzeichnet, dass sich Lernende nicht nur mit ursprünglich external bestimmten Lernzielen und Aufgabenstellungen identifizieren, sondern diese auch in Einklang mit persönlichen Zielen und Werten bringen und ohne Konflikte mit anderen persönlichen Zielen in ihr Selbstkonzept integrieren (Krapp et al., 2014; Schiefele & Schaffner, 2020). Die integrierte Regulation weist eine hohe Ähnlichkeit zur intrinsischen Motivation auf, allerdings wird sie als extrinsische Motivationsform betrachtet, da die Handlung nicht eigeninitiiert ist (wie bei der intrinsischen Motivation), sondern ursprünglich von außen vorgegeben wurde (Deci & Ryan, 2000; Willems, 2011).

Personenzentrierte Untersuchungen, in denen die wechselseitigen Einflüsse verschiedener Motivationsformen in Form von latenten Profilen betrachtet werden, zeigen, dass fremdbestimmte und selbstbestimmte Motivationsformen sich nicht gegenseitig ausschließen (Ratelle et al., 2007; Willems & Dreiling, 2022; Wormington et al., 2012; Dreiling & Willems, im Druck). Vielmehr können sie innerhalb von Lernenden zeitgleich auftreten. Im Hinblick auf den Lernerfolg ist das Verhältnis dieser beiden Motivationsqualitäten entscheidend: Die Studien weisen dabei übereinstimmend nach, dass Lernende mit einem Motivationsprofil, das durch ein hohes Maß an selbstbestimmter Motivation gekennzeichnet ist, über höhere Ausprägungen lernrelevanter Merkmale (wie z.B. Schulleistung, Prüfungsangst, Schulzufriedenheit, Anstrengungsbereitschaft, Verarbeitungstiefe) verfügen als Schülerinnen und Schüler mit einem Motivationsprofil, bei dem die fremdbestimmte Motivation höher ausgeprägt ist (Ratelle et al., 2007; Vansteenkiste et al. 2009; Wormington et al. 2012).

Aus pädagogischer Sicht kann die Entwicklung oder Transformation von eher fremdbestimmten Formen der extrinsischen Motivation hin zu stärker selbstbestimmten Formen der Lernmotivation, die mit Prozessen der Internalisierung und Integration von Handlungszielen verknüpft sind (Ryan & Deci, 2000), „von außen“ gefördert werden, z.B. durch die Gestaltung der Lernumgebung und ein unterstützendes Lehrkraftverhalten. Die Wirkung der Lernumgebung auf die Entwicklung der intrinsischen oder selbstbestimmten Motivation vom Lernenden wird im Rahmen der SDT durch die Erfüllung der grundlegenden Bedürfnisse, den *basic needs* erklärt (Deci & Ryan, 1993; Deci & Ryan, 2002; Ryan & Deci, 2000):

- Das *Autonomieerleben* äußert sich im Bestreben des Lernenden, über die Ziele, Inhalte und Anforderungen von Lernaufgaben mitbestimmen zu können.
- Das *Kompetenzerleben* äußert sich durch den Wunsch des Lernenden, die Anforderungen von Aufgabenstellungen aus eigener Kraft bewältigen zu können und noch fehlende Kompetenzen erwerben zu können.

- Schließlich hat der Lernende das Bestreben nach *sozialer Eingebundenheit*, also dem Gefühl, von seinen Mitschülerinnen und Mitschülern sowie seiner Lehrkraft akzeptiert und anerkannt zu sein (Krapp et al., 2014).

Lernkontexte unterscheiden sich darin, inwiefern sie die Erfüllung der drei Grundbedürfnisse unterstützen. Eine Lernumgebung, in der Lernende sich selbst als kompetent erleben, sich dabei in ihren Wahlmöglichkeiten und Entscheidungen autonom erleben können und sich gleichzeitig als Teil einer Lerngruppe wahrnehmen, ermöglicht es dem Lernenden, sich mit Inhalten und Lernzielen zu identifizieren und schließlich selbstbestimmte Formen der Motivation zu entwickeln (Rakoczy et al., 2022). Entsprechend finden sich empirische Belege dafür, dass Schülerinnen und Schüler, die eine hohe Ausprägung von Kompetenz- und Autonomieunterstützung durch die Lehrkraft wahrnehmen und sich zur Klasse zugehörig fühlen, auch über eine höhere selbstbestimmte Motivation berichten (Lazarides et al., 2015; Ntoumanis et al., 2009; Rakoczy et al., 2022; Rakoczy, Klieme & Pauli, 2008).

Des Weiteren konnte die *vermittelnde* Rolle der (wahrgenommenen) Erfüllung der Bedürfnisse nach Kompetenz und Autonomie für den Zusammenhang zwischen Unterrichtsmerkmalen und der selbstbestimmten Motivation von Lernenden in unterschiedlichen Fächern bereits empirisch nachgewiesen werden (z.B. Mouratidis et al., 2013; Rakoczy et al., 2013; Rakoczy et al., 2022; Ruzek et al., 2016). Dabei wird das motivationsrelevante Kompetenz- und Autonomieerleben als Mediatorvariablen in die statistischen Analysen einbezogen (Jansen et al., 2022; Lederberger, 2015; Lepper et al., 2022; Rakoczy, 2008; Rakoczy et al. 2013; Rakoczy et al., 2022; Willem, 2011). Diese Modellierung basiert auf Annahmen von theoretischen Angebots-Nutzungs-Modellen, in denen neben kognitiven Lernaktivitäten auch das motivationsrelevante Selbstbestimmungserleben der Angebotsnutzung zugeordnet werden, die wiederum den Zusammenhang zwischen der Unterrichtsqualität und dem Lernen mediiert (Klieme & Rakoczy, 2008; Seidel, 2014; Vieluf et al., 2020).

3.3 Empirische Befunde zur Entwicklung der Lernmotivation

Trotz der Bedeutung der intrinsischen Motivation für das Lernen zeigen nationale und internationale Studien wiederholt, dass die durchschnittliche fachspezifische Lernmotivation von Schülerinnen und Schülern, aber auch die allgemeine schulbezogene Motivation im Laufe der Schulzeit sinkt (*national*: u.a. McElvany et al., 2008; Lazarides et al., 2019; Spinath & Spinath, 2005; Spinath & Steinmayr, 2008; *international*: u.a. Corpus et al., 2009; Gnamb & Hanfstingl, 2015; Gottfried et al., 2001; Gottfried et al., 2007; Jacobs et al., 2002; Lepper et al., 2005; Otis et al., 2005).

Veränderungen in der Ausprägung der intrinsischen Lernmotivation können vor allem in Übergangsphasen beobachtet werden – dabei ist die Abnahme der intrinsischen Motivation insbesondere für den Übergang von der Grundschule in die Sekundarstufe I gut belegt (Dietrich et al., 2015). In einer internationalen Studie in den USA berichten Otis et al. (2005) außerdem von einem Motivationsabfall beim Übergang von der *Junior High School* (Klassenstufe 8) in die *Senior High School* (Klassenstufe 9 und 10), was in etwa vergleichbar mit dem Übergang von der Sekundarstufe I in die Sekundarstufe II ist. Gleichzeitig zeigt sich, dass die Entwicklung der Lernmotivation mit höherem Alter – ab ca. 16 Jahren – durch Stagnation geprägt ist, d.h. deutliche Veränderungen treten hier nicht mehr auf (Gottfried et al., 2001; Gottfried et al., 2007; Jacobs et al., 2002; Spinath & Steinmayr, 2012). Gottfried et al. (2007) weisen nach, dass sich die mathematikbezogene, intrinsische Motivation von Schülerinnen und Schülern ab dem Alter von 16 Jahren nicht mehr so stark verändert; im Vergleich zur starken Abnahme in den jüngeren Jahren. Dieser Trend wurde auch für den Deutschunterricht belegt: In einer Oberstufenstudie von Spinath und Steinmayr (2012), in der bei Schülerinnen und Schülern zu Beginn und am Ende der 11. Jahrgangsstufe u.a. die Motivation für das Lernen im Fach Deutsch erfasst wurde, zeigt sich in Übereinstimmung mit den genannten internationalen Befunden, dass sich die durchschnittliche intrinsische Motivation in diesem Zeitraum nicht mehr bedeutsam verändert.

Neben diesen Befunden, die eher die durchschnittliche Entwicklung der Motivation bei allen Schülerinnen und Schülern untersuchen, finden sich auch Belege, die Unterschiede zwischen unterschiedlichen Gruppen von Schülerinnen und Schülern aufzeigen. So treten differenzielle Entwicklungsverläufe in der Lernmotivation in Abhängigkeit des Geschlechts auf. Im Fach Deutsch entwickelt sich die Motivation im Zeitverlauf bei Jungen tendenziell ungünstiger als bei Mädchen (Schwabe et al., 2021). Außerdem weist eine Vielzahl von Studien mit einer *personenzentrierten* Forschungsrichtung nach, dass Motivationsverläufe über die Zeit in Abhängigkeit des jeweiligen „Motivationsprofils“ bzw. „Motivationstyps“ der Schülerinnen und Schüler unterschiedlich verlaufen (Corpus & Wormington, 2014; Hayenga & Corpus, 2010; Lazarides et al., 2020; Marcoulides et al., 2008; Dreiling & Willems, im Druck). Spezifische Befunde für die Oberstufe verdeutlichen dabei eine hohe Stabilität der Motivationsprofile, was konkret bedeutet, dass Schülerinnen und Schüler ihr Motivationsprofil im Laufe der Schulzeit nicht mehr wesentlich wechseln (Lazarides et al., 2020; Marcoulides et al., 2008; Dreiling & Willems, im Druck). Auf der anderen Seite treten für einzelne Oberstufenschülerinnen und -schüler auch noch systematische Veränderungen in der Profilverzögerlichkeit auf, wobei sie eher in Profile mit ungünstigen Ausprägungen in motivationalen Merkmalen wechseln (Dreiling & Willems, im Druck). Diese Schülerinnen und Schüler – auch als Profilwechsler oder *mover* bezeichnet – folgen also dem Trend der Abnahme der intrinsischen Motivation (Dreiling & Willems, im Druck). In der Studie konnten außerdem geschlechtsspezifische Unterschiede in der Motivationsentwicklung nachgewiesen werden. Die Geschlechtsunterschiede in der Profilverzögerlichkeit verstärkten sich

über die Zeit: Mädchen waren im Vergleich zu Jungen weniger wahrscheinlich dazu geneigt, in ein Profil mit geringerer intrinsischer Motivation zu wechseln und sie konnten mit höherer Wahrscheinlichkeit über die Zeit ein Profil mit hoher intrinsischer Motivation aufrechterhalten (Dreiling & Willems, im Druck). Auch in einer anderen personenzentrierten Untersuchung wurde beschrieben, dass Schülerinnen im Laufe der Sekundarstufe II mit höherer Wahrscheinlichkeit in Profile mit hoher Motivation in Bezug auf sprachliche Fächer wechseln als Schüler (Lazarides, Dicke et al., 2021).

Darüber hinaus konnten *interindividuelle* Unterschiede in der *intraindividuellen* Motivationsveränderung über die Zeit in längsschnittlich angelegten *variablenzentrierten* Untersuchungen nachgewiesen werden. Diese verdeutlichen, dass die Veränderung der Lernmotivation nicht bei allen Schülerinnen und Schülern gleichermaßen auftritt (Gottfried et al., 2007; Gottfried et al., 2017; Spinath & Steinmayr, 2008; Weidinger et al., 2015). Aus methodischer Sicht kommen dabei häufig Formen der längsschnittlichen Strukturgleichungsmodellierung, wie *Latent Change Modeling* (LCM) oder *Latent Growth Modeling* (LGM), zur Anwendung. Gottfried et al. (2007) untersuchten die Motivationsentwicklung bei Kindern und Jugendlichen im Alter von 9 bis 17 Jahren im Fach Mathematik. Mittels LGM konnten sie aufzeigen, dass die intrinsische Motivation im Laufe der Schulzeit sinkt. Allerdings verweisen die signifikanten Varianzen in der Ausgangsmotivation und der latenten Veränderungsvariable auf Unterschiede zwischen den Schülerinnen und Schülern. Vertiefende Analysen zeigten, dass die Mathematikleistung der Schülerinnen und Schüler ein bedeutsamer Prädiktor der Motivationsveränderung im Laufe der Schulzeit ist. Die Schülerinnen und Schüler entwickelten sich also in Abhängigkeit ihrer Leistung im Fach Mathematik unterschiedlich. Geringe Mathematikleistungen scheinen mit einer Abnahme in der intrinsischen Motivation einherzugehen (Gottfried et al., 2007).

Auch für das Fach Deutsch konnten interindividuelle Unterschiede in der zeitlichen Entwicklung der intrinsischen Motivation für Schülerinnen und Schüler nachgewiesen werden; allerdings beziehen sich die Ergebnisse auf die Primarstufe (Weidinger et al., 2015). Weidinger et al. (2015) berichten, dass die durchschnittliche intrinsische Motivation vom Beginn der 3. bis zum Ende der 4. Klassenstufe zwar sinkt, aber mittels LGM weisen sie auch Unterschiede zwischen den Kindern in der Motivationsabnahme nach. In der Untersuchung konnten die Autorinnen nur geringe Effekte der Deutschnote auf die Veränderung der intrinsischen Motivation beobachten.

Zusammenfassend zeigt der Forschungsstand, dass die intrinsische Motivation ab dem Alter von 16 Jahren stabil bleibt. Dieser Befund bezieht sich aber auf die mittlere Ausprägung der intrinsischen Motivation. Ausgehend von differenzierten Befunden kann davon ausgegangen werden, dass interindividuelle Unterschiede in intraindividuellen Veränderungen der intrinsischen Motivation auch in den höheren Altersstufen auftreten. Das bedeutet, dass sich die intrinsische Motivation nicht bei allen Schülerinnen und Schülern gleichermaßen entwickelt. Bislang ist noch nicht

ausreichend geklärt, welche Merkmale der Schülerinnen und Schüler und des Unterrichts diese Unterschiede in der Motivationsveränderung erklären können. Der Forschungsstand gibt Hinweise darauf, dass das Geschlecht und der Leistungsstand in Form der Deutschnote relevante Prädiktoren interindividueller Unterschiede in der Motivationsveränderung darstellen. Der Fokus bisheriger Studien liegt nach wie vor auf der Untersuchung der Motivationsveränderung in der Primar- oder Mittelstufe im Mathematikunterricht. Der spezifische Beitrag der vorliegenden Arbeit liegt in der Analyse interindividueller Unterschiede in intraindividuellen Veränderungen der intrinsischen Motivation bei Schülerinnen und Schülern im Deutschunterricht im Laufe der Sekundarstufe II.

3.4 Forschungsüberblick zum Zusammenhang zwischen Feedback und Motivation

3.4.1 Die Bedeutung von Feedback für die Entwicklung der Lernmotivation

Der in Abschnitt 3.3 berichtete Forschungsstand gibt Hinweise darauf, dass es interindividuelle Unterschiede in der Veränderung der intrinsischen Motivation gibt (u.a. Weidinger et al., 2015). Bei einigen Schülerinnen und Schülern ist die beobachtete Abnahme der intrinsischen Motivation somit stärker ausgeprägt als bei anderen, während bei einzelnen Schülerinnen und Schülern die Abnahme nicht auftritt. Aufgrund der Bedeutung der intrinsischen Motivation für das Lernen und die Leistung erscheint es notwendig, Merkmale des Unterrichts zu identifizieren, die diese interindividuellen Unterschiede in der Veränderung der intrinsischen Motivation erklären. Auf der Basis der SDT kann angenommen werden, dass die Entwicklung der intrinsischen oder selbstbestimmten Motivation damit zusammenhängt, inwieweit die Lernumwelt die Wahrnehmung der Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit unterstützt (Ryan & Deci, 2002; Schiefele, 2014). Vor diesem Hintergrund kann die Motivationsabnahme im Laufe der Schulzeit damit begründet werden, dass das mit steigendem Alter wachsende Bedürfnis nach Selbstbestimmung nicht zu der restriktiven Lernumwelt der Schule passt, in der extrinsische Anreize (z.B. Noten) eine größere Rolle einnehmen und der lehrkraftzentrierte Unterricht wenig Raum für die Entfaltung der Selbstbestimmung bietet (Schiefele & Schaffner, 2020). In Übereinstimmung mit diesen Annahmen demonstrieren Gnams und Hanfstingl (2016) in einer längsschnittlichen Studie mit Schülerinnen und Schülern zwischen 11 und 16 Jahren, dass die Abnahme im motivationsrelevanten Erleben der Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Bezogenheit mit der Abnahme der intrinsischen Motivation im Laufe der Schulzeit zusammenhängt. Demzufolge kann theoretisch davon ausgegangen werden, dass sich Merkmale der Lernumgebung positiv auf die Entwicklung der

intrinsischen Motivation auswirken und dass dieser Zusammenhang über die individuelle Wahrnehmung der Grundbedürfnisse vermittelt wird.

Motivationsförderliche Wirkungen von Feedback können mit der *Cognitive Evaluation Theory* (CET; Deci et al., 1999) erklärt werden. Die CET ist eine Subtheorie der Selbstbestimmungstheorie über die Determinanten der intrinsischen Motivation. Nach der CET wird die intrinsische Motivation durch das Erleben von Kompetenz gefördert. Die Theorie postuliert weiterhin, dass die intrinsische Motivation durch Bedingungen der Lernumwelt beeinflusst werden kann. Besonders förderlich wirkt sich ein informierendes Feedback auf die Motivation aus, da es durch die Information über die individuellen Leistungen der Lernenden das grundlegende Bedürfnis nach Kompetenz unterstützt (Deci & Ryan, 2000; Deci et al., 1999, 2001; Ryan & Deci, 2000). Das Feedback sollte dabei informierend gestaltet sein, indem es aufzeigt, wie nahe Lernende dem Lernziel sind und wie sie eine Aufgabe effektiver erledigen können (Klieme et al., 2010; Lipowsky, 2020; Rakoczy et al., 2013; Ryan & Deci, 2002). „Kontrollierendes Feedback hingegen betont, wie sich Lernende (hätten) verhalten sollen“ (Klieme et al., 2010, S. 66). Es führt häufig zu mehr Überwachung und Wettbewerb, was als Bedrohung des Bedürfnisses nach Autonomie wahrgenommen werden und die intrinsische Motivation beeinträchtigen kann (Hattie & Clarke, 2019; Klieme et al., 2010).

Übereinstimmend zeigen empirische Studien, dass informierendes Feedback in Bezug auf den Leistungsstand die intrinsische Motivation erhöht und dass dieser Effekt über das wahrgenommene Kompetenzerleben vermittelt wird (Gniewosz et al., 2014; Vallerand & Reid, 1984). Im Gegensatz dazu führt negatives Feedback, das dem Lernenden eine geringe Kompetenz rückmeldet, zu einer Abnahme der intrinsischen Motivation (Deci & Ryan, 2000; Weidinger et al., 2016). Das positive Feedback wird dabei häufig über eine gute Note und negatives Feedback über eine schlechte Note operationalisiert (Weidinger et al., 2015, 2016). Einschränkend ist festzuhalten, dass Noten als Leistungsrückmeldung den Lernenden nur wenig Informationen zum Lernfortschritt in Verbindung mit einem angestrebten Lernziel aufzeigt und auch keine Hinweise für die erfolgreiche und eigenständige Aufgabebearbeitung liefert (Rakoczy et al., 2013). So konnte Weidinger et al. (2015) geringe direkte und keine über das wahrgenommene Kompetenzerleben vermittelten Effekte von Noten (als Form der Leistungsrückmeldung) auf die Veränderung der intrinsischen Motivation bei Grundschülerinnen und -schülern im Deutschunterricht nachweisen. Die Autorinnen schlussfolgern, dass Noten nur eine untergeordnete Rolle in der Erklärung von differenziellen Entwicklungen der intrinsischen Lernmotivation in der Grundschule spielen und dass andere Formen von Feedback für die Veränderung der intrinsischen Lernmotivation eine höhere Relevanz einnehmen (vgl. auch Ryan & Deci, 2020). Empirische Befunde weisen dahingehend nach, dass sich prozess-orientiertes Feedback – vermittelt über das Kompetenzerleben – positiv auf die Entwicklung der intrinsischen Motivation auswirkt (Hondrich et al., 2018; Kiemer et al., 2015; Levesque et al., 2004; Pat-El et al., 2012; Rakoczy et al., 2013; Rakoczy, Klieme, Bürgermeister & Harks, 2008). So berichten

Hondrich et al. (2018) im Detail u.a. über direkte Effekte auf die intrinsische Motivation und das Kompetenzerleben von einem prozessorientierten Feedback, welches Schülerinnen und Schüler im naturwissenschaftlichen Grundschulunterricht erhalten. Außerdem mediiert das Kompetenzerleben die Wirkung von Feedback auf die intrinsische Motivation. Auch in der Studie von Rakoczy et al. (2013) wird der Effekt von prozessorientiertem Feedback auf das Interesse der Schülerinnen und Schüler im Mathematikunterricht der Sekundarstufe I über die wahrgenommene Kompetenzunterstützung vermittelt. Einzig in der Studie von Pat-El et al. (2012) wird das Feedback aus der subjektiven Sicht von Schülerinnen und Schülern erfasst, um diese Mediationseffekte empirisch nachzuweisen. Vergleichbare Studien für den Deutschunterricht oder die Sekundarstufe II liegen allerdings bisher nicht vor.

Vor dem Hintergrund dieser empirischen Forschungsbefunde untersucht die vorliegende Arbeit den Effekt von „informativen“ Dimensionen von Feedback auf die Entwicklung der intrinsischen Lernmotivation, sowie abgeleitet von der Selbstbestimmungstheorie die medierende Wirkung des Kompetenzerlebens.

3.4.2 Reziproke Effekte von Feedback und Lernmotivation

Mit Blick auf die Forschung zum Zusammenhang zwischen Feedback und Motivation zeigt sich, dass die Zusammenhänge häufig eindirektional untersucht werden, wobei Feedback typischerweise als ein Prädiktor der Motivationsentwicklung modelliert wird. Aus theoretischer und praktischer Perspektive sollten vielmehr wechselseitige Einflüsse der beiden Konstrukte in Betracht gezogen werden: So kann davon ausgegangen werden, dass nicht nur Feedback auf die Motivationsentwicklung der Schülerinnen und Schüler wirkt. Ebenso ist denkbar, dass auch die Schülerinnen und Schüler einer Klasse mit ihrer individuellen Motivation (u.a.) die Gestaltung von Feedback und dessen Wahrnehmungs- sowie Verarbeitungsprozesse beeinflussen (Willems & Dreiling, 2022). Auf dieser Basis kann angenommen werden, dass durch die individuellen Lernvoraussetzungen (wie die Motivation) von Lernenden sogenannte „Mikro-Lernumwelten“ (Jurik et al., 2015, S. 696) entstehen (vgl. auch Seidel et al. 2016), die dazu führen, dass Schülerinnen und Schüler innerhalb einer Klasse den gleichen Unterricht (und das Feedback) unterschiedlich wahrnehmen. Vermittelt über die unterschiedliche Nutzung des Unterrichtsangebots (z.B. Kompetenzerleben) ergeben sich differenzielle Wirkungen auf die Lernentwicklungen. Diese Annahme wird auch theoretisch gestützt: In Angebots-Nutzungs-Modellen der pädagogisch-psychologischen Unterrichtsforschung wird hervorgehoben, dass Unterricht kein eindirektionales Geschehen ist, sondern wechselseitige Interaktionen zwischen Unterricht und Lernenden beinhaltet (Lipowsky, 2020). Dieses reziproke Verhältnis wird in verschiedenen Angebots-Nutzungs-Modellen (Kleickmann et al., 2019; Lipowsky, 2020; Reusser & Pauli, 2010; Seidel,

2014; Vieluf et al., 2020) durch beidseitige Pfeile zwischen dem Feld „Unterrichtsangebot“ und dem Feld „Wahrnehmung und Nutzung“ des Unterrichtsangebots durch die Lernenden abgebildet (vgl. Abbildung 5 in Abschnitt 2.5.1).

Entsprechend verweisen aktuelle empirische Studien darauf, dass die Qualität von Unterrichtsmerkmalen wie Feedback nicht nur ein bedeutsamer Prädiktor der Motivation von Schülerinnen und Schülern ist. Die Wahrnehmung der Unterrichts- und Feedbackqualität wird ihrerseits durch die Eingangsmotivation und andere individuelle Lernvoraussetzungen der Schülerinnen und Schüler beeinflusst (Iglar et al., 2019; Jurik et al., 2015; Ruelmann et al., 2021; Schenke et al., 2018; Willems, 2022a; Willems & Dreiling, 2022). Dabei wurde insbesondere herausgestellt, dass die Qualität von Unterrichtsmerkmalen wie Feedback umso positiver eingeschätzt wird, je höher motivationale Merkmale (z.B. intrinsische Motivation, Fachinteresse, Selbstkonzept) ausgeprägt sind. Dies könnte ein Hinweis auf interindividuelle Unterschiede in den Unterrichtswahrnehmungen sein oder aber auch auf eine an die Bedürfnisse und Lernvoraussetzungen der Schülerinnen und Schüler angepasste Unterrichtsgestaltung der Lehrkraft zurückzuführen sein (Willems & Dreiling, 2022). Jedoch gibt es bislang nur wenige Studien, die diese reziproken Zusammenhänge zwischen Unterrichtswahrnehmungen und der Motivation von Schülerinnen und Schülern simultan betrachten. Insbesondere fehlt es an Studien, die auf der Individualebene untersuchen, inwiefern Veränderungen der subjektiven Unterrichtswahrnehmung von Schülerinnen und Schülern mit der Veränderung ihrer individuellen Motivation zusammenhängen (Ruzek & Schenke, 2019). Eine solche Betrachtung von intraindividuellen Veränderungen erfordert längsschnittliche Daten zur individuellen Unterrichtswahrnehmung und Motivation von einzelnen Schülerinnen und Schülern (Ruzek & Schenke, 2019).

Die Studie von Lazarides et al. (2019) widmet sich dieser Fragestellung in einer längsschnittlichen Studie, in der Schülerinnen und Schüler zu Beginn der fünften Klasse und nach einem Schuljahr im Mathematikunterricht zur wahrgenommenen Unterstützung der Lehrkraft und ihrem Interesse am Fach Mathematik befragt wurden. Im Rahmen eines *Latent-Change*-Modells können die Autoren lediglich einen eindirektionalen Effekt von der wahrgenommenen Unterstützung zu Beginn der fünften Klasse auf die Stärke der Veränderung des Interesses im Laufe des Schuljahres nachweisen. Demnach führt eine hohe Ausprägung der Unterstützung zu einer geringeren Abnahme des Interesses. Entgegen ihren Annahmen können sie keinen signifikanten Effekt vom Eingangsinteresse zu Beginn der Sekundarstufe I auf die Stärke der Veränderung der wahrgenommenen Unterstützung aufzeigen. Anders als erwartet lassen sich in der Studie auch keine korrelativen Zusammenhänge zwischen Veränderungen des Interesses und Veränderungen der wahrgenommenen Unterstützung bestätigen.

Maulana et al. (2016) untersuchen ebenfalls Zusammenhänge zwischen Veränderungen von Merkmalen der Unterrichtsqualität (wahrgenommene Klarheit, das Klassenmanagement und die Steuerung durch die Mathematik- und Englischlehrkräfte) und der Motivation der Schülerinnen und Schüler (Selbstwirksamkeit und

intrinsischer Wert). Anders als Lazarides et al. (2019) bestätigen Maulana et al. (2016) die Annahme, dass Veränderungen der Motivation der Schülerinnen und Schüler mit Veränderungen der wahrgenommenen Unterrichtsqualität zusammenhängen und dass eine Abnahme der wahrgenommenen Unterrichtsqualität mit einer Abnahme der Motivation einhergeht. Auch Ruzek & Schenke (2019) untersuchen in einer längsschnittlich angelegten Studie mit drei Messzeitpunkten über ein Schuljahr verteilt, inwiefern die schülerindividuelle Unterrichtswahrnehmung (u.a. Autonomieunterstützung) mit der selbsteingeschätzten Motivation (u.a. Leistungsmotivation) zusammenhängt. Die Stichprobe setzt sich aus 961 Schülerinnen und Schülern zusammen, die in unterschiedlichen Fächern (u.a. Mathematik, Englisch) in den Klassenstufen 6 bis 12 an US-amerikanischen Schulen unterrichtet werden. Auf der Basis einer komplexen Strukturgleichungsanalyse (*Random Intercept-Cross Lagged Panel Model*) können sie keine Belege dafür finden, dass Veränderungen in den individuellen Unterrichtswahrnehmungen mit Veränderungen in den individuellen motivationalen Merkmalen zusammenhängen.

Zusammenfassend gibt es bislang nur wenige Studien, die sich mit den komplexen Interaktionen zwischen der wahrgenommenen Unterrichtsgestaltung (z.B. Feedback) und der Motivation von Lernenden beschäftigen, obschon auf Basis von theoretischen Angebots-Nutzungs-Modellen angenommen werden kann, dass nicht nur (die aus Schülersicht erfassten) Merkmale der Lernumgebung die (motivationalen) Zielkriterien des Unterrichts beeinflussen, sondern auch Schülerinnen und Schüler in Abhängigkeit ihrer individuellen (motivationalen) Lernvoraussetzungen ihre Lernumgebung individuell wahrnehmen und nutzen. Mit Blick auf den bisherigen Forschungsstand zeigt sich, dass die Befunde zum Zusammenspiel dieser Merkmale bislang fragmentiert sind und weitere Forschung zu unterschiedlichen Komponenten der Unterrichtsqualität mit einem längsschnittlichen Studiendesign notwendig ist, um Veränderungen der Konstrukte erfassen zu können.

4 Zusammenfassung und Rahmenmodell

Das übergreifende Ziel dieser Arbeit ist es, die Wahrnehmung von Feedback aus der Sicht von Schülerinnen und Schülern im Deutschunterricht der gymnasialen Oberstufe zu analysieren sowie Zusammenhänge zwischen verschiedenen Dimensionen von wahrgenommener Feedbackqualität und der intrinsischen Lernmotivation – als Lernvoraussetzung der Schülerinnen und Schüler einerseits und Zielkriterium des Unterrichts andererseits – zu untersuchen. Vor diesem Hintergrund werden an dieser Stelle die theoretischen Ausführungen aus Kapitel 2 und 3 kurz zusammengefasst und das Rahmenmodell, das den eigenen empirischen Analysen zugrunde liegt, vorgestellt.

In den vorangegangenen Kapiteln konnte verdeutlicht werden, dass leistungsbezogenes Feedback, welches Schülerinnen und Schüler im Unterricht von ihrer Lehrkraft erhalten, ein wichtiges Merkmal eines guten Unterrichts (Hattie, 2009; Lipowsky, 2020) und eines qualitätvollen Unterrichtsgesprächs darstellt (Denn, 2021; Pauli, 2006, 2010; Pauli & Reusser, 2018; Pehmer et al., 2015). Konzeptuell werden dabei mit dem Begriff (Lehrer-)Feedback „jede Art von Rückmeldung verstanden, die sich auf die Leistung oder das Verständnis des Lernenden bezieht, diesen über die Richtigkeit seiner Antwort bzw. seiner Aufgabenlösung informiert [...] oder ihm inhaltliche und/ oder strategische Hilfen und Informationen zu seinem Bearbeitungsprozess zur Verfügung stellt“ (Lipowsky, 2020, S. 83). In Bezug auf

die inhaltliche Gestaltung von Feedback im schulischen Unterricht wurden anhand der Feedback-Klassifizierung von Narciss (2006) zunächst die Unterscheidung zwischen einer einfachen und elaborierten Komponente von Feedback getroffen. Zusätzlich konnte anhand der Ausführungen von Hattie und Timperley (2007) deutlich werden, dass Feedback ein mehrdimensionales Konstrukt darstellt und die Qualität von Feedback anhand der Ausprägung unterschiedlicher Dimensionen beschrieben werden kann: Als lernförderlich gilt Feedback, wenn es nicht nur über den aktuellen Leistungsstand oder die richtige Antwort informiert (*ergebnisorientierte Dimension*), sondern auch Informationen zur Verbesserung des Lernprozesses oder der Antwort liefert (*prozessorientierte Dimension*) und die Lernenden zur eigenständigen Kontrolle des Lernprozesses oder der Antwort anregt (*selbstregulationsorientierte Dimension*) (vgl. Hattie & Timperley, 2007; Lipowsky, 2020; Willems & Dreiling, 2022). In der Forschung zu Feedback im Unterrichtsgespräch (Denn, 2021; Pauli, 2006, 2010; Pauli & Reusser, 2018; Pehmer et al., 2015) wird außerdem argumentiert, dass ein qualitätsvolles Feedback nicht nur Informationen von der Lehrkraft enthält, sondern auch zur Kontrolle und Bewertung des Beitrags durch den Lernenden selbst oder die gesamte Klasse anregt (*dialogorientierte Dimension*; vgl. Willems & Dreiling, 2022).

Die dargestellte empirische Befundlage in Kapitel 2.3 verdeutlicht die Bedeutung von elaborierten Dimensionen von Feedback für die Leistungs- (Hattie, 2009; Harks et al., 2014; Kluger & DeNisi, 1996; Narciss & Huth, 2006; Wisniewski et al., 2020) und Motivationsentwicklung von Lernenden (Jurik et al., 2014; Kiemer et al., 2015; Rakoczy et al., 2013; Rakoczy et al., 2019; Rakoczy, Klieme, Bürgermeister & Harks, 2008). Mit Blick auf den gegenwärtigen Forschungsstand zur Ausprägung von Feedback im Unterricht zeigte sich in den dargestellten Videostudien, dass die Erteilung von ergebnisorientiertem Feedback die häufigste Form der Rückmeldung darstellt (vgl. z.B. Denn et al., 2015; Kobarg & Seidel, 2007; Lotz, 2016; Pohlmann-Rother et al., 2020). Dagegen wurden elaborierte Dimensionen von Feedback sehr viel seltener im Unterricht vergeben. Vor dem Hintergrund der in Abschnitt 2.5.1 skizzierten theoretischen Modelle zur Wirksamkeit von Feedback (Lipnevich et al., 2016; Strijbos & Müller, 2014) und Unterricht (Helmke, 2015; Seidel, 2014; Vieluf et al., 2020) sowie der dargestellten Forschungslage in Abschnitt 2.5.2 (u.a. Brooks et al., 2019; Hoya, 2019; Willems & Dreiling, 2022) kommt es nicht nur darauf an, wie das Feedback der Lehrkraft aus „objektiver“ Sicht gestaltet ist und wie sich dieses „objektiv“ gestaltete Feedback auf das Lernverhalten der Schülerinnen und Schüler auswirkt, sondern auch, wie es von den Lernenden als Adressatinnen und Adressaten des Feedbacks subjektiv wahrgenommen und interpretiert wird. Aktuell liegen mehrere Befunde zur Wahrnehmung von Feedback aus der Sicht von Schülerinnen und Schülern vor. Allerdings unterscheiden sich die Studien in der Operationalisierung der Feedbackwahrnehmung. Die meisten Studien erfassen, wie Lernende auf Feedback im Allgemeinen reagieren und operationalisieren unterschiedliche Merkmale der Akzeptanz und Nützlichkeit von Feedback (Brooks et al., 2019; Lipnevich et al., 2021; Rakoczy et al., 2013; Strijbos et al., 2021). In dieser Arbeit

wird theoretisch argumentiert, dass Schülerinnen und Schüler zunächst die *Ausprägung* von Feedback wahrnehmen müssen, bevor sie die Nützlichkeit und Akzeptanz bewerten und das Feedback für ihre eigenen Lernhandlungen nutzen. Mit Blick auf den empirischen Stand der Forschung wurde bislang mit wenigen Ausnahmen (vgl. z.B. Dreiling et al., 2018; Hoya, 2019; Sortkær, 2019; Willems & Dreiling, 2022) kaum untersucht, wie Schülerinnen und Schüler die Ausprägung von Feedback der Lehrkraft im Unterricht retrospektiv einschätzen und inwiefern unterschiedliche Dimensionen der Feedbackqualität aus der Sicht von Schülerinnen und Schülern beschrieben werden können. Auf der Basis von Befunden der pädagogisch-psychologischen Unterrichtsforschung kann davon ausgegangen werden, dass Schülereinschätzungen zur Unterrichtsqualität dazu genutzt werden können, um das Unterrichtsgeschehen differenziert zu beschreiben und die Trennung von unterschiedlichen Dimensionen der Unterrichtsqualität empirisch nachweisen zu können (Göllner et al., 2016). Außerdem stehen Schüleraussagen zur Unterrichtswahrnehmung nachweislich im Zusammenhang mit kognitiven und motivational-affektiven Zielkriterien des Unterrichts (vgl. z.B. Fauth et al., 2014; Stahns et al., 2020).

Anhand der in Abschnitt 2.5.3 beschriebenen Befunde zu interindividuellen Unterschieden in der Unterrichts- und Feedbackwahrnehmung im Fach Deutsch (vgl. z.B. Jurik et al., 2015; Hoya, 2021; Stang & McElvany, 2020; Willems & Dreiling, 2022) wurde deutlich, dass das von der Lehrkraft erteilte Feedback durch die individuellen Merkmale und Lernvoraussetzungen unterschiedlich (und anders als vielleicht von der Lehrkraft ursprünglich intendiert) wahrgenommen und interpretiert werden kann, sodass Lernende von ein und demselben Feedback unterschiedlich profitieren können. Somit kommt auch der Untersuchung der differenziellen Wahrnehmung von Feedback aus der Sicht von Schülerinnen und Schülern eine hohe Relevanz zu, um differenzielle Wirkungen von Feedback als Lernangebot besser zu verstehen.

Die Ergebnisse der in Abschnitt 2.5.5 dargestellten längsschnittlichen Untersuchungen verdeutlichen zudem, dass sich das Lehrkraftverhalten in Bezug auf unterschiedliche Unterrichtsmerkmale aus der Sicht von Schülerinnen und Schülern im Laufe der Schulzeit verändern kann. Obwohl Befunde in Bezug auf die Veränderung der Feedbackwahrnehmung fehlen, können die Ergebnisse zu längsschnittlichen Entwicklungen von ähnlichen Konstrukten (z.B. Unterstützung) ein Hinweis darauf sein, dass eine Untersuchung der Veränderung des Feedbackverhaltens aus der Sicht der Schülerinnen und Schüler inhaltlich aufschlussreich sein kann (De Wit et al., 2010; Lazarides, Fauth et al., 2021; Way et al., 2007). Eine Veränderung der Qualität des Unterrichts – im Speziellen des Feedbacks – ist auch in der Sekundarstufe II anzunehmen, da sich durch die Leistungsdifferenzierung in den Kursen auch Herausforderungen hinsichtlich der Anpassung des unterrichtlichen Handelns der Lehrkraft ergeben (vgl. Abschnitt 2.5.4). Neben der Analyse der intraindividuellen Veränderung der Feedbackwahrnehmung scheint es auch bedeutsam, das Ge-

schlecht und das Leistungsniveau als Prädiktoren von interindividuellen Unterschieden in der Veränderung der Wahrnehmung einzubeziehen (De Wit et al., 2010; Maulana et al., 2016).

Im Hinblick auf die Zielsetzung der Arbeit, die prädiktive Bedeutsamkeit der wahrgenommenen Feedbackqualität für die Motivationsentwicklung zu untersuchen, wurde im Rahmen des Kapitels 3 der Zusammenhang zwischen (wahrgenommenem) Feedback und der Motivation von Schülerinnen und Schülern in den Blick genommen. Als motivationstheoretische Grundlage dient dabei die SDT (Ryan & Deci, 2002, 2017). Die dargestellten Befunde nationaler und internationaler Studien zeigen, dass die durchschnittliche Lernmotivation von Schülerinnen und Schülern im Laufe der Schulzeit sinkt (Corpus et al., 2009; Gnambs & Hanfstingl, 2015; Gottfried et al., 2001; McElvany et al., 2008; Lazarides et al., 2019; Otis et al., 2005; Spinath & Steinmayr, 2005). Darüber hinaus konnten interindividuelle Unterschiede in der Motivationsveränderung über die Zeit nachgewiesen werden, die demonstrieren, dass die Entwicklung nicht bei allen Schülerinnen und Schülern gleichermaßen auftritt (Gottfried et al., 2007; Gottfried et al., 2017; Spinath & Steinmayr, 2008; Weidinger et al., 2015). Auf der Basis der SDT kann angenommen werden, dass diese Veränderung der Motivation im Laufe der Schulzeit damit zusammenhängt, inwieweit der Unterricht die Wahrnehmung der Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit bei Schülerinnen und Schülern unterstützt (Ryan & Deci, 2002). Die positive Wirkung von leistungsbezogenem Feedback lässt sich dabei dadurch erklären, dass insbesondere das Kompetenzerleben und vermittelt darüber die intrinsische Lernmotivation gefördert wird (Deci & Ryan, 2000; Deci et al., 1999, 2001; Ryan & Deci, 2000). Die diesbezüglich vorgestellten empirischen Untersuchungen bestätigen diese Mediationseffekte (vgl. z.B. Hondrich et al., 2018; Kiemer et al., 2015; Pat-El et al., 2012; Rakoczy et al., 2013). Mit Blick auf den Forschungsstand wird jedoch deutlich, dass in den dargestellten Studien häufig die motivationalen Effekte eines von außen beobachteten Feedbacks im Rahmen von Videostudien oder in quasi-experimentellen Designs im Fokus steht. Demgegenüber liegen zum gegenwärtigen Zeitpunkt kaum Studien vor, in denen die direkten und indirekten Zusammenhänge zwischen unterschiedlichen, aus Schülersicht wahrgenommenen Dimensionen der Feedbackqualität und der Entwicklung der Motivation im Deutschunterricht (der Sekundarstufe II) untersucht werden. Da bislang auch selten wechselseitige Einflüsse von Unterrichtswahrnehmungen und motivationalen Merkmalen simultan modelliert wurden (u.a. Lazarides et al., 2019; Maulana et al., 2016; Ruzek & Schenke, 2019), erweitert die Studie die Analysen zu den Zusammenhängen der Veränderungen der beiden Konstrukte sowie zu den *kreuzverzögerten* Effekten eines Konstrukts zu Beginn des Schuljahres auf die zeitlich nachfolgende Veränderung des anderen Konstrukts im Verlauf des Schulhalbjahres.

Als Analysegrundlage der Arbeit dient ein Angebots-Nutzungs-Modell (vgl. Abbildung 7), in dem das Feedback der Lehrkraft als ein Angebot verstanden wird, das von den Schülerinnen und Schülern in Abhängigkeit ihrer individuellen Merkmale

differenziell wahrgenommen und genutzt wird. Analog zu Modellen der Feedbackforschung (Lipnevich et al., 2016; Lipnevich & Panadero, 2021; Srijbos et al., 2014) und der pädagogisch-psychologischen Unterrichtsforschung (Seidel, 2014; Vieluf et al., 2021) impliziert das Angebots-Nutzungs-Modell, dass der Lernerfolg neben der Qualität der Feedbackgestaltung im Wesentlichen auch von der Wahrnehmung und der Nutzung des Feedbacks abhängt. Die (motivationalen) Wirkungen resultieren aus der Interaktion von Feedback als Angebot und der Wahrnehmung sowie Nutzung durch die Schülerinnen und Schüler. Daher steht auch in dieser Arbeit die *individuelle Wahrnehmung* von Feedback aus der Sicht von Schülerinnen und Schülern im Vordergrund, wobei ein Ziel der Arbeit darin besteht, die individuelle Wahrnehmung differenziert über theoretisch postulierte Dimensionen der Feedbackqualität (Ergebnis, Prozess, Selbstregulation, Dialog) zu erfassen.

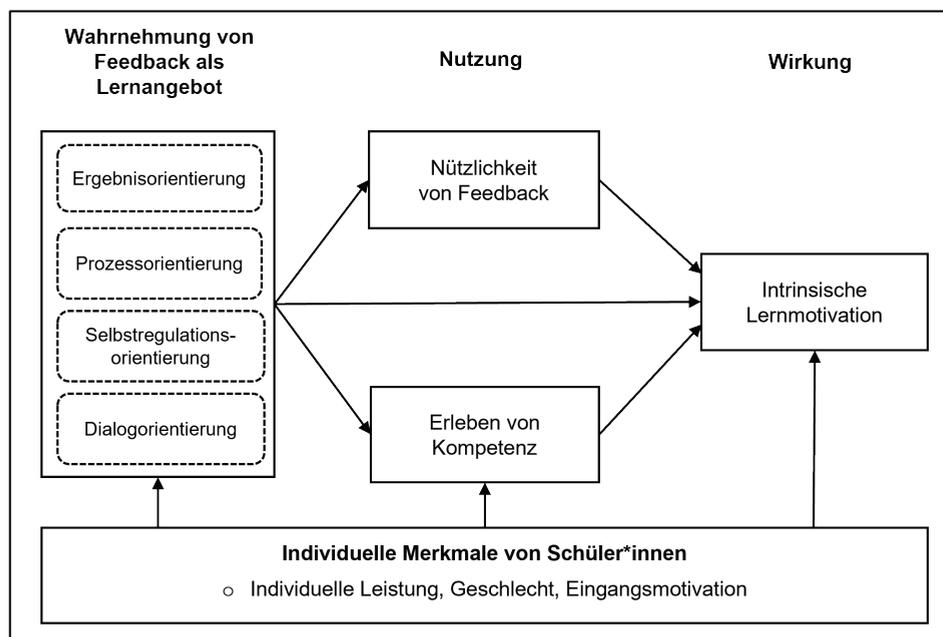


Abbildung 7: Theoretisches Rahmenmodell der Forschungsarbeit (eigene Darstellung)

Weil Feedback von Schülerinnen und Schülern in Abhängigkeit ihrer individuellen Merkmale in unterschiedlicher Weise wahrgenommen werden kann und sich dementsprechend auch differenzielle Wirkungen ergeben können (Hoya, 2021; Ruelmann et al., 2021; Sortkær, 2019; Willems & Dreiling, 2022), werden die Deutschnote sowie das Geschlecht der Schülerinnen und Schüler als relevante *individuelle Merkmale* in die Analysen einbezogen. Aus dem Zusammenspiel von Feedback als Angebot und den Merkmalen der Schülerinnen und Schüler können differenzielle Nutzungen und Wirkungen resultieren (Vieluf et al., 2020). In Anlehnung

an aktuelle Angebots-Nutzungs-Modelle (vgl. z.B. Kleickmann et al., 2019; Lipowsky, 2020; Vieluf et al., 2020) wird nicht nur die unidirektionale *Wirkung* der wahrgenommenen Feedbackqualität auf die intrinsische Motivation – als Zielkriterium von Unterricht – berücksichtigt, sondern auch deren komplexe Interaktion. In diesem Sinne wird die intrinsische Lernmotivation auch als relevante *individuelle Lernvoraussetzung* betrachtet, die Einfluss auf die Wahrnehmung der Feedbackqualität nehmen kann. Zudem werden auch indirekte Zusammenhänge zwischen der wahrgenommenen Feedbackqualität und der Motivation betrachtet, indem die individuellen Nutzungs- und Verarbeitungsprozesse von Schülerinnen und Schülern modelliert werden. Als Indikatoren der *individuellen Nutzung* von Feedback können – in Anlehnung an die theoretischen Modelle – sowohl kognitive Komponenten (operationalisiert über die Wahrnehmung der Nützlichkeit von Feedback, vgl. Lipnevich et al., 2016; Strijbos et al., 2021) als auch motivationale Komponenten (operationalisiert über das subjektive Kompetenzerleben, vgl. Klieme & Rakoczy, 2008; Rakoczy et al., 2013) betrachtet und methodisch als Mediatorvariablen in den Analysen zur Wirkung von Feedback auf die intrinsische Lernmotivation berücksichtigt werden.

Im nachfolgenden Kapitel werden die einzelnen Forschungsfragen und Hypothesen der empirischen Untersuchung konkretisiert und im theoretischen Rahmenmodell verortet.

5 Forschungsfragen und Hypothesen

Die vorliegende Arbeit knüpft an die umfangreichen Forschungsarbeiten an, die die Qualität und Wirksamkeit von Feedback im schulischen Unterricht untersuchen. Aufgrund der nachgewiesenen Bedeutsamkeit von Feedback für die Lernentwicklung von Schülerinnen und Schülern erscheint es notwendig, die Bedingungen zu untersuchen, unter denen Feedback seine förderlichen Wirkungen entfaltet. Vor dem Hintergrund der differenziellen Effekte von Feedback richtet die Forschung zunehmend ihren Blick auf die Wahrnehmung sowie auf die individuellen Merkmale von Schülerinnen und Schülern, die die Wirksamkeit von Feedback beeinflussen. Dabei ist zu berücksichtigen, dass es bisher an empirischen Untersuchungen mangelt, die systematisch untersuchen, wie Schülerinnen und Schüler die Ausprägung von Feedback der Lehrkraft im Unterricht retrospektiv einschätzen und inwiefern diese Einschätzung prädiktiv für die Motivationsentwicklung von Schülerinnen und Schülern ist. Diese Arbeit greift diese Forschungslücke auf. Anknüpfend an die empirischen Vorarbeiten von Dreiling et al. (2018) und Willems & Dreiling (2022) wird ein Modell der Feedbackqualität aus der Sicht von Schülerinnen und Schülern im Deutschunterricht der Oberstufe untersucht. Es soll geprüft werden, ob sich die Struktur des Modells für unterschiedliche Gruppen von Schülerinnen und Schülern (z.B. Geschlechtergruppen) und zu mehreren Zeitpunkten (z.B. im Laufe eines Schuljahres) in gleicher Weise abbilden lässt. Da bislang kaum Längsschnittstudien zur Veränderung von Wahrnehmungen von Unterrichtsqualitätsmerkmalen in der

gymnasialen Oberstufe vorliegen, wird unter Anwendung von *Latent-True-Change*-Modellen die (differenzielle) Veränderung der Feedbackwahrnehmung von Oberstufenschülerinnen und -schülern in Deutschkursen im Verlauf eines Schulhalbjahres untersucht.

Aus den Darstellungen des theoretischen und empirischen Forschungsstandes leiten sich konkret drei übergeordnete Forschungsfragen für die vorliegende Arbeit ab:

Hauptforschungsfrage I: Wie nehmen Schülerinnen und Schüler das Feedback der Lehrkraft im Deutschunterricht der gymnasialen Oberstufe zu Beginn und zur Mitte eines Schulhalbjahres wahr?

Hauptforschungsfrage II: In welchem Zusammenhang steht die wahrgenommene Feedbackqualität mit der intrinsischen Lernmotivation?

Hauptforschungsfrage III: Wird der Zusammenhang zwischen der wahrgenommenen Feedbackqualität und der Veränderung der intrinsischen Lernmotivation durch die wahrgenommene Nützlichkeit von Feedback und das Kompetenzerleben mediiert?

Die übergeordnete Forschungsfrage 1 bezieht sich dabei auf die differenzielle Wahrnehmung und Veränderung von Feedback als Unterrichtsmerkmal, während die Forschungsfragen 2 und 3 die direkten und indirekten Wirkungen von Feedback auf die intrinsische Lernmotivation als Zielkriterium des Unterrichts fokussieren. Abbildung 8 verankert die drei Hauptforschungsfragen in dem oben vorgestellten theoretischen Rahmenmodell.

Die aus jeder übergeordneten Forschungsfrage resultierenden Teilfragestellungen werden in den folgenden Abschnitten dargestellt. Dabei werden auch unter Bezugnahme auf den theoretischen und empirischen Hintergrund (Kapitel 2 und 3) die dazugehörigen Forschungshypothesen abgeleitet. Diese beziehen sich – wie dem theoretischen Teil der Arbeit zu entnehmen ist – auf vier verschiedene Dimensionen der wahrgenommenen Qualität von Feedback: Ergebnis-, Prozess-, Selbstregulations- und Dialogorientierung.

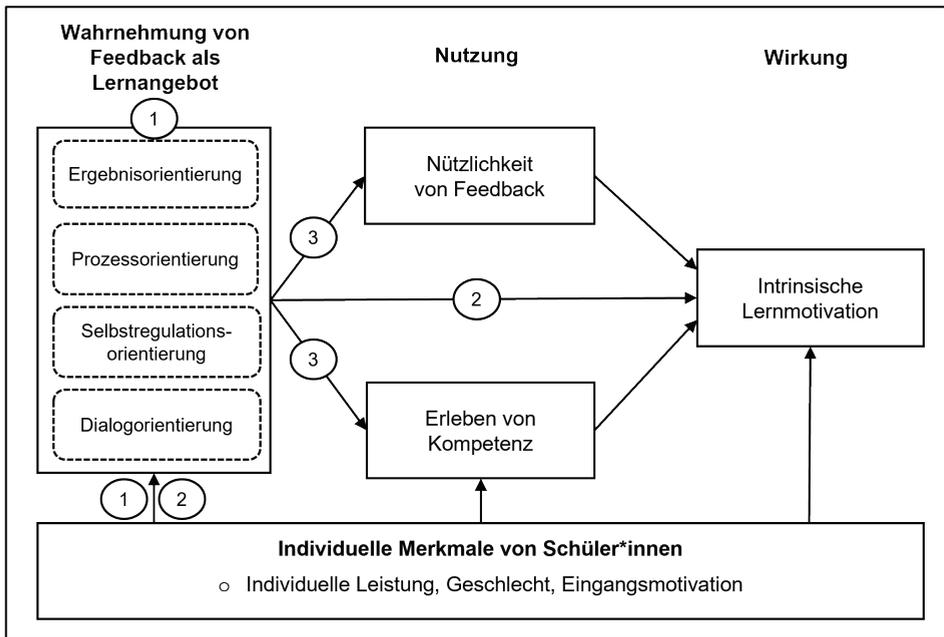


Abbildung 8: Verankerung der Forschungsfragen im theoretischen Rahmenmodell (eigene Darstellung)

5.1 Differenzielle Wahrnehmung von Feedback

Neben der Gestaltung von Feedback beeinflussen auch die individuelle Wahrnehmung und Nutzung von Feedback durch die Schülerinnen und Schüler dessen Wirkungen auf leistungsbezogene und motivationale Zielkriterien. Aus modelltheoretischer Sicht ist das Konstrukt der Feedbackqualität mehrdimensional konzipiert. Erste empirische Studien weisen dahingehend nach, dass die wahrgenommene Feedbackqualität anhand von vier Faktoren (Ergebnis-, Prozess-, Selbstregulations- und Dialogorientierung) differenziert abgebildet werden kann (Dreiling et al., 2018; Willems & Dreiling, 2022).

Faktorielle Struktur und Dimensionalität der wahrgenommenen Feedbackqualität

Aufbauend auf den modelltheoretischen Annahmen und ersten empirischen Untersuchungen soll im ersten Schritt die Dimensionalität und faktorielle Struktur der wahrgenommenen Feedbackqualität systematisch untersucht werden. Dabei soll aus methodischer Sicht geklärt werden, welches Verfahren (Generalfaktormodell, Modell mit korrelierten Faktoren, Faktormodell höherer Ordnung) zur Modellierung der Dimensionalität der wahrgenommenen Feedbackqualität geeignet ist. Daher lautet die erste Teilfragestellung:

- F1a: Inwieweit kann die faktorielle Struktur und Dimensionalität der Wahrnehmung der Feedbackqualität im Rahmen unterschiedlicher latenter Faktormodelle beschrieben werden?

Messinvarianz im Längsschnitt und im Multigruppenvergleich

Um Veränderungsprozesse zu untersuchen und differenzielle Wahrnehmungen zwischen Schülerinnen und Schülern unterschiedlichen Geschlechts herausstellen zu können, müssen in einem zweiten Schritt Äquivalenzannahmen überprüft und die Voraussetzungen für die Vergleichbarkeit des Feedbackkonstrukts über die verschiedenen Messzeitpunkte sowie über die Geschlechtergruppen geschaffen werden. Die nächsten zwei Teilfragestellungen adressieren demnach die Messinvarianz der Feedbackwahrnehmung:

- F1b: Inwieweit kann für das eingesetzte Instrument zur Erfassung der Wahrnehmung von Feedback starke faktorielle Messinvarianz über die verschiedenen Messzeitpunkte bestätigt werden?
- F1c: Inwieweit kann für das eingesetzte Instrument zur Erfassung der Wahrnehmung von Feedback starke faktorielle zwischen den Geschlechtergruppen bestätigt werden?

Veränderung der Feedbackwahrnehmung

In einem zweiten Schritt wird die Veränderbarkeit der Feedbackwahrnehmung über die Zeit in den Blick genommen. Dabei ist die folgende Fragestellung leitend:

- F1d: Inwieweit lassen sich eine Veränderung der Wahrnehmung von Feedback und interindividuelle Unterschiede in der Veränderung der Feedbackwahrnehmung im Laufe eines Schulhalbjahres beschreiben?

Bislang existieren nicht viele Studien zur Veränderung in der Qualität der Unterrichtswahrnehmung aus der Sicht von Schülerinnen und Schülern, da zum einen Schülereinschätzungen zur Unterrichtsqualität als stabil gelten (Göllner et al., 2016), zum anderen, da eine solche Untersuchung Daten zur Unterrichtswahrnehmung von Schülerinnen und Schülern zu mehreren Messzeitpunkten erfordert (Ruzek & Schenke, 2019). In Anbetracht der bisherigen Forschungsergebnisse, die Rückgänge der wahrgenommenen Qualität von Unterrichtsmerkmalen (insbesondere Unterstützung) im Laufe der Schulzeit nachweisen (Sekundarstufe I: Lazarides et al., 2019; Reddy et al., 2003; Oberstufe: De Wit et al., 2010), wird auch in der vorliegenden Studie erwartet, dass die wahrgenommene Feedbackqualität über das Schulhalbjahr hinweg

sinkt (H1d-1) und dass sich interindividuelle Unterschiede in der Veränderung der Feedbackwahrnehmung zeigen (H1d-2).

Prädiktoren differenzieller Wahrnehmungen

Theoretische Modelle und empirische Befunde geben Hinweise darauf, dass Lernende innerhalb gleicher Klassen das Feedback der Lehrkraft in unterschiedlicher Weise wahrnehmen und interpretieren. Daher werden in einem dritten Schritt auf Prädiktoren der differenziellen Wahrnehmung und Veränderung von Feedback fokussiert:

- F1e: Inwiefern beeinflusst das Geschlecht von Schülerinnen und Schülern interindividuelle Unterschiede im Ausgangsniveau und in der Veränderung der Wahrnehmung von Feedbackqualität?

In einigen Studien erwies sich das Geschlecht als signifikanter Prädiktor der Feedbackwahrnehmung im (Deutsch-) Unterricht (Carvalho et al., 2014; Havnes et al., 2012; Hoya, 2021; Lipnevich & Lopera-Oquendo, 2022; Sortkær, 2018, 2019). Außerdem zeigt sich, dass Mädchen die Qualität des Deutschunterrichts insgesamt positiver einschätzen als Jungen (Jurik et al., 2015; Wagner, 2008). Gleichzeitig entstehen auch Wahrnehmungsunterschiede in Abhängigkeit der Leistung (Hoya, 2021; Jurik et al., 2015; Wagner, 2008). Daher wird auch für die vorliegende Studie angenommen, *dass das Geschlecht – auch unter Berücksichtigung der Vorleistung – einen relevanten Prädiktor differenzieller Feedbackwahrnehmungen darstellt (H1e-1).*

Der Einfluss des Geschlechts auf interindividuelle Unterschiede in der *Veränderung* der Unterrichtswahrnehmung wurde hingegen seltener untersucht. Zudem erwiesen sich die Befunde als inkonsistent. Die Forschungsergebnisse von De Wit et al. (2010) und Maulana et al. (2016) liefern Hinweise für differenzielle Veränderungen in den Unterrichtswahrnehmungen (z.B. Unterstützung durch die Lehrkraft) in Abhängigkeit des Geschlechts. Entgegengesetzte Befunde berichten Reddy et al. (2003) sowie Way et al. (2007): Geschlechtsspezifische Unterschiede konnten im Ausgangsniveau, nicht aber in der Veränderung der Unterrichtswahrnehmung nachgewiesen werden. Allerdings liegen zum gegenwärtigen Zeitpunkt keine Studien vor, in denen interindividuelle Unterschiede in der Veränderung der Feedbackwahrnehmung untersucht werden. Außerdem bezieht sich lediglich die Studie von De Wit et al. (2010) auf die Oberstufe. Mit Blick auf die vorliegende Arbeit wird daher unter Bezugnahme auf theoretische Gesichtspunkte angenommen, *dass das Geschlecht der Schülerinnen und Schüler – auch unter Berücksichtigung der Vorleistung – einen relevanten Prädiktor interindividueller Unterschiede in der Veränderung der Feedbackwahrnehmung darstellt (H1e-2).*

5.2 Zusammenhang zwischen Feedback und Lernmotivation

Auf Basis der SDT kann angenommen werden, dass Unterschiede in der Entwicklung der Lernmotivation dadurch auftreten können, dass Lernumgebungen das Erleben von Autonomie, Kompetenz und sozialer Eingebundenheit in unterschiedlichem Maße unterstützen. Eine Veränderung der Lernumgebung (z.B. eine Abnahme oder Zunahme des Feedbacks der Lehrkraft) würde der Theorie zufolge auch zu einer Veränderung der selbstbestimmten Motivation führen. Gleichzeitig ist vor dem Hintergrund aktueller Angebots-Nutzungs-Modelle, in denen die Unterrichtswirkung als Resultat der Interaktion von Unterrichtsangebot und Schülermerkmalen beschrieben wird (Lipowsky, 2020; Reusser & Pauli, 2010; Vieluf et al., 2020), davon auszugehen, dass zum einen die Qualität von Feedback als Lernangebot die Motivation von Schülerinnen und Schülern beeinflusst. Zum anderen kann die Ausprägung motivationaler Merkmale als individuelle Lernvoraussetzung Unterschiede in der Wahrnehmung und Wirksamkeit von Feedback hervorrufen (Jurik et al., 2015; Ruelmann et al., 2021; Willems & Dreiling, 2022). Dies kann dazu führen, dass die Wahrnehmung und Wirkung von Feedback zwischen Schülerinnen und Schülern mit unterschiedlichen Ausprägungen in der Motivation variieren können. Die vorliegende Studie erlaubt es, das Zusammenspiel von Feedback und der Schülermotivation im Rahmen eines längsschnittlichen Strukturgleichungsmodells zu untersuchen und damit zentrale Annahmen von Angebots-Nutzungs-Modellen zu testen.

Zusammenhänge zwischen der Feedbackwahrnehmung und der Lernmotivation

Zunächst wird untersucht, ob sich eine Abnahme in der intrinsischen Lernmotivation im Laufe des Schulhalbjahres beobachten lässt und ob sich positive Zusammenhänge zwischen den intraindividuellen Veränderungen von Feedbackwahrnehmung und intrinsischer Lernmotivation zeigen. Daher lautet die Teilfragestellung:

- F2a: In welchem Zusammenhang stehen die wahrgenommene Feedbackqualität und die intrinsische Lernmotivation sowie die Veränderungen der beiden Konstrukte im Schulhalbjahr?

Bislang gibt es nur wenige Längsschnittstudien, die die Wirkrichtungen von Unterrichtswahrnehmungen und Motivation simultan in einem Modell untersuchen, wobei die gefundenen Ergebnisse ein fragmentiertes Bild liefern (Lazarides et al., 2019; Maulana et al., 2016; Ruzek & Schenke, 2019). Studien, die unidirektionale Wirkungen von Feedback und Motivation untersuchen, belegen zum einen positive Effekte von (wahrgenommenem) Feedback auf die Entwicklung der intrinsischen Motivation (Hondrich et al., 2018; Kiemer et al., 2015; Levesque et al., 2004; Pat-El et al., 2012; Rakoczy et al., 2013; Rakoczy, Klieme, Bürgermeister & Harks, 2008). Zum

anderen finden sich einzelne Hinweise dafür, dass auch die individuelle Motivation von Schülerinnen und Schülern ihre Wahrnehmung von Feedback beeinflusst (Ruelmann et al., 2021; Willems & Dreiling, 2022). Vor diesem Hintergrund wird auch für die vorliegende Untersuchung angenommen, *dass die wahrgenommene Feedbackqualität zu Beginn des Schuljahres die Veränderung in der intrinsischen Lernmotivation im Laufe des Schuljahres vorhersagt (H2b-1) und dass auch die intrinsische Lernmotivation zu Beginn des Schuljahres die Veränderung in der Wahrnehmung von Feedbackqualität im Laufe des Schuljahres vorhersagt (H2b-2).*

5.3 Indirekte Wirkungen von Feedback auf die Lernmotivation

Während die vorher genannten Fragestellungen analysieren, inwieweit die Wahrnehmung von Feedback mit der intrinsischen Lernmotivation in direkter Weise miteinander zusammenhängen, fokussieren die nachfolgenden Teilfragestellungen auf die mediiierende Funktion von Komponenten der *individuellen Nutzung*. Als Indikatoren der individuellen Nutzung von Feedback werden – in Anlehnung an theoretische Modelle – die Wahrnehmung der Nützlichkeit sowie das Kompetenzerleben betrachtet und methodisch als Mediatorvariablen in den Analysen zur Wirkung von Feedback auf die intrinsische Lernmotivation berücksichtigt. Die Teilfragestellungen lauten wie folgt:

- F3a: Inwiefern wird der Zusammenhang zwischen der wahrgenommenen Feedbackqualität und der Veränderung der intrinsischen Lernmotivation durch die individuell wahrgenommene Nützlichkeit von Feedback mediiert?
- F3b: Inwiefern wird der Zusammenhang zwischen der wahrgenommenen Feedbackqualität und der Veränderung der intrinsischen Lernmotivation durch das Kompetenzerleben mediiert?

Auf Basis theoretischer Modelle (vgl. Abschnitt 2.5.1) und empirischer Befunde (Harks et al., 2014; Rakoczy et al., 2013) wird erwartet, *dass die wahrgenommene Nützlichkeit von Feedback die Wirkung von wahrgenommener Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation vermitteln und erklären kann (H3a)*. Außerdem wird vor dem Hintergrund einzelner Forschungsbefunde (Gniewosz et al., 2014; Hondrich et al., 2018; Rakoczy et al., 2013; Vallerand & Reid, 1984) angenommen, *dass das Kompetenzerleben die Wirkung von wahrgenommener Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation vermitteln und erklären kann (H3b)*.

6 Methodisches Vorgehen

Für die empirische Prüfung der Fragestellungen werden Daten aus dem Projekt „Feedback im Kontext von Heterogenität“ (FeeHe) verwendet. Im Folgenden werden Hintergründe und Forschungsdesign des FeeHe-Projekts (Abschnitt 6.1) sowie die Stichprobe (Abschnitt 6.2), die auch dieser Forschungsarbeit zugrunde liegt, näher beschrieben. Nach der Beschreibung der verwendeten Erhebungsinstrumente (Abschnitt 6.3) werden die zentralen methodischen Anforderungen an die Auswertung der Daten diskutiert. Hierzu zählen der Umgang mit der hierarchischen Datenstruktur (Abschnitt 6.4) und die Behandlung fehlender Werte (Abschnitt 6.5). In Abschnitt 6.6 werden die statistischen Analyseverfahren zur Beantwortung der Forschungsfragen erläutert.

6.1 Die Messwiederholungsstudie FeeHe

Das Projekt FeeHe untersucht, inwieweit die Wahrnehmung und Wirkung von Feedback im Deutschunterricht der gymnasialen Oberstufe im Zusammenhang mit individuellen Lernvoraussetzungen der Schülerinnen und Schüler steht (Willems et al., 2020). Darüber hinaus stehen auch Fragen der differenziellen Wahrnehmung und Wirkung von Basisdimensionen der Unterrichtsqualität (Klassenführung, Schülerorientierung und kognitives Aktivierungspotenzial) im Deutschunterricht der

gymnasialen Oberstufe im Vordergrund von FeeHe. Das Projekt ordnet sich damit in bisherige Forschungsarbeiten der empirischen Schul- und Unterrichtsforschung ein, die zum einen die lernförderliche Wirkung von Feedback als Unterrichtsmerkmal zeigen und zum anderen die individuelle Wahrnehmung und Perspektive der Schülerinnen und Schüler in Bezug auf die Feedback- und Unterrichtsqualität in den Mittelpunkt rücken. Das Projekt wurde am Lehrstuhl für Empirische Bildungsforschung mit dem Schwerpunkt Schul- und Unterrichtsforschung des Instituts für Erziehungswissenschaft an der Georg-August-Universität Göttingen unter der Leitung von Prof. Dr. Ariane S. Willems durchgeführt.

FeeHe ist als Messwiederholungsstudie mit zwei Erhebungszeitpunkten – Beginn und Mitte eines Schuljahres – konzipiert. Die Eingangserhebung (MZP 1) fand zum Schuljahresbeginn 2017/18 und die Folgerhebung (MZP 2) zur Mitte des Schuljahres 2017/18 statt. Die an der Studie beteiligten Schulen wurden auf der Grundlage einer für das Schuljahr 2017/2018 vollständigen Liste von allgemeinbildenden Schulen mit einer gymnasialen Oberstufe im Raum Südniedersachsen zufällig ausgewählt. Für jede gezogene Schule (Erstschule) wurde eine zufällige Ersatzschule gezogen, die erst nach Absage einer Erstschule kontaktiert wurde. Die sich zur Teilnahme bereit erklärten Schulen entschieden selbst, welche Deutschkurse an der Befragung teilnehmen sollten. Zur Durchführung der Studie wurden vorab die Genehmigung der Niedersächsischen Landesschulbehörde sowie (bei minderjährigen Schülerinnen und Schülern) eine schriftliche Einverständniserklärung der Eltern eingeholt.

Zum ersten und zweiten Messzeitpunkt wurden die teilnehmenden Schülerinnen und Schüler zu ihrer Wahrnehmung von Feedback und zu ihrer Einschätzung der Basisdimensionen der Unterrichtsqualität sowie des motivationsunterstützenden Verhaltens (Erleben von Autonomie, Kompetenz und sozialer Eingebundenheit) befragt. Darüber hinaus wurden bei der Eingangsbefragung unterschiedliche Hintergrundvariablen (z.B. soziale Herkunft, Geschlecht, Sprachgebrauch) sowie motivationale und (meta-)kognitive Lerneingangsvoraussetzungen (z.B. intrinsische und extrinsische Lernmotivation, Zielorientierung, Fachinteresse, Selbstkonzept, Lernstrategien, Noten) der Schülerinnen und Schüler erhoben (Willems et al., 2020). Die motivationalen und (meta-)kognitiven Merkmale wurden auch zum zweiten Messzeitpunkt erhoben, um spätere Veränderungsmessungen durchführen zu können. Eine Übersicht über die einzelnen Messzeitpunkte und die erhobenen Konstrukte ist in Abbildung 9 dargestellt.

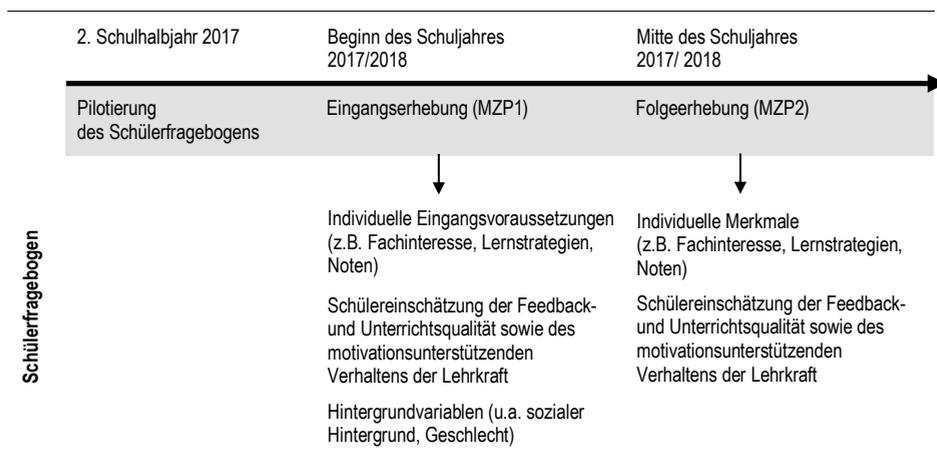


Abbildung 9: Übersicht über die Datenerhebung im FeeHe-Projekt (Willems et al., 2020, S. 10)

6.2 Beschreibung der Stichprobe

Im Folgenden wird die Gesamtstichprobe der FeeHe-Studie im Deutschunterricht der gymnasialen Oberstufe beschrieben.

An der Eingangsbefragung zum Schuljahresbeginn 2017/2018 (erster Messzeitpunkt, t_1) nahmen 807 Schülerinnen und Schüler aus 49 Deutschkursen der gymnasialen Oberstufe von 9 Schulen im Raum Südniedersachsen teil. Die Folgerhebung (zweiter Messzeitpunkt, t_2) wurde zur Mitte des Schuljahres mit 696 Schülerinnen und Schülern in 43 Deutschkursen (9 Schulen) durchgeführt.

Das durchschnittliche Alter lag zu t_1 bei 16.60 Jahren ($SD = .84$) und zu t_2 bei 17.17 ($SD = .90$). Etwa die Hälfte der Teilnehmenden war weiblich (t_1 : 53.2%, t_2 : 55.2%). Der *Migrationshintergrund* wurde über die Schülerangaben zum Sprachgebrauch zu Hause und zum Geburtsland der Eltern erfasst. In der Stichprobe der Oberstufenschülerinnen und -schüler gaben die meisten (81.1%) an, dass sie *immer* Deutsch zu Hause sprechen, 13.8 % sprechen *meistens* Deutsch, 4.7% sprechen *meistens eine andere Sprache und nur manchmal Deutsch* und nur 0.4% sprechen *nie* Deutsch zu Hause. Bei etwa einem Viertel (20.4%) der befragten Schülerinnen und Schüler ist mindestens ein Elternteil nicht in Deutschland geboren.

Als Indikatoren des *sozialen Hintergrunds* wurden der Bildungshintergrund der Eltern und die Anzahl der Bücher im Haushalt erfasst. Bei 68.3% der Schülerinnen und Schüler weist mindestens ein Elternteil das Abitur als höchsten Schulabschluss auf. Die meisten Schülerinnen und Schüler (34.4%) berichteten, dass sie über 200 Bücher besitzen, weniger Schülerinnen und Schüler (27.9%) gaben an, dass sie zwischen 26 und 100 Bücher besitzen. Zwischen 101 und 200 Bücher besitzen 26.2%

der Schülerinnen und Schüler. Die Anzahl der Schülerinnen und Schüler, die berichteten zwischen 11 und 25 Bücher zu besitzen lag bei 8.6% und zwischen 0 bis 10 Bücher bei lediglich 2.9%.

Zum ersten Erhebungszeitpunkt befanden sich 60.3% (t_2 : 47.8%) der Teilnehmenden in der Jahrgangsstufe 11 und 39.7 % (t_2 : 52.2%) in der Jahrgangsstufe 12. Die mittlere Kursgröße betrug zum ersten und zweiten Messzeitpunkt 16 Schülerinnen und Schüler ($SD_{t1} = 3.31$; $SD_{t2} = 3.67$). Ein Anteil von 54.1 % (t_2 : 52.2%) der Schülerinnen und Schüler besuchte zum ersten Messzeitpunkt einen Kurs mit einem erhöhten Anspruchsniveau und 45.9% (t_2 : 45.0%) einen Kurs mit einem grundlegenden Anspruchsniveau. Die deskriptiven Kennwerte der Stichprobe über beide Messzeitpunkte sind im Überblick in Tabelle 2 dargestellt.

Tabelle 2: Deskriptive Kennwerte zur Beschreibung der Stichproben zu beiden Messzeitpunkten

	MZP 1 (Beginn des Schuljahres)	MZP 2 (Mitte des Schuljahres)
<i>Anzahl Schüler*innen</i>	807	696
<i>Anzahl Schüler*innen nach Schulform</i>		
Gymnasium	755	662
Gesamtschule	52	34
<i>Jahrgangsstufe (in %)</i>		
11	487 (60.3%)	333 (47.8%)
12	320 (39.7%)	363 (52.2%)
<i>Anzahl Deutschkurse</i>	49	43
<i>Kursgröße</i>	$MW = 16$ ($SD = 3.46$)	$MW = 16$ ($SD = 3.67$)
<i>Kursniveau (in %)</i>		
gA	367 (45.9%)	312 (45.0%)
eA	433 (54.1%)	382 (55.0%)
<i>Geschlecht (in %)</i>		
Weiblich	429 (53.2%)	384 (55.2%)
Männlich	378 (46.8%)	312 (44.8%)
<i>Alter</i>	$MW = 16.70$ ($SD = .84$)	$MW = 17.17$ ($SD = .90$)

Anmerkungen. w = weiblich, m = männlich; gA = grundlegendes Anforderungsniveau, eA = erhöhtes Anforderungsniveau.

6.3 Erhebungsinstrumente

Im Folgenden werden die eingesetzten Erhebungsinstrumente zur Erfassung der in der vorliegenden Arbeit relevanten Konstrukte beschrieben. Insgesamt wurden vier Konstrukte, die im Rahmen des FeeHe-Projekts im Schülerfragebogen erfasst wurden, verwendet: (1) die Qualität von Feedback (Abschnitt 6.3.1), (2) die wahrgenommene Nützlichkeit von Feedback (Abschnitt 6.3.2), (3) das wahrgenommene Kompetenzerleben (Abschnitt 6.3.3) und (4) die intrinsische Lernmotivation (Abschnitt 6.3.4). Zusätzlich wurden die Vorleistung und das Geschlecht der Schülerinnen und Schüler als Hintergrundvariablen in die Analysen einbezogen (Abschnitt 6.3.5).

6.3.1 Qualität von Feedback

Die Qualität von Feedback wurde zu Beginn des Schuljahres (MZP 1) und zur Mitte des Schuljahres (MZP 2) aus der Sicht von Schülerinnen und Schülern erhoben. Die Schülerinnen und Schüler wurden gebeten einzuschätzen, inwiefern die Aussagen zum Feedback, das ihnen die Deutschlehrkraft nach einem Beitrag im Unterrichtsgespräch mündlich erteilt, auf sie zutreffen.

Zur differenzierten Erfassung der wahrgenommenen Qualität von Feedback wurden insgesamt 16 Items eingesetzt (vgl. Tabelle 3), die theoretisch vier Dimensionen der Feedbackqualität erfassen: (1) ergebnisorientierte Dimension (Items 1-4), (2) prozessorientierte Dimension (Items 5-8), (3) selbstregulationsorientierte Dimension (Items 9-12) und (4) dialogorientierte Dimension (Items 13-16). Die Skalen 1-3 bilden die von Hattie & Timperley (2007) theoretisch angenommenen Dimensionen von Feedback im Unterricht ab, während die vierte Skala eine speziell für das öffentliche Unterrichtsgespräch relevante Dimension beschreibt (vgl. Pauli, 2006). Das Instrument zur Feedbackqualität wurde im Rahmen des Projekts FeeDO (*Feedback im Deutschunterricht der gymnasialen Oberstufe*) entwickelt und in zwei Teilstudien mit Oberstufenschülerinnen und -schülern in Deutschkursen empirisch überprüft (Dreiling et al., 2018). Die 4-Faktor-Struktur der Feedbackqualität ließ sich dabei empirisch abbilden. Im Rahmen der FeeHe-Studie wurde das Instrument weiterentwickelt (Willems et al., 2020) und empirisch validiert (Willems & Dreiling, 2022). In Tabelle 3 sind die Items zur wahrgenommenen Feedbackqualität mit den deskriptiven Kennwerten aufgeführt.

Tabelle 3: Items zur Erfassung der Feedbackqualität mit deskriptiven Kennwerten

Nr.	Itemwortlaut	MW_{t_1}/MW_{t_2}	SD_{t_1}/SD_{t_2}
	Die Deutschlehrkraft ...		
1	macht durch ihre Rückmeldung deutlich, was sie von mir erwartet.	2.94/2.84	0.78/0.73
2	macht durch ihre Rückmeldung deutlich, was das Ziel der Aufgabenbearbeitung ist.	3.14/2.99	0.71/0.75
3	macht durch ihre Rückmeldung deutlich, ob ich auf dem richtigen Weg bin.	3.36/3.22	0.71/0.66
4	macht durch ihre Rückmeldung deutlich, was noch fehlt, damit ich ihre Erwartungen erfülle.	2.90/2.90	0.75/0.70
5	erklärt mir eine Frage oder Aufgabenstellung nochmals, wenn meine Antwort nicht stimmt.	2.92/2.97	0.82/0.80
6	gibt mir konkrete Denkanstöße, die mir dabei helfen, auf die richtige Antwort zu kommen.	2.97/3.02	0.79/0.78
7	verdeutlicht mir, was ich noch nicht verstanden habe, um mir zu helfen, auf die richtige Antwort zu kommen.	2.80/2.82	0.75/0.75
8	stellt Rückfragen, die mir dabei helfen, meinen Beitrag zu überdenken.	3.08/3.11	0.73/0.73
9	lässt mir ausreichend Zeit, meine Antwort zu überdenken.	2.88/2.88	0.75/0.77
10	gibt mir die Möglichkeit, mich selbst zu verbessern.	3.22/3.18	0.71/0.74
11	regt mich dazu an, meine Antwort alleine kritisch zu prüfen.	2.62/2.75	0.77/0.79
12	hinterfragt meinen Beitrag kritisch, so dass ich mich selbst korrigieren kann.	2.91/3.02	0.73/0.74
13	stellt meinen Beitrag zur Diskussion in den Kurs.	2.98/3.11	0.75/0.75
14	regt den ganzen Kurs dazu an, mir Tipps zu geben, damit wir gemeinsam auf die richtige Antwort kommen.	2.59/2.77	0.88/0.87
15	animiert den ganzen Kurs dazu, meinen Beitrag zu bewerten (z.B. „Was meinen denn die anderen dazu?“).	3.16/3.16	0.77/0.80
16	greift meinen Beitrag auf und lässt den Kurs darüber sprechen.	3.03/3.14	0.75/0.73

Anmerkungen. Instruktion: „Im Folgenden möchten wir Sie bitten, einige Fragen zu Ihrem Deutschunterricht zu beantworten. Die Fragen beziehen sich auf mündliche Rückmeldungen, die Ihre Lehrkraft Ihnen gibt, wenn Sie im Unterricht auf Fragen antworten oder sich mit einem Beitrag am Klassengespräch beteiligen“; Antwortformat: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft zu.

6.3.2 Wahrgenommene Nützlichkeit von Feedback

Für die Erfassung der wahrgenommenen Nützlichkeit von Feedback aus Schülersicht wurde im Rahmen des FeeHe-Projekts eine neue Skala entwickelt, die sich an bereits bestehende Skalen anlehnt und aus dem Englischen adaptiert wurde (Steelman et al., 2004; Strijbos et al., 2010; King et al., 2009; Rowe & Wood, 2008; vgl. hierzu ausführlich Willems et al., 2020). Die Skala umfasst insgesamt vier Items, die auf einer vierstufigen Antwortskala erhoben wurden. In Tabelle 4 sind die Items zur wahrgenommenen Nützlichkeit mit den deskriptiven Kennwerten aufgeführt. In den Analysen dieser Arbeit wurde die Einschätzung der Nützlichkeit von Feedback zum zweiten Messzeitpunkt berücksichtigt.

Tabelle 4: Items zur Erfassung der wahrgenommenen Nützlichkeit mit deskriptiven Kennwerten

Nr.	Itemwortlaut	MW_{t_2}	SD_{t_2}
	Die Rückmeldungen ...		
1	sind wichtig für mich, um mich zu verbessern.	3.18	0.86
2	spornen mich an.	2.70	0.93
3	sind nützlich für mich.	3.02	0.88
4	zeigen mir, wo meine Stärken und Schwächen liegen.	2.80	0.87

Anmerkungen. Instruktion: „Im Folgenden möchten wir Sie bitten, einige Fragen zu Ihrem Deutschunterricht zu beantworten. Die Fragen beziehen sich auf mündliche Rückmeldungen, die Ihre Lehrkraft Ihnen gibt, wenn Sie im Unterricht auf Fragen antworten oder sich mit einem Beitrag am Klassengespräch beteiligen“; Antwortformat: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft zu.

6.3.3 Wahrgenommenes Kompetenzerleben

In den Analysen wurde das aus Schülersicht wahrgenommene Kompetenzerleben zum Messzeitpunkt 2 berücksichtigt. Im Projekt FeeHe wurden vor dem Hintergrund der Selbstbestimmungstheorie (Deci & Ryan, 2000; Ryan & Deci, 2002, 2017) Skalen zum motivationsrelevanten Erleben (von Kompetenz, Autonomie und sozialer Eingebundenheit) weiterentwickelt, die sich an bereits bestehende Instrumente anlehnen (Gagné, 2003; Ilardi et al., 1993; Willems, 2011; Williams & Deci, 1996; vgl. hierzu ausführlich Willems et al., 2020). Die Skala zum wahrgenommenen Kompetenzerleben umfasst insgesamt sechs Items, die auf einer vierstufigen Antwortskala erhoben wurden. In Tabelle 5 sind die Items mit den deskriptiven Kennwerten aufgeführt.

Tabelle 5: Items zur Erfassung des wahrgenommenen Kompetenzerlebens mit deskriptiven Kennwerten

Nr.	Itemwortlaut	MW_{t_2}	SD_{t_2}
	Im Deutschunterricht ...		
1	bin ich zufrieden mit meinen Leistungen.	2.66	0.90
2	fühle ich mich den Anforderungen des Unterrichts gewachsen.	3.07	0.71
3	kann ich auch schwierige Aufgaben gut lösen.	2.83	0.76
4	finden meine Leistungen Anerkennung bei meinen Mitschüler/innen.	2.71	0.76
5	kann ich zeigen, was ich kann.	2.61	0.86
6	vermittelt meine Lehrkraft mir den Eindruck, dass sie mit meinen Leistungen zufrieden ist.	2.80	0.82

Anmerkungen. Instruktion: „Im Folgenden wollen wir mehr über Ihre Einstellung und Ihr Lernverhalten im Fach Deutsch erfahren“; Antwortformat: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft zu.

6.3.4 Intrinsische Lernmotivation

Die intrinsische Lernmotivation wurde als individuelles Merkmal der Schülerinnen und Schüler zu Beginn des Schuljahres und nach einem halben Jahr erhoben. Theoretische Grundlage (vgl. Abschnitt 3.1 und 3.2) bildete die Selbstbestimmungstheorie (Deci & Ryan, 2000; Ryan & Deci, 2002, 2017). Für die vorliegende Studie wurden dazu Items aus etablierten Instrumenten für den Kontext des Deutschunterrichts der Oberstufe adaptiert (Ryan & Connell 1989; Thomas & Müller 2011; vgl. hierzu ausführlich Willems et al., 2020). In Tabelle 6 sind die sechs Items der Subskala *Intrinsische Lernmotivation* mit ihren deskriptiven Kennwerten aufgeführt.

Tabelle 6: Items zur Erfassung der intrinsischen Lernmotivation mit deskriptiven Kennwerten

Nr.	Itemwortlaut	MW_{t_1}/MW_{t_2}	SD_{t_1}/SD_{t_2}
	In der Regel lerne ich für Deutsch, ...		
1	weil es mir Spaß macht.	2.16/2.12	0.96/0.98
2	weil ich mich gerne mit diesem Fach beschäftige.	2.24/2.18	0.98/0.96
3	weil ich gerne neue Dinge in Deutsch dazulerne.	2.40/2.30	0.96/0.97
4	weil ich mich für das Fach interessiere.	2.39/2.91	1.03/0.99

Nr.	Itemwortlaut	MW_{t_1}/MW_{t_2}	SD_{t_1}/SD_{t_2}
	In der Regel lerne ich für Deutsch, ...		
5	weil ich die Themen und Inhalte spannend finde.	2.40/2.25	0.94/0.92
6	weil mir die Inhalte des Faches wichtig sind.	2.50/2.34	0.98/0.97

Anmerkungen. Instruktion: „Im Folgenden wollen wir mehr über Ihre Einstellung und Ihr Lernverhalten im Fach Deutsch erfahren“; Antwortformat: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu, 4 = trifft zu.

6.3.5 Hintergrundvariablen

In den Analysen wurden potenzielle Einflüsse des Geschlechts und der Vorleistung der Schülerinnen und Schüler auf die wahrgenommene Feedbackqualität und die intrinsische Motivation untersucht.

Geschlecht

Das Geschlecht der Schülerinnen und Schüler wurde über die Angaben der Lehrkraft in der Schülerteilnahmeliste ermittelt und ging als dichotome Variable (0 = männlich, 1 = weiblich) in die Analysen ein.

Vorleistung

Die Vorleistung wurde über die Deutschnote zu Beginn des Schuljahres erhoben. Hierzu machten die Deutschlehrkräfte Angaben über die letzte Zeugnisnote der Schülerinnen und Schüler in ihrem Kurs. Für Schülerinnen und Schüler, die zu Beginn des Schuljahres gerade in die Jahrgangsstufe 11 eingetreten sind, wurde die Note im letzten Zeugnis (Klasse 10) auf der herkömmlichen Notenskala von 1 (sehr gut) bis 6 (ungenügend) erfasst. Für Schülerinnen und Schüler in Jahrgangsstufe 12 wurde die Deutschnote im letzten Zeugnis (Klasse 11) mit dem Punktesystem der Oberstufe erhoben (15 bis 13 Notenpunkte = sehr gut, 12 bis 10 Notenpunkte = gut, 9 bis 7 Notenpunkte = befriedigend, 6 bis 4 Notenpunkte = ausreichend, 3 bis 1 Notenpunkte = mangelhaft, 0 Notenpunkte = ungenügend).

Für die Analysen des FeeHe-Projekts wurden die Notenpunkte in Ziffernnoten umgewandelt (15 bis 13 Notenpunkte = Note 1, 12 bis 10 Notenpunkte = Note 2 usw.). Mit der Note 1 wurden 9.7% der Schülerinnen und Schüler bewertet, 29.1% mit der Note 2, 40.5% mit der Note 3, 19.4% mit der Note 4 und 1.3% mit der Note 5. Im Rahmen der Analysen dieser Arbeit wurde zur einfacheren Interpretation der Ergebnisse die Notenvariable umkodiert, sodass niedrige Werte eine befriedigende bis ungenügende Leistung und hohe Werte eine gute bis sehr gute Leistung abbilden.

6.4 Hierarchische Struktur der Daten

Die Daten der FeeHe-Studie weisen eine Mehrebenenstruktur auf: Die Schülerinnen und Schüler dieser Stichprobe sind innerhalb von Klassen (und Schulen) geschachtelt. Im Vergleich zu einer einfachen Zufallsstichprobe sind die Ergebnisse dieser *Klumpenstichprobe* oder *Clusterstichprobe* ungenauer, wenn die Schülerinnen und Schüler innerhalb einer Klasse (eines Clusters) sich ähnlicher sind als Schülerinnen und Schüler aus unterschiedlichen Klassen (Döring & Bortz, 2016). Wird dieser sog. *Designeffekt* bei der Modellschätzung nicht berücksichtigt, kann dies zu einer Unterschätzung der Standardfehler und damit zu einer liberalen Signifikanztestung führen (Julian, 2001; Raudenbush & Bryk, 2002).

Für die Analysen der Arbeit sind die Unterschiede in den Merkmalsausprägungen zwischen der Individual- und Klassenebene nicht explizit Teil der Hypothesen. Außerdem interessieren nicht die über verschiedene Schülerinnen und Schüler gemittelten Beurteilungen der Feedbackqualität (Klassenebene). Vielmehr sind die interindividuellen Unterschiede in den Wahrnehmungen der Feedbackqualität und deren Zusammenhänge mit den individuellen Ausprägungen und Entwicklungen der Motivation der einzelnen Schülerinnen und Schüler von Bedeutung. Im Vordergrund stehen also die individuellen Wahrnehmungen der Feedbackqualität, sodass die Individualebene die geeignete Analyseebene darstellt (Lüdtke et al., 2009).

Eine Möglichkeit, die geclusterte Datenstruktur dennoch zu berücksichtigen, besteht in der Korrektur der Standardfehler. Bei dem in der vorliegenden Arbeit verwendeten Statistikprogramm Mplus wird die Mehrebenenstruktur mittels der Analyseoption *type=complex* und durch die Verwendung der Klassen-ID als Clustervariable bei der Berechnung der Standardfehler und der Modellanpassung berücksichtigt (Muthén & Satorra, 1995; Watermann & Baumert, 2006).

6.5 Umgang mit fehlenden Werten

Der angemessene Umgang mit fehlenden Werten in der FeeHe-Studie stellt eine besondere Herausforderung dar. In Längsschnittstudien ist man häufig mit dem Problem konfrontiert, dass Personen an einem oder an mehreren Messzeitpunkten nicht an der Erhebung teilnehmen (Lüdtke & Robitzsch, 2011), wodurch fehlende Daten für ganze Messzeitpunkte entstehen (*unit nonresponse* bzw. *wave nonresponse*, vgl. Schafer & Graham, 2002; Spieß, 2010). In der FeeHe-Studie basieren die Daten zum ersten Messzeitpunkt auf $N = 807$ und zum zweiten Messzeitpunkte auf $N = 696$. Diese sog. *unit nonresponse* bzw. *wave nonresponse* ist auf unterschiedliche Gründe wie Krankheit oder Kurswechsel einzelner Schülerinnen und Schüler zurückzuführen. Außerdem konnten insgesamt sechs Deutschkurse aufgrund von Termenschwierigkeiten oder eines Lehrerwechsels nicht mehr an der Folgerhebung teilnehmen.

Demgegenüber ist der Ausfall bei einzelnen Variablen, die durch die fehlende Beantwortung einiger Items entstanden sind (*item nonresponse*, vgl. Schafer & Graham, 2002; Spieß, 2010), zu beiden Messzeitpunkten als gering einzustufen. Hier liegt der Anteil fehlender Werte für die in den Analysen einbezogenen Skalen auf Itemebene im Durchschnitt bei 0.6% (MZP 1) bzw. 0.8% (MZP 2), wobei der höchste Anteil fehlender Werte für ein Item bei 3.5% (MZP 1) bzw. 2.2% (MZP 2) liegt (vgl. Tabelle 7). Somit übersteigt keines der Items den als kritisch angesehenen Schwellenwert von 5% fehlender Werte (Lüdke et al., 2007).

Mit fehlenden Werten sind unterschiedliche Probleme verbunden: Erstens führen sie aufgrund der reduzierten Stichprobengröße zu einem Verlust an Effizienz bei der Parameterschätzung, zweitens ist der Umgang mit den Daten erschwert, weil die statistischen Verfahren üblicherweise vollständige Datenmatrizen erwarten und drittens besteht die Gefahr verzerrter Parameterschätzungen, weil sich Personen mit gültigen und Personen mit fehlenden Werten systematisch in den Variablenausprägungen unterscheiden (Lüdke et al., 2007; Lüdke & Robitzsch, 2011).

Bevor darauf eingegangen wird, wie in der vorliegenden Arbeit fehlende Werte behandelt werden, sollen kurz die verschiedenen Entstehungsformen fehlender Werte im Datensatz erläutert werden (Lüdke et al., 2007): Wenn fehlende Werte vollständig zufällig und unsystematisch entstehen, liegt *Missing Completely at Random* (MCAR) vor. Die Annahme von MCAR ist, dass das Fehlen von Werten auf einer Variable (z.B. Leistung) unabhängig von allen interessierenden Variablen ist – seien deren Werte beobachtet oder nicht. Dies ist beispielsweise der Fall, wenn das Fehlen von Werten im Leistungstest von Schülerinnen und Schülern weder mit anderen Variablen wie dem sozioökonomischen Status noch mit der Leistung selbst zusammenhängt, d.h. Schülerinnen und Schüler mit und ohne fehlende Werte im Leistungstest unterscheiden sich weder systematisch in ihrem sozioökonomischen Status noch in ihrer Leistung. Eine weniger restriktive und damit schwächere Annahme als MCAR ist *Missing at Random* (MAR). Konkret meint dies, dass das Auftreten von fehlenden Werten im Leistungstest vollständig von beobachteten Variablen im Datensatz abhängt, nicht aber von Variablen, deren Werte fehlen. Beispielsweise kann es vorkommen, dass systematisch Schülerinnen und Schüler mit einem niedrigen sozioökonomischen Status einzelne Aufgaben im Leistungstest nicht lösen. Das Fehlen der Werte im Leistungstest hängt aber nicht systematisch von der Leistung selbst ab. Solche potenziellen Einflussvariablen können im Modell statistisch kontrolliert werden, um korrekte Parameterschätzungen zu erhalten. Fehlen die Daten nicht zufällig, spricht man von *Missing not at Random* (MNAR). In diesem Fall hängt das Auftreten der fehlenden Werte auch nach Kontrolle von unterschiedlichen Hintergrundvariablen von der fehlenden Ausprägung in der Variable selbst ab. Fehlen beispielsweise auch nach Berücksichtigung des sozioökonomischen Status systematisch Werte von leistungsschwachen Schülerinnen und Schülern im Leistungstest, dann sind diese fehlenden Werte nicht zufällig entstanden und können somit nicht für korrekte Parameterschätzungen genutzt werden (Dietrich, 2019).

Mit fehlenden Werten im Datensatz kann je nach Entstehungsform der fehlenden Werte und statistischer Analyseverfahren unterschiedlich umgegangen werden. Dabei wird zwischen klassischen Ansätzen (z.B. fallweiser Ausschluss) und modernen Ansätzen unterschieden. Letztere werden weiterhin unterteilt in a) imputationsbasierte Verfahren (Multiple Imputation), in denen fehlende Werte durch sinnvolle Schätzungen (sog. *plausible values*) ersetzt werden, sowie b) modellbasierte Verfahren (*Full-Information-Maximum-Likelihood* (FIML)), in denen die Modellschätzung und der Umgang mit fehlenden Werten simultan erfolgen (Enders, 2001; Lüdtke et al., 2007; Schafer & Graham, 2002). Klassische Ausschlussverfahren gelten als problematisch, da der Ausschluss von Fällen aus dem Datensatz u.a. mit einer Verringerung der Teststärke und zu verzerrten Parameterschätzungen führen kann. In Simulationsstudien konnte die Überlegenheit des FIML-Ansatzes gegenüber dem Ausschluss von Daten nachgewiesen werden (Enders, 2001; Enders & Bandalos, 2001). FIML ist eine modellbasierte Methode zum Umgang mit fehlenden Werten, die insbesondere für die (längsschnittliche) Strukturgleichungsmodellierung empfohlen wird (Lüdtke et al., 2007; Schafer & Graham, 2002). Der FIML-Algorithmus zeichnet sich dadurch aus, dass die Behandlung fehlender Werte und die Modellschätzung in einem Schritt vorgenommen werden. Dabei erfolgt die Parameterschätzung für die Fälle mit fehlenden Werten auf Basis der beobachteten Daten unter der Annahme von MAR und multivariater Normalverteilung (Lüdtke et al., 2007). Aber auch bei nicht-normalverteilten Daten sind die FIML-Parameterschätzungen weniger verzerrt und effizienter als die zum Vergleich herangezogenen Ergebnisse anderer Techniken des listenweisen oder paarweisen Ausschlusses fehlender Werte (Enders, 2001). Gleichzeitig können die Standardfehler bei nicht normalverteilten Daten verzerrt sein – hier bietet sich aber eine Korrektur über einen robusten Schätzer (i.d.R. MLR-Schätzer bei kontinuierlichen Variablen bzw. WLSMV-Schätzer bei kategorialen Variablen) an (Reinecke, 2014). In Strukturgleichungsprogrammen wie Mplus ist das FIML-Verfahren inzwischen standardmäßig implementiert, sodass die Parameterschätzung von Modellen mit unvollständigen Daten unproblematisch ist. In der vorliegenden Arbeit wurde daher auch auf das FIML-Verfahren zur Behandlung der fehlenden Werte zurückgegriffen. Dieses wurde einer multiplen Imputation der Daten vorgezogen, da Maximum-Likelihood-Verfahren im Vergleich zur multiplen Imputation u.a. als effizienter und voraussetzungsärmer gelten (Allison, 2012).

Tabelle 7: Anteil fehlender Werte auf Itemebene (Item-Nonresponse)

Skala/Item	MZP 1	MZP 2
<i>Skala: EO Feedback</i>		
fbbev3	0.5	0.6
fbbev4	0.1	0.3
fbbev6	0.2	0.9
fbpo1	0.5	0.3
<i>Skala: PO Feedback</i>		
fbpo5	0.5	0.4
fbpo6	0.6	0.9
fbpo7	0.6	0.4
fbpo9	0.9	0.4
<i>Skala: SO Feedback</i>		
fbse4	0.9	0.4
fbse6	0.7	1.1
fbse8	0.6	0.6
fbse9	0.4	0.9
<i>Skala: DO Feedback</i>		
fbdi1	0.9	0.4
fbdi3	0.2	0.4
fbdi4	0.4	0.6
fbdi6	0.2	0.1
<i>Skala: Nützlichkeit</i>		
fbnut1	0.7	0.7
fbnut2	1.2	0.9
fbnut3	1.1	1.0
fbnut5	1.0	0.7
<i>Skala: Kompetenzerleben</i>		
komp1	0.7	1.4
komp2	0.9	1.0
komp3	1.4	1.9
komp4	3.5	2.2
komp5	2.0	1.4
komp6	1.6	1.9
<i>Skala: Intrinsische Lernmotivation</i>		
intrin1	0.5	1.0
intrin2	0.6	1.0
intrin3	0.5	1.1
intrin4	0.6	1.0
intrin5	0.4	1.0
intrin6	0.7	1.0

Anmerkungen. EO = Ergebnisorientierte Dimension; PO = Prozessorientierte Dimension; SO = Selbstregulationsorientierte Dimension; DO = Dialogorientierte Dimension.

6.6 Statistische Analysemethoden

Im folgenden Abschnitt erfolgt eine ausführliche Darstellung der in der vorliegenden Arbeit durchgeführten Analysen. Sämtliche Analysen wurden mit der Software Mplus in der Version 8.4 (Muthén & Muthén, 1998-2019) durchgeführt.

6.6.1 Konfirmatorische Faktorenanalyse

Insgesamt wurden alle verwendeten latenten Variablen zunächst im Rahmen von konfirmatorischen Faktoranalysen (*Confirmatory Factor Analysis*, CFA; vgl. z.B. Brown, 2015) auf ihre Dimensionalität und faktorielle Validität überprüft. Die CFA ist – im Vergleich zur explorativen Faktorenanalyse – ein hypothesengeleitetes Verfahren, bei dem die Anzahl der Faktoren und die Zuordnung der Indikatorvariablen zu den Faktoren a priori festgelegt wird (Brown, 2015). Die Grundlage der CFA ist die Spezifikation von Messmodellen, in denen systematische Zusammenhänge zwischen den manifesten Indikatorvariablen durch latente – und damit messfehlerbereinigte – Faktoren erklärt werden (Faktoren 1. Ordnung; vgl. Geiser, 2011). Bei systematischen, hohen Korrelationen zwischen den Faktoren 1. Ordnung wird zusätzlich geprüft, inwiefern (a) ein Generalfaktormodell, in dem den gemessenen Merkmalen ein gemeinsamer Faktor zugrunde liegt (z.B. *g*-Faktor der Feedbackwahrnehmung), und (b) ein Faktormodell 2. Ordnung die Datenstruktur angemessen abbilden. Bei (b) wird davon ausgegangen, dass hinter den Faktoren noch ein übergeordneter gemeinsamer Faktor steht (Gäde, Schermelleh-Engel & Brandt, 2020; Geiser, 2011).

Die Beurteilung der Modellpassung von Messmodellen erfolgt über den χ^2 -Differenztest und den CFI-Differenztest (vgl. Abschnitt 6.6.7), sofern die Modelle hierarchisch ineinander geschachtelt sind (z.B. Vergleich des Faktormodells 1. und 2. Ordnung). Nicht geschachtelte Modelle können über die gängigen Modellgütekriterien (vgl. Abschnitt 6.6.6) miteinander verglichen werden.

Im Anschluss an die Überprüfung der Faktorstruktur wird die Zuverlässigkeit der verwendeten Skalen anhand der modellbasierten Reliabilitätskoeffizienten bestimmt, die im Gegensatz zu den klassischen Methoden der Reliabilitätsschätzung im Rahmen der CFA anhand der Modellparameter geschätzt werden (Raykov & Marcoulides, 2011; Schermelleh-Engel & Gäde, 2020). Als Reliabilitätskoeffizienten werden neben Cronbachs Alpha, welcher auf der strengen Annahme von essentieller Tau-Äquivalenz (gleiche Faktorladungen) als Stufe der Messäquivalenz der Itemvariablen beruht, auch McDonalds Omega berichtet, der lediglich Tau-Kongenerität der Itemvariablen (unterschiedlich hohe Faktorladungen) voraussetzt (Schermelleh-Engel & Gäde, 2020). Für die Interpretation der Höhe der Reliabilitätskoeffizienten gilt, dass Werte $> .70$ als akzeptabel, Werte zwischen $.80$ und $.90$ als gut und Werte $> .90$ als sehr gut eingestuft werden (Gäde, Schermelleh-Engel & Werner, 2020; Schermelleh-Engel & Gäde, 2020).

Deskriptive Statistiken und Intraklassenkorrelation

Im Anschluss an die Überprüfung der Faktorenstruktur und der Reliabilitäten der verwendeten Skalen wurden die deskriptiven Statistiken berechnet. Als Maß für den Grad der Abhängigkeiten der Schülereinschätzungen von der Zugehörigkeit zu einer Klasse wurde für jedes latente Konstrukt außerdem die Intraklassenkorrelation (ICC) ermittelt. Die ICC kann in Mplus mittels der Analyseoption *type=twolevel* und unter der Verwendung der Klassen-ID als Clustervariable (vgl. Abschnitt 6.4) automatisch berechnet werden (Geiser, 2011). Die ICC kann Werte zwischen 0 und 1 annehmen. Hohe Werte für die ICC bedeuten, dass ein hoher Anteil der Gesamtvarianz in der entsprechenden Variable (z.B. Feedbackwahrnehmung) auf die Zugehörigkeit der Schülerinnen und Schüler zu einer Klasse zurückgeführt werden kann.

In Kombination mit dem Befehl *basic* erhält man zusätzlich die deskriptiven Statistiken der Variablen (Geiser, 2011). Da in Mplus zwar Mittelwerte und Varianzen, jedoch keine Standardabweichungen ausgegeben werden, werden diese aus den gegebenen Kennwerten berechnet. Die Standardabweichung (s) ist definiert als die Quadratwurzel der Varianz (s^2):

$$s = \sqrt{s^2} \quad (1)$$

6.6.2 Messinvarianzanalysen

In den Analysen der vorliegenden Arbeit sollen Unterschiede in den Ausprägungen der latenten Konstrukte (hier: Feedback und intrinsische Lernmotivation) zwischen verschiedenen Gruppen (Schülerinnen und Schüler) verglichen werden. Aus längsschnittlicher Perspektive soll außerdem die Stabilität und Veränderung der latenten Konstrukte über die zwei Messzeitpunkte hinweg überprüft werden. Diese Analysen setzen das Vorliegen der Invarianz der Messungen voraus – d.h., dass die verwendeten Skalen über die Messzeitpunkte hinweg bzw. in den Gruppen dasselbe Merkmal in gleicher Weise (= invariant) messen und dabei vergleichbare psychometrische Eigenschaften aufweisen (Gäde, Schermelleh-Engel & Brandt, 2020). Für die Analysen dieser Arbeit muss beispielsweise für das Feedback-Modell sichergestellt werden, dass die Feedback-Faktoren zu jedem Messzeitpunkt und in jeder Gruppe (also sowohl bei Schülerinnen als auch bei Schülern) in Bezug auf die Messmodelle vergleichbar sind. Mit dem Vorliegen von Messinvarianz wird nachgewiesen, dass beispielsweise Mittelwertveränderungen über verschiedene Messzeitpunkte hinweg bzw. Mittelwertunterschiede zwischen den Gruppen „wahre“, d.h. durch die zeitliche Veränderung oder die Gruppenzugehörigkeit begründete und damit systematische Unterschiede sind.

Die Prüfung der Messinvarianz für ein Messinstrument über Gruppen bzw. Messzeitpunkte erfolgt im Rahmen der konfirmatorischen Faktorenanalyse. In der Forschungspraxis hat sich ein schrittweises Vorgehen etabliert, bei dem Schritt für Schritt verschiedene Ebenen der Messinvarianz überprüft werden (Schulte et al.,

2013; Steenkamp & Baumgartner, 1998). Dabei wird der von Brown (2015) vorgeschlagene *Step-Up*-Ansatz (im Unterschied zum *Step-Down*-Ansatz) gewählt, bei dem mit der am wenigsten restriktiven Form der Messinvarianz begonnen wird (d.h. konfigurale Invarianz). Anschließend werden in der Modellierung der Messmodelle sukzessive verschiedene Restriktionen vorgenommen, die einen höheren Grad der Messinvarianz abbilden. Die folgenden Ebenen der Messinvarianz werden unterschieden (Meredith, 1993; vgl. auch Kline, 2016):

- *Konfigurale Messinvarianz (Basismodell)* ist die am wenigsten restriktive und damit schwächste Form der Messinvarianz. Sie ist bereits gegeben, wenn die Faktorstruktur (d.h. die Anzahl der Faktoren und das Ladungsmuster) in den Gruppen bzw. über die Messzeitpunkte hinweg unverändert bleibt. Dabei können die spezifischen Parameter über die Zeit hinweg bzw. in den Gruppen noch frei variieren.
- *Schwache faktorielle Messinvarianz (auch metrische Messinvarianz genannt)* liegt vor, wenn die Faktorladungen über die verschiedenen Gruppen bzw. über die Messzeitpunkte hinweg konstant bleiben.
- *Starke faktorielle Messinvarianz (auch skalare Messinvarianz genannt)* kann angenommen werden, wenn darüber hinaus auch die Item-Intercepts (bzw. bei ordinalen/kategorialen Variablen die Schwellenwerte) identisch sind. Wenn wenigstens starke faktorielle Invarianz gegeben ist, können die latenten Mittelwerte der Variablen zwischen den Gruppen bzw. die latenten Mittelwertveränderungen über die Zeit sinnvoll miteinander verglichen werden (Kleinke et al., 2017).
- *Strikte faktorielle Messinvarianz (auch Messinvarianz des Messfehlers genannt)* liegt vor, wenn zusätzlich zu den Faktorladungen und Intercepts auch die Residualvarianzen der manifesten Indikatoren invariant sind (bzw. bei ordinalen/kategorialen Variablen die Faktorladungen, Schwellenwerte und Residualvarianzen). Diese restriktivste Form der Messinvarianz wird in der Forschungspraxis aber nur selten erreicht (Meredith, 1993) und wird daher nur selten geprüft (Kleinke et al., 2017).

Da vollständige starke faktorielle Messinvarianz in der Forschungspraxis selten erreicht werden kann, ermöglicht auch das Vorliegen von *partieller Messinvarianz* einen Vergleich von latenten Mittelwerten und Beziehungen zwischen latenten Variablen (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). Partielle Messinvarianz liegt vor, wenn sich ein Teil der Modellparameter zwischen den Gruppen bzw. über die Zeit hinweg bedeutsam unterscheidet, die restlichen Parameter hingegen invariant modelliert werden. Für sinnvolle Vergleiche der latenten Mittelwerte sollten dabei jedoch mindestens zwei Items pro Faktor über vollständige schwache und starke faktorielle Messinvarianz verfügen (Byrne et al., 1989).

Welche Form der Messinvarianz in den Daten vorliegt, wird über Vergleiche der unterschiedlich restriktiven Messmodelle geprüft. Als Kriterien für den Modellvergleich zwischen den restriktiveren Formen der Messinvarianz und dem konfiguralen Basismodell dient zunächst die Beurteilung der gängigen Fit-Indizes (RMSEA, CFI, TLI; vgl. Abschnitt 6.6.6). Generell gilt, dass das sparsamere Modell mit den besten Fit-Werten gewählt wird (Geiser, 2011). Zur inferenzstatistischen Absicherung der Frage, welches Modell besser an die Daten passt, wird der χ^2 -Differenztest herangezogen. Aufgrund der Stichprobensensitivität des χ^2 -Differenztests wird empfohlen, zusätzlich den CFI-Differenztest zu verwenden (Chen, 2007; Cheung & Rensvold, 2002; Meade et al., 2008).

6.6.3 Modellierung und Analyse von Veränderung über die Zeit

Um die Veränderung der Wahrnehmung in den Feedbackdimensionen und der Einschätzung der intrinsischen Lernmotivation der Schülerinnen und Schüler über die Zeit sowie Unterschiede zwischen den Schülerinnen und Schülern in diesen intraindividuellen Veränderungen zu untersuchen, wurden *Latent-True-Change-Modelle* (LTC-Modell; auch *True-Intraindividual-Change-(TIC)-Models* genannt; Steyer et al., 1997; Steyer et al., 2000) geschätzt. Das LTC-Modell ist ein Strukturgleichungsansatz, bei dem durch die Modellierung einer Differenzvariable interindividuelle Unterschiede (d.h. Unterschiede in einer Variable zwischen Personen) in intraindividuellen Veränderungen (d.h. Veränderungen einer Person in einer Variable zu verschiedenen Zeitpunkten) auf latenter, d.h. messfehlerbereinigter Ebene untersucht werden (Dietrich, 2019; Geiser, 2011; Reuter et al., 2010). Der wesentliche Unterschied dieses Ansatzes gegenüber autoregressiven Modellen ist, dass Veränderungen direkt durch latente Differenzvariablen modelliert werden (vgl. Geiser, 2011; Gniewosz & Gniewosz, 2018).

Der Ausgangspunkt eines LTC-Modells ist ein longitudinales (Latent-State-)Modell im Rahmen einer CFA (Geiser, 2011; Anwendung bei Reuter et al., 2010), das für die einzelnen Konstrukte zunächst separat modelliert wird. In Abbildung 10a ist exemplarisch ein längsschnittliches Messmodell mit zwei Erhebungszeitpunkten und mit je zwei manifesten Indikatoren dargestellt. Als Voraussetzung der längsschnittlichen Messmodelle muss mindestens starke faktorielle Messinvarianz (Faktorladungen und Intercepts invariant) vorliegen (vgl. Abschnitt 6.6.2).

Die latenten Konstrukte wurden über *Item-Parcels* (Bandalos 2002, 2008; Little et al. 2013) für jeden Messzeitpunkt modelliert. Bei diesem Verfahren werden einzelne Items über Mittelwertbildung zu mehreren Päckchen zusammengefasst, die auch als *Item-Parcels* bezeichnet werden (Schermelleh-Engel & Gåde, 2020). Diese Methode schlagen Raykov und Marcoulides (2011) für Variablen mit zwei bis vier Antwortkategorien vor, die streng genommen nicht als kontinuierlich, sondern als geordnet kategorial eingestuft werden. Anstelle der einzelnen kategorialen Variablen können diese *Parcels* im Strukturgleichungsmodell aufgrund ihrer größeren Anzahl an Ab-

stufungen als annähernd kontinuierliche Indikatoren verwendet werden (Schermelleh-Engel & Gädde, 2020). Gleichzeitig ermöglicht die Methode des *Item-Parceling* es, die Anzahl der Parameter zu reduzieren und trotzdem stabile Parameterschätzungen zu erreichen (Little et al., 2002; Little et al., 2013). Die *Item-Parcels* wurden dabei unter Verwendung des *Balancing*-Ansatzes gebildet, indem die Items mit hohen und niedrigen Faktorladungen gleichmäßig auf die *Parcels* verteilt werden (Little et al., 2002; Little et al., 2013).

Zur Skalierung der latenten Faktoren wurde die *Effects Coding* Methode (Little et al., 2006) verwendet. Hierbei wird der Durchschnitt aller Faktorladungen auf 1 und der Durchschnitt der Intercepts auf 0 fixiert. Dadurch wird die Varianz der latenten Variable durch den Durchschnitt der Indikator-Varianzen gebildet und der Mittelwert der latenten Variable als Durchschnitt der Indikator-Mittelwerte geschätzt. Der *Effect Coding* Ansatz ist somit dadurch gekennzeichnet, dass die latente Variable dieselbe (unstandardisierte) Metrik aufweist wie der Durchschnitt aller manifesten Indikatoren (Brown, 2015). Im Unterschied dazu wird bei der stärker verbreiteten Methode des Referenzindikators die Ladung eines ausgewählten Indikators auf 1 fixiert, sodass die Varianz und der Mittelwert der latenten Variable durch die Varianz und den Mittelwert des Referenzindikators bestimmt wird (Brown, 2015).

In dem längsschnittlichen Messmodell wurde der indikatorspezifische Effekt über die Spezifikation von Residualkorrelationen modelliert (Brown, 2015; Geiser, 2011). Hierzu wurden im Modell Kovarianzen bzw. Korrelationen zwischen Messfehlern derselben Items über die verschiedenen Messzeitpunkte zugelassen (Reinecke et al., 2011). So wird beispielsweise angenommen, dass der Zusammenhang zwischen den Messfehlern der Items zu einer Variable zu jedem Messzeitpunkt nicht auf die Zusammenhänge im Modell zurückzuführen ist, sondern durch die im Längsschnitt wiederholte Messung gleicher Itemformulierungen oder durch andere Einflussfaktoren, die nicht kontrolliert wurden. Das Zulassen der Messfehlerkorrelationen führt zu einer Verbesserung der Modellpassung an die Längsschnittdaten.

In Abbildung 10b ist ein LTC-Modell dargestellt. Die grundlegende Idee des LTC-Modells ist, dass die latente Variable zum zweiten Messzeitpunkt (State 2) zerlegt wird in eine latente Ausgangsvariable zum ersten Messzeitpunkt (State 1) und eine latente Veränderungsvariable (Differenz State 2–State 1), die die Veränderung (Abnahme oder Zunahme) vom ersten Messzeitpunkt zum zweiten Messzeitpunkt darstellt (Reuter et al., 2010):

$$S_2 = S_1 + (S_2 - S_1) \quad (2)$$

Diese „Reformulierung“ des längsschnittlichen Messmodells wird in Abbildung 10b gezeigt. Zur Spezifikation der latenten Veränderungsvariable (in Abbildung 10b als Δ gekennzeichnet) wird die von Steyer et al. (2000) vorgeschlagene Parametrisierung gewählt, bei der die Veränderungsvariable direkt mit den manifesten Indikatoren verbunden ist (vgl. auch Gniewosz & Gniewosz, 2018; Reuter et al., 2010).

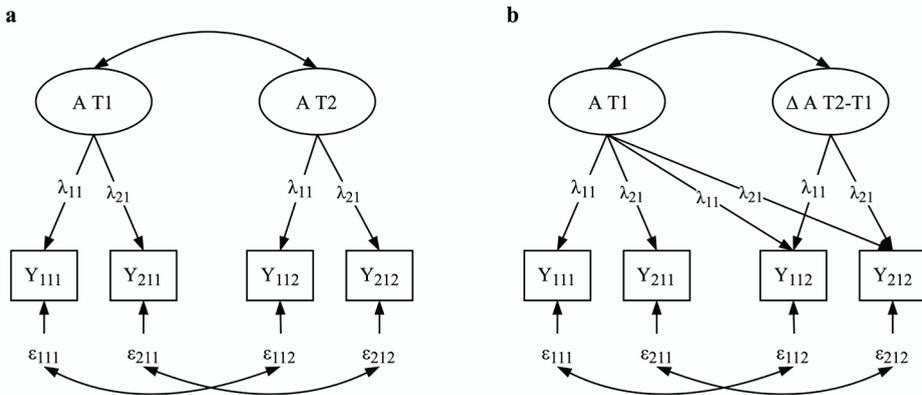


Abbildung 10: Schematisches längsschnittliches (Latent-State-)Modell und Latent-True-Change-Modell für ein Merkmal A.

Anmerkungen. a: Längsschnittliches (Latent-State-)Modell mit Residualkorrelationen. b: Latent-True-Change-Modell mit Residualkorrelationen. Δ = Latente Veränderungsvariable.

Wie in Abbildung 10b dargestellt, gibt es im LTC-Modell pro Konstrukt somit zwei latente Variablen: die latente Ausgangsvariable T1, die den Ausgangswert einer latenten Variable für den ersten Messzeitpunkt abbildet, sowie eine latente Differenzvariable T2 – T1 für die Veränderung über die Zeit. Durchschnittliche Veränderungen lassen sich direkt am Mittelwert der Veränderungsvariable ablesen. Ein positiver Wert der Veränderungsvariable entspricht einer Zunahme und ein negativer Wert einer Abnahme. Interindividuelle Unterschiede im Ausgangswert und in den intraindividuellen Veränderungen zeigen sich in der Varianz der Ausgangs- bzw. Veränderungsvariable (Steyer et al., 2009). Die Kovarianz zwischen der Ausgangs- und Veränderungsvariable ist ein Maß dafür, ob sich Personen je nach Ausgangslage unterschiedlich verändern.

Die latente Differenzvariable (hier: Veränderung der Feedbackqualität und der intrinsischen Lernmotivation) wurde in den sich anschließenden Zusammenhangs- und Mediatoranalysen als abhängige (endogene) Variable eingesetzt, um interindividuelle Unterschiede in intraindividuellen Veränderungen durch Prädiktorvariablen (hier: Geschlecht und Vorleistung) vorherzusagen (Geiser, 2011).

Analyse der Zusammenhänge zwischen den Veränderungen der Konstrukte

Um die Veränderung der wahrgenommenen Feedbackqualität mit der Veränderung der intrinsischen Lernmotivation in Beziehung zu setzen, wurden bivariate Latent-True-Change-Modelle (Dietrich, 2019; Ferrer & McArdle, 2010) geschätzt.

Abbildung 11 zeigt das schematische bivariate LTC-Modell für den Zusammenhang zwischen zwei Merkmalen A und B.

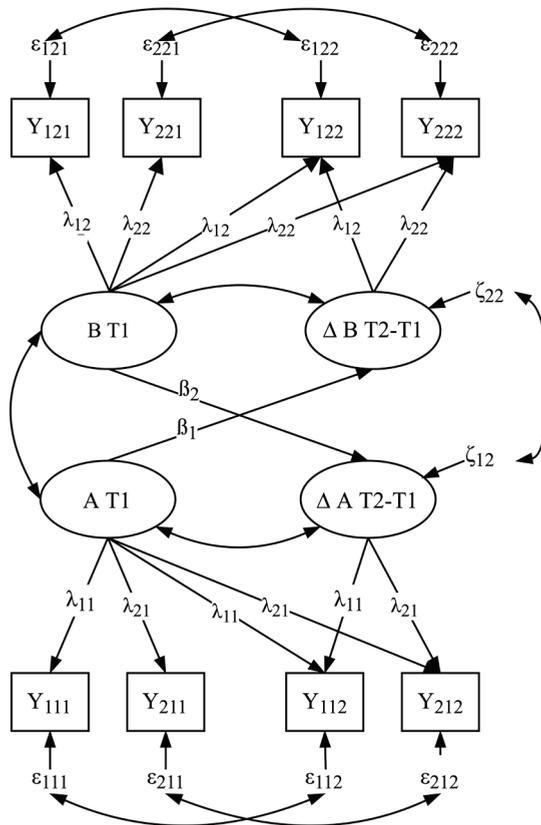


Abbildung 11: Schematisches bivariates Latent-True-Change-Modell für den Zusammenhang zwischen zwei Variablen A und B.

Insgesamt wurden vier bivariate Veränderungsmodelle spezifiziert, die Zusammenhänge zwischen den zeitlichen Veränderungen der (1) Wahrnehmung der ergebnisorientierten Dimension und der intrinsischen Lernmotivation, (2) Wahrnehmung der prozessorientierten Dimension und der intrinsischen Lernmotivation, (3) Wahrnehmung der selbstregulationsorientierten Dimension und der intrinsischen Lernmotivation und schließlich (4) Wahrnehmung der dialogorientierten Dimension und der intrinsischen Lernmotivation analysiert. Hierzu werden die spezifizierten LTC-Modelle der Konstrukte zu je einem Modell zusammengeführt (Grimm et al., 2012). Dabei werden folgende Zusammenhänge zwischen den Konstrukten modelliert: (i) Korrelationen zwischen dem Ausgangswert der Feedbackwahrnehmung

und der intrinsischen Lernmotivation zum ersten Messzeitpunkt und (ii) Korrelationen zwischen den Veränderungswerten der beiden Variablen. Zusätzlich dazu werden kreuzverzögerte Effekte (*cross-lagged effects*) modelliert, um interindividuelle Unterschiede in der Veränderung eines Konstrukts durch die zeitlich vorgeordnete Ausgangsvariable vorherzusagen. Dabei steht ein kreuzverzögerter Pfad für den Effekt von der Feedbackwahrnehmung zum ersten Messzeitpunkt auf die Veränderung der intrinsischen Lernmotivation und ein weiterer Pfad für den Effekt von der intrinsischen Lernmotivation zum ersten Messzeitpunkt auf die Veränderung der Feedbackwahrnehmung.

6.6.4 Mediationsmodelle

Die vermittelnde Rolle der wahrgenommenen Nützlichkeit von Feedback sowie des Kompetenzerlebens beim Zusammenspiel zwischen den wahrgenommenen Dimensionen der Feedbackqualität und der Veränderung der intrinsischen Lernmotivation wurde im Rahmen von Mediatormodellen untersucht. Nach McKinnon (2008) kann eine Mediation definiert werden als „*the addition of a third variable to an $X \rightarrow Y$ relation so that the causal sequences can be modeled such as X causes the mediator, M , and M causes Y , that is, $X \rightarrow M \rightarrow Y$* “ (McKinnon, 2008, S. 314). Ein einfaches Mediationsmodell mit einer Mediatorvariable sowie ein Mediationsmodell mit mehreren Mediatoren sind in Abbildung 12 dargestellt. In dem Modell wird deutlich, dass der totale Effekt von einer unabhängigen Variable X auf eine abhängige Variable Y zerlegt wird in einen direkten Effekt von X auf Y und einen indirekten (mediierenden) Effekt von X auf Y über einen oder mehrere Mediatorvariablen M (McKinnon, 2008). Diesem einfachen Mediationsmodell liegt zugrunde, dass M durch X vorhergesagt wird (Pfad a) und Y sowohl durch X (Pfad c) als auch durch M (Pfad b) vorhergesagt wird.

Der indirekte Effekt von X auf Y berechnet sich aus dem Produkt der Regressionskoeffizienten a und b (indirekter Effekt = $a \times b$). Der totale Effekt c berechnet sich aus der Summe des direkten Effekts (c) und aller indirekten Effekte:

$$c' = (a \times b) + c. \quad (3)$$

Der totale Effekt in einem Mediationsmodell entspricht statistisch dem Effekt von X auf Y in einem bivariaten Regressionsmodell ohne Berücksichtigung von Mediatorvariablen. Dabei muss nach jüngeren Forschungsarbeiten kein statistisch signifikanter Effekt von X auf Y (totaler Effekt c) ohne Kontrolle von M vorliegen (McKinnon, 2008; Rucker et al., 2011; Zhao et al., 2010.). X kann statistisch signifikante indirekte und direkte Effekte auf Y ausüben, die aber in gegenläufiger Einflussrichtung wirken und sich im totalen Effekt gegenseitig aufheben. Dies kann dazu führen, dass der totale Effekt verringert wird oder gänzlich verschwindet (sog. Suppressor-Effekte, vgl. Preacher & Hayes, 2008; Urban & Mayerl, 2007; Zhao et al., 2010). Somit wird der Einfluss von X auf Y in einer Modellschätzung, in der

lediglich die Regression von Y auf X betrachtet wird, unterschätzt. Folglich kann die Zerlegung des totalen Effekts in einen direkten und indirekten Effekt dazu beitragen, die Bedeutung von X bei der Varianzaufklärung von Y angemessener zu spezifizieren (Urban & Mayerl, 2007).

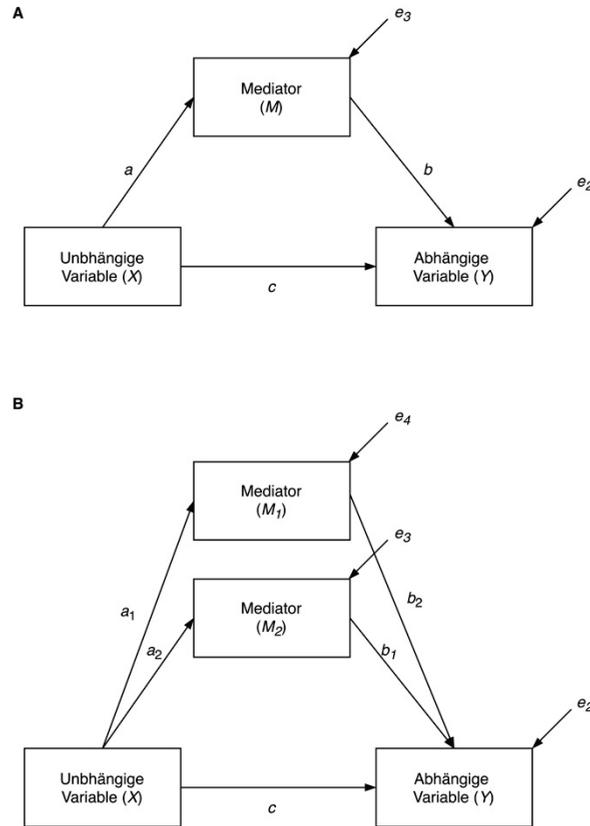


Abbildung 12: Schematische Darstellung eines einfachen Mediationsmodells (A) sowie eines erweiterten Modells mit einer multiplen Mediation (B).

In den Mediationsanalysen der vorliegenden Arbeit wurde die wahrgenommene Feedbackdimension als unabhängige Variable und die wahrgenommene Nützlichkeit von Feedback sowie das wahrgenommene Kompetenzerleben als Mediatorvariablen spezifiziert. Als abhängige Variable wurde die latente Differenzvariable eingesetzt, die die Veränderung der intrinsischen Lernmotivation im Latent-True-Change-Modell abbildet.

Um die Entstehung von Multikollinearitätseffekten¹ zu verhindern, die aufgrund der hohen Korrelationen zwischen den Feedbackdimensionen auftreten können, wurden die Effekte der einzelnen Feedbackdimensionen als unabhängige Variablen getrennt voneinander modelliert.

Für die Mediatoranalysen wurde die in Mplus implementierte *model indirect* Option angewendet. Da herkömmliche Verfahren der Signifikanztestungen von indirekten Effekten häufig nicht zuverlässig sind und zu verzerrten Ergebnissen führen können (Geiser, 2011; MacKinnon, 2008), empfehlen MacKinnon et al. (2004) zur Signifikanzprüfung die Betrachtung von Konfidenzintervallen und hierbei insbesondere die sogenannte Bias-korrigierte Bootstrap-Methode (Geiser, 2011; MacKinnon et al., 2004; MacKinnon, 2008). Nach dieser Methode ist ein indirekter Effekt auf dem 5%-Niveau signifikant, wenn der Wert Null nicht in das geschätzte 95%-Bootstrap-Konfidenzintervall um diesen Effekt fällt (Geiser, 2011). Die Konfidenzintervalle wurden dabei mit Bias-Korrektur auf der Basis von 10.000 Bootstrap-Stichproben geschätzt; dies erlaubt eine hinreichend genaue Schätzung der Grenzen des Konfidenzintervalls (Geiser, 2011). Die Bias-korrigierte Bootstrap-Methode ist in Mplus implementiert und kann über den *output*-Befehl *interval (bcbootstrap)* angefordert werden (Geiser, 2011).

6.6.5 Parameterschätzung

Zur Parameterschätzung im Rahmen der konfirmatorischen Faktorenanalysen und der Strukturgleichungsanalysen stehen unterschiedliche Schätzverfahren zur Verfügung. Die am häufigsten verwendete Schätzmethode ist die Maximum-Likelihood-Methode (ML-Methode), die bei großen Stichproben zu effizienten und erwartungstreuen Parameterschätzungen führt (Gäde, Schermelleh-Engel & Brandt, 2020). Die ML-Methode setzt allerdings voraus, dass die beobachteten Variablen normalverteilt sind und ein kontinuierliches Skalenniveau aufweisen (Rhemtulla et al., 2012). Bei kontinuierlichen Variablen, die die Annahme der Normalverteilung verletzen, kann die robuste ML-Methode (MLR) verwendet werden, bei der die Standardfehler mithilfe eines sog. „Sandwich-Schätzers“ korrigiert werden (Gäde, Schermelleh-Engel & Brandt, 2020). ML und MLR sind auch auf Datensätze mit fehlenden Werten anwendbar, sofern die fehlenden Daten MCAR bzw. MAR (vgl. Kapitel 6.5) sind (Gäde, Schermelleh-Engel & Brandt, 2020; Reinecke, 2014).

Wenn allerdings kategoriale Variablen (z.B. Likert-Skalen mit einer begrenzten Anzahl an geordneten Antwortkategorien) vorliegen, dann kann die Verwendung der ML-Schätzmethode zu verzerrten Parameterschätzungen sowie zu inkorrekten Standardfehlern und Modell-Fit-Tests führen, insbesondere dann, wenn Variablen weniger als fünf Kategorien oder eine starke Abweichung von der Normalverteilung

¹ Multikollinearität wird als „die gegenseitige lineare Abhängigkeit“ (Urban & Mayerl, 2014, S. 44) von zwei oder mehreren Prädiktoren bezeichnet. Eine hohe Multikollinearität kann zu Verzerrungen der Parameterschätzungen und zu einer Überschätzung der Standardfehler der Schätzwerte führen (Urban & Mayerl, 2014).

(Schiefe > 2) aufweisen (Muthén & Kaplan, 1985; Rhemtulla et al., 2012; West et al., 1995). In der Forschungspraxis werden ordinalskalierte Variablen mit geordneten Antwortkategorien häufig als kontinuierliche Variablen behandelt und Methoden für kontinuierliche Variablen mit ML-Schätzer genutzt (Byrne, 2012), wenn davon ausgegangen wird, dass mit mehreren beobachteten kategorialen Indikatoren ein zugrunde liegendes kontinuierliches latentes Konstrukt erfasst werden soll (Eid & Schmidt, 2014). In der Literatur gibt es Hinweise dafür, dass kategoriale Variablen als kontinuierliche Variablen behandelt werden können, wenn diese ausreichend Kategorien aufweisen und annähernd normalverteilt sind (Babakus et al., 1987; Muthén & Kaplan, 1985). Bentler und Chou (1987) argumentieren sogar, dass bei normal verteilten kategorialen Variablen auch Methoden für kontinuierliche Variablen genutzt werden können, wenn die Variablen mindestens vier Abstufungen aufweisen. Wird das kategoriale Skalenniveau explizit berücksichtigt, verbessert sich die Genauigkeit der Parameterschätzungen (Kühnel, 1993). So konnten Rhemtulla et al. (2012) in einer Simulationsstudie zeigen, dass bei Variablen mit weniger als fünf Antwortkategorien WLS (*weighted least squares*)-Verfahren mit robuster Schätzung der Standardfehler und der χ^2 -Statistik zu effizienteren Schätzergebnissen führen. In Mplus ist für kategoriale Variablen standardmäßig der WLSMV-Schätzer (WLSMV = *weighted least squares estimator with standard errors and mean- and variance adjusted chi-square test statistics*) implementiert, der auch bei kleinen Stichprobengrößen relativ stabile Modelle schätzen kann und zudem auch noch bei schief verteilten kategorialen Variablen unverzerrte Parameterschätzungen liefert (Urban & Mayerl, 2014). In der vorliegenden Arbeit findet dieses Verfahren bei den Modellgeschätzungen mit kategorialen Variablen Anwendung.

6.6.6 Modellgütekriterien

Zur Beurteilung der Qualität eines geschätzten Modells wurden unterschiedliche statistische Gütekriterien (Fit-Indizes) herangezogen (vgl. Hu & Bentler, 1999; Schermelleh-Engel et al., 2003; West et al., 2012).

Die Übereinstimmung zwischen der theoretisch spezifizierten Modellstruktur und der empirisch beobachteten Datenstruktur wird klassischerweise mit dem χ^2 -Anpassungstest geprüft. Der χ^2 -Wert ist umso kleiner, je kleiner die Differenz zwischen der geschätzten Stichprobenkovarianzmatrix und der Populationskovarianzmatrix ist, also je besser das aus der Stichprobe geschätzte Modell in der Grundgesamtheit zu den Daten passt (Reinecke & Pöge, 2010). Ein nicht signifikanter χ^2 -Test gibt an, dass sich die modellbestimmte und die empirisch beobachtete Datenstruktur nicht unterscheiden – das Modell also zu den empirischen Daten passt (Geiser, 2011; Urban & Mayerl, 2014). Außerdem gilt, dass das Modell mit dem geringeren χ^2 -Wert zu bevorzugen ist (Gäde, Schermelleh-Engel & Brandt, 2020). Zusätzlich zum χ^2 -Anpassungstest wird empfohlen das Verhältnis von modellspezifischen Freiheitsgraden zum χ^2 -Wert (χ^2/df) als Fit-Index heranzuziehen, wobei Werte ≤ 3 auf eine akzeptable Modellgüte und Werte ≤ 2 auf eine hohe Modellgüte

hinweisen (Schermelleh-Engel et al., 2003). Allerdings ist zu berücksichtigen, dass der χ^2 -Anpassungstest an mehrere Voraussetzungen geknüpft ist, die mit empirisch gewonnenen Daten kaum erfüllt werden (Reinecke & Pöge, 2010). Hierzu gehören beispielsweise die Multinomialverteilung der manifesten Variablen und eine hinreichende Stichprobengröße (Reinecke & Pöge, 2010). Nicht normalverteilte Variablen führen zu überschätzten χ^2 -Werten und zu einer Verzerrung der Standardfehler (Reinecke & Pöge, 2010). Bei großen Fallzahlen wird der Test oftmals auch bei sehr guter Passung an die Daten signifikant und bei kleinen Fallzahlen signalisiert der Test eine inflationiert gute Modellanpassung (Reinecke & Pöge, 2010; Schermelleh-Engel et al., 2003; Urban & Mayerl, 2014).

Aufgrund der Sensitivität des χ^2 -Anpassungstests sollten auch weitere Fit-Indizes berücksichtigt werden. Als komparative Fit-Indizes (auch als inkrementelle Fit-Indizes bezeichnet, vgl. z.B. Hu & Bentler, 1999) dienen der *Comparative Fit Index* (CFI) sowie der *Tucker-Lewis Index* (TLI). Beide vergleichen den Fit des geschätzten Modells mit dem Fit eines sog. Nullmodells bzw. Unabhängigkeitsmodells (mit unkorrelierten und messfehlerfreien Indikatorvariablen) und sind Maße dafür, wieviel besser das untersuchte Modell zu den Daten passt. Der CFI kann Werte zwischen 0 und 1 annehmen, für den TLI sind theoretisch auch Werte über 1 möglich. Für eine hohe Modellgüte sollten der CFI und der TLI einen Wert $\geq .97$ und für eine akzeptable Modellgüte einen Wert $\geq .95$ annehmen (Schermelleh-Engel et al., 2003; West et al., 2012).

Der *Root Mean Square Error of Approximation* (RMSEA) ist ein Maß zur Beurteilung der approximativen Passung von Modell und Daten. Werte des RMSEA kleiner als .05 können als kleine Differenz und damit als guter Modell-Fit interpretiert werden; Werte zwischen .05 und .08 können als mittlere Differenz bewertet werden und verweisen damit auf einen akzeptablen Modell-Fit; Werte größer als .08 können als große Differenz interpretiert werden und zeigen damit einen unbefriedigenden Modell-Fit (Reinecke & Pöge, 2010). Neben der Höhe des RMSEA sollte auch sein 90%-Konfidenzintervall in der Modellbeurteilung berücksichtigt werden (Geiser, 2011; Urban & Mayerl, 2014). Die geschätzten Grenzwerte des 90%-Konfidenzintervalls informieren darüber, wie präzise die RMSEA-Schätzung ist. Idealerweise sollte das 90%-Konfidenzintervall am unteren Ende einen Wert von 0.00 (mindestens jedoch kleiner als 0.05) und am oberen Ende einen Wert unter 0.08 aufweisen (Urban & Mayerl, 2014). Zusätzlich liegt als Beurteilungskriterium noch der sog. *test of close fit*, bei dem die Nullhypothese besagt, dass $RMSEA \leq .05$ sein soll (Geiser, 2011; Urban & Mayerl, 2014). Ein nicht signifikanter *test of close fit* besagt somit, dass das Modell einen ausreichend guten Fit aufweist (*close fit*).

Schließlich wird noch der Wert des *Standardized Root Mean Square Residual* (SRMR) als absoluter Fit-Index zur Evaluation der Modellgüte herangezogen. Der SRMR bestimmt die standardisierte Abweichung zwischen der beobachteten Datenstruktur und der geschätzten Modellstruktur. Werte $\leq .05$ weisen auf eine hohe Modellgüte und Werte $\leq .10$ auf eine akzeptable Modellgüte hin.

In Tabelle 8 sind die wichtigsten Fit-Indizes zur Beurteilung der Modellgüte und die diesbezüglichen Cut-Off-Werte zusammenfassend dargestellt.

Tabelle 8: Übersicht zentraler Modellgütekriterien (Fit-Indizes)

Fit-Index	Guter Modell-Fit	Akzeptabler Modell-Fit
χ^2	Modell mit der niedrigeren χ^2 -Statistik wird vorgezogen	
χ^2/df	$\chi^2/df \leq 2$	$\chi^2/df \leq 3$
CFI	CFI $\geq .97$	CFI $\geq .95$
TLI	TLI $\geq .97$	TLI $\geq .95$
RMSEA	RMSEA $\leq .05$	RMSEA $\leq .08$
p (RMSEA $< .05$)	sollte nicht signifikant sein	
90%-Konfidenzintervall (CI)	unterer Wert CI $< .05$ und oberer Wert CI $< .08$	
SRMR	SRMR $\leq .05$	SRMR $\leq .10$

Zur Einschätzung des Modell-Fits empfiehlt es sich, nicht nur einen der genannten Fit-Indizes heranzuziehen, da die Fit-Indizes unterschiedlich sensibel auf verschiedene Kriterien wie Stichprobengröße, Modellkomplexität oder die Verletzung der Normalverteilung reagieren (Brown, 2015; Hu & Bentler, 1999; Weiber & Mühlhaus, 2014). Stattdessen sollte zur Modellevaluation eine Kombination verschiedener Gütekriterien verwendet werden. Als günstige Kombinationen bei großen Stichproben ($N > 250$) gelten der SRMR in Verbindung mit CFI oder TLI (Hu & Bentler, 1999).

Im Rahmen der konfirmatorischen Faktoranalysen werden neben den Modellgütekriterien auch die Parameterschätzungen beurteilt. Die standardisierten Faktorladungen sollten statistisch signifikant von 0 verschieden sein und einen Wert $\geq .50$ aufweisen (Urban & Mayerl, 2014; Weiber & Mühlhaus, 2014). Zudem werden die geschätzten R^2 -Werte als Maß der Indikatorreliabilität betrachtet (Geiser, 2011; Weiber & Mühlhaus, 2014). Diese Werte geben den Anteil der Varianz eines manifesten Indikators an, der durch das latente Konstrukt erklärt wird (Weiber & Mühlhaus, 2014). Die Indikatorreliabilität sollte nach Weiber und Mühlhaus (2014) Werte $\geq .40$ annehmen.

6.6.7 Modellvergleiche

Für den Vergleich von zwei hierarchisch geschachtelten Daten, bei dem geprüft wird, ob sich der Modell-Fit eines restringierten Modells (mit weniger frei geschätzten Parametern) von dem Modell-Fit eines unrestringierten Modells signifikant unterscheidet, wurde der χ^2 -Differenztest herangezogen (Hu & Bentler, 1999). Ein nicht signifikanter χ^2 -Differenztest bedeutet, dass das restringierte, sparsamere Modell nicht schlechter auf die Daten passt als das unrestringierte Modell und daher beibehalten werden sollte. Bei einem signifikanten χ^2 -Differenztest sollte das restringierte Modell verworfen werden, da es den Modell-Fit im Vergleich zum unrestringierten Modell substanziell verschlechtert. Dieser Modellvergleich wurde beispielweise bei den Invarianzanalysen bzw. Gruppenvergleichen in der vorliegenden Arbeit verwendet.

Bei der Verwendung des χ^2 -Differenztests müssen die Schätzverfahren berücksichtigt werden. Wenn in den Daten eine multivariate Normalverteilung gegeben ist und die χ^2 -Werte somit mittels der ML-Schätzmethode ermittelt wurden, lässt sich der χ^2 -Differenztest einfach durchführen (Eid & Schmidt, 2014). Basiert der Modellvergleich wie im vorliegenden Fall jedoch auf χ^2 -Statistiken, die auf Grundlage einer robusten Schätzmethode (z.B. MLR bei kontinuierlichen Variablen bzw. WLSMV bei kategorialen Variablen) gewonnen wurden (vgl. Abschnitt 6.6.5), muss ein skaliertes χ^2 -Differenztest ($\Delta\chi^2_{\text{korrr}}$ mit Δdf), der die Abweichung von der Normalverteilung korrigiert, durchgeführt werden (Satorra & Bentler, 2001, 2010). Dieser robuste χ^2 -Differenztest ist in Mplus implementiert und kann somit automatisiert ermittelt werden.

Da der χ^2 -Differenztest jedoch abhängig von der Normalverteilungsannahme und der Stichprobengröße ist, empfehlen Cheung und Rensvold (2002) sowie Meade et al. (2008), den CFI-Differenztest zur Entscheidung der Modellpassung heranzuziehen (Chen, 2007; Cheung & Rensvold, 2002). Dabei sollte die Differenz des CFI zwischen dem Basismodell und dem jeweiligen Modell mit Restriktionen nicht mehr als .01 betragen (Chen, 2007; Cheung & Rensvold, 2002).

7 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der Datenanalyse berichtet, um die in Kapitel 5 abgeleiteten Hypothesen zu prüfen. Die Ergebnisdarstellung gliedert sich dabei entlang der aufgestellten Forschungsfragen. Die Abschnitte 7.1 bis 7.3 widmen sich der *Hauptforschungsfrage I*: Nach der Prüfung der faktoriellen Struktur des Konstrukts der wahrgenommenen Feedbackqualität und der Zusammenfassung der deskriptiven Befunde zur wahrgenommenen Feedbackqualität im Deutschunterricht (Abschnitt 7.1), werden in Abschnitt 7.2 die Ergebnisse hinsichtlich der Messinvarianz der eingesetzten Instrumente zur Erfassung der wahrgenommenen Feedbackqualität berichtet. In Abschnitt 7.3 werden im Rahmen eines Latent-True-Change-Modells die Ergebnisse zur (differenziellen) Veränderung der wahrgenommenen Feedbackqualität über die Zeit präsentiert. Abschnitt 7.4 werden die Ergebnisse zur *Hauptforschungsfrage II* in Bezug auf die reziproken Zusammenhänge zwischen den Dimensionen der Feedbackqualität und der intrinsischen Lernmotivation sowie zwischen den Veränderungen der beiden Konstrukte im Laufe des Schulhalbjahres dargestellt. Schließlich widmet sich Abschnitt 7.5 der *Hauptforschungsfrage III* hinsichtlich der indirekten Effekte von wahrgenommener Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation.

7.1 Wahrnehmung von Feedback im Deutschunterricht

In den nachstehenden Abschnitten wird die (längsschnittliche) Faktorenstruktur des Konstrukts der Feedbackqualität zu beiden Messzeitpunkten separat sowie im Rahmen eines gemeinsamen, längsschnittlichen Strukturmodells geprüft (*Teilfragestellung F1a*).

7.1.1 Faktorenstruktur der Skalen zur Feedbackwahrnehmung

Zur Überprüfung der faktoriellen Struktur der wahrgenommenen Feedbackqualität in der vorliegenden Stichprobe wurden drei Modelle jeweils separat für beide Messzeitpunkte spezifiziert (s. Abbildung 13).

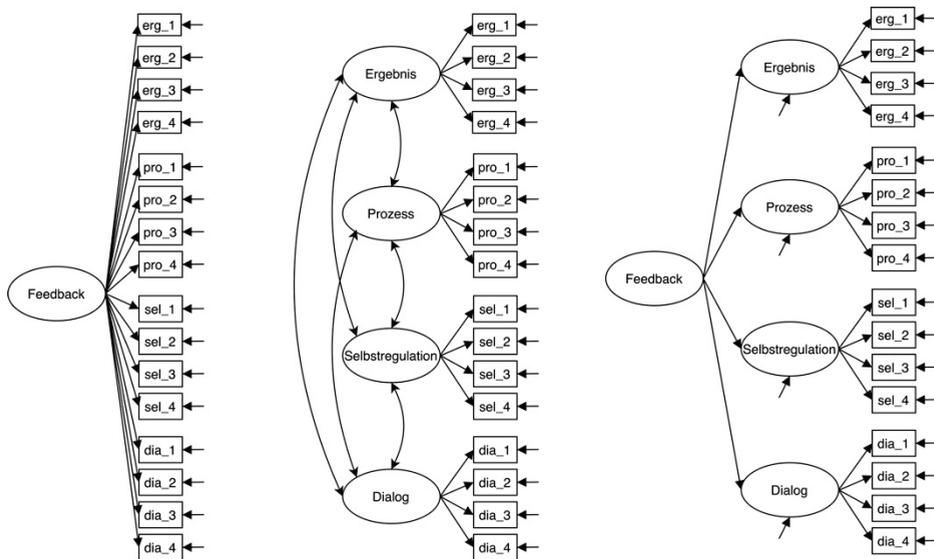


Abbildung 13: Schematische Darstellung des 1-Faktor-Modells (links), 4-Faktormodells mit korrelierten Faktoren (Mitte) und Faktormodells 2. Ordnung (rechts)

Im ersten Schritt wurde ein 1-Faktormodell (Generalfaktor- oder *g*-Faktormodell) geprüft (Modell I), in dem alle 16 Items der Feedbackqualität auf einem einzigen Faktor laden. Bei einem guten Modell-Fit müsste auf inhaltlicher Ebene davon ausgegangen werden, dass nicht zwischen ergebnisorientierter, prozessorientierter, selbstregulationsorientierter und dialogorientierter Dimension unterschieden werden kann, sondern dass sich die Feedbackqualität anhand einer Dimension abbilden lässt. Das zweite Modell (II) unterscheidet zwischen vier korrelierten Faktoren als voneinander abgrenzbare Subdimensionen der Feedbackqualität (Ergebnis, Prozess, Selbstregulation und Dialog).

Abschließend wurde in Modell III noch ein Faktormodell 2. Ordnung überprüft, in welchem die Feedback-Faktoren 1. Ordnung auf einem Faktor 2. Ordnung laden. Dabei wurde theoretisch angenommen, dass den vier Dimensionen ein gemeinsamer Faktor zugrunde liegt und darüber hinaus noch ein Anteil spezifischer Varianz der einzelnen Feedbackdimensionen vorliegt (Geiser, 2011). Die Fit-Statistiken für die spezifizierten Modelle sind jeweils separat für beide Messzeitpunkte in Tabelle 9 dargestellt.

Tabelle 9: Ergebnisse der konfirmatorischen Faktorenanalyse zur Feedbackqualität zu Beginn und zur Mitte des Schuljahres (MZP 1 und MZP 2)

Modell	Faktoren	MZP	χ^2	df	χ^2/df	RMSEA [90% CI]	C_{Fit}	CFI	TLI	SRMR
I	1	1	557.75*	104	5.36	.07 [.07, .08]	<.001	.90	.88	.06
I	1	2	486.02*	103	4.72	.08 [.07, .08]	<.001	.91	.90	.06
II	4	1	248.84*	98	2.54	.04 [.04, .05]	.94	.97	.96	.04
II	4	2	220.13*	98	2.25	.05 [.04, .05]	.87	.97	.97	.03
III	4/1	1	250.93*	100	2.51	.04 [.04, .05]	.95	.97	.96	.04
III	4/1	2	210.60*	100	2.12	.04 [.03, .05]	.95	.98	.97	.04

Anmerkungen. Modell I: Generalfaktormodell; Modell II: 4-Faktormodell; Modell III: Faktormodell 2. Ordnung. * $p < .001$.

Bei Vergleich der Modell-Fit-Statistiken zeigt sich, dass das geschätzte Generalfaktormodell (Modell I) zu beiden Messzeitpunkten die Daten am schlechtesten repliziert. Die Werte für den CFI, TLI und SRMR liegen außerhalb des geforderten Wertebereichs, sodass dieses Modell verworfen werden muss. Modell II und III zeigen zwar einen signifikanten χ^2 -Wert, allerdings verweisen die anderen Fit-Indizes auf eine gute Anpassungsgüte des 4-Faktormodells und des Faktormodells 2. Ordnung. Der nach Satorra und Bentler (2001) korrigierte χ^2 -Differenztest für den Modellvergleich zeigt für Messzeitpunkt 1 eine signifikante Differenz zwischen Modell II und Modell III ($t_1: \Delta\chi^2_{\text{kor}} = 7.01, \Delta df = 2, p = .03$), wodurch das Modell 2. Ordnung (Modell III) verworfen werden müsste. Der Modellvergleich zu Messzeitpunkt 2 hingegen zeigt keine bedeutsame Verschlechterung durch das restringiertere Modell 2. Ordnung ($t_2: \Delta\chi^2_{\text{kor}} = 0.65, \Delta df = 2, p = .72$), wodurch es angenommen werden müsste. Nach inhaltlichen Gesichtspunkten entspricht das Modell mit

vier korrelierten Faktoren 1. Ordnung eher dem Forschungsziel der Arbeit, die Dimensionen der Feedbackqualität einzeln in die Analysen einführen zu können und damit möglichst differenzierte Aussagen über die spezifischen Effekte der einzelnen Dimensionen treffen zu können. Nach Abwägung dieser inhaltlichen Kriterien fällt die Entscheidung für das Strukturmodell mit vier korrelierten Faktoren und gegen das Modell 2. Ordnung.

Die standardisierten Faktorladungen und die geschätzten R^2 -Werte als Maße für die Reliabilitäts-Schätzungen der manifesten Indikatoren (Geiser, 2011) können der Tabelle 10 entnommen werden. Alle Indikatoren verfügen über signifikante ($p < .001$) standardisierte Faktorladungen, die über dem Schwellenwert von .50 liegen ($t_1: .51 \leq \lambda \leq .77$; $t_2: .66 \leq \lambda \leq .87$). Die R^2 -Werte deuten auf überwiegend hohe Indikatorreliabilitäten hin.

Tabelle 10: Standardisierte Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im 4-Faktormodell der Feedbackqualität zu Beginn und zur Mitte des Schuljahres (MZP 1 und 2)

Skala /Item	MZP 1		MZP 2	
	λ	R^2	λ	R^2
<i>Ergebnisorientierte Dimension</i>				
fbev3	.66	.43	.82	.68
fbev4	.69	.47	.82	.67
fbev6	.71	.51	.73	.54
fbpo1	.69	.47	.75	.56
<i>Prozessorientierte Dimension</i>				
fbpo5	.62	.39	.66	.44
fbpo6	.75	.56	.80	.65
fbpo7	.71	.51	.81	.65
fbpo9	.70	.49	.71	.50
<i>Selbstregulationsorientierte Dimension</i>				
fbse4	.51	.26	.67	.45
fbse6	.71	.51	.77	.59
fbse8	.62	.38	.70	.49
fbse9	.75	.56	.74	.55
<i>Dialogorientierte Dimension</i>				
fbdi1	.65	.43	.73	.53
fbdi3	.71	.50	.74	.55
fbdi4	.64	.41	.72	.52
fbdi6	.77	.57	.87	.76

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

In Tabelle 11 sind die Interkorrelationen der latenten Faktoren des 4-Faktormodells dargestellt. Die Korrelationen liegen im Bereich $.62 \leq r \leq .84$ (t_1) bzw. $.59 \leq r \leq .83$ (t_2), wobei die Korrelation zwischen der ergebnisorientierten und der prozessorientierten Dimension am höchsten ausfällt.

Tabelle 11: Latente Interkorrelationen der Faktoren der Feedbackskalen zu Beginn (t_1) und zur Mitte (t_2) des Schuljahres

	(1)	(2)	(3)	(4)
(1) EO	-	.84	.78	.62
(2) PO	.83	-	.83	.68
(3) SO	.69	.82	-	.65
(4) DO	.59	.63	.64	-

Anmerkungen. Unterhalb der Diagonalen sind die Korrelationen zu t_1 und oberhalb zu t_2 ; EO = Ergebnisorientierte Dimension; PO = Prozessorientierte Dimension; SO = Selbstregulationsorientierte Dimension; DO = Dialogorientierte Dimension; alle Korrelationskoeffizienten sind auf einem Niveau von $p < .001$ signifikant.

7.1.2 Deskriptive Ergebnisse

In Tabelle 12 sind zum einen die Mittelwerte, Standardabweichungen und Intraklassenkorrelationen (ICC) der manifesten Subskalen der Feedbackqualität sowie auf latenter Ebene die modellbasierten Reliabilitätskoeffizienten (Cronbachs α bzw. McDonalds ω) dargestellt. Die deskriptiven Werte geben Auskunft darüber, wie das Feedback im Unterrichtsgespräch durch die Schülerinnen und Schüler wahrgenommen wird. Ein hoher Wert repräsentiert dabei eine hohe Ausprägung der Feedbackdimension. Die Mittelwerte der Subskalen liegen alle über dem theoretischen Mittelwert von 2.5. Zu Beginn des Schuljahres ist die ergebnisorientierte Dimension von Feedback am höchsten ausgeprägt ($MW = 3.09$, $SD = 0.54$), während zur Mitte des Schuljahres die dialogorientierte Dimension von Feedback die höchste Ausprägung aufweist ($MW = 3.05$, $SD = 0.61$). Auf deskriptiver Ebene zeigt sich außerdem, dass die Ausprägung der ergebnisorientierten Dimension im Laufe des Schulhalbjahres sinkt, während die berichteten Ausprägungen der anderen drei Dimensionen steigen.

Tabelle 12: Skalenmittelwerte (*MW*), Standardabweichungen (*SD*), Reliabilitäten (α und ω) und Intraklassenkorrelationen (ICC) der Feedbackdimensionen zu Beginn und zur Mitte des Schuljahres

Skala	Beginn des Schuljahres						Mitte des Schuljahres					
	<i>N</i>	<i>MW</i>	<i>SD</i>	α	ω	ICC	<i>N</i>	<i>MW</i>	<i>SD</i>	α	ω	ICC
EO	807	3.09	0.54	.72	.72	.27	697	2.98	0.56	.80	.80	.25
PO	805	2.95	0.57	.74	.74	.19	696	2.98	0.59	.77	.77	.21
SO	805	2.91	0.53	.68	.68	.12	696	2.96	0.58	.76	.76	.19
DO	807	2.94	0.58	.72	.72	.16	697	3.05	0.61	.78	.77	.21

Anmerkungen. EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension; SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension.

Die Reliabilitäten (α und ω) der Subskalen variieren zwischen .68 und .72 (t_1) bzw. .76 und .80 (t_2) und sind damit – mit Ausnahme der Skala Selbstregulationsorientierte Dimension zu t_1 – als zufriedenstellend bis gut zu bewerten. Abschließend ist noch die Intraklassenkorrelation (ICC) angegeben. Dieser gibt den prozentualen Anteil der Gesamtvarianz an, der auf Unterschiede zwischen den Klassen zurückzuführen ist (Lüdtke et al., 2006). Die Ausprägungen für die ICC zeigen, dass zwischen 12% bzw. 19% (selbstregulationsorientierte Dimension zu t_1 bzw. t_2) und 27% bzw. 25% (ergebnisorientierte Dimension zu t_1 bzw. t_2) der Varianz in den Schülerwahrnehmungen auf die Zugehörigkeit zu einem Deutschkurs und nicht auf individuelle Wahrnehmungstendenzen einzelner Schülerinnen und Schüler zurückgehen. Die ICC-Werte sind vergleichbar mit Werten in Studien, in denen Unterrichtsmerkmale untersucht werden (Fauth et al., 2014; Iglar et al., 2019; Lüdtke et al., 2006; Praetorius et al., 2018; Wagner et al., 2016; Willems, 2011).

7.1.3 Modellierung der Feedbackqualität im Längsschnitt

Als Ausgangspunkt der längsschnittlichen Analyse der Feedbackqualität wurde ein unrestriktives, longitudinales Latent-State-Modell (LS-Modell) geschätzt, in welchem noch keine Annahmen über zeitlich invariante Parameter getroffen werden. Hierzu wurden die 4-Faktor-Modelle von Messzeitpunkt 1 und Messzeitpunkt 2 zu einem LS-Modell verbunden. Die Korrelationen zwischen den latenten (State-)Faktoren wurden frei geschätzt. Ebenso wurden Residualkorrelationen zwischen Indikatoren verschiedener Messzeitpunkte zugelassen, um hiermit indikatorspezifische Effekte abzubilden (vgl. Abschnitt 6.6.3). Das Modell weist einen guten Modell-Fit auf ($\chi^2 = 602.01$, $df = 420$, $p(\chi^2) < .001$, CFI = .98, TLI = .97, RMSEA [90% CI] = .02 [.02, .03], $C_{Fit} = 1.00$, SRMR = .04). Die standardisierten Faktorladungen der Indikatoren der Feedbackwahrnehmung liegen im Bereich $\lambda_{t1}/\lambda_{t2} = .53/.67$ und $\lambda_{t1}/\lambda_{t2} = .76/.87$ ($p < .001$) (vgl. Tabelle 13). Die geschätzten R^2 -Werte als Maße für die Reliabilitäts-Schätzungen der manifesten Indikatoren (Geiser, 2011) liegen

für den ersten Messzeitpunkt im Bereich $.28 < R^2 < .58$ ($p < .001$) und für den zweiten Messzeitpunkt im Bereich $.45 < R^2 < .76$ ($p < .001$).

Tabelle 13: Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im Latent-State-Modell der Feedbackqualität

Skala/Item	MZP 1		MZP 2	
	λ	R^2	λ	R^2
<i>Ergebnisorientierte Dimension</i>				
fbev3	.66	.43	.82	.68
fbev4	.67	.45	.81	.65
fbev6	.73	.54	.75	.56
fbpo1	.68	.46	.75	.56
<i>Prozessorientierte Dimension</i>				
fbpo5	.61	.37	.67	.45
fbpo6	.76	.58	.81	.66
fbpo7	.71	.51	.80	.64
fbpo9	.70	.49	.70	.49
<i>Selbstregulationsorientierte Dimension</i>				
fbse4	.53	.28	.68	.46
fbse6	.72	.52	.77	.59
fbse8	.60	.36	.70	.49
fbse9	.74	.55	.74	.55
<i>Dialogorientierte Dimension</i>				
fbdi1	.65	.42	.72	.52
fbdi3	.72	.52	.77	.59
fbdi4	.64	.42	.70	.50
fbdi6	.75	.56	.87	.76

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

Die Korrelationen zwischen den State-Faktoren untereinander fallen durchschnittlich bis hoch aus, was auf eine vergleichsweise hohe zeitliche Stabilität interindividueller Unterschiede hinsichtlich der wahrgenommenen Feedbackqualität schließen lässt (Geiser, 2011): Ergebnisorientierte Dimension: $r_{t12} = .71, p < .001$; prozessorientierte Dimension: $r_{t12} = .64, p < .001$; selbstregulationsorientierte Dimension: $r_{t12} = .65, p < .001$; dialogorientierte Dimension: $r_{t12} = .59, p < .001$ (vgl. Tabelle 14).

Tabelle 14: Latente Faktor-Interkorrelationen der Dimensionen der Feedbackqualität im Latent-State-Modell

Latente Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) EO t ₁	1.00	.71	.83	.59	.69	.54	.59	.53
(2) EO t ₂		1.00	.64	.84	.58	.78	.44	.62
(3) PO t ₁			1.00	.64	.83	.61	.63	.52
(4) PO t ₂				1.00	.61	.83	.48	.67
(5) SO t ₁					1.00	.65	.64	.50
(6) SO t ₂						1.00	.46	.64
(7) DO t ₁							1.00	.59
(8) DO t ₂								1.00

Anmerkungen. EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension; SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension. Dargestellt sind standardisierte Korrelationen; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

Längsschnittliches Parcel-Strukturmodell

Das LS-Modell weist eine hohe Komplexität auf und enthält eine große Anzahl an zu schätzenden Parametern. Mit Blick auf die noch geplanten Analysen, in denen das Modell um Prädiktoren erweitert wird, empfiehlt es sich ein sparsameres Modell zu entwickeln. Um Modellsparbarkeit zu erreichen, wurde die Methode des *Item-Parceling* gewählt (vgl. auch Abschnitt 6.6.3). Mit dem *Item-Parceling* ist es möglich, die Anzahl der Parameter zu reduzieren und trotzdem stabile Parameterschätzungen zu erreichen (Little et al., 2002; Little et al., 2013). Um die *Item-Parcels* zu ermitteln, wurde der *balancing approach* gewählt, bei dem sich die Auswahl der Item-Paare an der Höhe der Faktorladungen orientiert. Die jeweils acht Items der Dimensionen der Feedbackqualität zu Beginn und zur Mitte des Schuljahres wurden nach den Prinzipien des *balancing approach* in je zwei Paare aufgeteilt, die aus dem Mittelwert der Einzelitems gebildet werden. Aus der Tabelle 15 geht die Zuordnung der Items zu den *Parcels* hervor. Die dargestellten Faktorladungen beziehen sich auf das oben beschriebene LS-Modell ohne *Item-Parcels*. Es ist zu erkennen, dass das erste Item-Paar aus dem Item mit der höchsten Faktorladung und dem Item mit der geringsten Faktorladung besteht. Das zweite Item-Paar setzt sich zusammen aus den Items mit der zweithöchsten und der zweitniedrigsten Ladung.

Tabelle 15: Übersicht der Item-Parcels mit den jeweiligen Items und Faktorladungen (λ) zu Beginn (t_1) und zur Mitte (t_2) des Schuljahres

Skala/Parcelnr.	Parcel	Itemkürzel
<i>Ergebnisorientierte Dimension zu Beginn des Schuljahres t_1</i>		
1	p11	afbev3 + afbev6
2	p21	afbev4 + afbp01
<i>Ergebnisorientierte Dimension zur Mitte des Schuljahres t_2</i>		
1	p12	bfbbev3 + bfbbev6
2	p22	bfbbev4 + bfbp01
<i>Prozessorientierte Dimension zu Beginn des Schuljahres t_1</i>		
3	p31	afbpo5 + afbpo6
4	p41	afbpo7 + afbpo9
<i>Prozessorientierte Dimension zur Mitte des Schuljahres t_2</i>		
3	p32	bfbpo5 + bfbpo6
4	p42	bfbpo7 + bfbpo9
<i>Selbstregulationsorientierte Dimension zu Beginn des Schuljahres t_1</i>		
5	p51	afbse4 + afbse9
6	p61	afbse6 + afbse8
<i>Selbstregulationsorientierte Dimension zur Mitte des Schuljahres t_2</i>		
5	p52	bfbse4 + bfbse9
6	p62	bfbse6 + bfbse8
<i>Dialogorientierte Dimension zu Beginn des Schuljahres t_1</i>		
7	p71	afbdi1 + afbdi3
8	p81	afbdi4 + afbdi6
<i>Dialogorientierte Dimension zur Mitte des Schuljahres t_2</i>		
7	p72	bfbdi1 + bfbdi3
8	p82	bfbdi4 + bfbdi6

In Abbildung 14 ist das längsschnittliche 4-Faktormodell der Feedbackqualität mit den *Item-Parcels* als manifeste Indikatoren der vier Feedbackdimensionen schematisch dargestellt. Wie im LS-Modell im vorherigen Abschnitt wurden auch hier Residualkorrelationen zwischen den gleichen *Parcels* über die Zeit zugelassen. Das Modell weist eine sehr gute Modellpassung auf ($\chi^2 = 75.96$, $df = 68$, $p(\chi^2) = .24$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .01 [.00, .03], $C_{Fit} = 1.00$, SRMR = .02). Die Faktorladungen in diesem Modell sind alle signifikant und liegen im Bereich $\lambda = .72$ und $\lambda = .83$ (vgl. Tabelle 16). Die Reliabilitäten der Indikatoren liegen über .51.

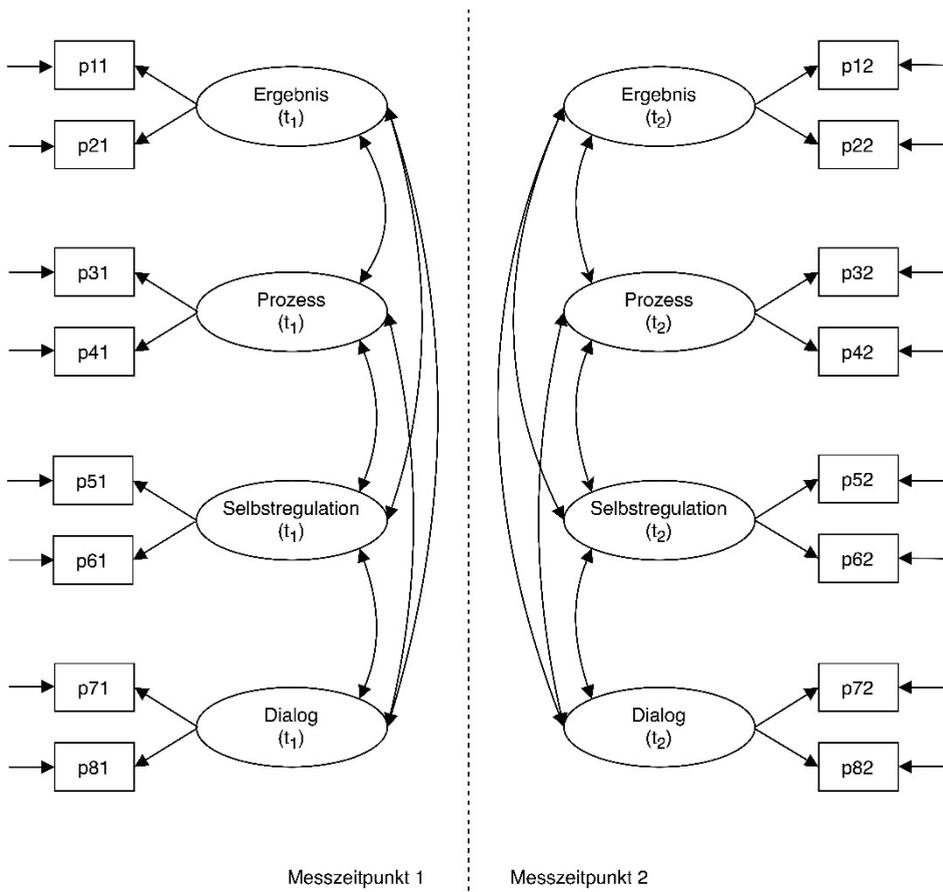


Abbildung 14: Schematische Darstellung des Latent-State-Modells mit Item-Parcels

Anmerkungen. Die Residualterme werden der besseren Übersichtlichkeit halber nur durch Pfeile dargestellt. Korrelationen zwischen den gleichen Konstrukten benachbarter Messzeitpunkte sind nicht dargestellt. Residualkorrelationen sind nicht dargestellt.

Tabelle 16: Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im Latent-State-Modell mit Item-Parcels

Skala/Parcel	λ	R^2
<i>Ergebnisorientierte Dimension t_1</i>		
p11	.76	.60
p21	.74	.70
<i>Ergebnisorientierte Dimension t_2</i>		
p12	.84	.55
p22	.82	.67
<i>Prozessorientierte Dimension t_1</i>		
p31	.72	.51
p41	.82	.62
<i>Prozessorientierte Dimension t_2</i>		
p32	.79	.66
p42	.82	.67
<i>Selbstregulationsorientierte Dimension t_1</i>		
p51	.75	.56
p61	.76	.67
<i>Selbstregulationsorientierte Dimension t_2</i>		
p52	.82	.58
p62	.83	.69
<i>Dialogorientierte Dimension t_1</i>		
p71	.77	.59
p81	.74	.68
<i>Dialogorientierte Dimension t_2</i>		
p72	.82	.54
p82	.77	.60

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

Tabelle 17 zeigt die Korrelationen zwischen den latenten Feedback-Faktoren im LS-Modell mit der *Parceling*-Methode. Die Korrelationen zwischen den State-Faktoren über die Zeit sind vergleichbar mit den Korrelationen im LS-Modell ohne *Item-Parcels*: Ergebnisorientierte Dimension: $r_{t12} = .69, p < .001$; prozessorientierte Dimension: $r_{t12} = .63, p < .001$; selbstregulationsorientierte Dimension: $r_{t12} = .62, p < .001$; dialogorientierte Dimension: $r_{t12} = .59, p < .001$. Insgesamt ist die Modellpassung des LS-Modells mit *Item-Parcels* als gut zu bewerten, bei gleichzeitiger Modellsparsamkeit. Das *Parcel*-Modell wird daher als Basismodell für die nachfolgenden Messinvarianzanalysen über die Zeit und weitergehenden Modellierungen im Längsschnitt herangezogen.

Tabelle 17: Latente Faktor-Interkorrelationen der Dimensionen der Feedbackqualität im Latent-State-Modell mit Item-Parcels

Latente Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) EO t ₁	1.00	.69	.81	.57	.66	.49	.60	.53
(2) EO t ₂		1.00	.61	.82	.55	.73	.43	.62
(3) PO t ₁			1.00	.63	.80	.56	.63	.50
(4) PO t ₂				1.00	.57	.80	.47	.66
(5) SO t ₁					1.00	.62	.62	.47
(6) SO t ₂						1.00	.43	.63
(7) DO t ₁							1.00	.59
(8) DO t ₂								1.00

Anmerkungen. Dargestellt sind die standardisierten Korrelationskoeffizienten; EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension; SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

7.2 Messinvarianz der Feedbackqualität

Im folgenden Abschnitt wird auf die weiteren Ergebnisse zur ersten Forschungsfrage eingegangen, in der untersucht werden soll, ob das Instrument zur Erfassung von Feedback messinvariant über die Zeit (*Teilfragestellung F1b*) und die Geschlechtergruppen (*Teilfragestellung F1c*) ist.

7.2.1 Messinvarianz über die Zeit

Voraussetzung für die Durchführung der Analysen zu den längsschnittlichen Fragestellungen ist die Prüfung der Frage, ob die Messungen der Konstrukte über die Zeit hinreichend vergleichbar sind. Hierzu wurde das längsschnittliche 4-Faktor-Parcel-Modell durch unterschiedliche Gleichheitsrestriktionen auf die verschiedenen Ebenen der Messinvarianz hin geprüft. Die Ergebnisse der Invarianzprüfungen sind Tabelle 18 (Modell-Fit-Indizes) und Tabelle 19 (Modellvergleiche) zu entnehmen.

Das Basismodell (Modell I), in dem die Gleichheit der Faktorenstruktur über die Zeit spezifiziert wird, wurde in Abschnitt 7.1.3 beschrieben. Es erfüllt die Voraussetzungen konfigurationaler Invarianz. Im Modell schwacher faktorieller Invarianz wurden die Faktorladungen über die Zeit gleichgesetzt (Modell II). Für das Modell liegen sehr gute Fit-Werte vor ($\chi^2 = 78.06$, $df = 72$, $p(\chi^2) = .29$, CFI = 1.00, TLI = 1.00, RMSEA = .01, SRMR = .02). Der Vergleich dieses restringierten Modells mit dem Basismodell (Modell I) mittels des korrigierten χ^2 -Differenztests zeigt keinen

signifikanten Unterschied ($\Delta\chi^2_{\text{korrr}} = 1.58$, $\Delta df = 4$, $p = .81$). Die CFI-Differenz von $-.001$ liegt unter der von Cheung & Rensvold (2002) beschriebenen Grenze von $.01$. Damit ist keine empirisch bedeutsame Abweichung vom Basismodell zu verzeichnen, weshalb das Modell schwacher faktorieller Invarianz beibehalten werden kann.

Tabelle 18: Modell-Fit-Indizes des 4-Faktor-Parcel-Modells der Feedbackqualität mit unterschiedlichem Grad der Messinvarianz

Modell	χ^2	df	p (χ^2)	RMSEA [90% CI]	CFit	SRMR	CFI	TLI
I	75.96	68	.24	.01 [.00, .03]	1.00	.02	1.00	1.00
II	78.06	72	.29	.01 [.00, .02]	1.00	.02	1.00	1.00
III	91.43	76	.11	.02 [.00, .03]	1.00	.02	1.00	1.00

Anmerkungen. I = Basismodell konfiguraler Messinvarianz, II = Modell schwacher faktorieller Messinvarianz, III = Modell starker faktorieller Messinvarianz.

Tabelle 19: Ergebnisse der Modellvergleiche

Modellvergleich	$\Delta\chi^2_{\text{korrr}}$	Δdf	p	ΔCFI
II vs. I	1.58	4	.81	-.001
III vs. I	16.22	8	.04	-.002
III vs. II	13.58	4	.01	-.001

Anmerkungen. I = Basismodell konfiguraler Messinvarianz, II = Modell schwacher faktorieller Messinvarianz, III = Modell starker faktorieller Messinvarianz.

Im darauffolgenden Modell starker faktorieller Messinvarianz (Modell III) wurde geprüft, ob zusätzlich zur Faktorstruktur und zu Faktorladungen auch die Intercepts über die Messzeitpunkte identisch sind. Die Modell-Fit-Statistiken weisen auf eine gute Passung an die Daten hin ($\chi^2 = 91.43$, $df = 76$, $p(\chi^2) = .11$, CFI = 1.00, TLI = 1.00, RMSEA = .02, SRMR = .02). Der Vergleich des restringierten Modells mit dem Basismodell mittels des korrigierten χ^2 -Differenztests legt nahe, das Modell zu Gunsten des Basismodells zu verwerfen ($\Delta\chi^2_{\text{korrr}} = 16.22$, $\Delta df = 8$, $p = .04$).

Auch der Vergleich mit dem Modell schwacher faktorieller Invarianz deutet auf eine signifikante Modellverschlechterung durch das Gleichsetzen der Intercepts hin ($\Delta\chi^2_{\text{korrr}} = 13.58$, $\Delta df = 4$, $p = .01$). Aufgrund der Stichprobensensitivität des χ^2 -Differenztests wird zusätzlich der CFI-Differenztest herangezogen. Die Differenz von $-.002$ (Modell III vs. I) bzw. $-.001$ (Modell II vs. I) liegt unter dem Cutoff-Wert von $\Delta CFI = .01$. Damit muss das Modell starker faktorieller Messinvarianz nicht verworfen werden, sodass Mittelwertvergleiche über die Zeit zulässig sind.

7.2.2 Messinvarianz über die Geschlechtergruppen

Um zu untersuchen, inwiefern sich die Mittelwerte in der Wahrnehmung der Feedbackdimensionen zwischen Schülerinnen und Schülern unterscheiden, muss zunächst geprüft werden, ob sich die Struktur des Konstrukts der wahrgenommenen Feedbackqualität für beide Gruppen abbilden lässt. Nur unter dieser Bedingung ist es zulässig, Mittelwerte zwischen den Gruppen zu vergleichen. Dies wurde durch die nachfolgende Prüfung auf Messinvarianz anhand eines multiplen Gruppenvergleichs im Rahmen von CFAs geklärt. Die Ergebnisse der Invarianzprüfung sind in Tabelle 20 (Fit-Indizes) und Tabelle 21 (Modellvergleiche) dargestellt.

Tabelle 20: Ergebnisse der Prüfung auf Messinvarianz der Feedbackqualität nach Geschlecht

Modell	χ^2	<i>df</i>	$p(\chi^2)$	RMSEA [90% CI]	C_{Fit}	SRMR	CFI	TLI
I weiblich	62. 22	68	.67	.00 [.00, .02]	1.16	.02	1.00	1.00
I männlich	90. 02	68	.04	.03 [.01, .04]	.99	.03	.99	.98
II	152 .13	136	.16	.02 [.00, .03]	1.00	.02	1.00	.99
III	167 .79	144	.09	.02 [.00, .03]	1.00	.04	1.00	.99
IV	182 .57	152	.05	.02 [.00, .03]	1.00	.04	.99	.99

Anmerkungen. I weiblich = Basismodell weibliche Stichprobe, I männlich = Basismodell männliche Stichprobe, II = Basismodell konfiguraler Messinvarianz, III = Modell schwacher faktorieller Messinvarianz, IV = Modell starker faktorieller Messinvarianz.

Tabelle 21: Ergebnisse der Modellvergleiche

Modellvergleich	$\Delta\chi^2_{\text{korr}}$	Δdf	<i>p</i>	ΔCFI
III vs. II	15.71	8	.05	-.002
IV vs. II	30.91	16	.01	-.003
IV vs. III	15.18	8	.05	-.001

Anmerkungen. II = Basismodell konfiguraler Messinvarianz, III = Modell schwacher faktorieller Messinvarianz, IV = Modell starker faktorieller Messinvarianz.

Im ersten Schritt wurde das Strukturmodell in beiden Gruppen separat überprüft (Modell I). Die Modell-Fit-Werte der CFAs sind gut bis sehr gut, sodass im nächsten Schritt das Baseline-Modell (Modell II) – in dem alle Parameter frei variieren

dürfen – im Rahmen eines multiplen Gruppenvergleichs für beide Gruppen simultan geschätzt werden kann. Die Fit-Werte können auch hier als gut bewertet werden und bestätigen damit die Annahme von konfigurationaler Messinvarianz, d.h. der Gleichheit der Faktorstruktur. Anschließend wurden zur Überprüfung eines Modells schwacher faktorieller Invarianz die Faktorladungen in den beiden Gruppen gleichgesetzt (Modell III). Der Modell-Fit dieses restringierten Modells ist sehr gut, allerdings fällt der χ^2 -Differenztest signifikant aus ($\Delta\chi^2_{\text{korrr}} = 15.71$, $\Delta df = 8$, $p = .05$). Daher wird zusätzlich der CFI-Differenztest zum Modellvergleich herangezogen. Der CFI-Wert liegt mit $-.002$ unter der vorgegebenen Cutoff-Grenze, sodass schwache faktorielle Messinvarianz angenommen werden kann. Um die Mittelwerte zwischen den Gruppen vergleichen zu können, muss starke faktorielle Messinvarianz vorliegen. Hierzu wurde das Modell IV spezifiziert, in dem zusätzlich zur Faktorstruktur und den Faktorladungen auch die Intercepts der manifesten Indikatoren in den beiden Gruppen gleich sind. Der Modell-Fit ist sehr gut. Allerdings zeigt der Vergleich mit dem Baseline-Modell und dem Modell schwacher faktorieller Invarianz mittels des korrigierten χ^2 -Differenztests eine (tendenziell) bedeutsame Verschlechterung (Modell IV vs. Modell II: $\Delta\chi^2_{\text{korrr}} = 30.91$, $\Delta df = 8$, $p = .01$; Modell IV vs. Modell III: $\Delta\chi^2_{\text{korrr}} = 15.18$, $\Delta df = 8$, $p = .05$). Da der CFI-Differenzwert unter der Cutoff-Grenze liegt ($\Delta CFI = -.003$ bzw. $-.001$), wird das Modell schwacher faktorieller bzw. starker faktorieller Invarianz nicht verworfen.

Geschlechtsspezifische Unterschiede in der wahrgenommenen Feedbackqualität

Nachdem die Messäquivalenz in den Dimensionen der Feedbackqualität zwischen den Gruppen (Schülerinnen und Schüler) nachgewiesen werden konnte, können schließlich die Ergebnisse des multiplen Gruppenvergleichs betrachtet werden und damit ein Teil der Hauptforschungsfrage 1 zur differenziellen Feedbackwahrnehmung in Abhängigkeit des Geschlechts beantwortet werden. Die Ergebnisse hierzu sind in Tabelle 22 dargestellt.

Die Referenzgruppe ist dabei die Gruppe der Schüler. Unterschiede werden mittels latent geschätzter, standardisierter Mittelwertdifferenzen in Skalenpunkten angegeben. Der Gruppenvergleich zeigt, dass zu beiden Messzeitpunkten keine signifikanten geschlechtsspezifischen Unterschiede hinsichtlich der Wahrnehmung der Dimensionen der Feedbackqualität vorliegen. Schülerinnen und Schüler unterscheiden sich demnach nicht bedeutsam in ihrer Wahrnehmung von Feedback.

Tabelle 22: Geschlechtsspezifische latente Mittelwertdifferenzen in der Feedbackwahrnehmung zu Beginn und zur Mitte des Schuljahres (MZP 1/MZP 2)

Skala	MZP	<i>MW</i>	<i>SE</i>	<i>p</i>
EO	1	0.07	0.11	.53
EO	2	0.05	0.11	.66
PO	1	0.04	0.09	.66
PO	2	0.08	0.10	.44
SO	1	-0.11	0.07	.11
SO	2	-0.12	0.10	.21
DO	1	-0.12	0.09	.16
DO	2	0.01	0.13	.94

Anmerkungen. Referenzgruppe: männlich; *n* (Schülerinnen) = 429, *n* (Schüler) = 348; EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension; SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension.

7.3 Differenzielle Veränderung der Feedbackwahrnehmung

Der nachfolgende Abschnitt widmet sich der Frage nach der Veränderung der Feedbackwahrnehmung im Verlauf eines Schulhalbjahres (*Teilfragestellung F1d*) und der Vorhersage von interindividuellen Unterschieden im Ausgangsniveau und in der Veränderung der Feedbackwahrnehmung in Abhängigkeit des Geschlechts (*Teilfragestellung F1e*). Hierzu werden zunächst die auf Basis des längsschnittlichen Latent-State-Modells geschätzten latenten Mittelwerte zu Beginn und zur Mitte des Schuljahres betrachtet (Abschnitt 7.3.1) und anschließend die latenten Veränderungen im Latent-True-Change-Modell (Abschnitt 7.3.2).

7.3.1 Deskriptive Entwicklungen

Um einen ersten Überblick über die Mittelwertveränderungen in den Feedbackdimensionen über die Zeit zu erhalten, werden die geschätzten latenten Mittelwerte der Dimensionen der Feedbackqualität zu Beginn und zur Mitte des Schuljahres und die Mittelwertdifferenzen über die beiden Messzeitpunkte betrachtet (vgl. Tabelle 23).

Tabelle 23: Latente Faktormittelwerte (MW), -varianzen (Var) der Dimensionen der Feedbackqualität im Latent-State-Modell sowie Mittelwertdifferenzen

Latente Variable	MZP 1		MZP 2		ΔMW
	$MW (SE)$	$Var (SE)$	$MW (SE)$	$Var (SE)$	
(1) EO	3.09 (0.05)	0.22 (0.03)	2.99 (0.04)	0.26 (0.02)	-0.10***
(2) PO	2.95 (0.04)	0.25 (0.02)	2.99 (0.05)	0.27 (0.03)	0.05
(3) SO	2.91 (0.03)	0.20 (0.02)	2.96 (0.04)	0.27 (0.03)	0.06
(4) DO	2.94 (0.04)	0.25 (0.02)	3.05 (0.05)	0.29 (0.03)	0.10***

Anmerkungen. EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension, SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension. ** $p < .01$, *** $p < .001$.

Es ist erkennbar, dass die latenten mittleren Ausprägungen in den Dimensionen der Feedbackqualität zu Beginn des Schuljahres über dem theoretischen Mittelwert von 2.5 liegen. Die Ausprägung in der ergebnisorientierten Dimension ist etwas höher als in den anderen drei Dimensionen², während die Ausprägung in der selbstregulationsorientierten Dimension zu MZP 1 am geringsten ausfällt³. Zur Mitte des Schuljahres liegen die mittleren Ausprägungen in den Dimensionen der Feedbackqualität weiterhin über dem theoretischen Mittelwert. Zur Mitte des Schuljahres ist die Ausprägung in der dialogorientierten Dimension etwas höher als in den anderen Dimensionen⁴, während die Wahrnehmung der selbstregulationsorientierten Dimension weiterhin am geringsten ausgeprägt ist.⁵ Anhand der Mittelwerte in

² Ein Modellvergleich mit sukzessiver Gleichsetzung der Mittelwerte lässt darauf schließen, dass sich die zu MZP 1 geschätzte mittlere Ausprägung in der ergebnisorientierten Dimension statistisch signifikant von der Ausprägung in der prozessorientierten Dimension ($\Delta\chi^2_{\text{kor}} = 33.43$, $\Delta df = 1$, $p < .001$), der selbstregulationsorientierten Dimension ($\Delta\chi^2_{\text{kor}} = 33.95$, $\Delta df = 1$, $p < .001$) und der dialogorientierten Dimension ($\Delta\chi^2_{\text{kor}} = 33.95$, $\Delta df = 1$, $p < .001$) unterscheidet.

³ Auch hier ergibt ein Modellvergleich mit sukzessiver Gleichsetzung der Mittelwerte der selbstregulationsorientierten Dimension und einer anderen Feedback-Dimension, dass sich die zu MZP 1 geschätzte mittlere Ausprägung in der selbstregulationsorientierten Dimension statistisch signifikant von der Ausprägung in der ergebnisorientierten Dimension ($\Delta\chi^2_{\text{kor}} = 33.95$, $\Delta df = 1$, $p < .001$) und der prozessorientierten Dimension ($\Delta\chi^2_{\text{kor}} = 4.39$, $\Delta df = 1$, $p = .04$) unterscheidet. Lediglich die Gleichsetzung der Mittelwerte in der selbstregulationsorientierten Dimension und dialogorientierten Dimension führt nicht zu einer signifikanten Modellverschlechterung ($\Delta\chi^2_{\text{kor}} = 2.29$, $\Delta df = 1$, $p = .13$), was darauf schließen lässt, dass sich die Mittelwerte in diesen Dimensionen der Feedbackqualität nicht bedeutsam unterscheiden.

⁴ Ein Modellvergleich zeigt, dass lediglich die Gleichsetzung der zu MZP 2 geschätzten Mittelwerte in der dialogorientierten Dimension und der selbstregulationsorientierten Dimension zu einer signifikanten Modellverschlechterung führt ($\Delta\chi^2_{\text{kor}} = 10.65$, $\Delta df = 1$, $p = .001$), was auf einen bedeutsamen Mittelwertunterschied zu MZP 2 schließen lässt.

⁵ Ein Modellvergleich mit sukzessiver Gleichsetzung der Mittelwerte zeigt, dass sich lediglich die Mittelwerte zwischen der selbstregulationsorientierten Dimension und der dialogorientierten Dimension signifikant unterscheiden (s.o.).

den Feedbackdimensionen ist erkennbar, dass sich die Wahrnehmungen kaum verändert haben. Die mittleren Ausprägungen in der Wahrnehmung der prozessorientierten ($MW_{t1} = 2.95$, $SE_{t1} = 0.04$, $MW_{t2} = 2.99$, $SE_{t2} = 0.05$) und selbstregulationsorientierten ($MW_{t1} = 2.91$, $SE_{t1} = 0.03$, $MW_{t2} = 2.96$, $SE_{t2} = 0.04$) Dimension sind nahezu konstant geblieben. Lediglich für die ergebnisorientierte Dimension ist eine leichte Abnahme ($MW_{t1} = 3.09$, $SE_{t1} = 0.05$, $MW_{t2} = 2.99$, $SE_{t2} = 0.04$) und für die dialogorientierte Dimension ein leichter Anstieg ($MW_{t1} = 2.94$, $SE_{t1} = 0.04$, $MW_{t2} = 3.05$, $SE_{t2} = 0.05$) in der mittleren Ausprägung zu verzeichnen. Abbildung 15 zeigt zusammenfassend die deskriptiven Entwicklungen der Feedbackdimensionen.

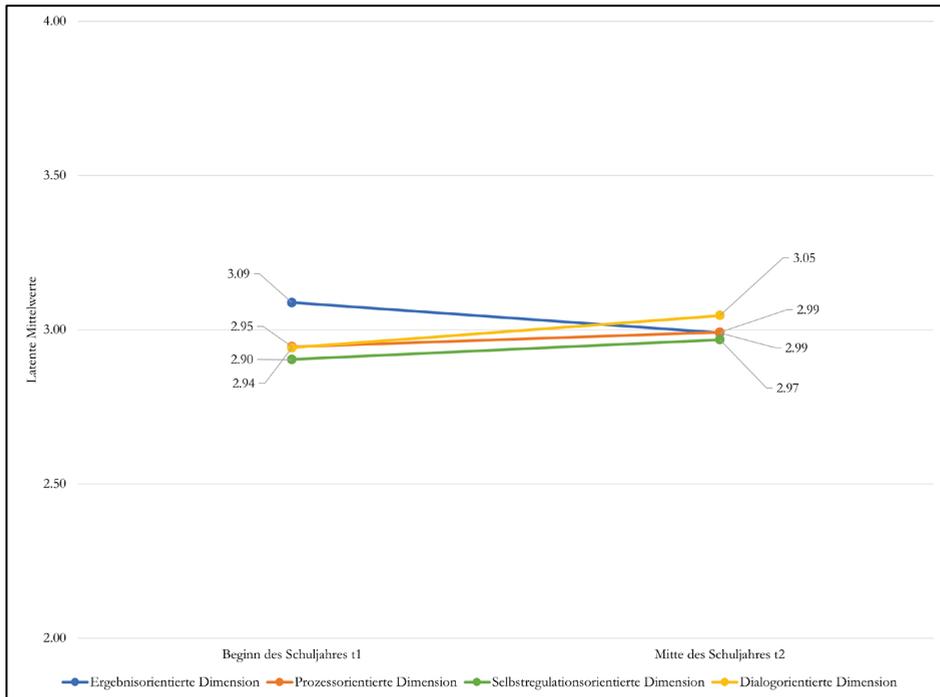


Abbildung 15: Veränderung der wahrgenommenen Dimensionen von Feedbackqualität

Um zu erkennen, ob sich die Mittelwerte in den Feedbackdimensionen zwischen den beiden Messzeitpunkten signifikant unterscheiden, wurden im LS-Modell sukzessive die Mittelwerte der vier Feedbackdimensionen über die Zeit gleichgesetzt und geprüft, ob diese Restriktionen zu einer Verschlechterung des Modell-Fits im Vergleich zum nicht restringierten LS-Modell mit frei geschätzten Mittelwerten führen. Eine Verschlechterung des Modell-Fits lässt auf eine bedeutsame Veränderung der jeweiligen Dimension der Feedbackqualität schließen (Geiser, 2011; vgl. Anwendung in Buff, 2014). Für die ergebnisorientierte Dimension der Feedbackqualität zeigt sich, dass das Modell mit gleichgesetzten Mittelwerten im Vergleich zum

Modell mit frei geschätzten Mittelwerten signifikant schlechter an die Daten passt ($\Delta\chi^2_{\text{korrr}} = 8.94$, $\Delta df = 1$, $p = .003$). Dies lässt auf eine signifikante Abnahme der Wahrnehmung der ergebnisorientierten Feedbackdimension schließen. Allerdings ist der absolute Fit des restriktiveren Modells immer noch als sehr gut zu bewerten ($\chi^2 = 112.81$, $df = 77$, $p(\chi^2) = .005$, CFI = .99, TLI = .99, RMSEA [90% CI] = .02 [.01, .03], $C_{\text{Fit}} = 1.00$, SRMR = .02), was darauf hindeutet, dass die Mittelwertunterschiede als eher gering einzustufen sind. Für die prozessorientierte Dimension weist der korrigierte χ^2 -Differenztest auf eine nicht signifikante Verschlechterung des Modell-Fits hin ($\Delta\chi^2_{\text{korrr}} = 2.06$, $\Delta df = 1$, $p = .15$). Dies deutet darauf hin, dass sich die Wahrnehmung der prozessorientierten Dimension nicht bedeutsam verändert. Auch die Gleichsetzung der Mittelwerte der selbstregulationsorientierten Dimension führt nicht zu einer signifikanten Modellverschlechterung ($\Delta\chi^2_{\text{korrr}} = 3.53$, $\Delta df = 1$, $p = .06$), d.h. auch die Wahrnehmung der selbstregulationsorientierten Dimension verändert sich nicht bedeutsam. Für die dialogorientierte Dimension wiederum verweist der χ^2 -Differenztest auf eine signifikante Verschlechterung des Modell-Fits ($\Delta\chi^2_{\text{korrr}} = 16.21$, $\Delta df = 1$, $p < .001$) und somit auf eine signifikante Zunahme der Wahrnehmung der dialogorientierten Dimension zum 2. Messzeitpunkt. Allerdings ist der absolute Fit des restriktiveren Modells mit gleichgesetzten Mittelwerten immer noch als sehr gut zu bewerten ($\chi^2 = 107.74$, $df = 77$, $p(\chi^2) = .01$, CFI = .99, TLI = .99, RMSEA [90% CI] = .02 [.01, .03], $C_{\text{Fit}} = 1.00$, SRMR = .03).

Die geschätzten latenten Faktorvarianzen fallen für alle Dimensionen signifikant aus ($p < .001$), d.h. es sind interindividuelle Unterschiede in den Ausprägungen der Dimensionen der Feedbackqualität zu beiden Messzeitpunkten vorhanden.

7.3.2 Veränderung im Latent-True-Change-Modell

Um die Veränderung in der Feedbackwahrnehmung über die zwei Messzeitpunkte zu modellieren, wurde ein Latent-True-Change-Modell (LTC-Modell) geschätzt. Hierzu wurde das im Abschnitt 7.1.3 beschriebene längsschnittliche Parcel-Modell zu einem LTC-Modell umgewandelt. Die Modellierung der Veränderung der Feedbackqualität erfolgte über die latenten Differenzen der Wahrnehmungsdimensionen zwischen den beiden Messzeitpunkten. Für die Annahme starker faktorieller Messinvarianz als Voraussetzung für die Längsschnittanalyse wurden die Faktorladungen und Intercepts über die Zeit gleichgesetzt (analog zum LS-Parcel-Modell in Abschnitt 7.1.3). Das LTC-Modell mit starker faktorieller Messinvarianz ist in Abbildung 16 dargestellt.

Der Fit des LTC-Modells mit invarianten Ladungen und Intercepts ist identisch zum Fit des LS-Modells mit starker faktorieller Invarianz (vgl. Tabelle 18): $\chi^2 = 91.43$, $df = 76$, $p(\chi^2) = .11$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .02 [.00, .03], $C_{\text{Fit}} = 1.00$, SRMR = .02. Die Faktorladungen im Modell betragen zwischen .64 und .81. Die Indikatorreliabilitäten liegen über .52 und damit über dem geforderten Schwellenwert.

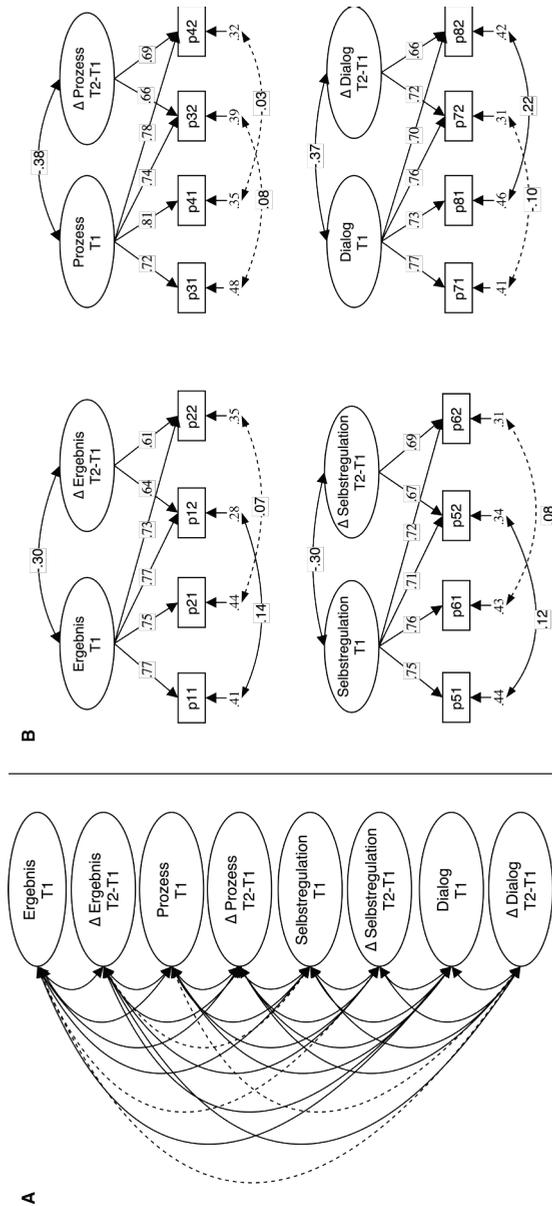


Abbildung 16: Latent-True-Change-Modell für die Veränderung der Feedbackwahrnehmung über die Zeit

Anmerkungen. A: Strukturmodell. B: Messmodelle mit autokorrelierten Fehlertermen. Dargestellt sind die standardisierten Parameterschätzungen. Nicht-signifikante Korrelationen sind gestrichelt dargestellt. Δ = Latente Veränderungsvariable.

In Tabelle 24 sind die Mittelwerte und Varianzen der latenten Variablen zum ersten Messzeitpunkt und der latenten Differenzvariablen dargestellt. Die latenten Differenzwerte entsprechen den berichteten Differenzen im längsschnittlichen Strukturmodell (vgl. Tabelle 23). Wie schon in Abschnitt 7.3.1 beschrieben, zeigen auch die Ergebnisse der LTC-Analyse eine leichte Abnahme in der mittleren Wahrnehmung der ergebnisorientierten Dimension vom Beginn des Schuljahres zur Mitte des Schuljahres. Diese Veränderung fällt statistisch signifikant aus ($MW_{t_1} = 3.09$, $SE = 0.05$, $\Delta MW_{t_2-t_1} = -0.10$, $p < .01$). Die mittleren Ausprägungen in der prozessorientierten Dimension, selbstregulationsorientierten Dimension und dialogorientierten Dimension nehmen etwas zu, wobei diese Veränderung nur für die selbstregulationsorientierte Dimension ($MW_{t_1} = 2.91$, $SE = 0.03$, $\Delta MW_{t_2-t_1} = 0.06$, $p = .05$) und dialogorientierte Dimension ($MW_{t_1} = 2.94$, $SE = 0.03$, $\Delta MW_{t_2-t_1} = 0.10$, $p < .001$) statistisch signifikant ausfällt. Allerdings zeigte das LS-Modell, in dem die Mittelwerte der State-Faktoren für beide Messzeitpunkte geschätzt wurden, dass die Mittelwertunterschiede über die Zeit – insbesondere in Bezug auf die selbstregulationsorientierte Dimension – als eher gering einzustufen sind.

Tabelle 24: Mittelwerte (MW) und Varianzen (Var) der latenten Faktoren im Latent-True-Change-Modell

Latente Variable	MW (SE)	Var (SE)
(1) EO t_1	3.09*** (0.05)	0.22*** (0.03)
(2) Δ EO t_2-t_1	-0.10** (0.03)	0.15*** (0.03)
(3) PO t_1	2.95*** (0.04)	0.25*** (0.02)
(4) Δ PO t_2-t_1	0.05 (0.03)	0.19*** (0.02)
(5) SO t_1	2.91*** (0.03)	0.20*** (0.02)
(6) Δ SO t_2-t_1	0.06* (0.03)	0.19*** (0.03)
(7) DO t_1	2.94*** (0.04)	0.25*** (0.02)
(8) Δ DO t_2-t_1	0.10*** (0.03)	0.22*** (0.03)

Anmerkungen. Dargestellt sind die unstandardisierten Parameterschätzungen; EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension; SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension; Δ = Latente Veränderungsvariable. * $p < .05$, ** $p < .01$, *** $p < .001$.

Tabelle 25 zeigt die Korrelationen der Veränderungsvariablen mit den Ausgangswerten und die Korrelationen zwischen den Veränderungsvariablen im LTC-Modell. Die latenten Korrelationen zwischen dem Ausgangswert und dem Veränderungswert innerhalb einer Dimension der Feedbackqualität fallen statistisch bedeutsam aus (Ergebnisdimension/ Selbstregulationsdimension: $r_{12} = -.30$, $p < .001$;

Prozessdimension/ Dialogdimension: $r_{t12} = -.37, p < .001$). Die negativen Korrelationskoeffizienten deuten darauf hin, dass ein hoher Ausgangswert mit einer schwächeren Veränderung in der Wahrnehmung der jeweiligen Dimension bzw. ein geringer Ausgangswert mit einer stärkeren Veränderung in der Wahrnehmung einhergeht. Das heißt, dass sich Schülerinnen und Schüler mit einer hohen Ausprägung in der Wahrnehmung von Feedback zu Beginn des Schulhalbjahres weniger in ihrer Wahrnehmung im Verlauf des Schulhalbjahres verändern als Schülerinnen und Schüler mit einer anfänglich geringen Feedbackwahrnehmung. Die positiven Korrelationen zwischen den Veränderungsvariablen Δ ergebnisorientiert t_2-t_1 , Δ prozessorientiert t_2-t_1 , Δ selbstregulationsorientiert t_2-t_1 sowie Δ dialogorientiert t_2-t_1 deuten darauf hin, dass eine Veränderung in einer Dimension der Feedbackqualität auch eine Veränderung in einer anderen Dimension zufolge hat. Anders ausgedrückt: Die Abnahme bzw. Zunahme in der Wahrnehmung der einen Dimension geht mit einer Abnahme bzw. Zunahme der Wahrnehmung in der anderen Dimension einher. Die Korrelationen zwischen dem Ausgangswert einer Dimension und dem Veränderungswert einer anderen Dimension der Feedbackqualität fallen nur schwach (z.B. ergebnisorientiert t_1 und Δ prozessorientiert t_2-t_1 : $r = -.23, p < .001$) oder nicht signifikant aus (z.B. ergebnisorientiert t_1 und Δ selbstregulationsorientiert t_2-t_1 : $r = -.10, n.s.$), was auf einen geringen Zusammenhang hindeutet.

Tabelle 25: Interkorrelationen der latenten Faktoren im Latent-True-Change-Modell

Latente Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) EO t_1	1.00	-.30***	.81***	-.23***	.66***	-.10	.60***	-.02
(2) Δ EO t_2-t_1		1.00	-.16*	.65***	-.07	.52***	-.15*	.36***
(3) PO t_1			1.00	-.37***	.80***	-.16**	.63***	-.09
(4) Δ PO t_2-t_1				1.00	-.22*	.62***	-.15**	.42***
(5) SO t_1					1.00	-.30***	.62***	-.11*
(6) Δ SO t_2-t_1						1.00	-.14*	.45***
(7) DO t_1							1.00	-.37***
(8) Δ DO t_2-t_1								1.00

Anmerkungen. EO = ergebnisorientierte Dimension; PO = prozessorientierte Dimension; SO = selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension; dargestellt sind standardisierte Korrelationen; Δ = Latente Veränderungsvariable; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

7.3.3 Effekte auf die Veränderung der Feedbackwahrnehmung

In diesem Abschnitt erfolgt die Prüfung der Hypothese, in der davon ausgegangen wird, dass das Geschlecht und die Deutschnote der Schülerinnen und Schüler prädiktiv für interindividuelle Unterschiede in der intraindividuellen Veränderung der Wahrnehmung von Feedback sind (*Teilfragestellung F1e*). Hierzu wurde ein Strukturgleichungsmodell spezifiziert, in dem direkte Effekte der Hintergrundvariablen auf die Veränderungen der Feedbackdimensionen (unter Kontrolle der Ausgangswahrnehmung zu Messzeitpunkt 1) modelliert wurden. Korrelationen zwischen den unabhängigen Variablen wurden zugelassen. Das so spezifizierte Modell weist eine sehr gute Datenanpassung auf ($\chi^2 = 113.50$, $df = 92$, $p(\chi^2) = .06$, CFI = 1.00, TLI = .99, RMSEA [90% CI] = .02 [.00, .03], $C_{Fit} = 1.00$, SRMR = .02). Der χ^2 -Test fällt nicht signifikant aus und auch die anderen Fit-Indizes weisen auf einen sehr gut Modell-Fit hin.

Die Ergebnisse der Strukturgleichungsanalysen sind in Tabelle 26 (standardisierte Regressionskoeffizienten) und Tabelle 27 (unstandardisierte Regressionskoeffizienten) dargestellt. Die Ergebnisse zeigen signifikante Zusammenhänge des Geschlechts und der Deutschnote mit den Feedbackdimensionen zu Beginn des Schuljahres: Die Deutschleistung hängt positiv mit der Wahrnehmung des ergebnisorientierten Feedbacks ($\beta = .14$, $SE = 0.05$, $p = .01$), des prozessorientierten Feedbacks ($\beta = .17$, $SE = 0.05$, $p = .001$), des selbstregulationsorientierten Feedbacks ($\beta = .18$, $SE = 0.05$, $p < .001$) und des dialogorientierten Feedbacks ($\beta = .14$, $SE = 0.05$, $p = .01$) zusammen. Demzufolge geht eine bessere Leistung mit einer positiveren Wahrnehmung der Feedbackqualität einher. Unter Kontrolle der Vorleistung wird auch der Einfluss des Geschlechts auf die Wahrnehmung des selbstregulationsorientierten Feedbacks ($\beta = -.09$, $SE = 0.04$, $p = .02$) und des dialogorientierten Feedbacks ($\beta = -.10$, $SE = 0.04$, $p = .03$) signifikant: Bei gleicher Deutschnote nehmen Schülerinnen die Ausprägung des selbstregulationsorientierten und dialogorientierten Feedbacks geringer wahr als Schüler. Die Modelle erklären allerdings lediglich einen geringen Anteil der Varianz in der Wahrnehmung des selbstregulationsorientierten Feedbacks ($R^2 = .03$, $p = .05$) und des dialogorientierten Feedbacks ($R^2 = .02$, $p = .12$) zu Beginn des Schuljahres.

Entgegen den Erwartungen können in diesem Modell keine statistisch signifikanten Effekte der Hintergrundmerkmale auf die Veränderung der Feedbackwahrnehmung im Schulhalbjahr nachgewiesen werden – weder für das Geschlecht ($.01 \leq \beta \leq .09$, n. s.) noch die Deutschnote der Schülerinnen und Schüler ($-.09 \leq \beta \leq -.05$, n. s.).

Tabelle 26: Standardisierte Regressionskoeffizienten in den Latent-True-Change-Modellen mit Geschlecht und Deutschnote als Prädiktoren der Wahrnehmungsveränderung von Feedback

	Ergebnis		Prozess		Selbstregulation		Dialog	
	Level	Change	Level	Change	Level	Change	Level	Change
Prädiktoren	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)
Geschlecht	.00 (0.06)	.01 (0.05)	-.02 (0.05)	.04 (0.06)	-.10** (0.04)	.00 (0.06)	-.10* (0.04)	.09 (0.06)
Note	.14** (0.05)	-.09 (0.06)	.17*** (0.05)	-.09 (0.05)	.18*** (0.05)	-.05 (0.07)	.14** (0.05)	-.07 (0.07)
R ²	.02	.01	.03	.01	.03	.00	.02	.01

Anmerkungen. Ergebnis = Ergebnisorientierte Dimension der Feedbackqualität; Prozess = Prozessorientierte Dimension der Feedbackqualität; Selbstregulation = Selbstregulationorientierte Dimension der Feedbackqualität; Dialog = Dialogorientierte Dimension der Feedbackqualität; das Geschlecht (1 = weiblich) ist eine dichotomisierte Variable; die Deutschnote wurde rekodiert (hohe Werte entsprechen einer guten Note). * $p < .05$; ** $p < .01$; *** $p < .001$.

Tabelle 27: Unstandardisierte Regressionskoeffizienten in den Latent-True-Change-Modellen mit Geschlecht und Deutschnote als Prädiktoren der Wahrnehmungsveränderung von Feedback

	Ergebnis		Prozess		Selbstregulation		Dialog	
	Level	Change	Level	Change	Level	Change	Level	Change
Prädiktoren	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)
Geschlecht	.00 (0.06)	.01 (0.04)	-.02 (0.05)	.04 (0.05)	-.09** (0.03)	.00 (0.05)	-.10* (0.05)	.08 (0.06)
Note	.07** (0.03)	-.04 (0.02)	.09** (0.03)	-.04 (0.03)	.09*** (0.03)	-.02 (0.03)	.08** (0.03)	-.03 (0.03)
R ²	.02	.01	.03	.01	.03	.00	.02	.01

Anmerkungen. Ergebnis = Ergebnisorientierte Dimension der Feedbackqualität; Prozess = Prozessorientierte Dimension der Feedbackqualität; Selbstregulation = Selbstregulationorientierte Dimension der Feedbackqualität; Dialog = Dialogorientierte Dimension der Feedbackqualität; das Geschlecht (1 = weiblich) ist eine dichotomisierte Variable; die Deutschnote wurde rekodiert (hohe Werte entsprechen einer guten Note). * $p < .05$; ** $p < .01$; *** $p < .001$.

7.4 Zusammenhänge zwischen Feedback und Lernmotivation

Im Folgenden werden die reziproken Zusammenhänge zwischen den Dimensionen von wahrgenommener Feedbackqualität und der intrinsischen Lernmotivation sowie ihren Veränderungsvariablen betrachtet (*Hauptforschungsfrage II*). Hier wird die Skala zur Erfassung der intrinsischen Lernmotivation zunächst einer faktorenanalytischen Prüfung zu beiden Messzeitpunkten separat sowie im Längsschnitt unterzogen. Dabei werden auch die deskriptiven Befunde sowie die Ergebnisse zu den Messinvarianzanalysen berichtet. Im Anschluss daran wird das Latent-True-Change-Modell zur Modellierung der Veränderung der intrinsischen Motivation spezifiziert.

7.4.1 Faktorenstruktur und deskriptive Statistik der Skala Lernmotivation

Im Rahmen einer konfirmatorischen Faktorenanalyse wurde die Eindimensionalität der Skala zur Erfassung der intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres geprüft. Die Modellgütekriterien sind Tabelle 28 zu entnehmen. Der Modell-Fit fällt im Hinblick auf den CFI, TLI und SRMR gut aus, da die gängigen Cut-Off-Werte nicht überschritten werden. Der χ^2 -Test, der Quotient aus χ^2/df und der RMSEA weisen hingegen auf eine nicht akzeptable Modellpassung hin. Die Modellierung einer Residualkorrelation (Kline, 2016) zwischen den Items 1 und 2 (aufgrund der benachbarten Platzierung im Fragebogen) führt zu einer substantziellen Verbesserung des Modell-Fits zu MZP 1.

Tabelle 28: Ergebnisse der konfirmatorischen Faktorenanalyse zur intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres (MZP 1 und MZP 2)

Modell	MZP	χ^2	df	χ^2/df	RMSEA [90% CI]	C_{Fit}	CFI	TLI	SRMR
IM	1	90.04*	9	10.00	.10 [.09, .13]	.00	1.00	.99	.02
IM (mit Methodenfaktor)	1	39.48*	8	4.94	.07 [.05, .09]	.06	1.00	1.00	.01
IM	2	71.50*	9	7.94	.10 [.08, .13]	.00	1.00	1.00	.02
IM (mit Methodenfaktor)	2	63.33*	8	7.92	.10 [.08, .13]	.00	1.00	1.00	.02

Anmerkungen. IM = Intrinsische Lernmotivation. * $p < .001$.

Die standardisierten Parameterschätzungen sind Tabelle 29 zu entnehmen. Alle Indikatoren verfügen über signifikante ($p < .001$) und hohe standardisierte Faktorladungen, die für t_1 im Bereich $\lambda = .74$ und $\lambda = .96$ und für t_2 im Bereich $\lambda = .78$ und $\lambda = .96$ liegen. Die R^2 -Werte deuten auf überwiegend hohe Indikatorreliabilitäten hin.

Die Reliabilitätsanalysen zu Beginn und zur Mitte des Schuljahres ergaben sehr gute Werte von $\alpha_{t1} = .93/\omega_{t1} = .93$ und $\alpha_{t2} = .94/\omega_{t2} = .93$ (vgl. Tabelle 30). Die deskriptiven Statistiken in Tabelle 30 zeigen, dass die intrinsische Lernmotivation zu Beginn und zur Mitte des Schuljahres unter dem theoretischen Mittelwert der Skala von 2.5 liegt. Auf deskriptiver Ebene zeigt sich außerdem, dass die intrinsische Lernmotivation im Verlauf des Schulhalbjahres sinkt: Zum ersten Messzeitpunkt berichten die Schülerinnen und Schüler im Mittel über eine höhere intrinsische Lernmotivation ($MW = 2.35$, $SD = 0.83$) als zum zweiten Messzeitpunkt ($MW = 2.25$, $SD = 0.84$). Die ICC-Koeffizienten zeigen, dass ein Varianzanteil von 16% (t_1) bzw. 21% (t_2) in den Schülereinschätzungen auf Unterschiede zwischen den Deutschkursen in der intrinsischen Lernmotivation zurückzuführen ist. Der Rest geht auf Unterschiede zwischen den Schülerinnen und Schülern selbst zurück.

Tabelle 29: Standardisierte Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im eindimensionalen Modell der intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres (MZP 1 und 2)

Skala/ Item	MZP 1		MZP 2	
	λ	R^2	λ	R^2
intrin1	.88	.78	.91	.82
intrin2	.94	.88	.96	.92
intrin3	.84	.70	.90	.82
intrin4	.96	.93	.94	.89
intrin5	.81	.65	.81	.65
intrin6	.74	.55	.78	.61

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

Tabelle 30: Skalenmittelwerte (MW), Standardabweichungen (SD), Reliabilitäten (α und ω) und Intraklassenkorrelationen (ICC) der Skala zur intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres

MZP	N	MW	SD	α	ω	ICC
1	804	2.35	0.83	.93	.93	.16
2	626	2.25	0.84	.94	.93	.21

7.4.2 Modellierung der Lernmotivation im Längsschnitt

Das eindimensionale Messmodell zur Modellierung der intrinsischen Lernmotivation im Längsschnitt ist in Abbildung 17 dargestellt. Die Modellgütekriterien weisen auf eine gute Anpassung an die empirischen Daten hin: $\chi^2 = 148.72$, $df = 45$, $p(\chi^2) < .001$, CFI = 1.00, TLI = 1.00; RMSEA [90% CI] = .05 [.04, .06], CFI_{it} = .26, SRMR = .02. Die hohen Korrelationen von $r_{t12} = .83$ ($p < .001$) zwischen den latenten (State-)Faktoren der intrinsischen Lernmotivation weisen auf eine hohe zeitliche Stabilität interindividueller Unterschiede hinsichtlich der intrinsischen Lernmotivation (vgl. Geiser, 2011). Die standardisierten Faktorladungen der Indikatoren der intrinsischen Lernmotivation liegen im Bereich $\lambda_{t1} / \lambda_{t2} = .74 / .77$ und $\lambda_{t1} / \lambda_{t2} = .96 / .96$ ($p < .001$) (vgl. Abbildung 17). Die geschätzten R^2 -Werte für die manifesten Indikatoren (vgl. Tabelle 31) sind als überwiegend hoch zu bewerten ($t_1: .54 < R^2 < .91$; $t_2: .60 < R^2 < .92$, $p < .001$).

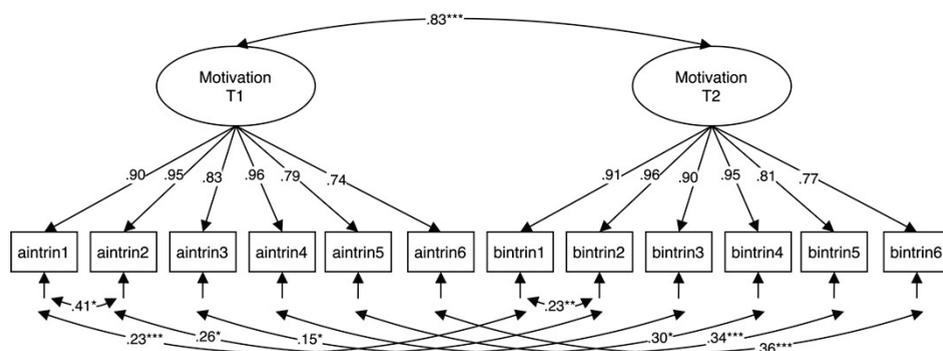


Abbildung 17: Eindimensionales Messmodell der intrinsischen Lernmotivation im Längsschnitt

Anmerkungen. Dargestellt sind standardisierte Parameterschätzungen; die Faktorladungen sind auf einem Niveau von $p < .001$ signifikant. * $p < .05$; ** $p < .01$; *** $p < .001$.

Tabelle 31: Reliabilitäts-Schätzungen (R^2) der Indikatoren im längsschnittlichen Messmodell der intrinsischen Lernmotivation

Item	MZP 1	MZP 2
intrin1	.81	.83
intrin2	.90	.92
intrin3	.69	.81
intrin4	.91	.90
intrin5	.63	.69
intrin6	.54	.60

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

Längsschnittliches Parcel-Messmodell

Mit Blick auf die anstehenden statistischen Analysen, in denen in die Strukturgleichungsmodelle mehrere latente Konstrukte simultan eingehen, wurde zur Reduktion der Modellkomplexität das *Item-Parceling* angewandt (vgl. Abschnitt 6.6.3). Um die *Item-Parcels* zu ermitteln, wurde der *balancing approach* gewählt. Im Rahmen der konfirmatorischen Faktorenanalysen zur Modellierung der intrinsischen Lernmotivation wurde eine Residualkorrelation spezifiziert, welche beim *Parceling* berücksichtigt wird, damit der bestehende Varianzanteil bestehen bleibt (Little et al., 2013). Die beiden Items *intrin1* und *intrin2* wurden daher vor der Anwendung des *balancing approach* einem *Parcel* zugeordnet. Das zweite *Parcel* besteht im Sinne des *balancing approach* aus dem Item mit der höchsten Faktorladung und dem Item mit der niedrigsten Faktorladung. Das Item mit der zweithöchsten Faktorladung und das Item mit der zweitniedrigsten Ladung bilden das dritte *Parcel*. Eine Übersicht über die gebildeten *Item-Parcels* ist in Tabelle 32 dargestellt.

Tabelle 32: Übersicht der Item-Parcels mit den jeweiligen Items zu Beginn (t_1) und zur Mitte (t_2) des Schuljahres

Skala/Parcelnr.	Parcel	Itemkürzel
<i>Intrinsische Lernmotivation zu Beginn des Schuljahres (t_1)</i>		
1	pm11	aintrin1 + aintrin2
2	pm21	aintrin4 + aintrin6
3	pm31	aintrin3 + aintrin5
<i>Intrinsische Lernmotivation zur Mitte des Schuljahres (t_2)</i>		
1	pm12	bintrin1 + bintrin2
2	pm22	bintrin4 + bintrin6
3	pm32	bintrin3 + bintrin5

Das Modell mit Item-Parcels weist eine sehr gute Modellpassung auf ($\chi^2 = 12.20$, $df = 5$, $p(\chi^2) = .03$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .04 [.01, .07], CFit = 0.62, SRMR = .01). Die Faktorladungen in diesem Modell sind alle signifikant ($p < .001$) und liegen über .89 (vgl. Abbildung 18).

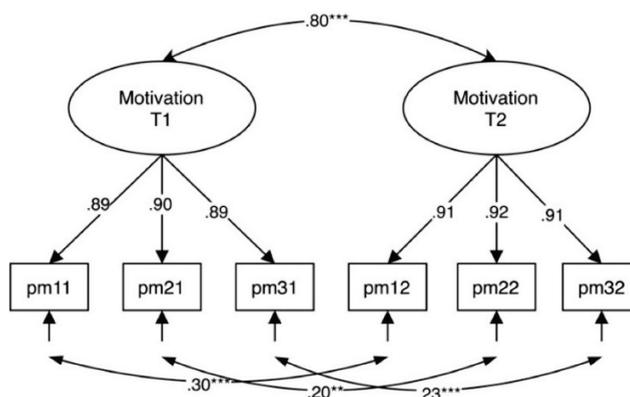


Abbildung 18: Längsschnittliches Messmodell der intrinsischen Lernmotivation mit Item-Parcels

Anmerkungen. Dargestellt sind standardisierte Parameterschätzungen; die Faktorladungen sind auf einem Niveau von $p < .001$ signifikant. * $p < .05$; ** $p < .01$; *** $p < .001$.

Die Inter-Faktorkorrelation liegt bei .80, was einer hohen zeitlichen Stabilität inter-individueller Unterschiede in der intrinsischen Lernmotivation entspricht. Die Reliabilitäten der Indikatoren liegen über .78 und sind somit als gut zu bewerten (vgl. Tabelle 33).

Tabelle 33: Reliabilitäts-Schätzungen (R^2) der Indikatoren der intrinsischen Lernmotivation im längsschnittlichen Messmodell mit Item-Parcels

Skala/Parcel	R^2
<i>Intrinsische Lernmotivation t_1</i>	
pm11	.79
pm21	.81
pm31	.78
<i>Intrinsische Lernmotivation t_2</i>	
pm12	.82
pm22	.85
pm32	.83

Anmerkungen. Die Angaben sind auf dem Niveau von $p < .001$ signifikant.

7.4.3 Messinvarianz der Lernmotivation über die Zeit

Um zu überprüfen, ob die Skala zur Messung der intrinsischen Lernmotivation über die beiden Messzeitpunkte hinweg hinreichend vergleichbar ist, wurden Messinvarianzanalysen durchgeführt. Die Ergebnisse der Invarianzprüfungen sind Tabelle 34 (Modell-Fit-Indizes) und Tabelle 35 (Modellvergleiche) zu entnehmen.

Tabelle 34: Modell-Fit-Indizes des eindimensionalen Parcel-Modells der intrinsischen Lernmotivation mit unterschiedlichem Grad der Messinvarianz über die Zeit

Modell	χ^2	df	$p(\chi^2)$	RMSEA [90% CI]	C_{Fit}	SRMR	CFI	TLI
I	12.20	5	.03	.04 [.01, .07]	.62	.01	1.00	1.00
II	14.36	7	.05	.04 [.01, .06]	.78	.01	1.00	1.00
III	28.95	9	< .001	.05 [.03, .07]	.39	.02	1.00	1.00

Anmerkungen. I = Basismodell konfiguraler Messinvarianz, II = Modell schwacher faktorieller Messinvarianz, III = Modell starker faktorieller Messinvarianz.

Tabelle 35: Ergebnisse der Modellvergleiche

Modellvergleich	$\Delta\chi^2_{\text{korr}}$	Δdf	p	ΔCFI
II vs. I	1.64	2	.44	.000
III vs. I	17.80	4	.001	-.002
III vs. II	16.21	2	< .001	.00

Anmerkungen. I = Basismodell konfiguraler Messinvarianz, II = Modell schwacher faktorieller Messinvarianz, III = Modell starker faktorieller Messinvarianz.

Das Basismodell (Modell I), das die Annahme konfiguraler Messinvarianz prüft, wurde in Abschnitt 7.4.2 beschrieben. Für das Modell liegen gute Fit-Werte vor ($\chi^2 = 12.20$, $df = 5$, $p(\chi^2) = .03$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .04 [.01, .07], $C_{Fit} = .62$, SRMR = .01), sodass angenommen werden kann, dass die Faktorstruktur über die beiden Messzeitpunkten identisch ist. Auch für das Modell schwacher faktorieller Invarianz, in dem die Faktorladungen über die Zeit gleichgesetzt wurden, liegen gute Fit-Werte vor ($\chi^2 = 14.36$, $df = 7$; $p(\chi^2) = .05$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .04 [.01, .06], $C_{Fit} = .78$, SRMR = .01). Die Modellrestriktionen führen nicht zu einer bedeutsamen Modellverschlechterung gegenüber dem Modell I, da der χ^2 -Differenztests nicht signifikant ausfällt ($\Delta\chi^2_{\text{korr}} = 1.64$, $\Delta df = 2$, $p = .439$) und der CFI-Differenzwert unter der Grenze von .01 liegt ($\Delta CFI = .000$). Die Befunde bestätigen somit die Stabilität der Faktorladungen über die

Zeit. Im Modell III wurden zusätzlich zur Faktorstruktur und zu den Faktorladungen auch die Intercepts über die Messzeitpunkte gleichgesetzt (starke faktorielle Messinvarianz). Die Modell-Fit-Statistiken weisen auf eine leichte Verschlechterung der Anpassungsgüte hin ($\chi^2 = 28.95$, $df = 9$, $p(\chi^2) < .001$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .05 [.03, .07], CFI_{fit} = .39, SRMR = .02), gleichwohl die Werte von RMSEA, SRMR, CFI und TLI immer noch als gut zu bewerten sind. Der Vergleich des restringierten Modells mit dem Basismodell und dem Modell schwacher faktorieller Invarianz mittels des korrigierten χ^2 -Differenztests legt nahe, das Modell zu Gunsten des Basismodells ($\Delta\chi^2_{\text{korrr}} = 17.80$, $\Delta df = 4$, $p = .001$) bzw. des Modells schwacher faktorieller Invarianz ($\Delta\chi^2_{\text{korrr}} = 16.21$, $\Delta df = 2$, $p < .001$) zu verwerfen. Aufgrund der Stichprobensensitivität des χ^2 -Differenztests wird zusätzlich der CFI-Differenztest herangezogen. Nach den Kriterien von Cheung und Rensvold (2002) fällt die CFI-Differenz von -.002 nicht bedeutsam aus, sodass die Annahme starker faktorieller Messinvarianz nicht abgelehnt werden muss. Die Ergebnisse bestätigen somit auch die Stabilität der Intercepts über die Zeit, sodass Mittelwertvergleiche über die Zeit zulässig sind.

7.4.4 Messinvarianz der Lernmotivation über die Geschlechtergruppen

Nachdem gezeigt wurde, dass die intrinsische Lernmotivation zum Ende und zur Mitte des Schuljahres das gleiche Konstrukt abbildet, wird im folgenden Abschnitt geprüft, ob sich die Struktur des Konstrukts auch für beide Geschlechter gleichermaßen abbilden lässt. Nur unter dieser Bedingung ist es zulässig, Mittelwerte und Regressionskoeffizienten zwischen den Gruppen zu vergleichen. Hierfür muss starke faktorielle Messinvarianz vorliegen. Dies wird durch die nachfolgende Prüfung auf Messinvarianz anhand eines multiplen Gruppenvergleichs im Rahmen von CFAs geklärt. Die Ergebnisse der Invarianzprüfung sind in Tabelle 36 (Fit-Indizes) und Tabelle 37 (Modellvergleiche) dargestellt.

Tabelle 36: Ergebnisse der Prüfung auf Messinvarianz der intrinsischen Lernmotivation nach Geschlecht

Modell	χ^2	df	$p(\chi^2)$	RMSEA [90% CI]	C _{Fit}	SRMR	CFI	TLI
I weiblich	10.01	5	.08	.05 [.00, .09]	.46	.01	1.00	.99
I männlich	4.07	5	.54	.00 [.00, .06]	.87	.01	1.00	1.00
II	15.27	10	.12	.04 [.00, .07]	.71	.01	1.00	1.00
III	27.44	14	.02	.05 [.02, .08]	.49	.04	1.00	.99
IV	39.59	18	< .01	.05 [.03, .08]	.34	.03	.99	.99

Anmerkungen. I_{weiblich} = Basismodell weibliche Stichprobe, I_{männlich} = Basismodell männliche Stichprobe, II = Basismodell konfiguraler Messinvarianz, III = Modell schwacher faktorieller Messinvarianz, IV = Modell starker faktorieller Messinvarianz.

Tabelle 37: Ergebnisse der Modellvergleiche

Modellvergleich	$\Delta\chi^2_{\text{korr}}$	Δdf	p	ΔCFI
III vs. II	11.56	4	.021	-.002
IV vs. II	23.03	8	.003	-.005
IV vs. III	11.48	4	.022	-.003

Anmerkungen. II = Basismodell konfiguraler Messinvarianz, III = Modell schwacher faktorieller Messinvarianz, IV = Modell starker faktorieller Messinvarianz.

Im ersten Schritt wurde das Messmodell für beide Gruppen separat geprüft (Modell I). Die Fit-Werte der CFAs sind gut bis sehr gut, sodass im nächsten Schritt das Baseline-Modell (Modell II) – in dem alle Parameter frei variieren dürfen – im Rahmen eines multiplen Gruppenvergleichs für beide Gruppen simultan geschätzt werden kann. Die Fit-Werte können auch hier als gut bewertet werden und bestätigen damit die Annahme von konfiguraler Messinvarianz, d.h. der Gleichheit der Faktorstruktur. Anschließend wurden zur Überprüfung eines Modells schwacher faktorieller Invarianz die Faktorladungen in den beiden Gruppen gleichgesetzt (Modell III). Der Modell-Fit dieses restringierten Modells ist gut bis sehr gut. Der Vergleich mit dem Baseline-Modell mittels des korrigierten χ^2 -Differenztests zeigt eine signifikante Modellverschlechterung ($\Delta\chi^2_{\text{korr}} = 11.56$, $\Delta df = 4$, $p = .021$). Die CFI-Differenz ($\Delta CFI = -.002$) liegt allerdings unter dem Cutoff-Wert von .01, sodass die Hypothese über die schwache faktorielle Invarianz nicht verworfen werden muss.

Um die Mittelwerte zwischen den Gruppen vergleichen zu können, muss starke faktorielle Messinvarianz vorliegen. Hierzu wurde das Modell IV spezifiziert, in dem zusätzlich zur Faktorstruktur und den Faktorladungen auch die Intercepts der manifesten Indikatoren in den beiden Gruppen gleich sind. Der Modell-Fit ist sehr gut, allerdings fällt der χ^2 -Differenztests signifikant aus ($\Delta\chi^2_{\text{korr}} = 23.03$, $\Delta df = 4$, $p = .003$). Daher wird zusätzlich der CFI-Differenztest zum Modellvergleich herangezogen. Der CFI-Differenzwert liegt mit $\Delta CFI = -.005$ unter der vorgegebenen Cutoff-Grenze, sodass starke faktorielle Messinvarianz angenommen werden kann.

Unterschiede zwischen Schülerinnen und Schülern in der intrinsischen Lernmotivation

Nachdem Messäquivalenz in der intrinsischen Lernmotivation zwischen den Gruppen nachgewiesen werden konnte, sind abschließend die Ergebnisse des Mittelwertvergleichs zu Beginn und zur Mitte des Schuljahres in Tabelle 38 dokumentiert. Die Referenzgruppe ist dabei die Gruppe der Schüler. Unterschiede werden mittels latent geschätzter, standardisierter Mittelwertdifferenzen in Skalenpunkten angeben.

Tabelle 38: Geschlechtsspezifische latente Mittelwertdifferenzen in der intrinsischen Lernmotivation zu Beginn (t_1) und zur Mitte des Schuljahres (t_2)

Skala	ΔMW	SE	p
Intrinsische Lernmotivation t_1	0.66	0.11	< .001
Intrinsische Lernmotivation t_2	0.67	0.01	< .001

Anmerkungen. Referenzgruppe: männlich; t_1 : n (Schülerinnen) = 427, n (Schüler) = 376; t_2 : n (Schülerinnen) = 348, n (Schüler) = 278.

Die Analysen belegen signifikante geschlechtsspezifische Unterschiede in der intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres. Schülerinnen weisen zu Beginn des Schuljahres eine signifikant höhere intrinsische Lernmotivation auf als Schüler ($\Delta MW = 0.66$, $SE = 0.11$, $p < .001$). Auch zur Mitte des Schuljahres bestehen signifikante Unterschiede in der intrinsischen Lernmotivation in Abhängigkeit des Geschlechts. Schülerinnen weisen weiterhin eine höhere intrinsische Lernmotivation auf als Schüler ($\Delta MW = 0.67$, $SE = 0.10$, $p < .001$).

7.4.5 Differenzielle Veränderung der Lernmotivation

Im nachfolgenden Abschnitt wird die Veränderung der intrinsischen Lernmotivation im Verlauf eines Schulhalbjahres beschrieben. Hierzu werden zunächst die auf Basis des längsschnittlichen Messmodells geschätzten latenten Mittelwerte zu Beginn und zur Mitte des Schuljahres und anschließend die latente Veränderung im Latent-True-Change-Modell betrachtet. Abschließend wird der Einfluss des Geschlechts und der Deutschnote auf interindividuelle Unterschiede im Ausgangsniveau und der Veränderung der Lernmotivation untersucht.

Deskriptive Entwicklungen

Anhand der Werte in Tabelle 39 zeigt sich eine leichte Abnahme in der mittleren intrinsischen Lernmotivation ($MW_{t_1} = 2.35$, $SE = 0.06$, $MW_{t_2} = 2.25$, $SE = 0.07$). Das Gleichsetzen der Mittelwerte über die Zeit führt zu einer signifikanten Modellverschlechterung im Vergleich zum Modell mit frei geschätzten Mittelwerten ($\chi^2 = 41.62$, $df = 10$, $p(\chi^2) < .001$, $CFI = 1.00$, $TLI = .99$, $RMSEA$ [90% CI] = .06 [.04, .08], $C_{Fit} = 0.13$, $SRMR = .03$, $\Delta\chi^2_{kor} = 7.77$, $\Delta df = 1$, $p = .01$). Dies lässt auf eine signifikante Abnahme der intrinsischen Lernmotivation schließen. Allerdings ist der absolute Fit des restriktiveren Modells immer noch als sehr gut zu bewerten, sodass die zeitliche Veränderung als gering einzustufen ist. Die Entwicklung der intrinsischen Lernmotivation im Verlauf des Schulhalbjahres ist in Abbildung 19 zusammenfassend dargestellt.

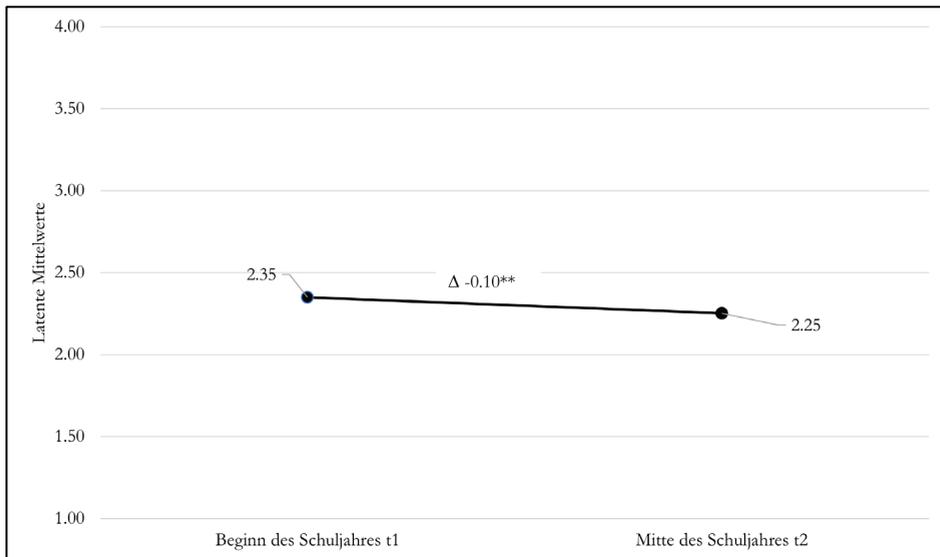


Abbildung 19: Veränderung der intrinsischen Lernmotivation

Anmerkung. ** $p < .01$

Veränderung im LTC-Modell

Zur Modellierung der Veränderung der intrinsischen Lernmotivation über die Zeit wurde – analog zur Veränderung der wahrgenommenen Feedbackqualität – ein Latent-True-Change-Modell geschätzt. Eine latente Differenzvariable bildet die latente Veränderung der intrinsischen Lernmotivation zwischen dem Beginn des Schuljahres und der Mitte des Schuljahres ab. Das LTC-Modell mit standardisierten Parameterschätzungen und der Annahme von starker faktorieller Messinvarianz ist in Abbildung 20 dargestellt. Das Modell verfügt über eine gute Anpassungsgüte an die Daten ($\chi^2 = 28.95$, $df = 9$, $p(\chi^2) < .001$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .05 [.03, .07], $C_{Fit} = 0.64$, SRMR = .02).

Auch die Ergebnisse der LTC-Analyse zeigen eine signifikante latente Mittelwertdifferenz ($\Delta MW = -0.10$, $SE = 0.03$, $p < .01$), die auf eine bedeutsame Abnahme der intrinsischen Lernmotivation hinweist (vgl. Tabelle 39). Darüber hinaus verdeutlicht die geschätzte Varianz der intrinsischen Lernmotivation zum ersten Messzeitpunkt ($Var = 0.63$, $SE = 0.03$, $p < .001$) und der Differenzvariable ($Var = 0.26$, $SE = 0.03$, $p < .001$), dass signifikante Unterschiede im Ausgangswert und im Veränderungswert vorliegen. Die Oberstufenschülerinnen und -schüler unterscheiden sich somit bedeutsam sowohl in ihrer intrinsischen Lernmotivation zu Beginn des Schuljahres als auch in ihrer Motivationsveränderung im Verlauf des Schulhalbjahres.

Tabelle 39: Latente Faktormittelwerte (MW), -varianzen (Var) der intrinsischen Lernmotivation zu beiden Messzeitpunkten (t) sowie Mittelwertdifferenzen

Latente Variable	MW (SE)	Var (SE)
Intrinsische Motivation t_1	2.35*** (0.06)	0.63*** (0.03)
Intrinsische Motivation t_2	2.25*** (0.07)	0.67*** (0.04)
Δ Intrinsische Motivation t_2-t_1	-0.10** (0.03)	0.26*** (0.03)

Anmerkung. Δ = Latente Veränderungsvariable. ** $p < .01$, *** $p < .001$.

Die latente Korrelation zwischen der Ausgangs- und der Veränderungsvariable der intrinsischen Lernmotivation im LTC-Modell (vgl. Abbildung 20) ist signifikant und beträgt $r = -.27$ ($p < .001$). Damit besteht ein mittelstarker negativer Zusammenhang, der so zu interpretieren ist, dass die Veränderung in der intrinsischen Lernmotivation umso stärker ist, je niedriger die intrinsische Lernmotivation zu Beginn des Schuljahres ausgeprägt ist. Konkret bedeutet dies, dass sich Schülerinnen und Schüler mit einer eher niedrigen intrinsischen Lernmotivation zu Beginn des Schuljahres stärker in der intrinsischen Lernmotivation im Verlauf des Schulhalbjahres verändern, während sich Schülerinnen und Schüler mit einer hohen intrinsischen Lernmotivation weniger verändern. Da der Veränderungswert im LTC-Modell negativ ist ($\Delta MW = -0.10$) heißt das, dass eine geringe Ausgangsmotivation mit einer stärkeren Abnahme in der intrinsischen Motivation einhergeht.

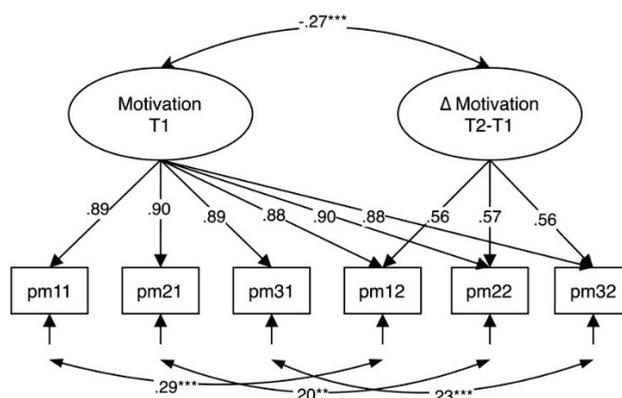


Abbildung 20: Latent-True-Change-Modell für die Veränderung der intrinsischen Lernmotivation

Anmerkungen. Dargestellt sind standardisierte Parameterschätzungen; die Faktorladungen sind auf einem Niveau von $p < .001$ signifikant. Δ = Latente Veränderungsvariable. * $p < .05$; ** $p < .01$; *** $p < .001$.

Effekte auf die Veränderung der intrinsischen Lernmotivation

Um zu untersuchen, welche Merkmale interindividuelle Unterschiede in der intra-individuellen Veränderung der intrinsischen Lernmotivation von Schülerinnen und Schülern erklären können, wurden das Geschlecht (1 = weiblich) und die Deutschnote (höhere Werte = gute Note) als Prädiktoren der Veränderung der intrinsischen Lernmotivation in das LTC-Modell aufgenommen. Die absoluten und inkrementellen Fit-Indizes verweisen im spezifizierten Modell auf eine gute Datenanpassung ($\chi^2 = 28.24$, $df = 17$, $p(\chi^2) = .04$, CFI = 1.00, TLI = .99, RMSEA [90% CI] = .04 [.01, .06], CFI_{fit} = 0.83, SRMR = .02). Obschon der χ^2 -Test signifikant ausfällt, liegen die anderen Fit-Indizes im Bereich der gängigen Cutoff-Werte. Die unstandardisierten und standardisierten Regressionskoeffizienten aus der Strukturgleichungsanalyse sind in Tabelle 40 dargestellt.

Tabelle 40: Standardisierte und unstandardisierte Regressionskoeffizienten im Latent-True-Change-Modell mit Geschlecht und Deutschnote als Prädiktoren der Motivationsveränderung

	Unstandardisierte Effekte		Standardisierte Effekte	
	Level	Change	Level	Change
Prädiktoren	B (SE)	B (SE)	β (SE)	β (SE)
Geschlecht	.41*** (0.09)	.07 (0.07)	.26*** (0.05)	.07 (0.06)
Note	.28*** (0.04)	-.07* (0.03)	.34*** (0.04)	-.12* (0.05)
R ²			.226	.015

Anmerkungen. Das Geschlecht ist eine dichotome Variable (1 = weiblich); die Deutschnote wurde umgepolt (hohe Werte entsprechen einer guten Note). * $p < .05$; ** $p < .01$; *** $p < .001$.

Zunächst können signifikante Effekte der Deutschleistung und des Geschlechts auf den Ausgangswert der intrinsischen Lernmotivation nachgewiesen werden: Eine höhere Deutschleistung der Schülerinnen und Schüler geht mit einer höheren intrinsischen Lernmotivation zu Beginn des Schuljahres einher ($\beta = .34$, $SE = 0.04$, $p < .001$). Auch unter Kontrolle der Deutschleistung zeigen Schülerinnen zu Beginn des Schuljahres eine höhere intrinsische Lernmotivation als Schüler ($\beta = .26$, $SE = 0.05$, $p < .001$). Das Modell erklärt einen statistisch bedeutsamen Anteil der Varianz in der intrinsischen Lernmotivation zu Beginn des Schuljahres ($R^2 = .23$, $p < .001$).

Weiterhin zeigen sich keine signifikanten Effekte des Geschlechts ($\beta = .07$, $SE = 0.06$, $p = .29$) auf die Veränderung der intrinsischen Lernmotivation. Demzufolge verändern sich Schülerinnen und Schüler – unabhängig von ihrem Geschlecht – in ähnlicher Weise in der intrinsischen Lernmotivation. Währenddessen können

signifikante negative Effekte der Deutschnote auf die Veränderung der intrinsischen Lernmotivation nachgewiesen werden ($\beta = -.12$, $SE = 0.05$, $p = .02$). Die intrinsische Motivation nimmt im Stichprobendurchschnitt über die Zeit ab; demnach ist eine höhere Leistung mit einer stärkeren Abnahme der Motivation assoziiert. Allerdings ist die Varianzaufklärung hinsichtlich der Veränderung der intrinsischen Lernmotivation nur sehr gering ($R^2 = .02$, $p = .25$).

7.4.6 Bivariate Latent-True-Change-Modelle

Zur Analyse der Zusammenhänge zwischen den Dimensionen der Feedbackqualität und der intrinsischen Lernmotivation über die Zeit (*Teilfragestellung F2a und Teilfragestellung F2b*) wurden sukzessive vier bivariate LTC-Modelle für jede der vier Feedbackdimensionen spezifiziert⁶. Die auf Grundlage der einzelnen Messmodelle ermittelten latenten Interkorrelationen der Variablen können den Tabelle 41 bis Tabelle 44 entnommen werden.

Tabelle 41: Latente Interkorrelationen der ergebnisorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell

Latente Variable	(1)	(2)	(3)	(4)
(1) EO t_1	1.00	-.29***	.30***	-.05
(2) Δ EO t_2-t_1		1.00	-.09	.16**
(3) IM t_1			1.00	-.27***
(4) Δ IM t_2-t_1				1.00

Anmerkungen. EO = Ergebnisorientierte Dimension; IM = Intrinsische Lernmotivation; Δ = Latente Veränderungsvariable; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

Tabelle 42: Latente Interkorrelationen der prozessorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell

Latente Variable	(1)	(2)	(3)	(4)
(1) PO t_1	1.00	-.08***	.13***	-.02
(2) Δ PO t_2-t_1		1.00	-.04*	.06***
(3) IM t_1			1.00	-.11***
(4) Δ IM t_2-t_1				1.00

Anmerkungen. PO = Prozessorientierte Dimension; IM = Intrinsische Lernmotivation; Δ = Latente Veränderungsvariable; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

⁶ Ein Modell, in dem alle vier Feedbackdimensionen simultan in die Analyse einbezogen wurden, konnte aufgrund des Vorliegens von Multikollinearität zwischen den Dimensionen von Feedback nicht geschätzt werden.

Tabelle 43: Latente Interkorrelationen der selbstregulationsorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell

Latente Variable	(1)	(2)	(3)	(4)
(1) SO t1	1.00	-.30***	.29***	-.12*
(2) ΔSO t2-t1		1.00	-.06	.32***
(3) IM t1			1.00	-.27***
(4) ΔIM t2-t1				1.00

Anmerkungen. SO = Selbstregulationsorientierte Dimension; IM = Intrinsische Lernmotivation; Δ = Latente Veränderungsvariable; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

Tabelle 44: Latente Interkorrelationen der dialogorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell

Latente Variable	(1)	(2)	(3)	(4)
(1) DO t1	1.00	-.38***	.25***	-.15*
(2) ΔDO t2-t1		1.00	-.04	.26***
(3) IM t1			1.00	-.27***
(4) ΔIM t2-t1				1.00

Anmerkungen. DO = Dialogorientierte Dimension; IM = Intrinsische Lernmotivation; Δ = Latente Veränderungsvariable; alle Korrelationen sind auf einem Niveau von $p < .001$ signifikant.

Abbildung 21 zeigt die vier bivariaten LTC-Modelle mit standardisierten Parameterschätzungen und der Annahme starker faktorieller Messinvarianz der Konstrukte. Die Abbildung liefert auch Informationen über die Anpassungsgüte der Modelle. In jedem bivariaten LTC-Modell wurden insgesamt zwei (Residual-)Korrelationen zugelassen: zwischen den Ausgangsvariablen der jeweiligen Feedbackdimension und der intrinsischen Lernmotivation zu Beginn des Schuljahres (t_1) sowie zwischen den Veränderungsvariablen der beiden Konstrukte im Verlauf des Schulhalbjahres (t_2-t_1). Darüber hinaus wurden zwei kreuzverzögerte Pfade miteinbezogen: der Effekt vom Ausgangswert der Feedbackdimension auf die Veränderung der intrinsischen Lernmotivation sowie der Effekt vom Ausgangswert der intrinsischen Lernmotivation auf die Veränderung der Feedbackwahrnehmung.

Die Analyse ergibt, dass die Modelle sehr gut auf die Daten passen (vgl. Abbildung 21). Der Quotient aus Chi-Quadrat und Freiheitsgraden sowie RMSEA und SRMR zeigen geringe Werte, die unter den Cutoff-Werten liegen und damit auf einen sehr guten Modell-Fit hinweisen. Auch die CFI- und TLI-Werte liegen innerhalb der üblichen Grenzen und weisen auf eine sehr gute Modellpassung hin.

Hinsichtlich der Zusammenhänge zwischen den Konstrukten (jeweilige Feedbackdimension und intrinsische Motivation) und der zeitlich parallelen Veränderungen im Schulhalbjahr (*Hypothese 2a-1* und *Hypothese 2a-2*) konnten in allen vier Modellen erwartungsgemäß statistisch signifikante Korrelationen zwischen den Ausgangswerten zu Beginn des Schuljahres ($.25 \leq r \leq .33$; $p < .001$) nachgewiesen werden. Das heißt, je positiver die Schülerinnen und Schüler das Feedback der Lehrkraft zu Beginn des Schuljahres wahrnehmen, desto höher ist auch ihre intrinsische Lernmotivation und umgekehrt. Auch die latenten Korrelationen der Veränderungswerte liegen im moderaten, positiven Bereich zwischen $.12 \leq r \leq .28$ ($p < .05$), wobei der höchste Zusammenhang zwischen der Veränderung des selbstregulationsorientierten Feedbacks und der Veränderung der intrinsischen Lernmotivation besteht. Demnach gehen Veränderungen in der Wahrnehmung der Feedbackdimensionen mit einer Veränderung der intrinsischen Lernmotivation im Verlauf des Schulhalbjahres einher.

Nun werden die kreuzverzögerten Pfadkoeffizienten in den einzelnen Modellen betrachtet (*Hypothese 2b-1* und *Hypothese 2b-2*). Im Einzelnen werden interindividuelle Unterschiede in der intraindividuellen Veränderung der intrinsischen Lernmotivation signifikant durch die Ausprägung des wahrgenommenen selbstregulationsorientierten Feedbacks ($\beta = -.12$, $SE = 0.07$, $p = .049$) und des dialogorientierten Feedbacks ($\beta = -.15$, $SE = 0.06$, $p = .021$) erklärt. Die Effekte sind negativ. Da die intrinsische Lernmotivation im Mittel über die Zeit abnimmt, sind die Befunde so zu interpretieren, dass eine höher ausgeprägte Wahrnehmung des selbstregulations- und dialogorientierten Feedbacks zu Beginn des Schuljahres die Abnahme der intrinsischen Motivation im Schulhalbjahr verstärkt. Allerdings fallen die Pfadkoeffizienten und auch die R^2 -Werte klein aus, was darauf hinweist, dass die kreuzverzögerten Effekte wenig zusätzliche Varianz erklären können. Entgegen der aufgestellten Hypothese zeigen sich keine signifikanten Effekte vom wahrgenommenen Ergebnisfeedback ($\beta = -.05$, $SE = 0.06$, $p = .42$) und Prozessfeedback ($\beta = -.07$, $SE = 0.06$, $p = .24$) zu Beginn des Schuljahres auf die Veränderung der intrinsischen Lernmotivation im Halbjahr.

Unterschiede in der Veränderung der Feedbackwahrnehmung durch die intrinsische Lernmotivation zu Beginn des Schuljahres können nur in Bezug auf die Prozessdimension von Feedback erklärt werden. Der kreuzverzögerte Effekt beträgt $\beta = -.12$ ($SE = 0.06$, $p = .036$), ist allerdings als sehr klein einzustufen und besitzt damit eine geringe Varianzaufklärung ($R^2 = 1.3\%$). Der negative Pfadkoeffizient ist so zu interpretieren, dass eine höhere intrinsische Lernmotivation zu Beginn des Schuljahres zu einer Abnahme der Prozessorientierung führt. Die kreuzverzögerten Effekte von der intrinsischen Lernmotivation auf die Veränderung der Wahrnehmung der Ergebnisdimension ($\beta = -.09$, $SE = 0.05$, $p = .08$), der Selbstregulationsdimension ($\beta = -.06$, $SE = 0.05$, $p = .20$) und der Dialogdimension ($\beta = -.04$, $SE = 0.07$, $p = .60$) fallen nicht statistisch signifikant aus, sodass die Annahmen in Bezug auf diese Zusammenhänge zu verwerfen sind.

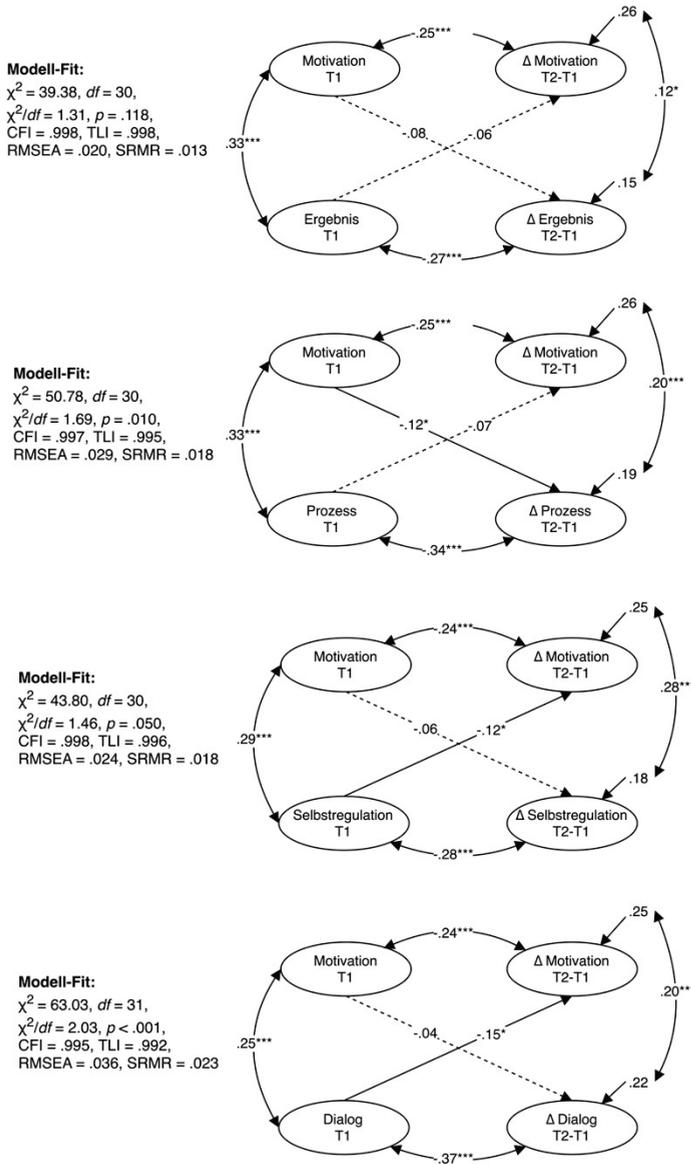


Abbildung 21: Bivariate Latent-True-Change-Modelle für den Zusammenhang zwischen Feedbackdimensionen und der intrinsischen Lernmotivation

Anmerkungen. Vereinfachte Darstellung ohne Messmodelle; Angabe der standardisierten Parameterschätzungen; nicht-signifikante Pfade sind gestrichelt dargestellt; Δ = Latente Veränderungsvariable. * $p < .05$, ** $p < .01$, *** $p < .001$.

7.5 Indirekte Wirkungen von Feedback auf die Lernmotivation

Im folgenden Abschnitt wird untersucht, ob die Dimensionen von wahrgenommenem Feedback einen indirekten Effekt auf die Veränderung der intrinsischen Lernmotivation über die Zeit ausüben und inwiefern eine vermittelnde Wirkung der wahrgenommenen Nützlichkeit sowie des wahrgenommenen Kompetenzerlebens vorliegt (*Hauptforschungsfrage III*). Dafür wird zunächst die faktorielle Struktur der Mediatorvariablen geprüft und im Anschluss werden die deskriptiven Kennwerte berichtet.

7.5.1 Faktorenstruktur und deskriptive Statistik der Skala Nützlichkeit

Im Rahmen einer konfirmatorischen Faktorenanalyse wurde die Eindimensionalität der Skala zur Erfassung der wahrgenommenen Nützlichkeit von Feedback geprüft. In Abbildung 22 ist das postulierte Ein-Faktormodell dargestellt.

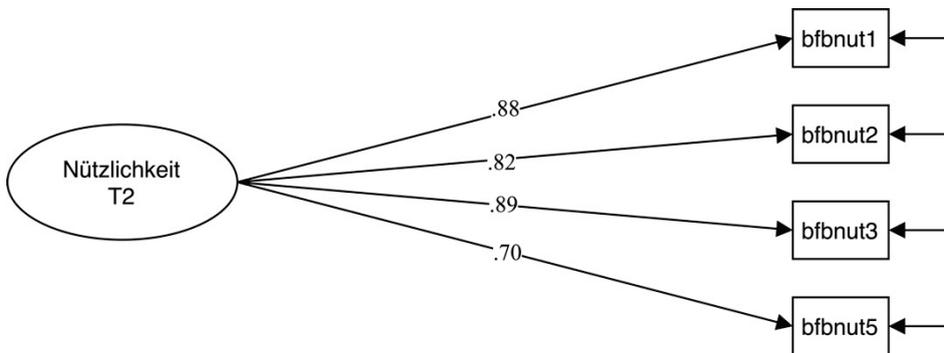


Abbildung 22: Ein-Faktormodell der Skala Wahrgenommene Nützlichkeit von Feedback

Für das Gesamtmodell fällt der χ^2 -Test signifikant aus, die übrigen Fit-Indizes weisen auf eine gute bis akzeptable Passung des Modells ($\chi^2 = 7.29$, $df = 2$, $p(\chi^2) = .03$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .07 [.02, .12], CFit = 0.24, SRMR = .01). Die Ergebnisse des Messmodells zeigen hohe und signifikante Faktorladungen ($\lambda \geq .70$). Die R^2 -Werte können der Tabelle 45 entnommen werden. Die Reliabilitätsanalysen ergeben gute Werte von $\alpha = .85/\omega = .85$. Die deskriptive Auswertung zeigt, dass die wahrgenommene Nützlichkeit von Feedback mit einem Mittelwert von $MW = 2.93$ ($SD = 0.54$) über dem theoretischen Mittelwert der Skala von 2.5 liegt. Die Ausprägung für die ICC zeigt, dass rund 20% der Varianz in den Schülerwahrnehmungen durch Unterschiede zwischen den Deutschkursen erklärt werden kann.

Tabelle 45: Reliabilitäts-Schätzungen (R^2) der Indikatoren im eindimensionalen Modell der wahrgenommenen Nützlichkeit zur Mitte des Schuljahres (t_2)

Item	R^2
bfbnut1	.78
bfbnut2	.68
bfbnut3	.79
bfbnut5	.48

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

7.5.2 Faktorenstruktur und deskriptive Statistik der Skala Kompetenzerleben

Die Prüfung der faktoriellen Validität der Skala Kompetenzerleben erfolgte auch hier im Rahmen einer konfirmatorischen Faktorenanalyse. Das spezifizierte Messmodell mit sechs manifesten Indikatoren führt zu einem unakzeptablen Modell-Fit ($\chi^2 = 260.33$, $df = 9$, $p(\chi^2) < .001$, CFI = .95, TLI = .91, RMSEA [90% CI] = .21 [.19, .23], CFI = 0.00, SRMR = .06). Insbesondere die Werte für den χ^2 -Test, den Quotienten aus χ^2/df und den RMSEA liegen außerhalb der geforderten Grenzbereiche. Daher werden sukzessive zwei Items (*bkomp4* „Im Deutschunterricht finden meine Leistungen Anerkennung bei meinen Mitschüler/innen“ und *bkomp6* „Im Deutschunterricht vermittelt meine Lehrkraft mir den Eindruck, dass sie mit meinen Leistungen zufrieden ist“) mit vergleichsweise geringeren Faktorladungen (*bkomp4*: $\lambda = .58$; *bkomp6*: $\lambda = .72$) und geringeren Indikatorreliabilitäten (*bkomp4*: $R^2 = .34$; *bkomp6*: $R^2 = .52$) ausgeschlossen. Der Ausschluss führt zu einer Verbesserung des Modell-Fits ($\chi^2 = 53.32$, $df = 2$, $p(\chi^2) < .001$, CFI = .99, TLI = .96, RMSEA [90% CI] = .20 [.16, .25], CFI = 0.00, SRMR = .03). Allerdings liegen die Werte für χ^2/df und RMSEA immer noch in einem nicht zufriedenstellenden Bereich. Die *Modification Indices* im Mplus Output zeigen an, dass das Zulassen der Residualkorrelation (Kline, 2016) zwischen den Indikatoren *bkomp2* und *bkomp3* (aufgrund der benachbarten Platzierung im Fragebogen) den χ^2 -Wert verringern würde. Die Modellierung dieser Residualkorrelation führt zu einer substanziellen Verbesserung des Modell-Fits ($\chi^2 = 2.50$, $df = 1$, $p(\chi^2) = .11$, CFI = 1.00, TLI = 1.00, RMSEA [90% CI] = .05 [.00, .13], CFI = 0.37, SRMR = .004). In Abbildung 23 ist das finale Ein-Faktormodell mit vier manifesten Indikatoren und spezifizierter Residualkorrelation dargestellt.

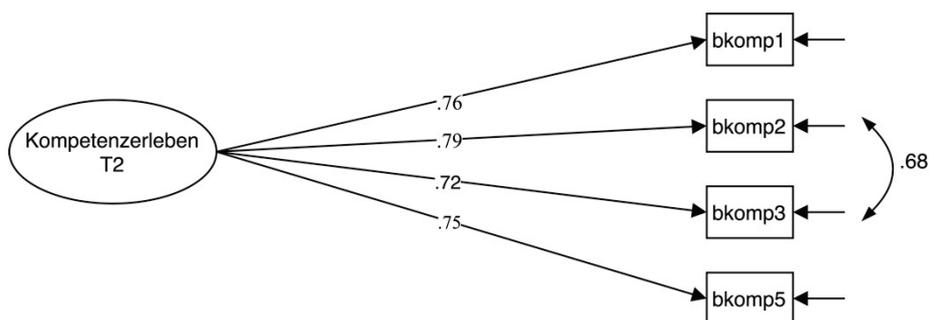


Abbildung 23: Ein-Faktormodell der Skala Wahrgenommenes Kompetenzerleben

Die Ergebnisse des Messmodells zeigen hohe und signifikante Faktorladungen ($\lambda \geq .72$). Die R^2 -Werte können der Tabelle 46 entnommen werden. Die Reliabilitätsanalysen ergeben gute Werte von $\alpha = .82/\omega = .80$. Hinsichtlich der deskriptiven Statistiken zeigt sich für die Variable ein Mittelwert von $MW = 2.79$ ($SD = 0.65$), der über dem theoretischen Mittelwert der Skala von 2.5 liegt. Der ICC-Koeffizient zeigt, dass ein Varianzanteil von 7% auf die Zugehörigkeit zu einem Deutschkurs zurückzuführen ist.

Tabelle 46: Reliabilitäts-Schätzungen (R^2) der Indikatoren im eindimensionalen Modell des wahrgenommenen Kompetenzerlebens zur Mitte des Schuljahres (t_2)

Item	R^2
bkomp1	.58
bkomp2	.63
bkomp3	.52
bkomp5	.56

Anmerkungen. Alle Angaben sind auf einem Niveau von $p < .001$ signifikant.

7.5.3 Mediationsanalysen

Die Prüfung der Hypothesen, in denen von der mediiierenden Bedeutung der wahrgenommenen Nützlichkeit von Feedback und des wahrgenommenen Kompetenzerlebens bei der Vorhersage der Veränderung der intrinsischen Lernmotivation ausgegangen wird, erfolgte im Rahmen von mehreren separaten Strukturgleichungsmodellen, in die die unterschiedlichen Ausgangswerte der Feedbackdimensionen (t_1) als unabhängige Variablen einzeln in das Mediationsmodell aufgenommen wurden. Das Geschlecht und die Deutschnote wurden als Kontrollvariablen in den Analysen mitberücksichtigt.

Zunächst wurde ein Gesamtmodell geschätzt, in das alle Variablen ohne Spezifizierung von direkten und indirekten Effekten aufgenommen wurden. Das Gesamtmodell weist eine gute Modellpassung auf ($\chi^2 = 689.53$, $df = 422$, $p(\chi^2) < .001$, $CFI = .96$, $TLI = .95$, $RMSEA [90\% CI] = .03 [.02, .03]$, $CFit = 1.00$, $SRMR = .04$). In Tabelle 47 sind die latenten Interkorrelationen zwischen den Variablen dargestellt. Es liegen signifikante positive Zusammenhänge zwischen den einzelnen Feedbackdimensionen und der wahrgenommenen Nützlichkeit ($.32 \leq r \leq .62$, $p < .001$) sowie dem wahrgenommenen Kompetenzerleben ($.27 \leq r \leq .34$, $p < .001$) vor. Die wahrgenommene Nützlichkeit korreliert signifikant positiv mit dem wahrgenommenen Kompetenzerleben ($r = .49$, $p < .001$). Die einzelnen Feedbackdimensionen korrelieren signifikant positiv mit der intrinsischen Lernmotivation zu Beginn des Schuljahres ($.24 \leq r \leq .33$, $p < .001$), nicht aber mit der Veränderung der Motivation ($r \leq -.07$, n. s.). Die wahrgenommene Nützlichkeit von Feedback und das Kompetenzerleben hängen sowohl mit der intrinsischen Lernmotivation zu Beginn des Schuljahres (Nützlichkeit: $r = .39$, Kompetenzerleben: $r = .49$, $p < .001$) als auch mit der Veränderung im Laufe des Schuljahres (Nützlichkeit: $r = .20$, Kompetenzerleben: $r = .20$, $p < .001$) signifikant positiv zusammen.

Tabelle 47: Interkorrelationen der untersuchten Variablen

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) EO t ₁	1.00	.82***	.69***	.59***	.62***	.33***	.31***	-.07	.04	.14***
(2) PO t ₁		1.00	.82***	.63***	.57***	.34***	.33***	-.05	.03	.14**
(3) SO t ₁			1.00	.64***	.49***	.32***	.31***	-.07	-.06	.14*
(4) DO t ₁				1.00	.32***	.27***	.24***	-.07	-.05	.13*
(5) NU t ₂					1.00	.49***	.39***	.20***	.08	.15*
(6) KO t ₂						1.00	.49***	.20**	.19***	.59***
(7) IM t ₁							1.00	-.28***	.32***	.40***
(8) ΔIM t ₂ -t ₁								1.00	.03	-.12
(9) Geschlecht									1.00	.27***
(10) Deutschnote										1.00

Anmerkungen. Dargestellt sind die standardisierten Korrelationskoeffizienten; EO = Ergebnisorientierte Dimension; PO = Prozessorientierte Dimension; SO = Selbstregulationsorientierte Dimension; DO = dialogorientierte Dimension; NU = Nützlichkeit; KO = Kompetenzerleben; IM = Intrinsische Lernmotivation; Δ = Latente Veränderungsvariable.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Das Mediationsmodell mit allen Feedbackdimensionen ist in Abbildung 24 mit standardisierten Parameterschätzungen dargestellt. Das Modell verfügt über eine gute Anpassungsgüte an die Daten (vgl. Tabelle 48). Da die einzelnen Feedbackdimensionen nicht gleichzeitig, sondern separat in das Mediationsmodell eingeschlossen wurden, werden in der Tabelle vier Modell-Fit-Statistiken berichtet. Zwar fallen die χ^2 -Werte signifikant aus und ein CFI-Wert liegt mit $CFI = .94$ knapp unter der geforderten Grenze von $.95$, allerdings liegen die anderen Fit-Indizes im Bereich der angestrebten Cutoff-Werte ($2.33 \leq \chi^2/df \leq 2.55$, $.04 \leq RMSEA \leq .04$, $.04 \leq SRMR \leq .04$, $.94 \leq CFI \leq .95$).

Tabelle 48: Fit-Indizes des Strukturgleichungsmodells für jede Dimension der Feedbackqualität

Modell	χ^2	<i>df</i>	$p(\chi^2)$	RMSEA [90% CI]	C_{Fit}	SRMR	CFI	TLI
Ergebnisorientiert	378.61	151	<.001	.04 [.04, .05]	.98	.04	.95	.93
Prozess	353.37	151	<.001	.04 [.04, .05]	1.00	.04	.95	.94
Selbstregulation	377.37	151	<.001	.04 [.04, .05]	.98	.04	.94	.93
Dialog	384.84	151	<.001	.04 [.04, .05]	.97	.05	.95	.93

Anmerkung. CI = *Confidence Interval*.

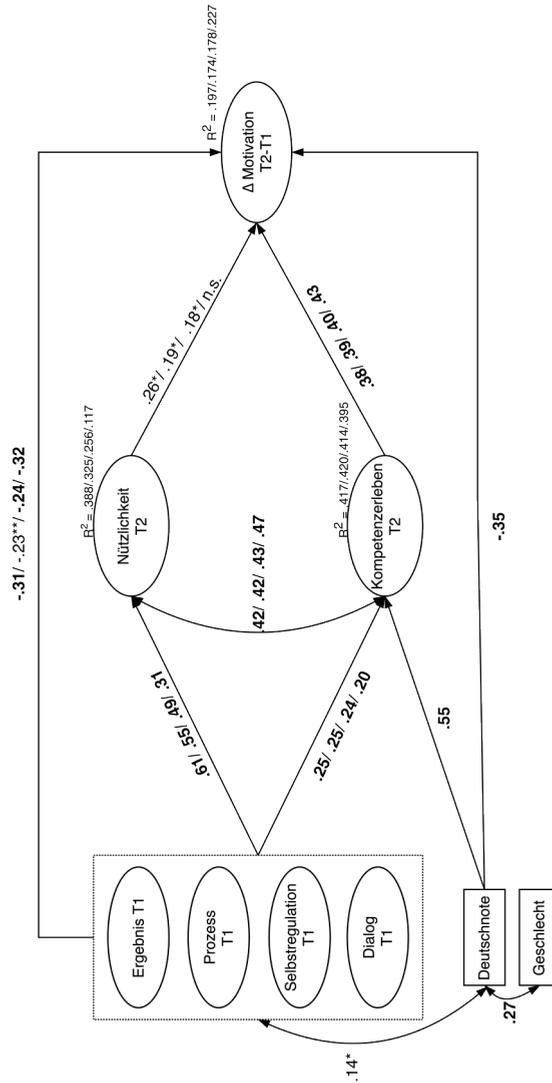


Abbildung 24: Strukturgleichungsmodell zur Vorhersage der Veränderung von intrinsischer Lernmotivation über die wahrgenommene Nützlichkeit und das Kompetenzerleben unter Kontrolle von Deutschnote und Geschlecht

Anmerkungen. Fett gedruckte Werte: $p < .001$, $**p < .01$, $*p < .05$; Geschlecht: 1 = weiblich, 0 = männlich, Notenstufen: 1 = sehr gut bis 6 = ungenügend; Angabe standardisierter Parameterschätzungen; Abgetragen sind signifikante Pfade; Δ = Latente Veränderungsvariable. Erste Angabe: Ergebnisorientierte Dimension, zweite Angabe: Prozessorientierte Dimension, dritte Angabe: selbstregulationsorientierte Dimension, vierte Angabe: dialogorientierte Dimension.

Einfluss der wahrgenommenen Feedbackqualität auf die Motivationsentwicklung

Zunächst werden die totalen Effekte betrachtet (vgl. Tabelle 49), die statistisch den direkten Effekten von den Dimensionen der wahrgenommenen Feedbackqualität auf die Motivationsentwicklung in einem Strukturgleichungsmodell ohne Berücksichtigung der wahrgenommenen Nützlichkeit und des Kompetenzerlebens als Mediatorvariablen entsprechen (vgl. Abschnitt 7.4.6).⁷ Es zeigt sich ein statistisch signifikanter totaler Effekt von der dialogorientierten Dimension auf die Veränderung der intrinsischen Motivation, allerdings fällt dieser negativ aus ($\beta = -.19$, $SE = .06$, $p = .003$, 95% CI [-.34, -.08]). Der schwach negative Effekt sagt aus, dass eine höher ausgeprägte Wahrnehmung des dialogorientierten Feedbacks zu Beginn des Schuljahres die Abnahme der intrinsischen Motivation im Schulhalbjahr verstärkt. Auch die ergebnisorientierte, prozessorientierte und selbstregulationsorientierte Dimension der Feedbackqualität üben einen schwach negativen totalen Effekt aus, der aber statistisch nicht signifikant ist (ergebnisorientiert t_1 : $\beta = -.06$, $SE = .07$, $p = .39$, 95% CI [-.16, .003]; prozessorientiert t_1 : $\beta = -.03$, $SE = .07$, $p = .71$, 95% CI [-.11, .07]; selbstregulationsorientiert t_1 : $\beta = -.05$, $SE = .06$, $p = .35$, 95% CI [-.11, .12]).

Im Einklang mit den Auswertungen im LTC-Modell (vgl. Abschnitt 7.4.6) können somit auch in den Mediationsanalysen keine signifikanten positiven Zusammenhänge zwischen den Feedbackdimensionen und der Motivationsentwicklung nachgewiesen werden. Die fehlenden totalen Effekte der Feedbackdimensionen können methodisch sog. Suppressor-Effekte darstellen (vgl. Abschnitt 6.6.4). Wie in Abbildung 24 zu erkennen, fallen die direkten Effekte von den Feedbackdimensionen auf die Motivationsveränderung signifikant negativ aus, während die indirekten Effekte positiv sind. Durch die gegenläufigen Richtungen der direkten und indirekten Effekte können sich die totalen Effekte in Summe gegenseitig aufheben. Bei alleiniger Betrachtung der bivariaten Beziehungen würde der Einfluss von Feedback auf die Motivationsentwicklung in den Mediationsmodellen somit unterschätzt werden.

Indirekte Effekte via wahrgenommene Nützlichkeit und Kompetenzerleben

Für die interessierenden Fragestellungen nach der mediiierenden Bedeutung der wahrgenommenen Nützlichkeit (*Teilfragestellung 3a*) und des wahrgenommenen Kompetenzerlebens (*Teilfragestellung 3b*) für die Motivationsentwicklung werden die indirekten Effekte im Strukturgleichungsmodell betrachtet.

Zunächst zeigt sich, dass alle vier Feedbackdimensionen zu t_1 einen statistisch signifikanten positiven Effekt auf die wahrgenommene Nützlichkeit zu t_2 haben (direkte Effekte: ergebnisorientiert t_1 : $\beta = .61$, $SE = .05$, $p < .001$; prozessorientiert t_1 : $\beta = .55$, $SE = .04$, $p < .001$; selbstregulationsorientiert t_1 : $\beta = .49$, $SE = .05$, $p < .001$; dialogorientiert t_1 : $\beta = .31$, $SE = .06$, $p < .001$), wobei der Effekt der

⁷ Die totalen Effekte sind in Abbildung 24 nicht ablesbar, da sie sich aus den jeweiligen direkten und indirekten Effekten zusammensetzen.

ergebnisorientierten Dimension auf die Nützlichkeit am stärksten ausfällt. Auch die direkten Effekte der Feedbackdimensionen auf das wahrgenommene Kompetenzerleben zu t_2 fallen signifikant positiv aus (direkte Effekte: ergebnisorientiert t_1 : $\beta = .25$, $SE = .05$, $p < .001$; prozessorientiert t_1 : $\beta = .25$, $SE = .05$, $p < .001$; selbstregulationsorientiert t_1 : $\beta = .24$, $SE = .05$, $p < .001$; dialogorientiert t_1 : $\beta = .19$, $SE = .06$, $p < .001$). Daraus folgt, dass mit einer hohen Ausprägung des wahrgenommenen Feedbacks eine höhere wahrgenommene Nützlichkeit und ein höheres wahrgenommenes Kompetenzerleben einhergehen. Die wahrgenommene Nützlichkeit von Feedback zu t_1 hat wiederum einen schwachen bis moderaten positiven Effekt auf die Veränderung der intrinsischen Motivation im Verlauf des Schulhalbjahres ($.15 \leq \beta \leq .26$, $.08 \leq SE \leq .11$, $p < .05$). Da die intrinsische Lernmotivation im Mittel abnimmt, bedeutet der positive Effekt, dass eine hohe Ausprägung in der wahrgenommenen Nützlichkeit die Abnahme in der intrinsischen Lernmotivation abmildert. Das wahrgenommene Kompetenzerleben zu t_1 hat einen moderat positiven Effekt auf die Veränderung der intrinsischen Motivation ($.43 \leq \beta \leq .38$, $SE = .12$, $p \leq .001$). Das bedeutet auch hier, dass eine hohe Ausprägung des wahrgenommenen Kompetenzerlebens zu einer verminderten Abnahme der intrinsischen Motivation führt.

Dementsprechend zeigen sich schwach positive indirekte Effekte auf die Veränderung der intrinsischen Motivation vermittelt über die wahrgenommene Nützlichkeit (indirekte Effekte: ergebnisorientiert t_1 : $\beta = .16$, $SE = .07$, $p = .02$, 95% CI [.14, .22]; prozessorientiert t_1 : $\beta = .11$, $SE = .05$, $p = .04$, 95% CI [.04, .15]; selbstregulationsorientiert t_1 : $\beta = .09$, $SE = .05$, $p = .06$, 95% CI [.04, .13]; dialogorientiert t_1 : $\beta = .05$, $SE = .03$, $p = .11$, 95% CI [.02, .07]) sowie schwach positive indirekte Effekte auf die Veränderung der intrinsischen Motivation vermittelt über das wahrgenommene Kompetenzerleben (indirekte Effekte: ergebnisorientiert t_1 : $\beta = .10$, $SE = .04$, $p = .01$, 95% CI [.04, .15]; prozessorientiert t_1 : $\beta = .10$, $SE = .04$, $p = .01$, 95% CI [.06, .18]; selbstregulationsorientiert t_1 : $\beta = .09$, $SE = .05$, $p = .002$, 95% CI [.03, .16]; dialogorientiert t_1 : $\beta = .08$, $SE = .06$, $p = .01$, 95% CI [.06, .18]). Mit der wahrgenommenen Nützlichkeit als vermittelnde Variable fällt der indirekte Effekt der ergebnisorientierten Dimension am stärksten aus. Mit dem Kompetenzerleben als vermittelnde Variable fallen die indirekten Effekte der ergebnis- und prozessorientierten Dimensionen am stärksten aus. Die Effekte bleiben auch unter Kontrolle des Geschlechts und der Vorleistung bestehen.

Gemäß der Bias-korrigierten Boot-Strap-Konfidenzintervalle sind die indirekten Effekte auf dem 5%-Niveau signifikant von Null verschieden. Mit den spezifizierten Modellen konnten zwischen 11.7% und 38.8% der Varianz in der Einschätzung der wahrgenommenen Nützlichkeit, zwischen 39.5% und 42.0% der Varianz in der Einschätzung des wahrgenommenen Kompetenzerlebens und zwischen 17.4% und 22.7% der Varianz in der Veränderung der intrinsischen Lernmotivation aufgeklärt werden.

Die indirekten sowie totalen Effekte der wahrgenommenen Dimensionen der Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation sind zusammenfassend in Tabelle 49 dargestellt.

Tabelle 49: Indirekte und totale Effekte der Dimensionen der Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation über die wahrgenommene Nützlichkeit und das wahrgenommene Kompetenzerleben

	Veränderung der intrinsischen Lernmotivation			
	β	<i>SE</i>	<i>p</i>	95% CI
<i>Ergebnisorientiert</i>				
Summe indirekt	.25	0.06	<.000	[.19, .33]
Direkt	-.31	0.10	.001	[-.45, -.22]
Total	-.06	0.07	.387	[-.16, .00]
<i>Prozessorientiert</i>				
Summe indirekt	.21	0.05	<.000	[.12, .25]
Direkt	-.23	0.08	.003	[-.35, -.11]
Total	-.03	0.07	.707	[-.11, .07]
<i>Selbstregulationsorientiert</i>				
Summe indirekt	.18	0.05	<.000	[.13, .24]
Direkt	-.24	0.07	<.001	[-.34, -.13]
Total	-.05	0.06	.349	[-.11, .12]
<i>Dialogorientiert</i>				
Summe indirekt	.13	0.04	.001	[.08, .16]
Direkt	-.32	0.06	<.001	[-.49, -.22]
Total	-.19	0.06	.003	[-.34, -.08]

Anmerkungen. Angegeben sind standardisierte Koeffizienten; signifikante Effekte sind fett gedruckt; CI = *Confidence Interval*.

8 Zusammenfassung und Diskussion der Ergebnisse

In den folgenden Abschnitten werden die zentralen Ergebnisse der Arbeit zusammengefasst und in Verbindung mit den aktuellen Forschungsbefunden diskutiert. In Abschnitt 8.1.1 erfolgt dies für die *Hauptforschungsfrage I*, in Abschnitt 8.1.2 für die *Hauptforschungsfrage II* und in Abschnitt 8.1.3 schließlich für die *Hauptforschungsfrage III*. Daran anknüpfend werden Limitationen der Arbeit diskutiert (Abschnitt 8.2) und Implikationen der Studie für die empirische Forschung (Abschnitt 8.3) sowie für die schulische Praxis (Abschnitt 8.4) abgeleitet.

8.1 Zusammenfassung der zentralen Befunde

8.1.1 Differenzielle Wahrnehmung von Feedback

Wie Schülerinnen und Schüler die Qualität ihres Unterrichts wahrnehmen ist von hoher Relevanz für ihren Lernprozess und Gegenstand verschiedener empirischer Studien (Fauth et al., 2014; Iglar et al., 2019; Kunter et al., 2007; Wagner et al., 2016; Willems, 2011, 2022a, 2022b). Ausgehend von den Annahmen etablierter Angebots-Nutzungs-Modelle ging diese Arbeit zunächst der Frage nach, wie Schülerinnen und Schüler Feedback – als ein relevantes Qualitätsmerkmal von Unterricht –

wahrnehmen und inwieweit das Geschlecht und die Leistung als individuelle Merkmale der Schülerinnen und Schüler differenzielle Feedbackwahrnehmungen und Wahrnehmungsveränderungen erklären (*Hauptforschungsfrage I*).

Faktorielle Struktur und Dimensionalität der wahrgenommenen Feedbackqualität im Quer- und Längsschnitt

Da es bisher an Instrumenten zur Erfassung der wahrgenommenen Feedbackqualität aus der Sicht von Schülerinnen und Schülern mangelt, fokussierten die *ersten Teilfragestellungen* auf forschungsmethodische Aspekte und zielten auf die Untersuchung der faktoriellen Struktur und Dimensionalität der wahrgenommenen Feedbackqualität sowie der Messinvarianz des Konstrukts über die Zeit und über Geschlechtergruppen hinweg ab. Die faktoranalytischen Auswertungen zur individuellen Wahrnehmung der Feedbackqualität aus der Sicht von Schülerinnen und Schülern (*Teilfragestellung 1a*) zeigten, dass ein Modell mit vier korrelierten Faktoren der wahrgenommenen Feedbackqualität sowohl gegenüber einem Generalfaktormodell als auch gegenüber einem Faktormodell höherer Ordnung zu bevorzugen ist. Somit stützt das empirisch vorgefundene Strukturmodell mit vier korrelierten Faktoren auch die theoretisch-konzeptuelle Annahme, dass Feedback kein eindimensionales, sondern ein komplexes, multidimensionales Konstrukt darstellt (Brooks et al., 2019; Hattie & Timperley, 2007; Strijbos et al., 2021). In den Analysen der vorliegenden Arbeit wurde daher zwischen vier unterschiedlichen Dimensionen der Feedbackqualität *Ergebnisorientierung*, *Prozessorientierung*, *Selbstregulationsorientierung* und *Dialogorientierung* differenziert.

Ergänzend zu den Befunden der konfirmatorischen Faktorenanalysen zur faktoriellen Struktur der wahrgenommenen Feedbackqualität wiesen die Ergebnisse auch auf hohe Korrelationen zwischen den einzelnen Dimensionen der Feedbackqualität hin (vgl. auch Willems & Dreiling, 2022), wobei die höchsten Zusammenhänge zwischen der ergebnis- und prozessorientierten Dimension sowie zwischen der prozess- und selbstregulationsorientierten Dimension bestehen. Aus inhaltlicher Sicht könnten die hohen Zusammenhänge darin begründet liegen, dass Lehrkräfte in der Unterrichtspraxis zeitgleich unterschiedliche Dimensionen von Feedback an die Schülerinnen und Schüler adressieren und beispielsweise mit einem Feedback zum Ergebnis der Aufgabenbearbeitung (*Ergebnisorientierung*) beginnen, um daran anschließend ein Feedback zum Prozess der Aufgabenbearbeitung (*Prozessorientierung*) zu erteilen. Diese Praxis spiegelt sich auch empirisch in Videostudien wider: Neben dem Befund, dass einfaches bzw. ergebnisorientiertes Feedback am häufigsten erteilt wird, zeigen die Analysen auch, dass in mehreren Fällen eine Kombination aus Feedback zum Ergebnis und zum Prozess der Aufgabenbearbeitung vorkommt (Lotz, 2016).

Die geringsten Zusammenhänge zeigten sich zwischen der dialogorientierten Dimension und den anderen drei Dimensionen, was den theoretischen Ansatz bekräftigt, dass auf dieser Ebene von Feedback der Fokus auf dem Feedbackdialog

zwischen der Lehrkraft mit der gesamten Klasse sowie zwischen den *peers* untereinander liegt (Orsmond et al., 2013), während auf den anderen Ebenen der Feedbackgestaltung der Fokus auf dem Feedbackdialog zwischen der Lehrkraft mit dem einzelnen Schüler bzw. der einzelnen Schülerin liegt.

Messinvarianz im Längsschnitt und im Multigruppenvergleich

Um die Vergleichbarkeit des Feedbackkonstrukts im Längsschnitt (*Teilfragestellung 1b*) und im Multigruppenvergleich (*Teilfragestellung 1c*) zu sichern, wurden die Feedbackskalen anschließend hinsichtlich ihrer Messinvarianz überprüft. Die längsschnittlichen konfirmatorischen Faktorenanalysen ergaben, dass für die identifizierte Faktorstruktur der wahrgenommenen Feedbackqualität starke faktorielle Messinvarianz über die Geschlechtergruppen sowie über die Messzeitpunkte hinweg angenommen werden kann. Dies bedeutet, dass sich die Messeigenschaften des Konstrukts zwischen den Messzeitpunkten nicht verändert haben und sich auch zwischen den Geschlechtergruppen nicht unterscheiden, sodass Mittelwerte, Varianzen und Regressionskoeffizienten in Bezug auf die Feedbackqualität miteinander verglichen werden können.

Geschlechtsspezifische Wahrnehmungen

Im Multigruppenvergleich konnten keine geschlechtsspezifischen Unterschiede in der Wahrnehmung der Feedbackqualität nachgewiesen werden. Schülerinnen und Schüler im Deutschunterricht der gymnasialen Oberstufe nehmen somit – unabhängig von ihrem Geschlecht – die Feedbackqualität in gleicher Weise wahr. Damit entsprechen die Befunde nicht den Ergebnissen vorangegangener Studien, die geschlechtsspezifische Unterschiede in der Wahrnehmung von Feedback im (Deutsch-)Unterricht nachweisen konnten (Carvalho et al., 2014; Chen et al., 2011; Havnes et al., 2012; Hoya, 2021; Sortkær, 2018, 2019). Die divergierenden Befunde könnten zum einen auf Unterschiede in der Operationalisierung von Feedback zurückgeführt werden. Während in dieser Arbeit leistungsbezogene Dimensionen der Feedbackqualität erfasst wurden, differenzierten die Studien von Hoya (2021) und Chen et al. (2011) zwischen positivem Feedback in Form von Lob der Leistungen oder des Verhaltens und negativem Feedback in Form von Ablehnung einer erbrachten Leistung oder eines Verhaltens. Diese Form der affektiven Tönung wurde in der vorliegenden Arbeit hingegen nicht berücksichtigt. Hinweise dafür, dass es geschlechtsspezifische Unterschiede in der Wahrnehmung von positivem und negativem Feedback gibt, aber nicht in der Wahrnehmung von leistungsbezogenen Dimensionen der Feedbackqualität, liefert die Studie von Denn et al. (2015). Allerdings handelt es sich um eine Videostudie, die zwar die Beschreibung der von außen beobachteten Feedbackgestaltung zulässt, aber keine Einblicke in die Innensicht der Schülerinnen und Schüler erlaubt.

Insgesamt konnten mit den vorliegenden Befunden keine Unterschiede in der Wahrnehmung der Feedbackqualität durch das Geschlecht der Schülerinnen und

Schüler erklärt werden, sodass die Frage aufkommt, ob Wahrnehmungsunterschiede stärker durch individuelle (kognitive und motivational-affektive) Lernvoraussetzungen erklärt werden können (vgl. Hoya, 2021). Erste Hinweise dahingehend liefert die Untersuchung von Willems und Dreiling (2022), in der nachgewiesen werden konnte, dass Schülerinnen und Schüler mit günstigen motivationalen Lernvoraussetzungen die Feedbackqualität höher einschätzen als jene mit vergleichsweise ungünstigen motivationalen Lernvoraussetzungen.

Differenzielle Veränderung der Feedbackwahrnehmung

Ausgehend von den deskriptiven Befunden (vgl. Abschnitt 7.3.1) liegt im Deutschunterricht der Oberstufe eine hohe Feedbackqualität vor. Vergleichbare Befunde zur (wahrgenommenen) Feedbackqualität im Deutschunterricht der Oberstufe sind bislang nicht vorhanden. In Studien zur (objektiven) Feedbackgestaltung in den unteren Jahrgangsstufen wurde deutlich, dass einfache Formen von Feedback im Unterricht häufiger erteilt werden als elaborierte (Denn et al., 2015; Krammer et al., 2010; Lotz, 2016; Pohlmann-Rother et al., 2020). Währenddessen erhalten Schülerinnen und Schüler der Oberstufe diesen Befunden zufolge häufig auch elaborierte Formen des Feedbacks. Dies könnte auch insgesamt die Bedeutung der Schülerperspektive stärken, die durchaus von der objektiveren Perspektive auf die Feedbackgestaltung abweichen kann (vgl. dazu auch Hoya, 2019). Denn während videobasierte Untersuchungen Einblicke in die tatsächlich vorliegende Feedbackgestaltung der Lehrkraft geben, können Schülerwahrnehmungen systematisch Aufschluss darüber geben, wie das Feedback bei den Adressatinnen und Adressaten wirklich ankommt und wie diese es für den Lernprozess verarbeiten (Willems & Dreiling, 2022).

Für die weiterführende Frage nach der Veränderung der Feedbackwahrnehmung im Laufe des Schulhalbjahres wurden Latent-True-Change-Analysen durchgeführt (vgl. Abschnitt 7.3.2). Hier zeigten sich signifikante Mittelwertveränderungen in der Wahrnehmung der ergebnis-, selbstregulations- und dialogorientierten Dimension der Feedbackqualität. Im Einzelnen nahm die wahrgenommene Ausprägung von Feedback zum Ergebnis im Laufe der Schulzeit ab, während die Ausprägungen von selbstregulationsorientiertem und dialogorientiertem Feedback zunahm. Beim Blick auf die einzelnen Mittelwerte zum ersten und zweiten Messzeitpunkt im Latent-State-Modell fiel aber auf, dass die Mittelwertunterschiede allenfalls gering ausfallen; für die prozessorientierte und selbstregulationsorientierte Dimension sind die mittleren Ausprägungen nahezu konstant geblieben. Vergleichsweise stärker fiel die Veränderung der Wahrnehmung der ergebnisorientierten ($\Delta MW_{t2-t1} = -0.10$) und dialogorientierten Dimension ($\Delta MW_{t2-t1} = 0.10$) aus. Die wahrgenommene Feedbackqualität kann sich im Laufe der Sekundarstufe II somit verändern. Dieser Befund korrespondiert zum Teil mit den Ergebnissen der Studie von De Wit et al. (2010), die eine Veränderung der wahrgenommenen Unterstüt-

zung beim Übergang in die Oberstufe beschreiben. Die Annahme, dass die wahrgenommene Feedbackqualität nach einem Schulhalbjahr sinkt (*Hypothese H1d-1*), konnte nur teilweise bestätigt werden. Eine Abnahme der Feedbackqualität ließ sich lediglich für die ergebnisorientierte Dimension feststellen. Tendenziell nahm die wahrgenommene Feedbackqualität etwas zu (selbstregulations- und dialogorientierte Dimension) oder veränderte sich nicht bedeutsam (prozessorientierte Dimension).

Gleichzeitig belegten die Befunde aber auch, dass die Veränderungen nicht für alle Schülerinnen und Schüler gleichermaßen ausfallen. Im Einzelnen konnte mit den Latent-True-Change-Analysen nachgewiesen werden, dass sich die Schülerinnen und Schüler im Deutschunterricht der gymnasialen Oberstufe nicht nur bedeutsam in der Wahrnehmung von Feedback zu Beginn des Schuljahres, sondern auch in ihrer Veränderung im Verlauf des Schulhalbjahres unterscheiden. Somit bestätigt sich die Annahme, der zufolge Schülerinnen und Schüler das Feedback nicht in gleicher Weise wahrnehmen und sich unterschiedlich in ihrer Wahrnehmung verändern (*Hypothese 1d-2*). Differenzielle Wahrnehmungs- und Veränderungsprozesse von Unterrichtsmerkmalen sind aufgrund von individuellen Merkmalen der Schülerinnen und Schüler sowie Merkmalen der Lernsituation zu erwarten und bestätigen auch vorangegangene Studienbefunde, die darauf verweisen, dass Schülerinnen und Schüler ihre Lernumwelt und dementsprechend auch die Qualität des Feedbacks ihrer Lehrkraft unterschiedlich wahrnehmen (Igeler et al., 2019; Jurik et al., 2015; Ruelmann et al., 2021; Schenke et al., 2018; Willems, 2022a; Willems & Dreiling, 2022).

Um interindividuelle Unterschiede in den Ausgangswerten der Wahrnehmung zu Beginn des Schuljahres und in der Veränderung der Wahrnehmung im Schulhalbjahr zu erklären, wurden in einem nächsten Schritt das Geschlecht und die Deutschnote (Leistungsindikator) als Prädiktoren in das LTC-Modell aufgenommen. Entgegen den Erwartungen (*Hypothese H1e-2*) konnten – ähnlich wie in den Untersuchungen von Reddy et al. (2003) und Way et al. (2007) – keine Effekte des Geschlechts und der Deutschnote auf interindividuelle Unterschiede in der intraindividuellen Veränderung der Feedbackwahrnehmung nachgewiesen werden. Allerdings erwies sich die Deutschnote prädiktiv für Wahrnehmungsunterschiede im Ausgangsniveau von ergebnisorientiertem, prozessorientiertem, selbstregulationsorientiertem und dialogorientiertem Feedback (*Hypothese H1e-1*). Den Befunden zufolge nehmen Schülerinnen und Schüler mit einer besseren Note wahr, dass sie diese Formen von Feedback häufiger erhalten als Schülerinnen und Schüler mit einer schlechteren Note im Fach Deutsch. Interessant an diesem Ergebnis ist außerdem, dass nach Einbezug der Deutschnote auch der Einfluss des Geschlechts auf Wahrnehmungsunterschieden im selbstregulationsorientierten und dialogorientierten Feedback zu Beginn des Schuljahres signifikant wurde. Im Vergleich zu Schülern berichten Schülerinnen demzufolge von einer geringeren Ausprägung des selbstregulationsorientierten und dialogorientierten Feedbacks. Denkbar wäre, dass es sich hier um einen Interaktionseffekt von Leistung und Geschlecht handelt:

Möglicherweise ist der Einfluss des Geschlechts unterschiedlich, je nachdem, ob es sich um leistungsschwache oder leistungsstarke Schülerinnen und Schüler handelt. Hoya (2021) belegt in seiner Studie signifikante Effekte des Geschlechts und der Leistung auf die Wahrnehmung von Feedback, allerdings kann er keine signifikanten Interaktionseffekte nachweisen.

Insgesamt bestätigen sich die Befunde vorangegangener Studien, die auf Unterschiede in der Unterrichtseinschätzung in Abhängigkeit des Leistungsniveaus hinweisen (Iglar et al., 2019; Hoya, 2021; Jurik et al., 2015; Sortkær, 2018; Wagner, 2008). Insbesondere der von Hoya (2021) nachgewiesene Befund, dass Wahrnehmungsunterschiede im erteilten Feedback stärker über die Leistung (und die Motivation) der Schülerinnen und Schüler und weniger über ihr Geschlecht erklärt werden können, bestätigt sich hier (vgl. auch Willems & Dreiling, 2022). Während in der vorliegenden Arbeit eine höhere Leistung mit einer höheren Feedbackwahrnehmung zusammenhing, ging in den Untersuchungen von Sortkaer (2018, 2019) umgekehrt eine geringere Leistung der Schülerinnen und Schüler mit einer höheren Ausprägung des wahrgenommenen Feedbacks einher, weshalb hier die Vermutung nahe lag, dass leistungsschwächere Schülerinnen und Schüler mehr Unterstützung in Form von Feedback erhalten. Mögliche methodische Ursachen für die unterschiedlichen Befunde könnten in der Operationalisierung von Feedback (als eindimensionales versus mehrdimensionales Konstrukt) sowie im Kontext der Studien (naturwissenschaftlicher Unterricht in der Sekundarstufe I versus Deutschunterricht in der Sekundarstufe II) liegen. Aus inhaltlicher Sicht wäre ein möglicher Grund für die divergenten Befunde, dass in den Analysen von Sortkaer (2018, 2019) die wahrgenommene Häufigkeit von Feedback im Unterricht im Fokus stand, während in der vorliegenden Arbeit die Wahrnehmung der Feedbackqualität im öffentlichen Unterrichtsgespräch erhoben wurde. Eine mögliche Begründung für den Befund, dass Schülerinnen und Schüler mit einer besseren Deutschleistung eine höhere Ausprägung von Feedback im Unterrichtsgespräch wahrnehmen, könnte sein, dass Schülerinnen und Schüler in Abhängigkeit von ihrer jeweiligen Leistung dasselbe Feedback unterschiedlich verarbeiten und bestimmten Formen von Feedback mehr Aufmerksamkeit widmen. Eine andere Erklärung wäre, dass Schülerinnen und Schüler im Unterrichtsgespräch tatsächlich ein interindividuell unterschiedliches Unterrichtsangebot erhalten (vgl. Decristan et al., 2020). Dies könnte zum einen darauf hinweisen, dass Lehrkräfte das Feedback im Sinne eines adaptiven Unterrichtshandelns an die Voraussetzungen der Schülerinnen und Schüler anpassen (Willems & Dreiling, 2022). Für den Kontext des Unterrichtsgesprächs ist zum anderen plausibel anzunehmen, dass die differenziellen Wahrnehmungen in Abhängigkeit der Leistung die unterschiedliche Beteiligung der Schülerinnen und Schüler am Unterrichtsgespräch widerspiegeln (vgl. Decristan et al., 2020). Wenn sich beispielsweise leistungsstarke Schülerinnen und Schüler aktiver am Unterrichtsgespräch beteiligen, erhalten sie vermutlich auch häufiger Feedback von der Lehrkraft auf ihre Beiträge, was sich wiederum in den differenziellen Wahrnehmungen nie-

derschlägt. Entsprechende empirische Belege finden sich in Studien, die nachweisen, dass leistungsstärkere Schülerinnen und Schüler häufiger am Unterrichtsgespräch beteiligt sind als ihre Mitschülerinnen und Mitschüler (Decristan et al., 2020; Jurik et al., 2013; Pielmeier et al., 2018). Anhand der Analysen von Jurik et al. (2014) zeigte sich außerdem, dass das Feedback der Lehrperson – als ein zentrales Merkmal des Unterrichtsgesprächs – die kognitiven Lernaktivitäten und die intrinsische Lernmotivation der Schülerinnen und Schüler positiv vorhersagte und das insbesondere Lernende mit einem „starken“ Profil (hoch ausgeprägte kognitive und motivational-affektive Lernvoraussetzungen) über höhere kognitive Lernaktivitäten und eine höhere intrinsische Lernmotivation berichteten. Offensichtlich profitieren Schülerinnen und Schüler also unterschiedlich vom Feedback der Lehrkraft im Unterrichtsgespräch in Abhängigkeit ihrer individuellen Lernvoraussetzungen. Vor diesem Hintergrund erscheint es für Lehrkräfte umso wichtiger, im Unterricht unterschiedliche Gelegenheiten für die aktive, verbale Beteiligung – auch von Schülerinnen und Schülern mit ungünstigen Lernvoraussetzungen – zu schaffen, damit alle Lernenden gleichermaßen Feedback erhalten und davon profitieren können (vgl. auch Jurik et al., 2014).

8.1.2 Zusammenhänge zwischen Feedback und Lernmotivation

In der Mehrzahl der Studien stand die Wirkung von Feedback auf kognitive und motivationale Merkmale im Fokus der Untersuchung. Basierend auf konzeptuellen Annahmen von Angebots-Nutzungs-Modellen (Lipowsky, 2020; Vieluf et al., 2020) und aufgrund jüngster Forschungsbefunde (Ruelmann et al., 2021; Willems & Dreiling, 2022) wurde aber davon ausgegangen, dass nicht nur Feedback auf die motivationalen Merkmale von Schülerinnen und Schülern wirkt, sondern auch die Wahrnehmung von Feedback durch die individuellen motivationalen Eingangsvoraussetzungen beeinflusst wird. Das Kernanliegen der vorliegenden Arbeit bestand daher darin, Wechselwirkungen von Unterrichtswahrnehmungen (Feedbackqualität) und der Motivation von Schülerinnen und Schülern zu untersuchen. Mit der Erforschung dieser reziproken Prozesse sollte ein Beitrag zur Überprüfung zentraler konzeptueller Annahmen des Angebots-Nutzungs-Modells geleistet werden. Somit erweitert die vorliegende Untersuchung den Forschungsstand zum Zusammenspiel der beiden Konstrukte über die Zeit, indem aus methodischer Sicht eine Längsschnittdanalyse auf latenter und damit messfehlerbereinigter Ebene durchgeführt wurde.

Ausgangspunkt der *Hauptforschungsfrage II* war zunächst die Spezifikation eines separaten Latent-True-Change-Modells für die Analyse der Veränderung der intrinsischen Lernmotivation (vgl. Kapitel 7.4.5) im Verlauf eines Schulhalbjahres. Die Analysen zeigten im Detail, dass die durchschnittliche intrinsische Lernmotivation der Schülerinnen und Schüler im Deutschunterricht der Oberstufe vom ersten zum zweiten Messzeitpunkt signifikant abnahm. Damit können die Ergebnisse der Oberstufenstudie von Spinath und Steinmayr (2012), die von einer Stabilität der

durchschnittlichen Motivation für das Lernen im Fach Deutsch berichten, nicht bestätigt werden. Allerdings bezogen sich die Ergebnisse von Spinath und Steinmayr auf die Entwicklung der mittleren Ausprägung der intrinsischen Motivation. Die LTC-Analyse der vorliegenden Arbeit erbrachte auch Belege für differenzielle Ausgangslagen und Entwicklungsverläufe der intrinsischen Lernmotivation. Die Oberstufenschülerinnen und -schüler unterscheiden sich nicht nur bedeutsam im Ausgangsniveau ihrer intrinsischen Lernmotivation zu Beginn des Schuljahres, sondern auch in ihrer Veränderung im Verlauf des Schulhalbjahres, d.h. die Entwicklung der intrinsischen Motivation (bzw. die Abnahme) vollzieht sich nicht bei allen Schülerinnen und Schülern gleichermaßen. Diese differenziellen Entwicklungen der Motivation bei Oberstufenschülerinnen und -schülern im Deutschunterricht konnten in ähnlicher Weise bereits mittels eines personenzentrierten Untersuchungsansatzes von Dreiling und Willems (2022) nachgewiesen werden. Vergleichbare Untersuchungsergebnisse finden sich in Bezug auf den Deutschunterricht im Primarbereich (Weidinger et al., 2015).

Vertiefende Analysen zu den Merkmalen, die interindividuelle Unterschiede in der Veränderung der intrinsischen Lernmotivation beeinflussen, erbrachten einen signifikant negativen Effekt der Deutschnote: Demzufolge fiel die Abnahme der Motivation im Schulhalbjahr bei leistungsstarken Schülerinnen und Schülern stärker aus als bei leistungsschwachen Schülerinnen und Schülern. Ein Effekt des Geschlechts auf Unterschiede in der Motivationsveränderung konnte in dieser Studie jedoch nicht nachgewiesen werden. Auch Weidinger et al. (2015) berichten in ihrer Untersuchung von einem direkten Effekt der Deutschnote auf die Veränderung der intrinsischen Motivation im Deutschunterricht der Primarstufe. Allerdings fiel der Effekt – wie in dieser Studie auch – nur gering aus, was darauf hindeutet, dass Noten allenfalls eine untergeordnete Rolle in der Erklärung differenzieller Entwicklungen der intrinsischen Motivation – auch im Deutschunterricht der Oberstufe – spielen (vgl. Weidinger et al., 2015) und dass andere Faktoren eine höhere Bedeutsamkeit aufweisen.

Für die Kernforschungsfrage dieser Studie nach dem Zusammenhang zwischen der Feedbackwahrnehmung und der intrinsischen Lernmotivation sowie den Veränderungen der Konstrukte über die Zeit wurden im nächsten Schritt bivariate Latent-True-Change-Modelle berechnet und folgende Pfade spezifiziert: (a) Zusammenhänge zwischen den Ausgangswerten der beiden Konstrukte zu Beginn des Schuljahres, (b) Zusammenhänge zwischen den parallelen Veränderungen der Konstrukte im Schulhalbjahr sowie (c) prädiktive Effekte des Ausgangswerts eines Konstrukts auf die Veränderung des jeweils anderen Konstrukts im Schulhalbjahr (kreuzverzögerte Effekte). Hinsichtlich der *Teilfragestellung F2a* zeigten die Ergebnisse, dass die wahrgenommenen Ausprägungen der Dimensionen der Feedbackqualität und die Ausgangsmotivation zu Beginn des Schuljahres positiv korrelieren. Das bedeutet, dass eine höhere Ausprägung der Feedbackqualität mit einer höheren intrinsischen Lernmotivation einhergeht und umgekehrt. *Hypothese 2a-1* konnte somit bestätigt werden. Auch die Veränderungen standen in einem moderat

positiven Zusammenhang. Eine Veränderung in der Wahrnehmung der Feedbackqualität geht demnach mit einer Veränderung der intrinsischen Lernmotivation im Verlauf des Schulhalbjahres einher und umgekehrt. Entsprechend konnte *Hypothese 2a-2* ebenso bestätigt werden. Die Befunde erweitern den bisherigen Forschungsstand zum dynamischen Zusammenspiel von Veränderungsprozessen der Unterrichtswahrnehmung und der individuellen Motivation von Schülerinnen und Schülern (Lazarides et al., 2019; Maulana et al., 2016; Ruzek & Schenke, 2019). Sie stützen zudem die theoretischen Annahmen der Selbstbestimmungstheorie, wonach eine Veränderung der Lernumgebung (hier: Feedback der Lehrkraft) die Entwicklung der intrinsischen Lernmotivation beeinflusst (Deci & Ryan, 2000; Deci et al., 1999, 2001; Ryan & Deci, 2000). Im Sinne einer reziproken Beziehung von Angebot und Merkmalen der Lernenden (u.a. Lipowsky, 2020; Vieluf et al., 2020), kann umgekehrt eine hohe intrinsische Lernmotivation dazu führen, dass die Qualität des Feedbacks von den Schülerinnen und Schülern positiver eingeschätzt wird (Ruelmann et al., 2021; Willems & Dreiling, 2022).

Allerdings kann aufgrund der korrelativen Befunde noch nicht auf gerichtete Beziehungen zwischen den Merkmalen geschlossen werden. Obschon die Konstrukte bzw. deren Veränderungen bedeutsam miteinander korrelieren, konnten hinsichtlich der Teilfragestellung F2b nicht für alle Feedbackdimensionen signifikante kreuzverzögerte Effekte nachgewiesen werden. Im Einzelnen konnten keine direkten Effekte der Ausgangswerte der ergebnis- und prozessorientierten Dimensionen auf die spätere Veränderung der Lernmotivation im Schulhalbjahr beobachtet werden. Demgegenüber zeigten sich (geringe) negative Effekte der Ausgangswerte der selbstregulationsorientierten und dialogorientierten Feedbackdimension auf die spätere Veränderung der intrinsischen Lernmotivation. Somit konnte *Hypothese 2b-1* nur teilweise bestätigt werden. Da die intrinsische Lernmotivation im Mittel abnimmt, wäre der negative Koeffizient so zu interpretieren, dass eine höher ausgeprägte Wahrnehmung des selbstregulations- und dialogorientierten Feedbacks zu Beginn des Schuljahres zu einem stärkeren Rückgang der intrinsischen Lernmotivation im Laufe des Schulhalbjahres führt. Dieser Befund erscheint zunächst widersprüchlich, da aus motivationspsychologischer Sicht und aufgrund empirischer Befunde (vgl. Abschnitt 3.4) zu erwarten wäre, dass sich eine hohe wahrgenommene Qualität von Feedback (im Sinne einer Selbstregulations- und Dialogorientierung) positiv auf die Veränderung der intrinsischen Lernmotivation auswirkt. Aus methodischer Sicht könnten die negativen Effekte von der Feedbackwahrnehmung auf die Veränderung der Motivation ein Hinweis für einen methodischen Artefakt sein: Durch diese Art der Veränderungsmodellierung würde bei geringen zeitübergreifenden Zusammenhängen – wie es hier der Fall war – fälschlicherweise ein signifikanter Effekt aufgezeigt werden (Christoph et al., 2014; Steyer et al., 2000). Die negativen Effekte auf die Veränderung der intrinsischen Lernmotivation sind daher mit Vorsicht zu interpretieren. Eher ist davon auszugehen, dass die wahrgenommene Qualität von Feedback die Motivationsveränderung nicht direkt beeinflusst. Es könnte vielmehr darauf ankommen, wie das Feedback im weiteren Prozess von

den Schülerinnen und Schülern verarbeitet wird – ob es beispielsweise als nützlich wahrgenommen wird oder ob es das motivationsrelevante Erleben von Kompetenz unterstützt (z.B. Hondrich et al., 2018; Rakoczy et al., 2013). Diese Annahmen wurden daher in der *Hauptforschungsfrage III* überprüft; die Ergebnisse dazu werden im nächsten Unterkapitel diskutiert.

Hinsichtlich der *Hypothese 2b-1* konnte ein (geringer) negativer Effekt des Ausgangsniveaus der intrinsischen Lernmotivation auf die Veränderung der Wahrnehmung von prozessorientiertem Feedback – nicht aber auf die Veränderung der anderen Feedbackdimensionen nachgewiesen werden, sodass sich auch diese Hypothese nur teilweise bestätigen ließ. Da die Ausprägung der prozessorientierten Dimension im Mittel zunimmt, war der negative Effekt so zu interpretieren, dass eine höhere intrinsische Lernmotivation zu Beginn des Schuljahres mit einer Abnahme der Ausprägung einhergeht. Dementsprechend führt eine geringere intrinsische Motivation bei Schülerinnen und Schülern zu einer stärkeren Zunahme von wahrgenommen prozessorientierten Feedback. Während also zu Beginn des Schuljahres eine höhere Motivation (und höhere Deutschleistung, vgl. Abschnitt 7.3.3) mit einer positiveren Wahrnehmung von (prozessorientiertem) Feedback einhergeht, zeichnet sich nach einem Schulhalbjahr ein anderer Trend ab: Es scheint, dass Lehrkräfte den wenig motivierten Schülerinnen und Schülern – die im Durchschnitt auch eine geringere Deutschleistung aufweisen – im Laufe des Schulhalbjahres mehr (prozessorientiertes) Feedback erteilen. Diese benötigen im Unterricht möglicherweise mehr Unterstützung in Form von präzisen Hinweisen und Strategien zur Bearbeitung einer Lernaufgabe bzw. Beantwortung einer Frage als ihre hoch motivierten (und leistungsstarken) Mitschülerinnen und -schüler (Sortkaer, 2019). Einschränkend ist allerdings festzuhalten, dass insgesamt nur geringe kreuzverzögerte Effekte festgestellt werden konnten. Dies lässt darauf schließen, dass die Effekte nur wenig zusätzliche Varianz in der Veränderung der intrinsischen Lernmotivation bzw. der Feedbackwahrnehmung erklären können.

Die geringen oder zum Teil ausgebliebenen kreuzverzögerten Effekte in der vorliegenden Studie können möglicherweise auf den in der Studie gewählten Zeitintervall von einem halben Schuljahr zwischen den beiden Messzeitpunkten zurückzuführen sein. In früheren Studien wurde die Entwicklung der Motivation im Zeitraum von einem Schuljahr betrachtet (z.B. Spinath & Steinmayr, 2012). Denkbar wäre aber auch, dass der hier gewählte Zeitraum zu lang ist, um reziproke Entwicklungszusammenhänge zwischen Unterrichtswahrnehmungen und der Motivation von Schülerinnen und Schülern nachweisen zu können. Um diese Frage zu klären, bedarf es zukünftig mehr längsschnittlicher Studien, die Einschätzungen der Schülerinnen und Schüler zur Unterrichtsqualität und ihrer Motivation zu mehreren Messzeitpunkten erheben und kürzere Zeitabstände zwischen den einzelnen Messungen realisieren (z.B. Patall et al., 2018; Rakoczy et al., 2022). Bei mehr als zwei Messzeitpunkten ist es zudem möglich, gezielte Hypothesen über die Form der Veränderung zu testen (z.B. lineare oder nicht lineare Veränderung) (Dietrich, 2019).

8.1.3 Indirekte Effekte von Feedback auf die Lernmotivation

Die *Hauptforschungsfrage III* fokussierte auf indirekte Effekte der Dimensionen der Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation. Dabei wurden zwei Mediatorvariablen der individuellen Nutzung aufgenommen: die wahrgenommene Nützlichkeit von Feedback und das subjektive Kompetenzerleben. Diese ließen sich wie folgt theoretisch begründen: Damit Feedback lernwirksam sein kann, muss es von den Lernenden zunächst in seiner Ausprägung wahrgenommen und als nützlich bewertet werden, um angemessene kognitive, motivational-affektive und verhaltensbezogene Reaktionen hervorzurufen (Lipnevich et al., 2016; Lipnevich & Panadero, 2021; Rakoczy et al., 2013; Rakoczy et al., 2019; Strijbos et al., 2010; Strijbos et al., 2021). Vor dem Hintergrund der Selbstbestimmungstheorie wirkt sich ein Feedback, das informierend gestaltet ist und das Kompetenzerleben von Lernenden unterstützt, positiv auf die intrinsische Motivation aus. In Anlehnung an theoretische Angebots-Nutzungs-Modelle (Klieme & Rakoczy, 2008; Seidel, 2014; Vieluf et al., 2020) kann das motivationsrelevante Erleben von Kompetenzunterstützung als Indikator der individuellen Nutzung des Lernangebots betrachtet und methodisch als Mediatorvariable in die Analyse der Wirkungsweise von Feedback einbezogen werden (Jansen et al., 2022; Ledergerber, 2015; Rakoczy, 2008; Rakoczy et al., 2013; Rakoczy et al., 2022; Willems, 2011).

Die Mediationsanalysen bestätigten diese theoretischen Annahmen. Im Einzelnen zeigte sich in Bezug auf die *Hypothese 3a*, dass – auch unter Kontrolle der Deutschsleistung und des Geschlechts der Schülerinnen und Schüler – höhere Ausprägungen in den Dimensionen der Feedbackqualität zu Beginn des Schuljahres mit einer höheren Ausprägung in der wahrgenommenen Nützlichkeit einhergingen, die sich wiederum positiv auf die Veränderung der intrinsischen Lernmotivation im Schulhalbjahr auswirkte. Entsprechend zeigten sich positive indirekte Effekte der wahrgenommenen Feedbackqualität auf die Motivationsveränderung, vermittelt über die wahrgenommene Nützlichkeit von Feedback. Ein qualitativ volles Feedback, das von Schülerinnen und Schülern als nützlich wahrgenommen wird, kann die Abnahme der intrinsischen Lernmotivation im Deutschunterricht sozusagen abmildern. Damit stehen die Befunde auch im Einklang mit den Studienergebnissen von Rakoczy et al. (2013), Rakoczy et al. (2019) sowie Harks et al. (2014), die anhand von Daten aus dem Co2Ca-Projekt positive Effekte der wahrgenommenen Nützlichkeit von Feedback auf die motivationale Entwicklung und einen positiven indirekten Effekt von (prozessorientiertem) Feedback auf die motivationale Entwicklung – vermittelt über die wahrgenommene Nützlichkeit – nachweisen konnten. Interessant an den dargestellten Befunden dieser Arbeit ist, dass die ergebnis- und prozessorientierte Dimension von Feedback stärker mit der wahrgenommenen Nützlichkeit korrelieren als mit der selbstregulations- und dialogorientierten Dimension. Denkbar wäre, dass die Schülerinnen und Schüler mit ergebnis- und prozessorientiertem Feedback vertrauter sind, weshalb diese mit einer höheren Nütz-

lichkeit assoziiert sind. Im Kontext der gymnasialen Oberstufe, in der die Schülerinnen und Schüler auf die Abiturprüfungen vorbereitet werden, kann es naheliegend sein, dass solche Formen von Feedback, die Informationen in Bezug auf die Lernziele (ergebnisorientiert) und spezifische Hinweise zur Verbesserung (prozessorientiert) enthalten, als hilfreicher empfunden werden.

Des Weiteren konnte auch *Hypothese 3b* bestätigt werden: So führten höhere Ausprägungen in den Dimensionen der Feedbackqualität zu einer positiveren Wahrnehmung des subjektiven Kompetenzerlebens, das wiederum die Veränderung der intrinsischen Lernmotivation im Schulhalbjahr positiv beeinflusste. Entsprechend zeigten sich auch hier positive indirekte Effekte der wahrgenommenen Feedbackqualität auf die Motivationsveränderung, vermittelt über das wahrgenommene Kompetenzerleben. Anders ausgedrückt: Ein qualitätsvolles Feedback, das das Kompetenzerleben von Schülerinnen und Schülern unterstützt, kann die Abnahme der intrinsischen Lernmotivation im Deutschunterricht abschwächen. Die vermittelnde Wirkung des Kompetenzerlebens wurde auch in anderen Studien bereits beschrieben (z.B. Gniewosz et al., 2014; Hondrich et al., 2018; Rakoczy et al., 2013). Insgesamt wird die motivationsförderliche Wirkung von Feedback als Unterrichtsmerkmal durch die Nutzung mediiert, die in dieser Studie über die wahrgenommene Nützlichkeit von Feedback und das motivationsrelevante Erleben von Kompetenz operationalisiert wurde.

Zusammenfassend wird anhand der dargestellten Befunde deutlich, dass Schülerinnen und Schüler mit günstigeren Lernvoraussetzungen (höhere Leistung und Eingangsmotivation) die Qualität von Feedback positiver wahrnehmen. Wenn Schülerinnen und Schüler über eine höhere Ausprägung von qualitativem Feedback berichten, dann geht dies mit einer positiveren Bewertung der Nützlichkeit und einem höheren Kompetenzerleben einher, was wiederum eine positive Motivationsentwicklung begünstigt. In dieser Hinsicht untermauern die Befunde eine differenzielle Wahrnehmung (und Nutzung) von Feedback als eine Ursache für differenzielle Motivationsentwicklungen. Tabelle 50 fasst die (teilweise) bestätigten oder verworfenen Hypothesen abschließend zusammen.

Tabelle 50: Übersicht über Hypothesen und empirische Befunde der Studie

		Bestätigt	Teilweise bestätigt	Nicht bestätigt
H1d-1	Die wahrgenommene Feedbackqualität sinkt über das Schulhalbjahr hinweg.		x	
H1d-2	Es zeigen sich interindividuelle Unterschiede in der Veränderung der Feedbackwahrnehmung.	x		
H1e-1	Das Geschlecht stellt – auch unter Berücksichtigung der Vorleistung – einen relevanten Prädiktor differenzieller Feedbackwahrnehmungen dar.		x	
H1e-2	Das Geschlecht stellt – auch unter Berücksichtigung der Vorleistung – einen relevanten Prädiktor interindividueller Unterschiede in der Veränderung der Feedbackwahrnehmung dar.			x
H2a-1	Je höher die wahrgenommene Feedbackqualität ausgeprägt ist, desto höher ist die intrinsische Lernmotivation.	x		
H2a-2	Je stärker sich die Feedbackwahrnehmung zur Mitte des Schuljahres verändert, desto stärker verändert sich die intrinsische Lernmotivation.		x	
H2b-1	Die wahrgenommene Feedbackqualität zu Beginn des Schuljahres sagt die Veränderung in der intrinsischen Lernmotivation im Laufe des Schulhalbjahres vorher.		x	
H2b-2	Die intrinsische Lernmotivation zu Beginn des Schuljahres sagt die Veränderung in der Wahrnehmung von Feedbackqualität im Laufe des Schuljahres vorher.		x	
H3a	Die wahrgenommene Nützlichkeit von Feedback kann die Wirkung von wahrgenommener Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation vermitteln und erklären.	x		
H3b	Das Kompetenzerleben kann die Wirkung von wahrgenommener Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation vermitteln und erklären.	x		

8.2 Grenzen der Studie

Bei der Interpretation der Befunde der vorliegenden Studie sind einige Einschränkungen zu beachten.

Zunächst beziehen sich die Ergebnisse auf das Fach Deutsch im Kontext der gymnasialen Oberstufe. Offen bleibt beispielsweise, ob sich das Modell der vier Dimensionen von Feedbackqualität auch aus der Sicht von Schülerinnen und Schülern in anderen Fächern der gymnasialen Oberstufe abbilden lässt und ob sich die berichteten (indirekten) Zusammenhänge zwischen der wahrgenommenen Feedbackqualität und der Motivation von Schülerinnen und Schülern auf andere Kontexte übertragen lassen. Neben dem Fokus auf den Deutschunterricht beziehen sich die Einschätzungen zur Feedbackqualität nur auf das öffentliche Unterrichtsgespräch, sodass keine Aussagen über das Feedback der Lehrkraft in anderen Phasen des Unterrichts (z.B. Schülerarbeitsphasen) getroffen werden können. Denkbar wäre, dass es Unterschiede in der wahrgenommenen Feedbackqualität zwischen lehrerzentrierten (z.B. öffentliches Unterrichtsgespräch) und schülerzentrierten (z.B. Schülerarbeitsphasen) Unterrichtsphasen gibt, die in dieser Studie allerdings nicht adressiert werden konnten. Die Videoanalysen der Studie von Kobarg und Seidel (2007) zeigten dahingehend, dass Lehrkräfte sowohl in Phasen des Unterrichtsgesprächs als auch in Schülerarbeitsphasen des Physikunterrichts nur selten prozessorientiertes Feedback erteilen.

Zudem gehen mit der hier fokussierten Schülerwahrnehmung Einschränkungen bei der Interpretation der Befunde einher. Da es sich um Selbstauskünfte der Schülerinnen und Schüler handelt, könnten die Befundmuster potenziell über- oder unterschätzt werden. Aufgrund des Erhebungsdesigns bleibt ungeklärt, ob die beschriebenen hohen Ausprägungen der Feedbackdimensionen sowie die empirisch nachgewiesenen Wahrnehmungsunterschiede ein Produkt individueller Verarbeitungsprozesse der Schülerinnen und Schüler sind oder die differenzielle Feedbackgestaltung der Lehrkraft widerspiegeln (vgl. auch Willems & Dreiling, 2022).

Eine weitere Limitation dieser Studie ist, dass die einzelnen Dimensionen der Feedbackqualität hoch miteinander korrelieren und fraglich ist, ob die Indikatoren trennscharf genug sind. Wie den Befunden der Korrelations- und Mediationsanalysen zu entnehmen ist, scheint es wichtig zu sein zwischen den unterschiedlichen Dimensionen zu differenzieren, da sie unterschiedlich mit der motivationalen Entwicklung der Schülerinnen und Schüler zusammenhängen. Außerdem zeigen sich auch unterschiedliche Effekte der Deutschleistung auf die Schülereinschätzungen der Feedbackdimensionen, was auch das Argument stärkt, die Feedbackdimensionen – trotz des hohen Anteils an geteilter Varianz – separat zu erfassen. In vertiefenden Analysen der FeeHe-Daten sollten weitere kognitive und motivational-afektive Zielkriterien des Unterrichts einbezogen werden, um zu untersuchen, welche spezifische Vorhersagekraft die Feedbackdimensionen zur Veränderung dieser Merkmale haben.

Eine weitere Einschränkung betrifft die Modellierung. Das in der vorliegenden Arbeit eingesetzte Verfahren der Strukturgleichungsmodellierung bietet gegenüber anderen statistischen Analysemethoden (z.B. manifeste Korrelations- oder Pfadanalysen) insbesondere den Vorteil, dass komplexe Zusammenhänge zwischen Variablen auf latenter und damit messfehlerbereinigter Ebene untersucht werden. Auch für Veränderungsmessungen stellen lineare Strukturgleichungsmodelle ein geeignetes Verfahren dar (vgl. Kapitel 6.6, s. auch Geiser, 2011). Gleichzeitig weisen auch Strukturgleichungsmodellierungen Grenzen auf. So ist es mit dieser Analyse-methode nur bedingt möglich Annahmen über kausale Beziehungen zwischen Variablen nachzuweisen. Neben theoretischen Herleitungen im Hinblick auf die vermutete Wirkungsbeziehung zwischen Variablen ist es darüber hinaus erforderlich, dass die unabhängige Variable, die als Ursache angenommen wird, der abhängigen Variablen zeitlich vorgelagert ist (Weber & Mühlhaus, 2014). Diese Bedingung erfordert ein längsschnittliches Design. In der vorliegenden Arbeit wurden diese Voraussetzungen nur bedingt erfüllt: Die Untersuchung der Wirkungsrichtungen zwischen der wahrgenommenen Feedbackqualität und der intrinsischen Lernmotivation konnte aufgrund der zeitlich wiederholten Erfassung beider Merkmale über die Modellierung von bivariaten Latent-True-Change-Modellen realisiert werden. Dadurch konnten vermutete Wirkungen von einzelnen Dimensionen der Feedbackqualität auf die Motivationsentwicklung, aber auch von der intrinsischen Lernmotivation als individuelle Lernvoraussetzung auf die Veränderung der wahrgenommenen Feedbackqualität nachgewiesen werden. Für das Mediationsmodell hingegen, mittels dessen auch Wirkungen der beiden Mediatorvariablen (wahrgenommene Nützlichkeit und Kompetenzerleben) auf die Motivationsentwicklung überprüft werden sollten, muss einschränkend festgehalten werden, dass die Mediatorvariablen der Motivationsvariable im Modell nicht zeitlich vorgelagert sind. Obschon aufgrund der Selbstbestimmungstheorie davon ausgegangen werden kann, dass die Entwicklung einer intrinsischen Motivationsqualität vom Kompetenzerleben bedingt wird, sind die kausalen Beziehungen hier mit Vorsicht zu interpretieren. Beispielsweise wäre es auch denkbar, dass Unterschiede in der intrinsischen Motivation von Schülerinnen und Schülern zu unterschiedlichen Wahrnehmungsprozessen in Bezug auf die Nützlichkeit von Feedback führen können. Letztlich sind randomisierte Experimentalstudien erforderlich, in denen Störvariablen möglichst ausgeschlossen werden und somit Annahmen über kausale Beziehungen zwischen Variablen nachgewiesen werden können.

Eine weitere Einschränkung ist zudem, dass sich die Operationalisierung der Lernmotivation auf die intrinsische Motivation beschränkte und die Formen der extrinsischen Motivation (external, introjiziert, identifiziert) nicht betrachtet wurden. Außerdem wurde – anknüpfend an vorherige Forschungsbefunde – nur das Kompetenzerleben als mögliche motivationsrelevante Mediatorvariable in der Analyse der indirekten Wirkungen von Feedback auf die Entwicklung der intrinsischen Motivation berücksichtigt. Die medierende Bedeutung des Autonomieerlebens wurde nicht untersucht, obgleich aufgrund bestehender Forschungsbefunde (z.B.

Mouratidis et al., 2010) angenommen werden könnte, dass bestimmte Formen der Feedbackqualität auch das Autonomieerleben und vermittelt darüber die intrinsische Motivation der Schülerinnen und Schüler positiv unterstützen. Diese Annahme sollte insbesondere für die Indikatoren der selbstregulations- und dialogorientierten Dimensionen von Feedbackqualität zukünftig überprüft werden. Indem auf diesen Ebenen von Feedback die Verantwortung für die Verbesserung eines Schülerbeitrags an die Lernenden selbst übertragen wird, ist anzunehmen, dass das Erleben von Autonomie positiv unterstützt wird.

8.3 Implikationen für die Forschung

Die Untersuchung leistet einen wichtigen Beitrag zur Theoriebildung, indem Erkenntnisse zum Zusammenspiel von Feedbackwahrnehmung, individueller Nutzung und intrinsischer Lernmotivation gewonnen wurden. Im Folgenden wird diskutiert, welche weiterführenden Studien sich aus den Ergebnissen dieser Arbeit ergeben.

Modellierung von Feedback

Ein Ergebnis der vorliegenden Studie ist, dass distinkte Dimensionen der Feedbackqualität aus der Sicht von Schülerinnen und Schülern erfasst werden können. Allerdings bestehen nach wie vor starke Zusammenhänge zwischen den Dimensionen der Feedbackqualität im Strukturmodell. Um diese Zusammenhänge besser abzubilden, könnten alternative Messmodelle in Betracht gezogen werden, die die allgemeine Wahrnehmung bzw. Gesamtbeurteilung der Feedbackqualität (Generalfaktor) von den spezifischen Wahrnehmungen der vier unterschiedlichen Dimensionen der Feedbackqualität (spezifische Faktoren) trennen. Dies kann in einem Bifaktormodell dargestellt werden (Reise, 2012; Anwendung bei Scherer et al., 2016). Mit diesem hierarchischen Modell kann überprüft werden, ob die Items auf einen Generalfaktor *Feedbackqualität* 1. Ordnung und vier unkorrelierten Faktoren 1. Ordnung laden, die den Dimensionen *Ergebnisorientierung*, *Prozessorientierung*, *Selbstregulationsorientierung* und *Dialogorientierung* entsprechen. Hierdurch lässt sich abschätzen, wie viel Varianz der Indikatorvariablen durch den Generalfaktor und wie viel Varianz durch einen spezifischen Residualfaktor erklärt wird, der unabhängig vom Generalfaktor ist (Gäde, Schermelleh-Engel & Brandt, 2020). Wird ein geringerer Anteil der Indikatorvarianz durch den Generalfaktor erklärt als durch einen bestimmten Residualfaktor, so wäre bestätigt, dass die Indikatoren besser durch vier eigenständige Faktoren (die vier Dimensionen der Feedbackqualität) als durch einen Generalfaktor der Feedbackqualität abgebildet werden können.

Um die Zusammenhänge im Messmodell als auch die konzeptuellen Überlappungen von Dimensionen der Unterrichtsqualität abbilden zu können, schlagen Scherer et al. (2016) noch eine weitere Form der Modellierung vor, in der Doppelloadungen von Items – also Ladungen eines Items auf mehrere Faktoren – zugelassen

werden: Die explorative Strukturgleichungsmodellierung (*exploratory structural equation modeling*, ESEM; vgl. Marsh et al., 2014) ist ein alternatives Verfahren zur konfirmatorischen Faktoranalyse. Scherer et al. (2016) erachten die Annahme der konfirmatorischen Faktoranalyse, in der Items eindeutig einem Faktor zugeordnet werden dürfen, als zu stark. In ihren Analysen zur Dimensionalität der Unterrichtsqualität auf Basis von Schülerwahrnehmungen erreichte das ESEM-Modell gegenüber dem traditionellen CFA-Modell einen deutlich besseren Modell-Fit. In zukünftigen Arbeiten könnten solche alternativen Formen der Modellierung in Bezug auf die Feedbackqualität in Betracht gezogen werden. In diesem Zusammenhang könnte auch geprüft werden, mit welchen faktoranalytischen Verfahren (traditionelle CFA, Bifaktormodell, ESEM) die Effekte der Dimensionen der Feedbackqualität auf die Lernergebnisse besser erfasst werden können (vgl. Scherer et al., 2016).

Feedback im Modell der drei Basisdimensionen

In der vorliegenden Studie wurde nachgewiesen, dass die Feedbackqualität (aus der Sicht der Schülerinnen und Schüler) ein multidimensionales Konstrukt darstellt und dass diese – vermittelt über die wahrgenommene Nützlichkeit und das motivationsrelevante Kompetenzerleben – die Entwicklung der intrinsischen Motivation von Schülerinnen und Schülern positiv unterstützt. Für die pädagogisch-psychologische Unterrichtsforschung, in der das generische Modell der drei Basisdimensionen zur Beschreibung der Unterrichtsqualität weit verbreitet ist (Fauth et al., 2014; Klieme et al., 2006; Klieme, 2019; Klieme & Rakoczy, 2008; Praetorius et al., 2018) und die darauf basierenden Messinstrumenten stets weiterentwickelt werden, sollte angesichts dieser Befunde künftig in Betracht gezogen werden, die zentralen Dimensionen von qualitativem Feedback noch systematischer als bisher in das Modell der drei Basisdimensionen einzubeziehen. Zwar sind in einigen Beschreibungen von einzelnen Basisdimensionen – wie der konstruktiven Unterstützung oder der kognitiven Aktivierung – durchaus auch Feedback als Qualitätsmerkmal enthalten (vgl. Göllner et al., 2021; Kunter & Trautwein, 2013; Praetorius et al., 2014). Allerdings wird Feedback hier eher als eindimensionales Konstrukt behandelt. Dass aber das Erteilen von Feedback nicht per se zu größerem Lernerfolg oder Verständnis der Lerninhalte bei den Lernenden führt, ist bereits unumstritten. Um die Forschungsbefunde dahingehend adäquat abzubilden, erscheint das Aufgreifen von Qualitätsaspekten der Feedbackgestaltung bei der Beschreibung der einzelnen Basisdimensionen unumgänglich. Beispielsweise könnten unterschiedliche Qualitätsdimensionen von Feedback als zusätzliche Subdimensionen berücksichtigt werden, indem auch elaborierte Aspekte wie Selbstregulations- und Dialogorientierung weiter ausdifferenziert werden. Um zentrale Dimensionen von Unterrichtsqualität valide zu erfassen, könnte zusätzlich darüber nachgedacht werden, Feedback als eigenständige Dimension von Unterrichtsqualität im Modell der drei Basisdimensionen abzubilden, wie das beispielsweise in der Konzeptualisierung von Charalambous und Praetorius (2020) geschieht (vgl. auch Praetorius & Gräsel, 2021) und in der Arbeit

von Buholzer et al. (2023) in Bezug auf das verwandte Konstrukt *Formatives Assessment* nahegelegt wird. In zukünftigen Forschungsarbeiten könnte daran anknüpfend untersucht werden, wie Feedback als eigenständige Qualitätsdimension von Unterrichtsqualität mit den drei Basisdimensionen zusammenhängt.

Effekte auf kognitive Zielkriterien

In der FeeHe-Studie liegen keine Daten zur Entwicklung der Leistungen und Kompetenzen der Schülerinnen und Schüler im Fach Deutsch der Oberstufe vor, sodass Wirkungen der wahrgenommenen Dimensionen von Feedbackqualität auf den Leistungs- oder Kompetenzzuwachs in der vorliegenden Arbeit nicht untersucht werden konnten. Für eine systematische Untersuchung des Modells der vier Feedbackdimensionen bedarf es weiterer Studien, die den Zusammenhang der Schüler-einschätzungen in Bezug auf die Feedbackqualität mit kognitiven – sowie weiteren motivational-affektiven – Zielkriterien von Unterricht differenziert untersuchen, und damit die Bedeutung der aus Schülersicht wahrgenommenen Feedbackqualität für die Lernprozesse von Schülerinnen und Schüler prüfen.

Berücksichtigung der Variabilität in der Feedbackwahrnehmung

Für Forschung im Bereich der Unterrichtswahrnehmung von Lernenden ist relevant, dass in der vorliegenden Studie bedeutsame interindividuelle Unterschiede in der Ausgangswahrnehmung zu Beginn des Schuljahres und in der Veränderung der Wahrnehmung von Feedback im Schulhalbjahr gefunden wurden. Ein Hauptergebnis dieser Studie ist demnach, dass Schülerinnen und Schüler nicht alle gleichermaßen das Feedback der Lehrkraft wahrnehmen und dass sich ihre Wahrnehmung über die Zeit unterschiedlich verändert. Allerdings konnte nur ein geringer Anteil der Gesamtvarianz in der differenziellen Wahrnehmung und Veränderung der Feedbackqualität erklärt werden. Das heißt, die Deutschleistung und die Eingangsmotivation sagten zwar zum Teil signifikante Unterschiede in der Feedbackwahrnehmung oder der Veränderung der Wahrnehmung voraus, aber interindividuelle Unterschiede werden zum größten Teil durch andere Faktoren beeinflusst. Es blieb also ein erheblicher Anteil an unerklärter Varianz. Zukünftige Forschung könnte dieses Desiderat aufgreifen und weitere Prädiktorvariablen untersuchen, die als potenziell einflussnehmend auf die Wahrnehmung der Qualität von Feedback und deren zeitliche Veränderung gelten. Aus den theoretischen Ausführungen von Strijbos und Müller (2014) ergibt sich, dass Merkmale der Lehrperson (z.B. die wahrgenommen Kompetenz) und weitere kognitive und motivational-affekte Merkmale der Lernenden (z.B. Selbstwirksamkeit, Attribution von Erfolg und Misserfolg, motivationale Zielorientierungen) Wahrnehmungsunterschiede hervorrufen können. In der Studie von Iglér et al. (2019) wurden Wahrnehmungsunterschiede in Merkmalen der Unterrichtsqualität untersucht und festgestellt, dass insbesondere die individuellen Schülermerkmale Vorwissen, kognitive Fähigkeiten, motivationale Orientie-

rung und emotionale Bindung sowie soziale Herkunft prädiktiv für die Schülereinschätzung hinsichtlich der Unterrichtsqualität sind. Zukünftige Untersuchungen könnten den Einfluss dieser Variablen auch auf die Einschätzung der Feedbackqualität untersuchen.

Aufgrund der nachgewiesenen Bedeutsamkeit individueller Merkmale und Lernvoraussetzungen von Schülerinnen und Schülern für die Wahrnehmung der Feedbackqualität muss zukünftige Forschung die Frage klären, wie solche individuellen Lernvoraussetzungen in Modellen berücksichtigt werden können, um der Annahme differenzieller Wahrnehmungsprozesse und Wirkungsweisen von Feedback noch stärker als bisher zu begegnen. Für die Variablen im Modell statistisch zu kontrollieren, erscheint nicht mehr ausreichend, um die Annahme zu prüfen, *für welche* Schülerinnen und Schüler eine bestimmte Form von Feedback besonders nützlich oder wirkungsvoll ist. Anknüpfend an die Studien von Rakoczy et al. (2013) und Weidinger et al. (2016) könnten individuelle Merkmale auch als mögliche *Moderatorvariablen* in den Analysen berücksichtigt werden. Den Befunden von Rakoczy et al. (2013) zufolge wird der Einfluss der Feedbackqualität auf die wahrgenommene Nützlichkeit durch die Ausprägung der Lernzielorientierung der Schülerinnen und Schüler moderiert. Dementsprechend nehmen lernzielorientierte Schülerinnen und Schüler, die bestrebt sind, die Lerninhalte zu verstehen, ein prozessorientiertes Feedback als nützlicher wahr als ein sozialvergleichendes Feedback (Rakoczy et al., 2013). Jüngste Studien im Bereich der personenzentrierten Forschung untersuchen die kognitiven und motivational-affektiven Merkmale von Schülerinnen und Schülern in Form von komplexen *Profilen*, die sich differenziell auf die Wahrnehmung der Unterrichts- und Feedbackqualität auswirken (Ruelmann et al., 2021; Seidel, 2006; Willems, 2022; Willems & Dreiling, 2022; Willems et al., im Druck). Aufgrund dessen könnte weiterführend untersucht werden, inwiefern sich in Abhängigkeit dieser Schülerprofile auch differenzielle Wirkungen der Feedbackqualität ergeben.

Implikationen in Bezug auf das Studiendesign

Da wechselseitige Einflüsse der wahrgenommenen Feedbackqualität und der Lernmotivation im Zeitraum von einem halben Schuljahr in der vorliegenden Arbeit nicht oder nur teilweise bestätigt werden konnten, erscheint es für zukünftige Forschungsarbeiten relevant, längsschnittliche Studiendesigns durchzuführen, in denen mehrere Messzeitpunkte sowie kürzere Zeitintervalle zwischen den Erhebungen der Unterrichtswahrnehmung und der Motivation von Schülerinnen und Schülern realisiert werden. So berichten Rakoczy et al. (2022), dass ein bedeutsamer Anteil der Varianz in den Einschätzungen der Unterrichtsqualität zwischen den Messzeitpunkten liegt. In der Studie von Patall et al. (2018) wurde die wahrgenommene Autonomieunterstützung sowie die Motivation und das Verhalten von Schülerinnen und Schülern der US-amerikanischen *High School* täglich während einer naturwissenschaftlichen Unterrichtseinheit erfasst sowie entsprechende Zusammenhänge untersucht. Die Analysen ergaben, dass die Einschätzungen von Lernenden zwischen

verschiedenen Messzeitpunkten variieren. Mittels Mehrebenenanalysen konnten die Autoren außerdem reziproke Effekte nachweisen: Zum einen konnten Veränderungen der selbstbestimmten Motivation und des Verhaltens durch das wahrgenommene autonomieunterstützende Unterrichtshandeln vorhergesagt werden, zum anderen beeinflussten auch die Motivation und das Verhalten der Schülerinnen und Schüler Veränderungen in ihrer Wahrnehmung des autonomieunterstützenden Lehrkraftverhaltens.

Insgesamt stärken die Befunde das Argument, nicht nur die Variabilität von Unterrichtseinschätzungen zwischen den einzelnen Schülerinnen und Schülern (*interindividuelle* Unterschiede), sondern auch zwischen den Messzeitpunkten (*intraindividuelle* Unterschiede) stärker zu berücksichtigen. Vor diesem Hintergrund sind auch retrospektiv erfasste und klassenweise aggregierte Urteile über Unterrichtswahrnehmung, Nutzung und Motivation im Kontext mehrebenenanalytischer Auswertungen kritisch zu hinterfragen (Rakoczy et al., 2022; vgl. hierzu auch kritisch Iglér et al., 2019; Willems & Dreiling, 2022). Daran anknüpfend sind auch weitere Studien notwendig, die untersuchen, in welchem Ausmaß die Variabilität in den Unterrichtseinschätzungen durch individuelle Wahrnehmungsprozesse der Schülerinnen und Schüler oder differenzielle Angebotsstrukturen im Unterricht entsteht (Rakoczy et al., 2022; vgl. auch Iglér et al., 2019; Willems & Dreiling, 2022). Um solche Fragen angemessen zu beantworten, erscheint es für die zukünftige Forschung gewinnbringend, die subjektive Erfassung der Schülerperspektive mit einer objektiven Beobachtung des unterrichtlichen Handelns zu kombinieren und wechselseitig aufeinander zu beziehen.

8.4 Implikationen für die Praxis

Ausgehend von den Befunden dieser Untersuchung lassen sich Implikationen sowohl für die pädagogische Praxis als auch für die universitäre Ausbildung von Lehrkräften ableiten.

Die vorliegende Studie demonstriert die positive, über die wahrgenommene Nützlichkeit und das Kompetenzerleben vermittelnde Wirkung von Rückmeldungen, die spezifische Informationen zum *Ergebnis* in Bezug auf konkrete Lernziele, weiterführende Hinweise zum *Lernprozess* enthalten und die zur *Selbstregulation* des Lernprozesses sowie zum *Dialog* bzw. Austausch mit den Mitschülerinnen und -schülern anregen. Die Ergebnisse zeigten außerdem, dass die ergebnis- und prozessorientierte Dimension von Feedback stärker mit der wahrgenommenen Nützlichkeit zusammenhängen als mit der selbstregulations- und dialogorientierten Dimension. Denkbar wäre, dass die Schülerinnen und Schüler mit ergebnis- und prozessorientiertem Feedback vertrauter sind und es deshalb als nützlicher bewerten.

Dies ist angesichts der gängigen eher summativen Leistungsbeurteilungspraxis (üblicherweise in Form einer am Ende einer Unterrichtseinheit erfolgten Klassenarbeit) nicht verwunderlich: Diese signalisiert Lehrenden sowie Lernenden, dass sie einen bestimmten Sollwert erreichen müssen, was zur Verengung von Unterrichtsprozessen (Klieme et al., 2010) und damit auch zu einer geringen Bereitschaft zur Umsetzung von vergleichsweise zeitintensiveren und voraussetzungsreicheren Formen von Feedback führen kann. Aufgrund der nachgewiesenen förderlichen Wirkungen sollte das Potenzial unterschiedlicher Feedbackformen im Unterricht stärker genutzt werden, indem einerseits Feedback als fester (und von Noten losgelöster) Bestandteil in den Lehr-Lernprozess eingebunden wird und andererseits auch bei einer summativen Feedbackkultur auf eine differenzierte Gestaltung der Rückmeldungen geachtet wird (Besser et al., 2017; Klieme et al., 2010).

In Bezug auf die Praxis zeigen die Ergebnisse, dass es beim Erteilen von Feedback im Unterricht entscheidend darauf ankommt, wie Schülerinnen und Schüler vor dem Hintergrund ihrer individuellen Merkmale das Feedback überhaupt wahrnehmen und verarbeiten. Die Befunde zur differenziellen Feedbackwahrnehmung unterstreichen, dass das Feedback der Lehrkraft nicht von allen Schülerinnen und Schülern gleichermaßen wahrgenommen wird und dementsprechend nicht für alle Schülerinnen und Schüler wirkungsvoll ist. Die Herausforderung besteht darin, dass das Feedback von den Schülerinnen und Schülern als nützlich für den Lernprozess bewertet wird und das Bedürfnis nach Kompetenz erfüllt, also den Schülerinnen und Schülern das Gefühl gibt, den Lernprozess aus eigener Kraft verbessern zu können. Im Hinblick auf die unterrichtliche Praxis kann es zielführend sein, dass Lehrkräfte ihre Schülerinnen und Schüler zur aktiven Beteiligung am Unterrichtsgespräch motivieren, sodass möglichst viele ein individuelles Feedback erhalten können (Decristan et al., 2020). Hier bedürfen auch gerade Schülerinnen und Schüler mit ungünstigeren Lernvoraussetzungen besondere Aufmerksamkeit, da diese seltener im Unterrichtsgespräch beteiligt werden und entsprechend seltener von Feedback profitieren (Decristan et al., 2020; Jurik et al., 2013). Zum anderen scheint es im Sinne einer individuellen Förderung im Unterricht allgemein darauf anzukommen, dass Schülerinnen und Schüler unterschiedliche, an ihre Lernvoraussetzungen angepasste Rückmeldungen erhalten. Auch im Kontext der gymnasialen Oberstufe bestehen trotz erfolgter Differenzierung durch die Kurswahlen immer noch Leistungs- (Nagy et al., 2006; Trautwein et al., 2010) und Motivationsunterschiede (Baumert & Köller, 2000; Dreiling & Willems, 2020) zwischen den Schülerinnen und Schülern innerhalb der Kurse, sodass auch hier Lehrkräfte mit einer heterogenen Schülerschaft konfrontiert sind, die Feedback unterschiedlich wahrnimmt und verarbeitet. Hierfür müssen die unterschiedlichen Voraussetzungen der Schülerinnen und Schüler im Unterricht zunächst sichtbar gemacht werden. In diesem Zusammenhang kann Feedback als dialogischer Prozess verstanden werden, der sich nicht nur von der Lehrkraft an die Schülerinnen und Schüler richtet, sondern auch umgekehrt (Hattie & Zierer, 2018). Lernende können der Lehrkraft Rückmeldung dar-

über geben, auf welchem Leistungsstand sie stehen und wo die Verständnisprobleme liegen. Das Feedback von Lernenden kann die Lehrkräfte dann im nächsten Schritt dazu nutzen, den Unterricht auf der Ebene des Lernprozesses und der Selbstregulation so anzupassen, dass die Lernaufgaben verstanden und bearbeitet werden können (Zierer et al., 2019).

Ein solches adaptives Unterrichtshandeln setzt entsprechende diagnostische Kompetenzen der Lehrkraft voraus, mit denen Lehrkräfte die Unterschiede in den Lernvoraussetzungen von Schülerinnen und Schülern feststellen und ihr Feedbackverhalten darauf anpassen. Damit rücken die professionellen Merkmale der Lehrkraft als eine Bedingung der Effektivität von Feedback in den Fokus (Besser et al., 2017), weshalb das Thema Feedback in der universitären Lehrkräfteausbildung fest verankert sein sollte. Neben diagnostischen Techniken, einem vertieften Verständnis fachlicher Lerninhalte und -prozesse sollte dabei auch ein differenziertes und systematisches Wissen über verschiedene Formen von lernförderlichem Feedback und die wirksame Umsetzung davon im Unterricht vermittelt werden (Degeling, 2019). Zu dem fachlichen, fachdidaktischen und pädagogisch-psychologischem Wissen kommt noch hinzu, dass es der Motivation bzw. Bereitschaft bedarf, die im Rahmen des Feedbackprozesses gewonnenen Informationen für die Weiterentwicklung des Unterrichts zu nutzen (Degeling, 2019; Zierer et al., 2019). Außerdem kommen ausgeprägte Überzeugungen bezüglich des Potentials von einem lernförderlichen (formativen) Feedback hinzu (Besser et al., 2015; Klieme et al., 2010). Im Bereich der Forschung zur Förderung professioneller Kompetenzen belegen die Ergebnisse die Wirksamkeit von Fortbildungen beim Aufbau des Wissens von Lehrkräften in Bezug auf Feedback (u.a. Besser et al., 2017). Weiterhin konnte im Rahmen des Forschungsprojekts ProFee das Potenzial einer (strukturierten) Videoanalyse zur Förderung der Handlungskompetenz von Lehramtsstudierenden beim Erteilen von Feedback demonstriert werden (Hess & Lipowsky, 2019).

Schließlich scheint es vor dem Hintergrund der wachsenden Bedeutung von digital geprägten Lehr-Lernangeboten (Eickelmann & Gerick, 2020) wichtig, die Möglichkeit von digital gestütztem Feedback zur Verbesserung der Unterrichtsqualität und zur Förderung der Lernergebnisse stärker in den Blick zu nehmen (Wisniewski, 2019). Eine Möglichkeit kann beispielweise darin bestehen, bestimmte Formen von Feedback (z.B. Informationen zur Richtigkeit) bei einer Aufgabenbearbeitung an ein digitales Tool auszulagern, um so kontinuierlich und individualisiert auf die Schülerinnen und Schüler mit ihren unterschiedlichen Lernvoraussetzungen eingehen zu können.

9 Verzeichnisse

9.1 Literaturverzeichnis

- Allison, P. D. (2012). Handling Missing Data by Maximum Likelihood. SAS Global Forum. Verfügbar unter <https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf> [Zugriff am 16.03.2023].
- Babakus, E., Ferguson, C. E. & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, 24(2), 222–228. <https://doi.org/10.1177/002224378702400209>
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9(1), 78–102. https://doi.org/10.1207/S15328007SEM0901_5
- Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(2), 211–240. <https://doi.org/10.1080/10705510801922340>
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A. & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238. <https://doi.org/10.3102/00346543061002213>

- Baumert, J. & Köller, O. (2000). Motivation, Fachwahlen, selbstreguliertes Lernen und Fachleistungen im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 2 Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 181–213). Leske + Budrich.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W. & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Waxmann.
- Bayer, S., Klieme, E. & Jude, N. (2016). Assessment and evaluation in educational contexts. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Eds.), *Assessing contexts of learning: Methodology of educational measurement and assessment* (pp. 469–488). Springer. https://doi.org/10.1007/978-3-319-45357-6_19
- Bentler, P. M. & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78–117. <https://doi.org/10.1177/0049124187016001004>
- Berger, R., Granzer, D., Looss, W. & Waack, S. (2013). „Warum fragt ihr nicht einfach uns?“. *Mit Schüler-Feedback lernwirksam unterrichten: Unterrichtsentwicklung nach Hattie*. Beltz.
- Berner, V.-D., Seitz-Stein, K., Segerer, R., Oesterlen, E. & Niklas, F. (2022). “Good” or “well calculated”? Effects of feedback on performance and self-concept of 5- to 7-year-old children in math. *Educational Psychology*, 42(3), 296–315. <https://doi.org/10.1080/01443410.2021.2001790>
- Besser, M., Leiss, D. & Klieme, E. (2015). Wirkung von Lehrerfortbildungen auf Expertise von Lehrkräften zu formativem Assessment im kompetenzorientierten Mathematikunterricht. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 110–122. <https://doi.org/10.1026/0049-8637/a000128>
- Besser, M., Leiss, D. & Schütze, B. (2017). Wirkung von Lehrerfortbildungen auf Überzeugungen von Lehrkräften zu lernförderlichem Assessment im Fach Mathematik. *mathematica didactica*, 40(2), 1–17. <https://doi.org/10.18716/ojs/md/2017.1229>
- Brooks, C., Huang, Y., Hattie, J., Carroll, A. & Burton, R. (2019). What is my next step? School students’ perceptions of feedback. *Frontiers in Education*, 4(96). <https://doi.org/10.3389/educ.2019.00096>
- Brophy, J. E. & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Eds.), *Handbook of research on teaching: A project of the American Educational Research Association* (pp. 328–375). MacMillan.

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research. Methodology in the social sciences*. The Guilford Press.
- Buff, A. (2014). Enjoyment of learning and its personal antecedents: Testing the change–change assumption of the control-value theory of achievement emotions. *Learning and Individual Differences, 31*, 21–29. <https://doi.org/10.1016/j.lindif.2013.12.007>
- Buff, A., Reusser, K. & Pauli, C. (2010). Die Qualität der Lernmotivation in Mathematik auf der Basis freier Äusserungen: Welches Bild präsentiert sich bei Deutschschweizer Schülerinnen und Schüler im 8. und 9. Schuljahr? In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 253–278). Waxmann.
- Buholzer, A., Baer, M., Ruelmann, M. & Zulliger, S. (2023). Formatives Assessment – eine vernachlässigte Dimension von Unterrichtsqualität? *Bildung und Erziehung, 76*(1), 43–60. <https://doi.org/10.13109/buer.2023.76.1.43>
- Burton, B. D., Lydon, J. E., D’Alessandro, D. U. & Koestner, R. (2006). The differential effects of intrinsic and identified motivation on well-being and performance: Prospective, experimental, and implicit approaches to self-determination theory. *Journal of Personality and Social Psychology, 91*(4), 750–762. <https://doi.org/10.1037/0022-3514.91.4.750>
- Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. <https://doi.org/10.3102/00346543065003245>
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge. <https://doi.org/10.4324/9780203807644>
- Byrne, B. M., Shavelson, R. J. & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466. <https://doi.org/10.1037/0033-2909.105.3.456>
- Carvalho, C., Santos, J., Conboy, J. & Martins, D. (2014). Teachers’ feedback: Exploring differences in students’ perceptions. *Procedia - Social and Behavioral Sciences, 159*, 169–173. <https://doi.org/10.1016/j.sbspro.2014.12.351>
- Cerasoli, C. P., Nicklin, J. M. & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980–1008. <https://doi.org/10.1037/a0035661>
- Charalambous, C. Y. & Praetorius, A.-K. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies in Educational Evaluation, 67*, 100894. <https://doi.org/10.1016/j.stueduc.2020.100894>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>

- Chen, Y.-H., Thompson, M. S., Kromrey, J. D. & Chang, G. H. (2011). Relations of student perceptions of teacher oral feedback with teacher expectancies and student self-concept. *The Journal of Experimental Education*, 79(4), 452–477. <https://doi.org/10.1080/00220973.2010.547888>
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Christoph, G., Gniewosz, B. & Reinders, H. (2014). How does community service promote prosocial behavior? Examining the role of agency and ideology experience. *International Journal of Behavioral Development*, 38(6), 499–508. <https://doi.org/10.1177/0165025414531465>
- Clausen, M. (2002). *Qualität von Unterricht: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Waxmann.
- Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Downer, J., Hamre, B. & Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade: Implications for children’s experiences and conducting classroom observations. *The Elementary School Journal*, 112(1), 16–37. <https://doi.org/10.1086/660682>
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. <https://doi.org/10.1037/0033-2909.125.6.627>
- Deci, E. L., Koestner, R. & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, 71(1), 1–27. <https://doi.org/10.3102/00346543071001001>
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223–238. <https://doi.org/10.25656/01:11173>
- Deci, E. L. & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Deci, E. L. & Ryan, R. M. (Eds.). (2002). *Handbook of self-determination research*. University of Rochester Press.
- Decristan, J., Fauth, B., Heide, E. L., Locher, F. M., Troll, B., Kurucz, C. & Kunter, M. (2020). Individuelle Beteiligung am Unterrichtsgespräch in Grundschulklassen: Wer ist (nicht) beteiligt und welche Konsequenzen hat das für den Lernerfolg? *Zeitschrift für Pädagogische Psychologie*, 34(3-4), 171–186. <https://doi.org/10.1024/1010-0652/a000251>

- Degeling, M. (2019). Feedback im Unterricht. Warum lernförderliches Feedback zu geben, eine hohe Kunst ist und wie sie sich in der Praxissemestervorbereitung und -begleitung anbahnen lässt. Vorschläge zur Diskussion. In M. Degeling, N. Franken, S. Freund, S. Greiten, D. Neuhaus & J. Schellenbach-Zell (Hrsg.), *Herausforderung Kohärenz: Praxisphasen in der universitären Lehrerbildung: Bildungswissenschaftliche und fachdidaktische Perspektiven* (S. 312–326). Klinkhardt. <https://doi.org/10.25656/01:17284>
- Denn, A.-K. (2021). *Interaktionen von Lehrpersonen mit Mädchen und Jungen im Mathematikunterricht der Grundschule: Geschlechtsspezifische Unterschiede und Zusammenhänge mit der Selbstkonzeptentwicklung*. Springer. <https://doi.org/10.1007/978-3-658-32372-1>
- Denn, A.-K., Lotz, M., Theurer, C. & Lipowsky, F. (2015). „Prima, Lisa. Richtig“ und „Psst, Max. Hör auf zu stören!“: Eine quantitative Studie zu Unterschieden im Feedbackverhalten von Lehrkräften gegenüber Mädchen und Jungen im Mathematikunterricht des zweiten Schuljahres. *GENDER*, 7(1), 29–47. <https://doi.org/10.3224/gender.v7i1.18155>
- De Wit, D. J., Karioja, K. & Rye, B. J. (2010). Student perceptions of diminished teacher and classmate support following the transition to high school: Are they related to declining attendance? *School Effectiveness and School Improvement*, 21(4), 451–472. <https://doi.org/10.1080/09243453.2010.532010>
- Diedrich, J., Schiepe-Tiska, A., Ziernwald, L., Tupac-Yupanqui, A., Weis, M., McElvany, N. & Reiss, K. (2019). Lesebezogene Schülermerkmale in PISA 2018: Motivation, Leseverhalten, Selbstkonzept und Lesestrategiewissen. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018. Grundbildung im internationalen Vergleich* (S. 81–109). Waxmann. <https://doi.org/10.31244/9783830991007>
- Dietrich, J. (2019). Methoden der Veränderungsmessung. In B. Kracke & P. Noack (Hrsg.), *Handbuch Entwicklungs- und Erziehungspsychologie* (S. 541–563). Springer. https://doi.org/10.1007/978-3-642-53968-8_32
- Dietrich, J., Dicke, A.-L., Kracke, B. & Noack, P. (2015). Teacher support and its influence on students' intrinsic value and effort: Dimensional comparison effects across subjects. *Learning and Instruction*, 39, 45–54. <https://doi.org/10.1016/j.learninstruc.2015.05.007>
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48(2), 262–286. <https://doi.org/10.25656/01:3833>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humantwissenschaften* (5. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-41089-5>

- Dreiling, K., Fliel, R. & Willems, A. S. (2018). FeeDo – Feedbackprozesse im Deutschunterricht der Oberstufe: Erste Befunde einer Pilotierungsstudie zur standardisierten Erfassung wahrgenommenen Feedbacks. In A. Krüger, F. Radisch, A. S. Willems, T. H. Häcker & M. Walm (Hrsg.), *Empirische Bildungsforschung im Kontext von Schule und Lehrer*innenbildung* (S. 194–205). Klinkhardt.
- Dreiling, K. & Willems, A. S. (im Druck). Geschlechtsspezifische Entwicklungen der Lernmotivation von Schülerinnen und Schülern im Deutschunterricht der gymnasialen Oberstufe: eine Anwendung von Latent-Transition-Analysen. In A. S. Willems (Hrsg.), „*Wieso? Weshalb? Warum?*“: *Motivation und Einstellung in (außer-)schulischen Bildungskontexten*. Universitätsverlag Göttingen.
- Dreiling, K. & Willems, A. S. (2020). Heterogene Lernvoraussetzungen - heterogene Kurszusammensetzungen? Geschlechtsspezifische Motivationsprofile als Lernausgangslagen in Deutschkursen der gymnasialen Oberstufe. In C. I. Keßler & S. Nonte (Hrsg.), *(Neue) Formen der Differenzierung: Schul- und Klassenprofilierung im Spannungsfeld gesellschaftlicher Teilhabe und sozialer Ungleichheit* (S. 159–179). Waxmann.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C. & Iver, D. M. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, 48(2), 90–101. <https://doi.org/10.1037/0003-066X.48.2.90>
- Eccles, J. S. & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. D. Steinberg (Eds.), *Handbook of adolescent psychology* (3th ed.). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470479193.adlpsy001013>
- Eckert, M., Dreiling, K. & Willems, A. S. (im Druck). Wie beeinflussen motivationsrelevante Überzeugungen das Kurswahlverhalten von Oberstufenschülerinnen und -schülern im Fach Deutsch? In A. S. Willems (Hrsg.), „*Wieso? Weshalb? Warum?*“: *Motivation und Einstellung in (außer-)schulischen Bildungskontexten*. Universitätsverlag Göttingen.
- Eickelmann, B. & Gerick, J. (2020). Lernen mit digitalen Medien: Zielsetzungen in Zeiten von Corona und unterbesonderer Berücksichtigung von sozialen Ungleichheiten. In D. Fickermann & B. Edelstein (Hrsg.), „*Langsam vermisst ich die Schule ...*“: *Schule während und nach der Corona-Pandemie* (S. 153–162). Waxmann. <https://doi.org/10.31244/9783830992318.09>
- Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion. Bachelorstudium Psychologie*. Hogrefe.
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6(4), 352–370. <https://doi.org/10.1037/1082-989X.6.4.352>

- Enders, C. K. & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430–457.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.
<https://doi.org/10.1016/j.learninstruc.2013.07.001>
- Ferrer, E. & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, 19(3), 149–154. <https://doi.org/10.1177/0963721410370300>
- Gäde, J. C., Schermelleh-Engel, K. & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 615–659). Springer.
https://doi.org/10.1007/978-3-662-61532-4_24
- Gäde, J. C., Schermelleh-Engel, K. & Werner, C. S. (2020). Klassische Methoden der Reliabilitätsschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 305–334). Springer.
https://doi.org/10.1007/978-3-662-61532-4_14
- Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion*, 27(3), 199–223.
<https://doi.org/10.1023/A:1025007614869>
- Gamlem, S. M. & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150–169.
<https://doi.org/10.1080/0969594X.2012.749212>
- Geiser, C. (2011). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* (2. Aufl.). VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-93192-0>
- Gerstenmaier, J. & Mandl, H. (1995). Wissenserwerb unter konstruktivistischer Perspektive. *Zeitschrift für Pädagogik*, 41, 867–888.
<https://doi.org/10.25656/01:10534>
- Gnamb, T. & Hanfstingl, B. (2016). The decline of academic motivation during adolescence: An accelerated longitudinal cohort analysis on the effect of psychological need satisfaction. *Educational Psychology*, 36(9), 1691–1705.
<https://doi.org/10.1080/01443410.2015.1113236>
- Gniewosz, B., Eccles, J. S. & Noack, P. (2014). Early adolescents' development of academic self-concept and intrinsic task value: The role of contextual feedback. *Journal of Research on Adolescence*, 25(3), 459–473.
<https://doi.org/10.1111/jora.12140>
- Gniewosz, B. & Gniewosz, G. (2018). How to model intraindividual change in cohort data using Mplus' multi-group approach. *International Journal of Behavioral Development*, 42(3), 373–380. <https://doi.org/10.1177/0165025417750282>

- Göllner, R., Wagner, W., Klieme, E., Lüdtke, O., Nagengast, B. & Trautwein U. (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. In Bundesministerium für Bildung und Forschung (Hrsg.), *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments*. (S. 63-82). Bundesministerium für Bildung und Forschung. <https://doi.org/10.25656/01:12674>
- Göllner, R., Fauth, B. & Wagner, W. (2021). Student ratings of teaching quality dimensions: Empirical findings and future directions. In W. Rollett, H. Bijlsma & S. Röhl (Eds.), *Student feedback on teaching in schools: Using student perceptions for the development of teaching and teachers* (pp. 111–122). Springer. https://doi.org/10.1007/978-3-030-75150-0_7
- Göllner, R., Wagner, W., Eccles, J. S. & Trautwein, U. (2018). Students' idiosyncratic perceptions of teaching quality in mathematics: A result of rater tendency alone or an expression of dyadic effects between students and teachers? *Journal of Educational Psychology*, 110(5), 709–725. <https://doi.org/10.1037/edu0000236>
- Gottfried, A. E., Fleming, J. S. & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93(1), 3–13. <https://doi.org/10.1037/0022-0663.93.1.3>
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H. & Wright Guerin, D. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. *International Journal of Behavioral Development*, 31(4), 317–327. <https://doi.org/10.1177/0165025407077752>
- Gottfried, A. E., Nylund-Gibson, K., Gottfried, A. W., Morovati, D. & Gonzalez, A. M. (2017). Trajectories from academic intrinsic motivation to need for cognition and educational attainment. *The Journal of Educational Research*, 110(6), 642–652. <https://doi.org/10.1080/00220671.2016.1171199>
- Goy, M., Valtin, R. & Hußmann, A. (2017). Leseselbstkonzept, Lesemotivation, Leseverhalten und Lesekompetenz. In A. Hussmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 143–175). Waxmann. <https://doi.org/10.25656/01:20755>
- Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B. & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(2), 268–292. <https://doi.org/10.1080/10705511.2012.659627>

- Gruber, H., Prenzel, M. & Schiefele, U. (2014). Spielräume für Veränderung durch Erziehung. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (6. Aufl., S. 115–137). Beltz.
- Guthrie, J. T. & Wigfield, A. (1999). How motivation fits into a science of reading. *Scientific Studies of Reading*, 3(3), 199–205.
https://doi.org/10.1207/s1532799xssr0303_1
- Harks, B., Rakoczy, K., Hattie, J., Besser, M. & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology*, 34(3), 269–290.
<https://doi.org/10.1080/01443410.2013.785384>
- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analysis relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
- Hattie, J. & Clarke, S. (2019). *Visible learning: Feedback*. Routledge.
<https://doi.org/10.4324/9780429485480>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hattie, J. & Zierer, K. (2018). *Kenne deinen Einfluss! „Visible Learning“ für die Unterrichtspraxis* (3. Aufl.). Schneider.
- Hattie, J. & Zierer, K. (2019). *Visible learning insights*. Routledge.
<https://doi.org/10.4324/9781351002226>
- Havnes, A., Smith, K., Dysthe, O. & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21–27. <https://doi.org/10.1016/j.stueduc.2012.04.001>
- Hayenga, A. O. & Henderlong Corpus, J. (2010). Profiles of intrinsic and extrinsic motivations: A person-centered approach to motivation and achievement in middle school. *Motivation and Emotion*, 34, 371–383.
<https://doi.org/10.1007/s11031-010-9181-x>
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerverprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (6. Aufl.). Kallmeyer.
- Henderlong Corpus, J., McClintic-Gilbert, M. S. & Hayenga, A. O. (2009). Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology*, 34, 154–166. <https://doi.org/10.1016/j.cedpsych.2009.01.001>
- Henderlong Corpus, J. & Wormington, S. V. (2014). Profiles of intrinsic and extrinsic motivations in elementary school: A longitudinal analysis. *The Journal of Experimental Education*, 82(4), 480–501.
<https://doi.org/10.1080/00220973.2013.876225>

- Hess, M. & Lipowsky, F. (2019). Zur Handlungskompetenz von Lehramtsstudierenden beim Erteilen von Feedback - Effekte der Strukturiertheit bei der Analyse eigener Videoaufnahmen. In M.-C. Vierbuchen & F. Bartels (Hrsg.), *Feedback in der Unterrichtspraxis: Schülerinnen und Schüler beim Lernen wirksam unterstützen* (S. 143–160). Verlag W. Kohlhammer.
- Hondrich, A. L., Decristan, J., Hertel, S. & Klieme, E. (2018). Formative assessment and intrinsic motivation: The mediating role of perceived competence. *Zeitschrift für Erziehungswissenschaft*, 21(4), 717–734. <https://doi.org/10.1007/s11618-018-0833-z>
- Howe, C. & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43(3), 325–356. <https://doi.org/10.1080/0305764X.2013.786024>
- Hoya, F. (2019). *Feedback aus der Sicht von Kindern und Lehrkräften: Die Relevanz der Erteilung und Wahrnehmung im Leseunterricht der Grundschule*. Springer VS. <https://doi.org/10.1007/978-3-658-23129-3>
- Hoya, F. (2021). Unterschiede in der Wahrnehmung positiven und negativen Feedbacks von Mädchen und Jungen im Leseunterricht der Grundschule. *Unterrichtswissenschaft*, 49(3), 423–441. <https://doi.org/10.1007/s42010-021-00102-1>
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huth, K. (2004). *Entwicklung und Evaluation von feblerspezifischem informativem tutoriellem Feedback (ITF) für die schriftliche Subtraktion*. Technische Universität Dresden. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:swb:14-1105354057406-47158> [Zugriff am 16.03.2023].
- Igler, J., Ohle-Peters, A. & McElvany, N. (2019). Mit den Augen eines Grundschulkindes: Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität. *Zeitschrift für Pädagogische Psychologie*, 33(3-4), 191–205. <https://doi.org/10.1024/1010-0652/a000243>
- Ilardi, B., Leone, D., Kasser, T. & Ryan, R. M. (1993). Employee and supervisor ratings of motivation: Main effects and discrepancies associated with job satisfaction and adjustment in a factory setting. *Journal of Applied Social Psychology*, 23, 1789–1805. <https://doi.org/10.1111/j.1559-1816.1993.tb01066.x>
- Ilgen, D. R., Fisher, C. D. & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371. <https://doi.org/10.1037/0021-9010.64.4.349>
- Irvine, J. J. (1986). Teacher–student interactions: Effects of student race, sex, and grade level. *Journal of Educational Psychology*, 78(1), 14–21. <https://doi.org/10.1037/0022-0663.78.1.14>

- Jacobs, B. (2002). *Aufgaben stellen und Feedback geben*. Verfügbar unter <https://doi.org/10.23668/psycharchives.8555> [Zugriff am 16.03.2023].
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S. & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527. <https://doi.org/10.1111/1467-8624.00421>
- Jansen, N. C., Decristan, J. & Fauth, B. (2022). Individuelle Nutzung unterrichtlicher Angebote – Zur Bedeutung von Lernvoraussetzungen und Unterrichtsbeteiligung. *Unterrichtswissenschaft*, 50(2), 157–183. <https://doi.org/10.1007/s42010-021-00141-8>
- Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, 8, 325–352.
- Jurik, V., Gröschner, A. & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, 23, 33–42. <https://doi.org/10.1016/j.learninstruc.2012.09.002>
- Jurik, V., Gröschner, A. & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences*, 32, 132–139. <https://doi.org/10.1016/j.lindif.2014.01.005>
- Jurik, V., Häusler, J., Stubben, S. & Seidel, T. (2015). Interaction: Erste Ergebnisse einer vergleichenden Videostudie im Deutsch- und Mathematikunterricht. *Zeitschrift für Pädagogik*, 61(5), 692–711. <https://doi.org/10.25656/01:15422>
- Karlen, Y. (2016). Perceived learning environments and metacognitive strategy knowledge at the upper secondary school level. *Journal for Educational Research Online*, 8, 212–232. <https://doi.org/10.25656/01:12442>
- Kiemer, K., Gröschner, A., Pehmer, A.-K. & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and Instruction*, 35, 94–103. <https://doi.org/10.1016/j.learninstruc.2014.10.003>
- King, P. E., Schrod, P. & Weisel, J. J. (2009). The instructional feedback orientation scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. *Communication Education*, 58(2), 235–261. <https://doi.org/10.1080/03634520802515705>
- Kleickmann, T., Praetorius, A.-K. & Riecke-Baulecke, T. (2019). Beurteilung von Unterrichtsqualität. In F. Zimmermann, J. Möller & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung: Schulische Diagnostik und Leistungsbeurteilung* (S. 207–223). Klett/Kallmeyer.
- Kleinke, K., Schlüter, E. & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus: Eine praktische Einführung* (2. Aufl.). De Gruyter. <https://doi.org/10.1515/9783486989458>

- Klieme, E. (2019). Unterrichtsqualität. In M. Haring, C. Rohlf's & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (S. 393–408). Waxmann.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D. & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht: Projekt Co2CA. *Zeitschrift für Pädagogik* (56. Beiheft), 64–74. <https://doi.org/10.25656/01:3361>
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht: Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Al-lolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik.: Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237. <https://doi.org/10.25656/01:4348>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Knickenberg, M. (2018). *Feedback und Attributionen im Grundschulunterricht: Bedeutung für Motivationen und Lesekompetenzen*. Springer VS. <https://doi.org/10.1007/978-3-658-22835-4>
- Kobarg, M. & Seidel, T. (2003). Prozessorientierte Lernbegleitung im Physikunterricht. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 151–200).
- Kobarg, M. & Seidel, T. (2007). Prozessorientierte Lernbegleitung - Videoanalysen im Physikunterricht der Sekundarstufe I. *Unterrichtswissenschaft*, 35(2), 148–168. <https://doi.org/10.25656/01:5490>
- Köller, O., Daniels, Z., Schnabel, K.-U. & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, 14(1), 26–37. <https://doi.org/10.1024//1010-0652.14.1.26>
- Kopp, B. & Mandl, H. (2014a). Aspekte der Feedbacknachricht. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 151–162). Waxmann.
- Kopp, B. & Mandl, H. (2014b). Lerntheoretische Grundlagen von Rückmeldungen. In H. Ditton (Hrsg.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 29–41). Waxmann.

- Krammer, K., Reusser, K. & Pauli, C. (2010). Individuelle Unterstützung der Schülerinnen und Schüler durch die Lehrperson während der Schülerarbeitsphasen. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 107–122). Waxmann.
- Krapp, A., Geyer, C. & Lewalter, D. (2014). Motivation und Emotion. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (6. Aufl., S. 193–219). Beltz.
- Krause, U.-M. (2007). *Feedback und kooperatives Lernen*. Waxmann.
- Krause, U.-M., Stark, R. & Mandl, H. (2004). Förderung des computerbasierten Wissenserwerbs durch kooperatives Lernen und eine Feedbackmaßnahme. *Zeitschrift für Pädagogische Psychologie*, 18(2), 125–136.
<https://doi.org/10.1024/1010-0652.18.2.125>
- Kühnel, S. M. (1993). Lassen sich ordinale Daten mit linearen Strukturgleichungsmodellen analysieren? *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 33, 29–51.
- Kulhavy, R. W. & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279–308.
<https://doi.org/10.1007/BF01320096>
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Waxmann.
- Kunter, M., Baumert, J. & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494–509. <https://doi.org/10.1016/j.learninstruc.2007.09.002>
- Kunter, M. & Trautwein, U. (2018). *Psychologie des Unterrichts*. Schöningh.
- Kyaruzi, F., Strijbos, J.-W., Ufer, S. & Brown, G. T. L. (2019). Students' formative assessment perceptions, feedback use and mathematics performance in secondary schools in Tanzania. *Assessment in Education: Principles, Policy & Practice*, 26(3), 278–302. <https://doi.org/10.1080/0969594X.2019.1593103>
- Labuhn, A. S., Zimmerman, B. J. & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173–194.
<https://doi.org/10.1007/s11409-010-9056-2>
- Lazarides, R., Dicke, A.-L., Rubach, C. & Eccles, J. S. (2020). Profiles of motivational beliefs in math: Exploring their development, relations to student-perceived classroom characteristics, and impact on future career aspirations and choices. *Journal of Educational Psychology*, 112(1), 70–92.
<https://doi.org/10.1037/edu0000368>
- Lazarides, R., Dicke, A.-L., Rubach, C., Oppermann, E. & Eccles, J. S. (2021). Motivational profiles across domains and academic choices within Eccles et al.'s situated expectancy-value theoretical framework. *Developmental Psychology*, 57(11), 1893–1909. <https://doi.org/10.1037/dev0001250>

- Lazarides, R., Fauth, B., Gaspard, H. & Göllner, R. (2021). Teacher self-efficacy and enthusiasm: Relations to changes in student-perceived teaching quality at the beginning of secondary education. *Learning and Instruction, 73*, 101435. <https://doi.org/10.1016/j.learninstruc.2020.101435>
- Lazarides, R., Gaspard, H. & Dicke, A.-L. (2019). Dynamics of classroom motivation: Teacher enthusiasm and the development of math interest and teacher support. *Learning and Instruction, 60*, 126–137. <https://doi.org/10.1016/j.learninstruc.2018.01.012>
- Lazarides, R., Ittel, A. & Juang, L. (2015). Wahrgenommene Unterrichtsgestaltung und Interesse im Fach Mathematik von Schülerinnen und Schülern. *Unterrichtswissenschaft, 43*(1), 67–82.
- Ledergerber, C. (2015). *Unterrichtskommunikation und motivational-emotionale Aspekte des Lernens: Eine videobasierte Analyse im Mathematikunterricht*. Waxmann.
- Lepper, C., Stang, J & McElvany, N. (2022). Bedeutung der wahrgenommenen Motivierungsqualität für intrinsische Motivation und Selbstkonzept von Grundschulkindern. *Unterrichtswissenschaft, 50*, 125–147. <https://doi.org/10.1007/s42010-021-00121-y>
- Lepper, M. R., Henderlong Corpus, J. & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184–196. <https://doi.org/10.1037/0022-0663.97.2.184>
- Levesque, C., Zuehlke, A. N., Stanek, L. R. & Ryan, R. M. (2004). Autonomy and competence in german and american university students: A comparative study based on self-determination theory. *Journal of Educational Psychology, 96*(1), 68–84. <https://doi.org/10.1037/0022-0663.96.1.68>
- Lipnevich, A. A., Berg, D. A. G. & Smith, J. K. (2016). Toward a model of student response to feedback. In G. Brown & L. Harris (Eds.), *Educational psychology handbook series. Handbook of human and social conditions in assessment* (pp. 169–185). Routledge.
- Lipnevich, A. A., Gjicali, K., Asil, M. & Smith, J. K. (2021). Development of a measure of receptivity to instructional feedback and examination of its links to personality. *Personality and Individual Differences, 169*, 110086. <https://doi.org/10.1016/j.paid.2020.110086>
- Lipnevich, A. A. & Lopera-Oquendo, C. (2022). Receptivity to instructional feedback: A validation study in the secondary school context in Singapore. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000733>
- Lipnevich, A. A. & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*, 720195. <https://doi.org/10.3389/educ.2021.720195>

- Lipnevich, A. A. & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319–333. <https://doi.org/10.1037/a0017841>
- Lipnevich, A. A. & Smith, J. K. (Eds.). (2018). *The Cambridge handbook of instructional feedback*. Cambridge University Press. <https://doi.org/10.1017/9781316832134>
- Lipnevich, A. A. & Smith, J. K. (2022). Student – feedback interaction model: Revised. *Studies in Educational Evaluation*, 75, 101208. <https://doi.org/10.1016/j.stueduc.2022.101208>
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (S. 47–70). Beltz. <https://doi.org/10.25656/01:7370>
- Lipowsky, F. (2020). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 69–118). Springer. https://doi.org/10.1007/978-3-662-61403-7_4
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1
- Little, T. D., Rhemtulla, M., Gibson, K. & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, 18(3), 285–300. <https://doi.org/10.1037/a0033266>
- Little, T. D., Slegers, D. W. & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3
- Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule: Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Springer VS. <https://doi.org/10.1007/978-3-658-10436-8>
- Lotz, M. & Lipowsky, F. (2015). Die Hattie-Studie und ihre Bedeutung für den Unterricht: Ein Blick auf ausgewählte Aspekte der Lehrer-Schüler-Interaktion. In G. Mehlhorn, F. Schulz & K. Schöppe (Hrsg.), *Begabungen entwickeln & Kreativität fördern* (S. 97–136). kopaed.
- Lüders, M. (2003). *Unterricht als Sprachspiel. Eine systematische und empirische Studie zum Unterrichtsbegriff und zur Unterrichtssprache*. Klinkhardt.
- Lüdtke, O. & Robitzsch, A. (2011). *Umgang mit fehlenden Daten in der empirischen Bildungsforschung*. *Enzyklopädie Erziehungswissenschaft Online*. Fachgebiet Methoden der empirischen erziehungswissenschaftlichen Forschung. Juventa.

- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, 34(2), 120–131. <https://doi.org/10.1016/j.cedpsych.2008.12.001>
- Lüdtke, O., Trautwein, U., Kunter, M. & Baumert, J. (2006). Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 85–96. <https://doi.org/10.1024/1010-0652.20.12.85>
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis. Multivariate applications series*. Routledge. <https://doi.org/10.4324/9780203809556>
- MacKinnon, D. P., Lockwood, C. M. & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99. https://doi.org/10.1207/s15327906mbr3901_4
- Mang, J., Seidl, L., Schiepe-Tiska, A., Tupac-Yupanqui, A., Ziernwald, L., Doroganova, A., Weis, M., Diedrich, J., Heine, J.-H., González Rodríguez, E. & Reiss, K. (2021). *PISA 2018 Skalenhandbuch: Dokumentation der Erhebungsinstrumente*. Waxmann. <https://doi.org/10.31244/9783830994961>
- Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). *PISA 2015 Skalenhandbuch: Dokumentation der Erhebungsinstrumente*. Waxmann. <https://doi.org/10.25656/01:17449>
- Marcoulides, G. A., Gottfried, A. E., Gottfried, A. W. & Oliver, P. H. (2008). A latent transition analysis of academic intrinsic motivation from childhood through adolescence. *Educational Research and Evaluation*, 14(5), 411–427. <https://doi.org/10.1080/13803610802337665>
- Marsh, H. W., Morin, A. J. S., Parker, P. D. & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Maulana, R., Opdenakker, M.-C. & Bosker, R. (2016). Teachers' instructional behaviors as important predictors of academic motivation: Changes and links across the school year. *Learning and Individual Differences*, 50, 147–156. <https://doi.org/10.1016/j.lindif.2016.07.019>

- Maulana, R., Opdenakker, M.-C., Stroet, K. & Bosker, R. (2013). Changes in teachers' involvement versus rejection and links with academic motivation during the first year of secondary education: A multilevel growth curve analysis. *Journal of Youth and Adolescence*, 42(9), 1348–1371.
<https://doi.org/10.1007/s10964-013-9921-9>
- McElvany, N., Kortenbruck, M. & Becker, M. (2008). Lesekompetenz und Lesemotivation: Entwicklung und Mediation des Zusammenhangs durch Leseverhalten. *Zeitschrift für Pädagogische Psychologie*, 22(3-4), 207–219.
<https://doi.org/10.1024/1010-0652.22.34.207>
- Meade, A. W., Johnson, E. C. & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568–592. <https://doi.org/10.1037/0021-9010.93.3.568>
- Mehan, H. (1981). Learning lessons: Social organization in the classroom. *Language in Society*, 10(2), 274–278. <https://doi.org/10.1017/S0047404500008708>
- Mercer, N. & Dawes, L. (2014). The study of talk between teachers and students, from the 1970s until the 2010s. *Oxford Review of Education*, 40(4), 430–445.
<https://doi.org/10.1080/03054985.2014.934087>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Möller, J. & Schiefele, U. (2004). Motivationale Grundlagen der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz: Vertiefende Analysen im Rahmen von PISA 2000* (S. 101–124). VS Verlag für Sozialwissenschaften.
https://doi.org/10.1007/978-3-322-81031-1_5
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Lawrence Erlbaum Associates Publishers.
- Mouratidis, A., Lens, W. & Vansteenkiste, M. (2010). How you provide corrective feedback makes a difference: The motivating role of communicating in an autonomy-supporting way. *Journal of Sport & Exercise Psychology*, 32(5), 619–637.
<https://doi.org/10.1123/jsep.32.5.619>
- Mouratidis, A., Vansteenkiste, M., Michou, A. & Lens, W. (2013). Perceived structure and achievement goals as predictors of students' self-regulated learning and affect and the mediating role of competence need satisfaction. *Learning and Individual Differences*, 23, 179–186. <https://doi.org/10.1016/j.lindif.2012.09.001>
- Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267–316.
<https://doi.org/10.2307/271070>

- Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171–189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, L. K. & Muthén, B. O. (1998-2019). *Mplus User's Guide* (8th ed.). Muthén & Muthén.
- Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J. & Eccles, J. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 115–143). American Psychological Association. <https://doi.org/10.1037/11706-004>
- Nagy, G., Trautwein, U., Baumert, J., Köller, O. & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, 12(4), 323–345. <https://doi.org/10.1080/13803610600765687>
- Narciss, S. (2006). *Informatives tutorielles Feedback: Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse. Pädagogische Psychologie und Entwicklungspsychologie*. Waxmann.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, J. van Merriënboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125–144). Erlbaum.
- Narciss, S. (2014). Modelle zu den Bedingungen und Wirkungen von Feedback in Lehr-Lernsituationen. In H. Dittton & A. Müller (Hrsg.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 43–82). Waxmann.
- Narciss, S. & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, 16(4), 310–322. <https://doi.org/10.1016/j.learninstruc.2006.07.003>
- Nicol, D. J. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. <https://doi.org/10.1080/02602931003786559>
- Nicol, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Ntoumanis, N., Barkoukis, V. & Thøgersen-Ntoumani, C. (2009). Developmental trajectories of motivation in physical education: Course, demographic differences, and antecedents. *Journal of Educational Psychology*, 101(3), 717–728. <https://doi.org/10.1037/a0014696>

- Orsmond, P., Maw, S. J., Park, J. R., Gomez, S. & Crook, A. C. (2013). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education*, 38(2), 240–252. <https://doi.org/10.1080/02602938.2011.625472>
- Otis, N., Grouzet, F. M. & Pelletier, L. G. (2005). Latent Motivational Change in an Academic Setting: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, 97(2), 170–183. <https://doi.org/10.1037/0022-0663.97.2.170>
- Panadero, E. & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, 35, 100416. <https://doi.org/10.1016/j.edurev.2021.100416>
- Patall, E. A., Steingut, R. R., Vasquez, A. C., Trimble, S. S., Pituch, K. A. & Freeman, J. L. (2018). Daily autonomy supporting or thwarting and students' motivation and engagement in the high school science classroom. *Journal of Educational Psychology*, 110(2), 269–288. <https://doi.org/10.1037/edu0000214>
- Pat-El, R., Tillema, H. & van Koppen, S. W. (2012). Effects of formative feedback on intrinsic motivation: Examining ethnic differences. *Learning and Individual Differences*, 22(4), 449–454. <https://doi.org/10.1016/j.lindif.2012.04.001>
- Pauli, C. (2006). Klassengespräch. In I. Hugener, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Teil 3: *Videoanalysen* (S. 124–147). GPPF.
- Pauli, C. (2010). Klassengespräche - Einführung des Denkens oder gemeinsame Wissenskonstruktion selbstbestimmt lernender Schülerinnen und Schüler? In T. Bohl, K. Kansteiner-Schänzlin, M. Kleinknecht, B. Kohler & A. Nold (Hrsg.), *Selbstbestimmung und Classroom-Management: Empirische Befunde und Entwicklungsstrategien zum guten Unterricht* (S. 145–161). Klinkhardt.
- Pauli, C. & Reusser, K. (2018). Unterrichtsgespräche führen – das Transversale und das Fachliche einer didaktischen Kernkompetenz. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 36(3), 365–377. <https://doi.org/10.36950/bzl.36.2018.9424>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879e891. <https://doi.org/10.3758/BRM.40.3.879>
- Pehmer, A.-K., Gröschner, A. & Seidel, T. (2015). Fostering and scaffolding student engagement in productive classroom discourse: Teachers' practice changes and reflections in light of teacher professional development. *Learning, Culture and Social Interaction*, 7, 12–27. <https://doi.org/10.1016/j.lcsi.2015.05.001>
- Pianta, R. C., La Paro, K. & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Paul H. Brooks.

- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
<https://doi.org/10.3102/0013189X09332374>
- Pielmeier, M., Huber, S. & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education*, 76, 255–266.
<https://doi.org/10.1016/j.tate.2018.01.002>
- Pintrich, P. R. & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037/0022-0663.82.1.33>
- Pohlmann-Rother, S., Kürzinger, A. & Lipowsky, F. (2020). Feedback im Anfangsunterricht der Grundschule – Eine Videostudie zum Feedbackverhalten von Lehrpersonen in der Domäne Schreiben. *Zeitschrift für Erziehungswissenschaft*, 23(3), 591–611. <https://doi.org/10.1007/s11618-020-00950-0>
- Praetorius, A.-K. & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? *Unterrichtswissenschaft*, 49(2), 167–188.
<https://doi.org/10.1007/s42010-021-00119-6>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education*, 50(3), 407–426.
<https://doi.org/10.1007/s11858-018-0918-4>
- Praetorius, A.-K., Lauermaun, F., Klassen, R. M., Dickhäuser, O., Janke, S. & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education*, 65, 241–254. <https://doi.org/10.1016/j.tate.2017.03.023>
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12. <https://doi.org/10.1016/j.learninstruc.2013.12.002>
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht: Unterricht aus der Perspektive von Lernenden und Beobachtern*. Waxmann.
- Rakoczy, K., Frick, U., Weiß-Wittstadt, S., Tallon, M. & Wagner, E. (2022). Einmal begeistert, immer begeistert? Eine Experience-Sampling Studie zur wahrgenommenen Unterrichtsqualität und Motivation von Schülerinnen und Schülern im Kunstunterricht. *Unterrichtswissenschaft*, 50(2), 211–236.
<https://doi.org/10.1007/s42010-022-00147-w>
- Rakoczy, K., Harks, B., Klieme, E., Blum, W. & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, 27, 63–73.
<https://doi.org/10.1016/j.learninstruc.2013.03.002>

- Rakoczy, K., Klieme, E., Bürgermeister, A. & Harks, B. (2008). The interplay between student evaluation and instruction: Grading and feedback in mathematics classrooms. *Zeitschrift für Psychologie*, 216(2), 111–124. <https://doi.org/10.1027/0044-3409.216.2.111>
- Rakoczy, K., Klieme, E. & Pauli, C. (2008). Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation. *Zeitschrift für Pädagogische Psychologie*, 22(1), 25–35. <https://doi.org/10.1024/1010-0652.22.1.25>
- Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B. & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, 60, 154–165. <https://doi.org/10.1016/j.learninstruc.2018.01.004>
- Ramaprasad, A. (1983). On the definition of feedback. *Arkalgud Behavioral Science*, 28(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S. & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented analysis. *Journal of Educational Psychology*, 99(4), 734–746. <https://doi.org/10.1037/0022-0663.99.4.734>
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications, Inc.
- Raykov, T. & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge. <https://doi.org/10.4324/9780203841624>
- Reddy, R., Rhodes, J. E. & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology*, 15(1), 119–138. <https://doi.org/10.1017/S0954579403000075>
- Reinecke, J. (2014). *Strukturgleichungsmodelle in den Sozialwissenschaften* (2. Aufl.). De Gruyter. <https://doi.org/10.1524/9783486854008>
- Reinecke, J., Kunadt, S. & Pollich, D. (2011). Längsschnitt- und Panelanalyse. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online*. Juventa.
- Reinecke, J. & Pöge, A. (2010). Strukturgleichungsmodelle. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 775–804). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92038-2_29
- Reinmann-Rothmeier, G. & Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie: Ein Lehrbuch* (5. Aufl., S. 613–658). Beltz.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>

- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 9–32). Waxmann.
- Reuter, T., Ziegelmann, J. P., Wiedemann, A. U., Geiser, C., Lippke, S., Schütz, B. & Schwarzer, R. (2010). Changes in intentions, planning, and self-efficacy predict changes in behaviors: An application of latent true change modeling. *Journal of Health Psychology, 15*(6), 935–947. <https://doi.org/10.1177/1359105309360071>
- Rhemtulla, M., Brosseau-Liard, P. É. & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. <https://doi.org/10.1037/a0029315>
- Richert, P. (2005). *Typische Sprachmuster der Lehrer-Schüler-Interaktion: Empirische Untersuchungen zur Feedbackkomponente in der unterrichtlichen Interaktion*. Klinkhardt.
- Rowe, A. & Wood, L. (2008). Student perceptions and preferences for feedback. *Asian Social Science, 4*(3), 78–88. <https://doi.org/10.5539/ass.v4n3p78>
- Rucker, D. D., Preacher, K. J., Tormala, Z. L. & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359–371. <https://doi.org/10.1111/j.1751-9004.2011.00355.x>
- Ruelmann, M., Torchetti, L., Zulliger, S., Buholzer, A. & Praetorius, A.-K. (2021). Kognitiv-motivationale Schüler*innenprofile und ihre Bedeutung für die Schüler*innenwahrnehmung der Lernunterstützung durch die Lehrperson. *Unterrichtswissenschaft, 49*(3), 395–422. <https://doi.org/10.1007/s42010-021-00100-3>
- Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y. & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction, 42*, 95–103. <https://doi.org/10.1016/j.learninstruc.2016.01.004>
- Ruzek, E. A. & Schenke, K. (2019). The tenuous link between classroom perceptions and motivation: A within-person longitudinal study. *Journal of Educational Psychology, 111*(5), 903–917. <https://doi.org/10.1037/edu0000323>
- Ryan, R. M. & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*(5), 749–761. <https://doi.org/10.1037/0022-3514.57.5.749>
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>

- Ryan, R. M. & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–33). University of Rochester Press.
- Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.
<https://doi.org/10.1521/978.14625/28806>
- Ryan, R.M., & Deci, E.L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860.
<https://doi.org/10.1016/j.cedpsych.2020.101860>
- Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514.
<https://doi.org/10.1007/BF02296192>
- Satorra, A. & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248.
<https://doi.org/10.1007/s11336-009-9135-y>
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Schenke, K., Ruzek, E., Lam, A. C., Karabenick, S. A. & Eccles, J. S. (2018). To the means and beyond: Understanding variation in students' perceptions of teacher emotional support. *Learning and Instruction*, *55*, 13–21.
<https://doi.org/10.1016/j.learninstruc.2018.02.003>
- Scherer, R., Nilsen, T. & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, *7*, 110. <https://doi.org/10.3389/fpsyg.2016.00110>
- Schermelleh-Engel, K. & Gåde, J. C. (2020). Modellbasierte Methoden der Reliabilitätsschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Lehrbuch. Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 335–368). Springer.
https://doi.org/10.1007/978-3-662-61532-4_15
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.
- Schiefele, U., Köller, O. & Schaffner, E. (2018). Intrinsische und extrinsische Motivation. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch pädagogische Psychologie* (5. Aufl., S. 309–319). Beltz.
- Schiefele, U. & Schaffner, E. (2020). Motivation. In E. Wild & J. Möller (Hrsg.), *Lehrbuch. Pädagogische Psychologie* (3. Aufl., S. 163–185). Springer.
https://doi.org/10.1007/978-3-662-61403-7_7

- Schiefele, U. & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen. Ein Überblick zu Ergebnissen der Forschung. *Zeitschrift für Pädagogische Psychologie*, 8(1), 1–13.
- Schiefele, U. (2014). Förderung von Interessen. In G. W. Lauth & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis* (2. Aufl., S. 251–261). Hogrefe.
- Schiefele, U. & Streblov, L. (2005). Intrinsische Motivation - Theorien und Befunde. In R. Vollmeyer & J. Brunstein (Hrsg.), *Motivationspsychologie und ihre Anwendung* (S. 39–58). Kohlhammer.
- Schulte, K., Nonte, S. & Schwippert, K. (2013). Die Überprüfung von Messinvarianz in international vergleichenden Schulleistungsstudien am Beispiel der Studie PIRLS. *Zeitschrift für Bildungsforschung*, 3(2), 99–118.
<https://doi.org/10.1007/s35834-013-0062-8>
- Schwabe, F., Schlitter, T., Iglar, J., Ohle-Peters, A., Teerling, A., Köller, O. & McElvany, N. (2021). Lesemotivation, Leseselbstkonzept und Leseverhalten am Ende der Grundschulzeit – Wirksamkeit und differenzielle Effekte der schulischen Teilnahme an einer bundesweiten Förderinitiative. *Zeitschrift für Pädagogische Psychologie*, 35(1), 53–70. <https://doi.org/10.1024/1010-0652/a000262>
- Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research*, 9(3), 253–271.
<https://doi.org/10.1007/s10984-006-9012-x>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik*, 60(6), 850–866. <https://doi.org/10.25656/01:14686>
- Seidel, T., Jurik, V., Häusler, J. & Stubben, S. (2016). Mikro-Umwelten im Klassenverband: Wie sich kognitive und motivational-affektive Schülervoraussetzungen auf die Wahrnehmung und das Verhalten im Fachunterricht auswirken. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 65–87). Waxmann.
- Seidel, T., Prenzel, M., Wittwer, J. & Schwindt, K. (2007). Unterricht in den Naturwissenschaften. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 147–180). Waxmann.
- Seidel, T., Rimmel, R. & Prenzel, M. (2003). Gelegenheitsstrukturen beim Klassengespräch und ihre Bedeutung für die Lernmotivation. Videoanalysen in Kombination mit Schülerselbsteinschätzungen. *Unterrichtswissenschaft*, 31(2), 142–165. <https://doi.org/10.25656/01:6776>

- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
<https://doi.org/10.3102/0034654307310317>
- Seidel, T. & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaft*, 20(Supplement 1), 1–21.
<https://doi.org/10.1007/s11618-017-0726-6>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Sortkær, B. (2018). Feedback for everybody? Variations in students' perception of feedback. In D. Reimer, B. Sortkær, M. Oskarsson, T. Nilsen, M. Rasmusson & K. Nissinen (Eds.), *Northern Lights on TIMSS and PISA 2018* (pp. 161–182). Nordic Council of Ministers.
- Sortkær, B. (2019). Feedback for everybody? Exploring the relationship between students' perceptions of feedback and students' socioeconomic status. *British Educational Research Journal*, 45, 717–735. <https://doi.org/10.1002/berj.3522>
- Spieß, M. (2010). Der Umgang mit fehlenden Werten. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 117–142). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92038-2_6
- Spinath, B. & Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, 15(2), 87–102. <https://doi.org/10.1016/j.learninstruc.2005.04.008>
- Spinath, B. & Steinmayr, R. (2008). Longitudinal analysis of intrinsic motivation and competence beliefs: Is there a relation over time? *Child Development*, 79(5), 1555–1569. <https://doi.org/10.1111/j.1467-8624.2008.01205.x>
- Spinath, B. & Steinmayr, R. (2012). The roles of competence beliefs and goal orientations for change in intrinsic motivation. *Journal of Educational Psychology*, 104(4), 1135–1148. <https://doi.org/10.1037/a0028115>
- Stahns, R., Rieser, S. & Hußmann, A. (2020). Können Viertklässlerinnen und Viertklässler Unterrichtsqualität valide einschätzen? Ergebnisse zum Fach Deutsch. *Unterrichtswissenschaft*, 48(4), 663–682.
<https://doi.org/10.1007/s42010-020-00084-6>
- Stang, J. & McElvany, N. (2020). Unterschiede in der Wahrnehmung der Qualität des Deutschunterrichts zwischen Grundschülerinnen und Grundschulern. *Zeitschrift für Pädagogische Psychologie*, 35(4), 261–273.
<https://doi.org/10.1024/1010-0652/a000275>
- Steelman, L. A., Levy, P. E. & Snell, A. F. (2004). The Feedback Environment Scale: Construct Definition, Measurement, and Validation. *Educational and Psychological Measurement*, 64(1), 165–184.
<https://doi.org/10.1177/0013164403258440>

- Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–90. <https://doi.org/10.1086/209528>
- Steyer, R., Eid, M. & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research*, 2(1), 21–33.
- Steyer, R., Partchev, I. & Shanahan, M. J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little, K. U. Schnabel & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 109–126). Routledge.
- Steyer, R., Wolf, A., Funke, F. & Partchev, I. (2009). Strukturgleichungsmodelle. In H. Holling (Hrsg.), *Enzyklopädie der Psychologie: Grundlagen und statistische Methoden der Evaluationsforschung* (S. 393–448). Hogrefe.
- Strijbos, J. W. & Müller, A. (2014). Personale Faktoren im Feedbackprozess. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 83–134). Waxmann.
- Strijbos, J. W., Narciss, S. & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291–303. <https://doi.org/10.1016/j.learninstruc.2009.08.008>
- Strijbos, J. W., Pat-El, R. & Narciss, S. (2021). Structural validity and invariance of the feedback perceptions questionnaire. *Studies in Educational Evaluation*, 68, 100980. <https://doi.org/10.1016/j.stueduc.2021.100980>
- Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S. & Koestner, R. (2014). A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemporary Educational Psychology*, 39(4), 342–358. <https://doi.org/10.1016/j.cedpsych.2014.08.002>
- Thomas, A. E. & Müller, F. H. (2011). Skalen zur motivationalen Regulation beim Lernen von Schülerinnen und Schülern: Skalen zur akademischen Selbstregulation von Schüler/innen SRQ-A [G] (überarbeitete Fassung). *Wissenschaftliche Beiträge aus dem Institut für Unterrichts- und Schulentwicklung Nr. 5*. Alpen-Adria-Universität.
- Thommen, D., Sieber, V., Grob, U. & Praetorius, A.-K. (2021). Teachers' motivational profiles and their longitudinal associations with teaching quality. *Learning and Instruction*, 76, 101514. <https://doi.org/10.1016/j.learninstruc.2021.101514>
- Thurlings, M., Vermeulen, M., Bastiaens, T. & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review*, 9, 1–15. <https://doi.org/10.1016/j.edurev.2012.11.004>

- Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O. & Maaz, K. (Hrsg.). (2010). *Schulleistungen von Abiturienten: Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand*. Waxmann. <https://doi.org/10.1007/978-3-531-92037-5>
- Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U. & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, 100(2), 460–472. <https://doi.org/10.1037/0022-0663.100.2.460>
- Urban, D. & Mayerl, J. (2007). *Mediator-Effekte in der Regressionsanalyse (direkte, indirekte und totale Effekte)*. Verfügbar unter https://www.sowi.uni-stuttgart.de/abteilungen/ses/forschung/regression/Mediator-Effekte_v1-3.pdf [Zugriff am 05.04.2023].
- Urban, D. & Mayerl, J. (2014). *Strukturgleichungsmodellierung. Ein Ratgeber für die Praxis*. Springer VS. <https://doi.org/10.1007/978-3-658-01919-8>
- Vallerand, R. J., Fortier, M. S. & Guay, F. (1997). Self-determination and persistence in a real-life setting: Toward a motivational model of high school dropout. *Journal of Personality and Social Psychology*, 72(5), 1161–1176. <https://doi.org/10.1037/0022-3514.72.5.1161>
- Vallerand, R. J. & Reid, G. (1984). On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport Psychology*, 6(1), 94–102. <https://doi.org/10.1123/jsp.6.1.94>
- van der Kleij, F. M. & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, 33(2), 345–373. <https://doi.org/10.1007/s11092-020-09331-x>
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K. & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, 101(3), 671–688. <https://doi.org/10.1037/a0015083>
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: ein kritischer Vergleich verschiedener Modellvarianten. In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundlagen und quantitative Modellierungen* (S. 63–80). Beltz Juventa. <https://doi.org/10.25656/01:25864>
- Voerman, L., Meijer, P. C., Korthagen, F. A. & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education*, 28(8), 1107–1115. <https://doi.org/10.1016/j.tate.2012.06.006>
- Vollmeyer, R. & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, 10(4), 293–309. [https://doi.org/10.1016/S0959-4752\(99\)00031-6](https://doi.org/10.1016/S0959-4752(99)00031-6)

- Vollmeyer, R. & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction, 15*(6), 589–602. <https://doi.org/10.1016/j.learninstruc.2005.08.001>
- Vries, J. de, Feskens, R., Keuning, J. & van der Kleij, F. (2022). Comparability of Feedback in PISA 2015 across Culturally Diverse Countries. *Education Sciences, 12*(2), 145. <https://doi.org/10.3390/educsci12020145>
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht - am Beispiel der Studie DESI (Deutsch Englisch Schülerleistungen International) der Kultusministerkonferenz*. Universität Koblenz-Landau. Verfügbar unter <https://kola.opus.hbz-nrw.de/frontdoor/index/index/docId/194> [Zugriff am 16.03.2023].
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U. & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction, 28*, 1–11. <https://doi.org/10.1016/j.learninstruc.2013.03.003>
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B. & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108*(5), 705–721. <https://doi.org/10.1037/edu0000075>
- Waldis, M., Grob, U., Pauli, C. & Reusser, K. (2010). Der schweizerische Mathematikunterricht aus der Sicht von Schülerinnen und Schülern und in der Perspektive hochinferenter Beobachterurteile. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität: Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 171–208). Waxmann.
- Walker, C. O., Greene, B. A. & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences, 16*, 1–12. <https://doi.org/10.1016/j.lindif.2005.06.004>
- Watermann, R. & Baumert, J. (2006). Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (S. 61–94). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90082-7_3
- Way, N., Reddy, R. & Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological and behavioral adjustment. *American Journal of Community Psychology, 40*(3-4), 194–213. <https://doi.org/10.1007/s10464-007-9143-y>

- Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Springer Gabler. <https://doi.org/10.1007/978-3-642-35012-2>
- Weidinger, A. F., Spinath, B. & Steinmayr, R. (2015). Zur Bedeutung von Grundschulnoten für die Veränderung von Intrinsischer Motivation und Fähigkeits-selbstkonzept in Deutsch. *Zeitschrift für Pädagogische Psychologie*, 29(3-4), 193–204. <https://doi.org/10.1024/1010-0652/a000160>
- Weidinger, A. F., Spinath, B. & Steinmayr, R. (2016). Why does intrinsic motivation decline following negative feedback? The mediating role of ability self-concept and its moderation by goal orientations. *Learning and Individual Differences*, 47, 117–128. <https://doi.org/10.1016/j.lindif.2016.01.003>
- West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Eds.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Sage Publications, Inc.
- West, S. G., Taylor, A. B. & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Eds.), *Handbook of structural equation modeling* (pp. 209–231). Guilford Press.
- Wigfield, A. & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, 30(1), 1–35. <https://doi.org/10.1016/j.dr.2009.12.001>
- Willems, A. S. (2011). *Bedingungen des situationalen Interesses im Mathematikunterricht - Eine mehrebenenanalytische Perspektive*. Waxmann.
- Willems, A. S. (2018). Lernmotivation und Interesse. In K. Sommer, J. Wambach-Laicher & P. Pfeifer (Hrsg.), *Konkrete Fachdidaktik Chemie: Grundlagen für das Lernen und Lehren im Chemieunterricht* (S. 195–205). Friedrich Aulis.
- Willems, A. S. (2022a). Individuelle Schüler*innenprofile des situationalen und dispositionalen Interesses und ihre Bedeutung für die Wahrnehmung der Unterrichtsqualität im Fach Mathematik. *Zeitschrift für Erziehungswissenschaft*, 25(2), 377–404. <https://doi.org/10.1007/s11618-022-01094-z>
- Willems, A. S. (2022b). Wie beeinflusst die Wahrnehmung der Unterrichtsqualität das situationale und individuelle Interesse von Schülerinnen und Schülern im Mathematikunterricht? *mathematica didactica*, 45, 1–24. <https://doi.org/10.18716/ojs/md/2022.1594>
- Willems, A. S. & Dreiling, K. (2022). Erklären individuelle Motivationsprofile von Schülerinnen und Schülern Unterschiede in ihrer Feedbackwahrnehmung im Deutschunterricht der gymnasialen Oberstufe? *Journal for Educational Research Online (JERO)*, 14(2).

- Willems, A. S., Dreiling, K. & Benning, A. (im Druck). Wie beeinflussen individuelle Motivationsprofile von Oberstufenschülerinnen und -schülern ihre Wahrnehmung der Unterrichtsqualität im Fach Deutsch? In A. S. Willems (Hrsg.), „*Wieso? Weshalb? Warum?*“: *Motivation und Einstellung in (außer-)schulischen Bildungskontexten*. Universitätsverlag Göttingen.
- Willems, A. S., Dreiling, K. & Eckert, M. (2020). *Skalendokumentation des Projekts FeeHe – Feedback im Kontext von Heterogenität*. Göttinger Beiträge zur erziehungswissenschaftlichen Forschung. <https://doi.org/10.17875/gup2020-1254>
- Williams, G. C. & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, 70(4), 767–779. <https://doi.org/10.1037/0022-3514.70.4.767>
- Winstone, N. E., Nash, R. A., Parker, M. & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>
- Wisniewski, B. (2019). Feedback digital: Möglichkeiten für eine Verbesserung der Lehrer-Schüler-Beziehung. *Friedrich Jahresheft*, 37, 58–61.
- Wisniewski, B., Zierer, K. & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Wollenschläger, M., Möller, J. & Harms, U. (2011). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken. *Zeitschrift für Pädagogische Psychologie*, 25(3), 197–202. <https://doi.org/10.1024/1010-0652/a000040>
- Wormington, S. V., Henderlong Corpus, J. & Anderson, K. G. (2012). A person-centered investigation of academic motivation and its correlates in high school. *Learning and Individual Differences*, 22, 429–438. <https://doi.org/10.1016/j.lindif.2012.03.004>
- Zhao, X., Lynch, J. G. & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206. <https://doi.org/10.1086/651257>
- Zierer, K., Wisniewski, B., Schatz, C., Weckend, D. & Helmke, A. (2019). Wie kann Feedback der Lernenden die Unterrichtsqualität verbessern? *Journal für LehrerInnenbildung*, 19(1), 26–40. https://doi.org/10.35468/10.35468/jlb-01-2019_02

9.2 Abbildungsverzeichnis

Abbildung 1: Ebenen des Feedbacks (nach Hattie und Timperley, 2007; vereinfachte Darstellung)	23
Abbildung 2: Dimensionen der Feedbackqualität (eigene Darstellung)	25
Abbildung 3: Interaktionales Rahmenmodell der Gestaltung und Verarbeitung von Feedback (nach Strijbos & Müller, 2014; modifiziert durch die Autorin)	34
Abbildung 4: Modell der Interaktion zwischen Feedback und Lernenden (nach Lipnevich et al., 2016 und Lipnevich & Smith, 2022; übersetzt und modifiziert durch die Autorin)	35
Abbildung 5: Vereinfachtes Angebots-Nutzungs-Modell (in Anlehnung an Kleickmann et al., 2019, S. 209)	36
Abbildung 6: Motivationsformen, Regulationsstile und erlebte Selbstbestimmung (Buff et al., 2010, S. 257 in Anlehnung an Ryan & Deci, 2000, 2002)	52
Abbildung 7: Theoretisches Rahmenmodell der Forschungsarbeit (eigene Darstellung)	67
Abbildung 8: Verankerung der Forschungsfragen im theoretischen Rahmenmodell (eigene Darstellung)	71
Abbildung 9: Übersicht über die Datenerhebung im FeeHe-Projekt (Willems et al., 2020, S. 10)	79
Abbildung 10: Schematisches längsschnittliches (Latent-State-)Modell und Latent-True-Change-Modell für ein Merkmal A	95
Abbildung 11: Schematisches bivariates Latent-True-Change-Modell für den Zusammenhang zwischen zwei Variablen A und B.	96
Abbildung 12: Schematische Darstellung eines einfachen Mediationsmodells (A) sowie eines erweiterten Modells mit einer multiplen Mediation (B).	98
Abbildung 13: Schematische Darstellung des 1-Faktor-Modells (links), 4-Faktormodells mit korrelierten Faktoren (Mitte) und Faktormodells 2. Ordnung (rechts)	106
Abbildung 14: Schematische Darstellung des Latent-State-Modells mit Item-Parcels	114
Abbildung 15: Veränderung der wahrgenommenen Dimensionen von Feedbackqualität	122
Abbildung 16: Latent-True-Change-Modell für die Veränderung der Feedbackwahrnehmung über die Zeit	124
Abbildung 17: Eindimensionales Messmodell der intrinsischen Lernmotivation im Längsschnitt	131
Abbildung 18: Längsschnittliches Messmodell der intrinsischen Lernmotivation mit Item-Parcels	133

Abbildung 19: Veränderung der intrinsischen Lernmotivation.....	138
Abbildung 20: Latent-True-Change-Modell für die Veränderung der intrinsischen Lernmotivation.....	139
Abbildung 21: Bivariate Latent-True-Change-Modelle für den Zusammenhang zwischen Feedbackdimensionen und der intrinsischen Lernmotivation.....	144
Abbildung 22: Ein-Faktormodell der Skala Wahrgenommene Nützlichkeit von Feedback	145
Abbildung 23: Ein-Faktormodell der Skala Wahrgenommenes Kompetenzerleben	147
Abbildung 24: Strukturgleichungsmodell zur Vorhersage der Veränderung von intrinsischer Lernmotivation über die wahrgenommene Nützlichkeit und das Kompetenzerleben unter Kontrolle von Deutschnote und Geschlecht	151

9.3 Tabellenverzeichnis

Tabelle 1:	Klassifikation unterschiedlicher Feedbackkomponenten (nach Narciss, 2006, S. 23).....	22
Tabelle 2:	Deskriptive Kennwerte zur Beschreibung der Stichproben zu beiden Messzeitpunkten.....	80
Tabelle 3:	Items zur Erfassung der Feedbackqualität mit deskriptiven Kennwerten.....	82
Tabelle 4:	Items zur Erfassung der wahrgenommenen Nützlichkeit mit deskriptiven Kennwerten.....	83
Tabelle 5:	Items zur Erfassung des wahrgenommenen Kompetenzerlebens mit deskriptiven Kennwerten.....	84
Tabelle 6:	Items zur Erfassung der intrinsischen Lernmotivation mit deskriptiven Kennwerten.....	84
Tabelle 7:	Anteil fehlender Werte auf Itemebene (Item-Nonresponse).....	89
Tabelle 8:	Übersicht zentraler Modellgütekriterien (Fit-Indizes).....	102
Tabelle 9:	Ergebnisse der konfirmatorischen Faktorenanalyse zur Feedbackqualität zu Beginn und zur Mitte des Schuljahres (MZP 1 und MZP 2).....	107
Tabelle 10:	Standardisierte Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im 4-Faktormodell der Feedbackwahrnehmung zu Beginn und zur Mitte des Schuljahres (MZP 1 und 2).....	108
Tabelle 11:	Latente Interkorrelationen der Faktoren der Feedbackskalen zu Beginn (t_1) und zur Mitte (t_2) des Schuljahres.....	109
Tabelle 12:	Skalenmittelwerte (MW), Standardabweichungen (SD), Reliabilitäten (α und ω) und Intraklassenkorrelationen (ICC) der Feedbackdimensionen zu Beginn und zur Mitte des Schuljahres.....	110
Tabelle 13:	Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im Latent-State-Modell der Feedbackwahrnehmung.....	111
Tabelle 14:	Latente Faktor-Interkorrelationen der Dimensionen der Feedbackqualität im Latent-State-Modell.....	112
Tabelle 15:	Übersicht der Item-Parcels mit den jeweiligen Items und Faktorladungen (λ) zu Beginn (t_1) und zur Mitte (t_2) des Schuljahres.....	113
Tabelle 16:	Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im Latent-State-Modell mit Item-Parcels.....	115
Tabelle 17:	Latente Faktor-Interkorrelationen der Dimensionen der Feedbackqualität im Latent-State-Modell mit Item-Parcels.....	116

Tabelle 18:	Modell-Fit-Indizes des 4-Faktor-Parcel-Modells der Feedbackqualität mit unterschiedlichem Grad der Messinvarianz	117
Tabelle 19:	Ergebnisse der Modellvergleiche	117
Tabelle 20:	Ergebnisse der Prüfung auf Messinvarianz der Feedbackqualität nach Geschlecht.....	118
Tabelle 21:	Ergebnisse der Modellvergleiche	118
Tabelle 22:	Geschlechtsspezifische latente Mittelwertdifferenzen in der Feedbackwahrnehmung zu Beginn und zur Mitte des Schuljahres (MZP 1/MZP 2)	120
Tabelle 23:	Latente Faktormittelwerte (MW), -varianzen (Var) der Dimensionen der Feedbackqualität im Latent-State-Modell sowie Mittelwertdifferenzen	121
Tabelle 24:	Mittelwerte (MW) und Varianzen (Var) der latenten Faktoren im Latent-True-Change-Modell	125
Tabelle 25:	Interkorrelationen der latenten Faktoren im Latent-True-Change-Modell.....	126
Tabelle 26:	Standardisierte Regressionskoeffizienten in den Latent-True-Change-Modellen mit Geschlecht und Deutschnote als Prädiktoren der Wahrnehmungsveränderung von Feedback.....	128
Tabelle 27:	Unstandardisierte Regressionskoeffizienten in den Latent-True-Change-Modellen mit Geschlecht und Deutschnote als Prädiktoren der Wahrnehmungsveränderung von Feedback.....	128
Tabelle 28:	Ergebnisse der konfirmatorischen Faktorenanalyse zur intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres (MZP 1 und MZP 2).....	129
Tabelle 29:	Standardisierte Faktorladungen (λ) und Reliabilitäts-Schätzungen (R^2) der Indikatoren im eindimensionalen Modell der intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres (MZP 1 und 2)	130
Tabelle 30:	Skalenmittelwerte (MW), Standardabweichungen (SD), Reliabilitäten (α und ω) und Intraklassenkorrelationen (ICC) der Skala zur intrinsischen Lernmotivation zu Beginn und zur Mitte des Schuljahres	130
Tabelle 31:	Reliabilitäts-Schätzungen (R^2) der Indikatoren im längsschnittlichen Messmodell der intrinsischen Lernmotivation	131
Tabelle 32:	Übersicht der Item-Parcels mit den jeweiligen Items zu Beginn (t_1) und zur Mitte (t_2) des Schuljahres	132
Tabelle 33:	Reliabilitäts-Schätzungen (R^2) der Indikatoren der intrinsischen Lernmotivation im längsschnittlichen Messmodell mit Item-Parcels.....	133

Tabelle 34:	Modell-Fit-Indizes des eindimensionalen Parcel-Modells der intrinsischen Lernmotivation mit unterschiedlichem Grad der Messinvarianz über die Zeit.....	134
Tabelle 35:	Ergebnisse der Modellvergleiche	134
Tabelle 36:	Ergebnisse der Prüfung auf Messinvarianz der intrinsischen Lernmotivation nach Geschlecht.....	135
Tabelle 37:	Ergebnisse der Modellvergleiche	136
Tabelle 38:	Geschlechtsspezifische latente Mittelwertdifferenzen in der intrinsischen Lernmotivation zu Beginn (t_1) und zur Mitte des Schuljahres (t_2).....	137
Tabelle 39:	Latente Faktormittelwerte (MW), -varianzen (Var) der intrinsischen Lernmotivation zu beiden Messzeitpunkten (t) sowie Mittelwertdifferenzen	139
Tabelle 40:	Standardisierte und unstandardisierte Regressionskoeffizienten im Latent-True-Change-Modell mit Geschlecht und Deutschnote als Prädiktoren der Motivationsveränderung.....	140
Tabelle 41:	Latente Interkorrelationen der ergebnisorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell.....	141
Tabelle 42:	Latente Interkorrelationen der prozessorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell.....	141
Tabelle 43:	Latente Interkorrelationen der selbstregulationsorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell.....	142
Tabelle 44:	Latente Interkorrelationen der dialogorientierten Feedbackdimension und der Lernmotivation im bivariaten LTC-Modell.....	142
Tabelle 45:	Reliabilitäts-Schätzungen (R^2) der Indikatoren im eindimensionalen Modell der wahrgenommenen Nützlichkeit zur Mitte des Schuljahres (t_2).....	146
Tabelle 46:	Reliabilitäts-Schätzungen (R^2) der Indikatoren im eindimensionalen Modell des wahrgenommenen Kompetenzerlebens zur Mitte des Schuljahres (t_2).....	147
Tabelle 47:	Interkorrelationen der untersuchten Variablen	149
Tabelle 48:	Fit-Indizes des Strukturgleichungsmodells für jede Dimension der Feedbackqualität	150
Tabelle 49:	Indirekte und totale Effekte der Dimensionen der Feedbackqualität auf die Veränderung der intrinsischen Lernmotivation über die wahrgenommene Nützlichkeit und das wahrgenommene Kompetenzerleben	154

Tabelle 50:	Übersicht über Hypothesen und empirische Befunde der Studie	167
-------------	---	-----

Das Feedback von Lehrkräften ist von hoher Bedeutung für die Lernentwicklung von Schülerinnen und Schülern. Für eine optimale Unterstützung der Lernprozesse ist es wichtig, wie Schülerinnen und Schüler das Feedback wahrnehmen und nutzen. Ausgehend von theoretischen Angebots–Nutzungs–Modellen widmet sich das vorliegende Buch der Frage, inwiefern Schülerinnen und Schüler der gymnasialen Oberstufe das Feedback von Lehrkräften im Deutschunterricht unterschiedlich wahrnehmen und wie sich diese Wahrnehmung auf die Motivationsentwicklung der Schülerinnen und Schüler auswirkt. Die Befunde aus Latent–Change–Modellen geben Aufschluss über die Stabilität und Veränderung der Feedbackwahrnehmung im Laufe eines Schulhalbjahres sowie zu Unterschieden in der Feedbackwahrnehmung zwischen Schülerinnen und Schülern. Darüber hinaus wird anhand der längsschnittlichen Analysen das Zusammenspiel zwischen unterschiedlichen Dimensionen der Feedbackqualität und der Entwicklung der Lernmotivation herausgestellt. Die Befunde erweitern den Forschungsstand zur Wahrnehmung von Feedback sowie zur Entstehung differenzieller Wirkungen auf die Motivation und eröffnen darüber hinaus vielfältige Anregungen für die pädagogische Praxis.