

# Gelingensbedingungen und Wirkungen beobachtungsbasierter Unterrichtsfeedbacks durch Schulleitungen an Lehrkräfte

Dissertation

zur Erlangung des Grades  
Doktor der Philosophie (Dr. phil.)

am Fachbereich  
Erziehungswissenschaft und Psychologie  
der Freien Universität Berlin

vorgelegt von  
M.A. Christopher Kellermann

Berlin, 2023

Erstgutachterin:

Prof. Dr. Felicitas Thiel, Freie Universität Berlin

Zweitgutachter:

Prof. Dr. Holger Gärtner, Freie Universität Berlin

Tag der Disputation: 24.04.2023

# Inhaltsverzeichnis

<b>Zusammenfassung</b> .....	<b>6</b>
<b>Abstract</b> .....	<b>8</b>
<b>1 Einleitung</b> .....	<b>11</b>
<b>2 Theoretischer Rahmen</b> .....	<b>15</b>
2.1 Zur veränderten Rolle der Schulleitung: Vom Schulmanager zum <i>Instructional Leader</i> .....	15
2.2 Beobachtungsbasierte Unterrichtsfeedbacks durch Schulleitungen .....	18
2.2.1 Beobachtungsbasierte Unterrichtsfeedbacks als Instrument der individualisierten Personalentwicklung in Deutschland.....	18
2.2.2 Beobachtungsbasierte Unterrichtsfeedbacks am Beispiel der US-amerikanischen <i>Teacher Evaluation Systems</i> .....	21
2.2.3 Zur Wirksamkeit beobachtungsbasierter Unterrichts- feedbacks .....	22
2.3 Wirksamkeitsbedingungen beobachtungsbasierter Unterrichtsfeedbacks durch Schulleitungen .....	25
2.3.1 Unterrichtsbeobachtung .....	28
2.3.2 Unterrichtsfeedback .....	29
2.3.3 Merkmale der Schulleitung .....	35
2.3.4 Feedbackverarbeitung .....	37
2.3.5 Feedbacknutzung .....	41
2.3.6 Merkmale der Lehrkraft .....	42
2.4 Feedbackintervention, Forschungsfragen und Teilstudien .....	48
2.4.1 Beschreibung der Feedbackintervention.....	48
2.4.2 Herleitung der Forschungsfragen.....	50
2.4.3 Überblick über die empirischen Teilstudien .....	52
<b>Literaturverzeichnis I</b> .....	<b>54</b>

<b>3</b>	<b>Teilstudie 1: Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interventionsstudie unter Berücksichtigung der Zielorientierungen von Lehrkräften.....</b>	<b>74</b>
3.1	Einleitung.....	76
3.2	Theoretischer Rahmen und Forschungsbefunde.....	77
3.3	Fragestellung und Hypothesen .....	82
3.4	Methodik.....	83
3.5	Ergebnisse.....	91
3.6	Diskussion .....	97
3.7	Literatur .....	100
<b>4</b>	<b>Teilstudie 2: The effect of school leaders’ feedback on teaching quality perceived by students – results of an intervention study.....</b>	<b>108</b>
<b>5</b>	<b>Teilstudie 3: Improvements in teaching quality after receiving feedback: What role do teachers’ goal orientations play for the effectiveness of feedback? .....</b>	<b>136</b>
5.1	Introduction.....	137
5.2	Method.....	143
5.3	Results .....	147
5.4	Discussion.....	153
5.5	References.....	158

<b>6</b>	<b>Gesamtdiskussion .....</b>	<b>167</b>
6.1	Zusammenfassung und Diskussion der Ergebnisse .....	168
6.2	Limitationen.....	178
6.3	Praktische Implikationen .....	182
6.4	Ausblick.....	184
6.5	Fazit .....	187
	<b>Literaturverzeichnis II.....</b>	<b>188</b>
	<b>Eigenanteil an Studien mit Co-Autorenschaft .....</b>	<b>200</b>
	<b>Selbstständigkeitserklärung .....</b>	<b>201</b>
	<b>Publikationsliste.....</b>	<b>202</b>

## **Zusammenfassung**

Schulleitungen müssen seit Einführung des neuen Steuerungsparadigmas vermehrt strategische Führungsaufgaben im Hinblick auf die innerschulische Qualitätssicherung und -entwicklung übernehmen. Im Fokus des Schulleitungshandelns steht dabei vor allem die Qualität der von Lehrkräften gestalteten unterrichtlichen Lerngelegenheiten, die für den Lernerfolg von Schülerinnen und Schülern ausschlaggebend sind. Unterrichtsbeobachtungen und -feedbacks gehören zwar zu den Führungsaktivitäten, über die Schulleitungen die Qualität der unterrichtlichen Lerngelegenheiten an ihren Schulen positiv beeinflussen können, hierzulande sind beobachtungsbasierte Unterrichtsfeedbacks durch Schulleitungen jedoch kaum verbreitet. Außerdem zeigen Studienbefunde, dass Unterrichtsfeedbacks nicht immer zu einer besseren Unterrichtsqualität bzw. einer Steigerung des Lernerfolgs von Schülerinnen und Schülern führen.

Die vorliegende Dissertation beschäftigt sich vor diesem Hintergrund mit den Gelingensbedingungen und Wirkungen von beobachtungsbasiertem Unterrichtsfeedback durch Schulleitungen. Ausgehend von einem theoretischen Feedbackmodell wurde eine Feedbackintervention entwickelt, die Schulleitungen sowohl bei der Unterrichtsbeobachtung als auch bei der Feedbackgestaltung unterstützen soll. In insgesamt drei Teilstudien wurde die Wirksamkeit der Intervention anhand der von Lehrkräften selbst eingeschätzten Weiterentwicklung ihrer Unterrichtskompetenz sowie anhand der von Schülerinnen und Schülern eingeschätzten Unterrichtsqualität überprüft. Weil die Wirksamkeit von Unterrichtsfeedbacks jedoch auch von der Feedbackverarbeitung und -nutzung durch die Lehrkräfte selbst abhängt, die wiederum von verschiedenen Merkmalen der Lehrkraft beeinflusst wird, wurde in den drei Teilstudien neben den Haupteffekten auch untersucht, ob die Zielorientierungen von Lehrkräften die Wirksamkeit der Feedbackintervention moderieren.

Zur Überprüfung der Interventionswirksamkeit wurden die Daten einer quasi-experimentellen Studie genutzt, die im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts „Unterrichtsbeurteilungen und -feedback durch Schulleitungen“ durchgeführt wurde. Während Lehrkräfte der Interventionsgruppe von ihren Schulleitungen im Unterricht beobachtet wurden und im Anschluss an die Beobachtungen ein Feedback zu insgesamt 14 Unterrichtsmerkmalen erhielten, bekamen Lehrkräfte

der Kontrollgruppe kein Feedback. Die Befunde zu den Haupteffekten der Feedback-intervention sind inkonsistent. Mit Blick auf die Weiterentwicklung der Unterrichtskompetenz zeigte sich, dass Lehrkräfte, die ein Feedback erhielten, ihre Kompetenzentwicklung im Mittel nicht positiver einschätzten als Lehrkräfte der Kontrollgruppe (Teilstudie 1). Auf die von Schülerinnen und Schülern eingeschätzte Unterrichtsqualität hatte die Intervention demgegenüber einen positiven Effekt. Bei drei der 14 Unterrichtsmerkmale schätzten die Schülerinnen und Schüler von Lehrkräften der Interventionsgruppe den Unterricht nach dem Feedback besser ein als Schülerinnen und Schüler von Lehrkräften der Kontrollgruppe, was darauf schließen lässt, dass die entwickelte Feedbackintervention wirksam ist (Teilstudie 2).

Auch die Befunde zu den differenziellen Effekten der Zielorientierungen von Lehrkräften ergeben kein einheitliches Bild. Hinsichtlich der Weiterentwicklung der Unterrichtskompetenz zeigte sich, dass sowohl Lehrkräfte mit hoch ausgeprägter Lernzielorientierung als auch Lehrkräfte mit hoch ausgeprägter Annäherungsleistungszielorientierung, die ein Feedback erhielten, ihre Kompetenzentwicklung positiver einschätzten als Lehrkräfte der Kontrollgruppe (Teilstudie 1). Für die Lernzielorientierung zeigten sich auch bei der von Schülerinnen und Schülern eingeschätzten Unterrichtsqualität bei zwei der 14 Unterrichtsmerkmale positive Interaktionseffekte. Die Annäherungsleistungszielorientierung wies demgegenüber bei nur einem Unterrichtsmerkmal einen positiven Interaktionseffekt auf und bei drei Unterrichtsmerkmalen negative Interaktionseffekte, was teilweise im Widerspruch zu den differenziellen Effekten der Annäherungsleistungszielorientierung auf die Kompetenzentwicklung der Lehrkräfte steht (Teilstudie 3).

Die Befunde zu den Haupteffekten der Feedbackintervention und die teilweise widersprüchlichen Befunde zu den differenziellen Effekten der Zielorientierungen werden abschließend integrierend diskutiert. Außerdem werden Limitationen der Arbeit aufgezeigt, Implikationen für die pädagogische Praxis abgeleitet und der sich ergebene weitere Forschungsbedarf skizziert.

**Schlagwörter:** Unterrichtsbeobachtungen · Feedback · Unterrichtsentwicklung · Unterrichtsqualität · Professionalisierung von Lehrkräften · Zielorientierungen · schulisches Führungsverhalten · Instructional Leadership

## **Abstract**

Due to Germany's new approach to educational governance, tasks and responsibilities of school leaders have changed significantly in recent years. School leaders nowadays are increasingly expected to function as instructional leaders. Instructional leadership emphasises the school leader's responsibility for the quality of instruction provided by teachers and, therefore, for student achievement. Classroom observations and feedback are discussed as a promising measure for improving the quality of teaching and in turn promoting student achievement. However, teachers in Germany do not often receive observation-based feedback from school leaders. In addition, feedback does not always lead to an improved teaching quality and better student achievement.

Against this background, this PhD thesis deals with the conditions observation-based feedback from school leaders has to meet in order to be effective. Based on a theoretical feedback model, an intervention was developed to assist school leaders in both classroom observations and feedback provision. The effectiveness of the intervention was examined in three empirical studies. Effects of the intervention were tested based on teachers self-reported competence development as well as students' perceptions of teaching quality. The effectiveness of feedback, however, also depends on the way teachers process and use the information provided to them. Feedback processing and usage is influenced by various teacher characteristics, including their goal orientations. Therefore, it was also examined whether teachers' goal orientations moderate the effectiveness of feedback.

To test the effectiveness of the intervention, data from a quasi-experimental study was used, which was conducted as part of the project "Classroom observations and feedback by school leaders" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). While school leaders conducted classroom observations and provided feedback in relation to 14 teaching characteristics to teachers of the intervention group, teachers of the control group did not receive feedback. Findings on the effectiveness of the intervention are inconclusive. On the one hand, teachers who received feedback on average did not report a higher competence development than teachers of the control group (Study 1). On the other hand, the intervention yielded an effect on changes in teaching quality as perceived by students. In relation to three of the 14 teaching char-

acteristics, intervention group students rated changes in teaching quality significantly better than control group students, providing first evidence for the effectiveness of the intervention (Study 2).

With respect to the question of whether teachers' goal orientations moderate the effectiveness of feedback, results also do not provide a clear picture. In relation to teachers' self-rated competence development, teachers with a pronounced learning goal orientation as well as teachers with a pronounced performance approach goal orientation who received feedback reported a higher competence development than teachers of the control group (Study 1). In relation to changes in teaching quality as perceived by students, for two of the 14 teaching characteristics examined, a positive interaction of teachers' learning goal orientation and school leaders' feedback could be detected, too. In contrast, a performance approach goal orientation had a positive impact on changes in teaching quality after receiving feedback in relation to only one teaching characteristic. With regard to three teaching characteristics, a negative interaction was found (Study 3).

In conclusion, results on the effectiveness of the feedback intervention and on the partially contradictory moderating effects of teachers' goal orientations are discussed. Furthermore, limitations of the present study are pointed out, practical implications are derived, and the need for further research is outlined.

**Keywords:** Classroom observations · teacher evaluation · feedback · school improvement · teaching quality · teachers' professional development · goal orientations · school leadership · instructional leadership

# 1

## Einleitung

# 1 Einleitung

Der Lernerfolg von Schülerinnen und Schülern hängt maßgeblich von den von Lehrkräften gestalteten Lerngelegenheiten, d.h. von der Qualität des angebotenen Unterrichts, ab (vgl. z. B. Aaronson et al., 2007; Kyriakides et al., 2013; Lipowsky, 2006; Muijs et al., 2014; Nye et al., 2004; Rivkin et al., 2005; Rockoff, 2004). Aus diesem Grund sollten Maßnahmen, die darauf abzielen, den Lernerfolg von Schülerinnen und Schülern zu steigern, bei den Lehrkräften und ihrer Unterrichtskompetenz ansetzen. Ein beobachtungs-basiertes Unterrichtsfeedback durch Schulleitungen an Lehrkräfte gilt vor allem im internationalen Forschungsraum als erfolgversprechende Maßnahme, die zu einer Weiterentwicklung der Unterrichtskompetenz von Lehrkräften beitragen und – vermittelt über eine bessere Unterrichtsqualität – zu einer Steigerung des Lernerfolgs von Schülerinnen und Schülern führen kann (vgl. z. B. Hill & Grossmann, 2013; Papay, 2012; Kraft & Gilmour, 2016; Ovando & Ramirez, 2007).

Empirische Studien, die die Effekte von Unterrichtsfeedbacks auf die Qualität der von Lehrkräften gestalteten unterrichtlichen Lerngelegenheiten und den Lernerfolg von Schülerinnen und Schülern untersuchen, deuten darauf hin, dass es sich bei einem beobachtungs-basierten Unterrichtsfeedback durch Schulleitungen an Lehrkräfte tatsächlich um eine wirksame Entwicklungsmaßnahme handelt. Andererseits zeigt sich aber auch, dass Unterrichtsfeedbacks nicht immer zu einer besseren Unterrichtsqualität und zu einer Steigerung des Lernerfolgs von Schülerinnen und Schülern führen (vgl. z. B. Kraft et al., 2018; Song et al., 2021; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). Dies steht im Einklang mit meta-analytischen Befunden aus der arbeitspsychologischen (vgl. Kluger & DeNisi, 1996; Smither et al., 2005a) sowie instruktionspsychologischen Feedbackforschung (vgl. Bangert-Drowns et al., 1991), die ebenfalls zeigen, dass nicht jedes Feedback per se wirksam ist.

Vor dem Hintergrund des inkonsistenten Forschungsstands beschäftigt sich die vorliegende Dissertation mit der Frage, warum Unterrichtsfeedbacks von Schulleitungen nicht immer wirksam sind und welche Bedingungen erfüllt sein müssen, damit die gewünschten Effekte eintreten. Zur theoretischen Einbettung wird ein Feedbackmodell entwickelt, das den Prozess eines beobachtungs-basierten Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte beschreibt. Die Modellentwicklung erfolgt vor allem auf der Grundlage von Theorien der angewandten arbeitspsychologischen Feedbackforschung

(vgl. Ilgen & Davis, 2000; Ilgen et al., 1979; Kluger & DeNisi, 1996; London & Smither, 2002; Smither et al., 2005a; Strijbos & Müller, 2014), die in der empirischen Schul- und Unterrichtsforschung bislang kaum rezipiert wurden. In drei Teilstudien werden ausgewählte Wirksamkeitsbedingungen zudem empirisch untersucht.

Der Fokus liegt dabei zum einen auf der *formalen Gestaltung des Unterrichtsfeedbacks durch die Schulleitung*, die als eine der zentralen Bedingungen der Feedbackwirksamkeit gelten kann. Obwohl Lehrkräfte die häufig geringe Qualität der Unterrichtsfeedbacks ihrer Schulleitungen bemängeln und auch Schulleitungen angeben, dass es ihnen schwerfalle, Lehrkräften ein qualitativ volles Feedback zu ihrem Unterricht zu geben (vgl. z. B. Kimball, 2003; Kimball & Milanowski, 2009; Kraft & Gilmour, 2016; Liu et al., 2015; Weisberg et al., 2009), haben sich bislang nur wenige Studien mit der Frage auseinandergesetzt, was ein qualitativ hochwertiges Unterrichtsfeedback auszeichnet (für eine Ausnahme siehe z. B. Cherasaro et al., 2016; Hunter & Springer, 2022; Kraft & Christian, 2022). Unter Bezugnahme auf Forschungsarbeiten aus der arbeits- und instruktionspsychologischen Feedbackforschung wurde daher ein Feedbackverfahren entwickelt, das Schulleitungen dabei unterstützen soll, Lehrkräften ein qualitativ hochwertiges Unterrichtsfeedback zu geben. Das Feedbackverfahren ist Teil einer Intervention, die im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts „Unterrichtsbeurteilungen und -feedback durch Schulleitungen“ entwickelt und deren Wirksamkeit in einer quasi-experimentellen Studie evaluiert wurde.<sup>1</sup> Die Haupteffekte der Feedbackintervention auf die Unterrichtskompetenz von Lehrkräften und die Qualität der von ihnen gestalteten unterrichtlichen Lerngelegenheiten werden in den Teilstudien 1 und 2 untersucht.

Neben der formalen Feedbackgestaltung durch die Schulleitung wird als weitere Wirksamkeitsbedingung die *Feedbackverarbeitung und -nutzung durch die Lehrkraft* näher in den Blick genommen, die von verschiedenen Merkmalen der Lehrkraft beeinflusst wird. Obwohl in der empirischen Schul- und Unterrichtsforschung erkannt wurde, dass die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks entscheidend von der Feedbackverarbeitung und -nutzung durch die Lehrkräfte selbst abhängt (vgl. z. B. Cher-

---

<sup>1</sup> Das DFG-Projekt „Unterrichtsbeurteilungen und -feedback durch Schulleitungen“ (Projektnummer 283241518) wurde im Zeitraum September 2016 bis Februar 2020 unter der Leitung von Prof. Dr. Felicitas Thiel und Prof. Dr. Holger Gärtner an der Freien Universität Berlin durchgeführt. Die Feedbackintervention und die quasi-experimentelle Studie werden in Kapitel 2.4 ausführlicher beschrieben.

asaro et al., 2016; Jiang et al., 2015; Kimball, 2003; Kraft & Gilmour, 2016), wurde bislang nur in Ansätzen beschrieben, welche emotionalen und kognitiven Verarbeitungsprozesse involviert sind und wie Lehrkräfte ein beobachtungsbasiertes Unterrichtsfeedback effektiv nutzen können. Darüber hinaus wurde bis dato in nur wenigen Studien untersucht, welche personalen Merkmale die Feedbackverarbeitung und -nutzung beeinflussen (für eine Ausnahme siehe Kraft & Christian, 2022; Tuytens & Devos, 2014). Die angewandte arbeitspsychologische Feedbackforschung hebt in diesem Zusammenhang insbesondere die Bedeutung von Zielorientierungen hervor (vgl. z. B. Ilgen & Davis, 2000; Smither et al., 2005a; Strijbos & Müller, 2014; VandeWalle et al., 2001). Auch die Forschung zur Lehrkräftemotivation zeigt, dass lernrelevante Einstellungen und Verhaltensweisen von Lehrkräften entscheidend von ihren Zielorientierungen abhängen (vgl. z. B. Butler, 2007; Janke et al., 2019; Nitsche et al., 2013a, 2013b), was die Bedeutung von Zielorientierungen für die Verarbeitung und Nutzung eines beobachtungsbasierten Unterrichtsfeedbacks durch Lehrkräfte unterstreicht. Neben den Haupteffekten der Feedbackintervention werden in den Teilstudien 1 und 3 daher auch die differenziellen Effekte der Zielorientierungen von Lehrkräften auf die Feedbackwirksamkeit untersucht.

Das Feedbackmodell, das als theoretische Rahmung der vorliegenden Arbeit dient, wird in Kapitel 2.3 beschrieben. Zunächst wird in Kapitel 2.1 auf die veränderte Rolle der Schulleitung im Kontext neuer Steuerung und das Führungskonzept *Instructional Leadership* eingegangen. In diesem Zusammenhang wird dargelegt, dass Unterrichtsbeobachtungen und -feedbacks zu den zentralen unterrichtsbezogenen Führungsaufgaben von Schulleitungen gezählt werden können. In Kapitel 2.2 werden die Rahmenbedingungen und die Verbreitung von beobachtungsbasiertem Unterrichtsfeedback in Deutschland und den USA beschrieben. Außerdem wird der Forschungsstand zur Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback aufgearbeitet. In Kapitel 2.4 wird die entwickelte Feedbackintervention schließlich beschrieben, die im Promotionsvorhaben bearbeiteten Forschungsfragen hergeleitet und die empirischen Teilstudien vorgestellt, die in den Kapiteln 3, 4 und 5 abgedruckt sind. In Kapitel 6 folgt eine abschließende Gesamtdiskussion der Befunde.

# 2

## Theoretischer Rahmen

## 2 Theoretischer Rahmen

### 2.1 Zur veränderten Rolle der Schulleitung: Vom Schulmanager zum *Instructional Leader*

Seit den 1990er Jahren hat sich ein Paradigmenwechsel in der Steuerung des deutschen Schulsystems vollzogen, der in der im Jahr 2006 erstmals verabschiedeten *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring* seinen programmatischen Ausdruck findet (vgl. KMK, 2006; für eine überarbeitete Fassung des Beschlusses siehe KMK, 2016). Die sogenannte neue Steuerung des Schulsystems (vgl. Altrichter & Maag Merki, 2016) ist zum einen durch eine Verlagerung von Entscheidungskompetenzen von der Ebene der Schulverwaltung auf die Ebene der Einzelschule gekennzeichnet (Stichwort *Schulautonomie*; vgl. Altrichter et al., 2016b; Rürup & Heinrich, 2007). Gleichzeitig geht das neue Steuerungsparadigma mit einer unter anderem an der Erreichung von Bildungsstandards (vgl. Klieme, 2004; Klieme et al., 2003) orientierten Outputsteuerung und einer erhöhten Rechenschaftspflicht der Einzelschule einher. Der Erfolg des schulischen Handelns bemisst sich seither vor allem an den (fachlichen) Leistungen von Schülerinnen und Schülern, wodurch das neue Steuerungsparadigma zu einer stärkeren Fokussierung auf die Qualität unterrichtlicher Lerngelegenheiten geführt hat (vgl. Altrichter et al., 2016a).

Aufgrund der gesteigerten Rechenschaftspflichten, vor allem aber wegen der erweiterten Gestaltungsspielräume, ist die Einzelschule im neuen Steuerungsparadigma zum zentralen Akteur der innerschulischen Qualitätssicherung und -entwicklung geworden. Ausschlaggebend für diese Entwicklung ist die Annahme, dass die Einzelschule die Leistungserbringung vor Ort am effektivsten und effizientesten planen und koordinieren kann (vgl. Thiel & Thillmann, 2012; Thillmann et al., 2015b). Damit hat sich auch die Rolle der Schulleitung grundlegend verändert: Zusätzlich zu operativen Managementaufgaben müssen Schulleitungen hierzulande vermehrt strategische Führungsaufgaben im Hinblick auf die Qualitätssicherung und -entwicklung an ihren Schulen übernehmen (vgl. Altrichter et al., 2019; Bonsen, 2016; Brauckmann & Eder, 2019; Brauckmann & Schwarz, 2015).

Dass Schulleitungen mehr Verantwortung zugewiesen werden sollte, unterstreichen auch die Befunde der – vor allem internationalen – Schulleitungsforschung, nach denen

das Schulleitungshandeln einen, wenn auch nur schwachen bis moderaten, positiven Effekt auf den Lernerfolg von Schülerinnen und Schülern hat. Einschlägige Meta-Analysen und Literaturreviews zeigen, dass Schulleitungen den Lernerfolg von Schülerinnen und Schülern dabei vorrangig indirekt, vermittelt über intervenierende Schul- und Klassenvariablen, beeinflussen (vgl. z. B. Hallinger & Heck, 1996, 1998; Marzano et al., 2005; Witziers et al., 2003). Leithwood et al. (2008, S. 28) kommen in einem narrativen Literaturreview sogar zu dem Schluss, dass „[s]chool leadership is second only to classroom teaching as an influence on student learning“. Wichtiger als der Befund, dass Schulleitungen den Lernerfolg von Schülerinnen und Schülern positiv beeinflussen, ist jedoch die Frage, wie Schulleitungen zum Lernerfolg von Schülerinnen und Schülern beitragen, d.h. welche Konzepte schulischen Führungshandelns und welche konkreten Führungsaktivitäten wirksam sind.

Ein Konzept schulischen Führungshandelns, das im Zusammenhang mit dem neuen Steuerungsparadigma und der veränderten Rolle der Schulleitung viel Aufmerksamkeit erfährt, ist das Führungskonzept *Instructional Leadership*. So schlussfolgert Hallinger (2005, S. 222) mit Blick auf die auch international zu beobachtenden gesteigerten Rechenschaftspflichten und erweiterten Gestaltungsspielräume der Einzelschule, dass „[p]rincipals find themselves at the nexus of accountability and school improvement with an increasingly explicit expectation they will function as instructional leaders“ (siehe auch Daniëls et al., 2019; Neumerski, 2013). Die Aufmerksamkeit, die dem Führungskonzept *Instructional Leadership* zuteilwird, ist dabei vor allem auf die mit dem neuen Steuerungsparadigma einhergehende Orientierung an den (fachlichen) Leistungen von Schülerinnen und Schülern zurückzuführen. Auch der Fokus von *Instructional Leadership* liegt auf dem “improvement of teaching and learning and (...) on the behavior of teachers as they engage in activities directly influencing student achievement“ (Daniëls et al., 2019, S. 114).

Das am häufigsten rezipierte theoretische Modell von *Instructional Leadership* stammt von Hallinger und Murphy (1985; siehe auch Hallinger 2003, 2005). Es unterscheidet zwischen drei Führungsdimensionen: (1) *Defining the schools mission*, (2) *Managing the instructional programme* sowie (3) *Promoting a positive school climate*. Die zweite Führungsdimension, *Managing the instructional programme*, betont die Verantwortung der Schulleitung für und ihre Einflussnahme auf das Lehren und Lernen an ihrer Schule. Hallinger (2005, S. 230) spricht in diesem Zusammenhang auch von „hands-on involvement in the classroom“. Unterrichtsbeobachtungen durch Schulleitungen und ein

auf die Beobachtungen folgendes entwicklungsbezogenes Feedback an die beobachteten Lehrkräfte können dieser Führungsdimension zugeordnet werden.

Hinsichtlich der oben aufgeworfenen Frage, welche Konzepte schulischen Führungshandelns wirksam sind, zeigt eine Meta-Analyse von Robinson et al. (2008), in der insgesamt 22 Primärstudien berücksichtigt wurden, dass *Instructional Leadership* einen positiven Effekt auf den Lernerfolg von Schülerinnen und Schülern hat – und wirksamer ist als andere Konzepte schulischen Führungshandelns. So ist der gemittelte Effekt von *Instructional Leadership* mit  $z = .42$  deutlich größer als der Effekt einer mehrere unterschiedliche Führungskonzepte umfassenden Kategorie ( $z = .30$ ) und fast viermal so groß wie der Effekt von *Transformational Leadership* ( $z = .11$ ).<sup>2</sup> Darüber hinaus liefert die Meta-Analyse erste Hinweise darauf, dass Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen zu den wirksamen unterrichtsbezogenen Führungsaktivitäten gezählt werden können: Robinson et al. (2008) haben nicht nur die Wirksamkeit verschiedener Führungskonzepte miteinander verglichen, sondern auch die Effekte unterschiedlicher Führungsdimensionen auf den Lernerfolg von Schülerinnen und Schülern untersucht. Durch ein induktives Vorgehen wurden ausgehend von den Primärstudien insgesamt fünf Führungsdimensionen ermittelt. Von besonderem Interesse für die vorliegende Arbeit ist die Dimension *Planning, coordinating, and evaluating teaching*, zu der Robinson et al. (2008, S. 656) unter anderem „[d]irect involvement in the support and evaluation of teaching through regular classroom visits and provision of formative and summative feedback to teachers“ zählen. Mit einer gemittelten Effektstärke von  $z = .42$  hat diese Führungsdimension den zweitgrößten Effekt aller gebildeten Führungsdimensionen auf den Lernerfolg von Schülerinnen und Schülern.

Zusammenfassend kann festgehalten werden, dass die Qualität unterrichtlicher Lerngelegenheiten seit Einführung des neuen Steuerungsparadigmas im Fokus des Schulleitungshandelns steht und dass Schulleitungen dementsprechend vermehrt unterrichtbezogene Führungsaktivitäten übernehmen müssen. Außerdem wurden erste empirische

---

<sup>2</sup> *Transformational Leadership* stellt neben *Instructional Leadership* das in der Schulleitungsforschung am häufigsten diskutierte Konzept schulischen Führungshandelns dar. Nach Leithwood und Jantzi (2005, S. 177) führen *Transformational Leader* „by developing a vision for the organization, developing commitments and trust among workers, and facilitating organizational learning“. Beim *Transformational Leadership* stehen Emotionen und Werte, symbolisches Verhalten und „helping making events meaningful for followers“ (Leithwood & Jantzi, 2005, S. 178) im Fokus.

Befunde präsentiert, die darauf hindeuten, dass Schulleitungen die Qualität unterrichtlicher Lehr- und Lernprozesse über ein beobachtungsbasiertes Unterrichtsfeedback positiv beeinflussen können. Das nächste Kapitel beschäftigt sich näher mit Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen und beschreibt deren Rahmenbedingungen und Verbreitung in Deutschland und den USA. Anschließend wird der Forschungsstand zur Wirksamkeit von beobachtungsbasierten Unterrichtsfeedbacks zusammenfassend dargestellt. Zunächst wird aber definiert, was in der vorliegenden Dissertation unter einem beobachtungsbasierten Unterrichtsfeedback verstanden wird.

## **2.2 Beobachtungsbasierte Unterrichtsfeedbacks durch Schulleitungen**

Müller und Ditton (2014) definieren Feedback unter Rückgriff auf in der arbeitspsychologischen (vgl. Ilgen et al., 1979; Kluger & DeNisi, 1996) und instruktionspsychologischen Feedbackforschung (vgl. Hattie & Timperley, 2007) weit verbreiteten Definitionen als „bewusste Rückmeldung von Informationen an eine Person zu ihrem vorherigen Verhalten“ (vgl. Müller & Ditton, 2014, S. 15). Hieran anknüpfend wird unter einem beobachtungsbasierten Unterrichtsfeedback in der vorliegenden Arbeit eine Rückmeldung von Informationen an eine Lehrkraft zu ihrem in einer beobachteten Unterrichtsstunde gezeigten Verhalten verstanden.

### **2.2.1 Beobachtungsbasierte Unterrichtsfeedbacks als Instrument der individualisierten Personalentwicklung in Deutschland**

In Deutschland sind Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen vor allem im Zusammenhang mit Maßnahmen der individualisierten Personalentwicklung vorgesehen. Im Gegensatz zu schulbezogenen Personalentwicklungsmaßnahmen wie der Planung schulinterner Fortbildungen oder der Förderung kooperativer Arbeitsformen, die auf das gesamte Lehrkräftekollegium ausgerichtet sind, zielen Maßnahmen der individualisierten Personalentwicklung auf die professionelle Weiterentwicklung einzelner Lehrkräfte ab. Zu den individualisierten Personalentwicklungsmaßnahmen werden vor allem dienstliche Beurteilungen<sup>3</sup> sowie Mitarbeiterinnen- und Mitarbeitergespräche zwischen

---

<sup>3</sup> Im Hinblick auf dienstliche Beurteilungen wird zwischen anlassbezogenen und regelmäßigen Beurteilungen unterschieden. Anlassbezogene Beurteilungen werden bei beschäftigungsrelevanten Entscheidungen durchgeführt, z. B. bei einer bevorstehenden Verbeamtung einer Lehrkraft oder bei der Besetzung einer Funktionsstelle. Regelmäßige Beurteilungen erfolgen demgegenüber turnusmäßig und dienen vor allem der professionellen Weiterentwicklung individueller Lehrkräfte (vgl. Thiel et al., 2022).

der Schulleitung und ihren Lehrkräften gezählt (vgl. Appius et al., 2012; Bach et al., 2020; Böckelmann & Mäder, 2018; Meetz, 2007; Thiel & Schewe, 2022; Thillmann et al., 2015a). Unterrichtsbeobachtungen können dabei als integraler Bestandteil von dienstlichen Beurteilungen gelten, weil nur Unterrichtsbeobachtungen einen unmittelbaren Einblick in die unterrichtliche Praxis von Lehrkräften erlauben und nur so eine Beurteilung ihrer Arbeitsleistung möglich wird. Mitarbeiterinnen- und Mitarbeitergespräche wiederum können ihr Potenzial für eine individualisierte Personalentwicklung nur dann voll entfalten, wenn sie um ein entwicklungsbezogenes Feedback an die beobachteten Lehrkräfte ergänzt werden (vgl. Thiel et al., 2022).

Eine aktuelle Analyse rechtlicher Dokumente zur schulischen Personalentwicklung in Deutschland zeigt, dass die Schulgesetze und nachgeordneten Verwaltungsvorschriften bzw. Verordnungen nahezu aller deutschen Bundesländer dienstliche Beurteilungen und/oder Mitarbeiterinnen- und Mitarbeitergespräche – und damit implizit auch Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen – vorsehen (vgl. Thiel et al., 2022; Thiel & Schewe, 2022). Gleichzeitig zeigen Studienbefunde jedoch, dass beobachtungs-basierte Unterrichtsfeedbacks durch Schulleitungen in Deutschland nur wenig verbreitet sind. So berichten Bach et al. (2020), die im Jahr 2013  $N = 658$  Lehrkräfte aus Berlin und Brandenburg befragt haben, dass 39 % dieser Lehrkräfte innerhalb von fünf Jahren kein einziges Mal von ihrer Schulleitung im Unterricht beobachtet wurden und 26 % kein Mitarbeiterinnen- und Mitarbeitergespräch mit ihrer Schulleitung geführt haben. Die Befunde von Bach et al. (2020) korrespondieren mit den Ergebnissen von Thillmann et al. (2015a), die aus derselben Studie hervorgegangen sind, denen jedoch keine Auskünfte von Lehrkräften, sondern die Befragungsdaten von  $N = 327$  Schulleitungen zugrunde liegen. Nur 51 % der befragten Schulleitungen gaben an, regelmäßige Unterrichtsbeobachtungen (mindestens einmal alle zwei Jahre) durchzuführen und 65 % gaben an, regelmäßige Mitarbeiterinnen- und Mitarbeitergespräche mit ihren Lehrkräften zu führen (siehe auch Thillmann et al., 2015b).

Die Ergebnisse einer von der *Gewerkschaft für Erziehung und Wissenschaft* (GEW) in Auftrag gegebenen, deutschlandweiten Befragung von  $N = 3.734$  Lehrkräften und  $N = 388$  Schulleitungen (vgl. Barfknecht & von Saldern, 2010), die parallel zum *Teaching and Learning International Survey* (TALIS) der *Organisation for Economic Co-operation and Development* (vgl. OECD, 2009) durchgeführt wurde, weisen in eine ähnliche Richtung: 55 % der befragten Lehrkräfte gaben an, selten bis nie ein Unterrichtsfeedback von ihrer Schulleitung zu erhalten. Auch 16 % der befragten Schulleitungen berichten

davon, die Arbeit ihrer Lehrkräfte nie zu beurteilen und 40 % gaben an, dies seltener als einmal in zwei Jahren zu tun (vgl. Barfknecht & von Saldern, 2010). Zu ähnlichen Ergebnissen kommt auch Schleicher (2012) auf der Grundlage von Daten des *Programme for International Student Assessment* (PISA) aus dem Jahr 2009, in dem Schulleitungen unter anderem zu ihren Führungsaktivitäten befragt wurden. Demnach werden nur 40 % aller deutschen Schülerinnen und Schüler an Schulen unterrichtet, an denen Schulleitungen nach eigener Auskunft Unterrichtsbeobachtungen durchführen und nur 53 % aller Schülerinnen und Schüler werden an Schulen unterrichtet, an denen Schulleitungen ihren Lehrkräften ein Feedback zu ihrem Unterricht geben. Sowohl im Hinblick auf die Unterrichtsbeobachtungen als auch hinsichtlich des Feedbacks liegen deutsche Schulen damit unterhalb des OECD-Durchschnitts (50 % bzw. 69 %).

Vor allem in den USA spielen Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen eine bedeutsame Rolle: Hier werden 95 % aller Schülerinnen und Schüler an Schulen unterrichtet, an denen Schulleitungen Unterrichtsbeobachtungen durchführen und die Lehrkräfte von 92 % aller Schülerinnen und Schüler erhalten von ihren Schulleitungen ein Feedback zu ihrem Unterricht (vgl. Schleicher, 2012). Dass Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen in den USA verbreiteter sind als hierzulande, ist vor allem auf die systematische Implementation von sogenannten *Teacher Evaluation Systems* (TES) zurückzuführen, die Schulleitungen dazu verpflichten, regelmäßige Unterrichtsbeobachtungen durchzuführen und ihren Lehrkräften ein entwicklungsbezogenes Feedback zu ihrem Unterricht zu geben. Während im deutschsprachigen Raum abgesehen von den zuvor zitierten Studien nach meinem Kenntnisstand keine weiteren Forschungsarbeiten zu beobachtungsbasiertem Unterrichtsfeedback durch Schulleitungen vorliegen, wurden die US-amerikanischen TES umfassend evaluiert. Weil die Befunde dieser Studien für die vorliegende Arbeit von besonderem Interesse sind, da sie Rückschlüsse auf die Wirksamkeitsbedingungen von beobachtungsbasiertem Unterrichtsfeedback zulassen, soll auf die TES nachfolgend näher eingegangen werden.<sup>4</sup>

---

<sup>4</sup> Auch in anderen Ländern gibt es Beurteilungsverfahren, die Schulleitungen dazu verpflichten, den Unterricht an ihren Schulen regelmäßig zu beobachten und ihren Lehrkräften ein Unterrichtsfeedback zu geben (für Belgien vgl. Delvaux et al., 2013; Tuytens & Devos, 2011, 2014; für Chile vgl. Taut & Sun, 2014; Taut et al., 2011; für China vgl. Chow et al., 2002; Mo et al., 1998; für Singapur vgl. Kelly et al., 2008; für Portugal vgl. Flores, 2012, 2018; für einen aktuellen Überblick siehe auch Manzi et al., 2022). Weil diese Beurteilungsverfahren in konzeptueller Hinsicht große Überschneidungen mit den US-amerikanischen TES aufweisen, soll auf sie nicht näher eingegangen werden. Empirische Befunde werden, sofern sie für die vorliegende Arbeit von Interesse sind, im weiteren Verlauf jedoch zitiert.

### 2.2.2 Beobachtungsbasierte Unterrichtsfeedbacks am Beispiel der US-amerikanischen *Teacher Evaluation Systems*

In den USA sind Unterrichtsbeurteilungen durch Schulleitungen seit den 1960er Jahren etabliert (vgl. Cohen & Goldhaber, 2016; Rowan & Raudenbush, 2016). Die frühen Beurteilungsverfahren setzten dabei vor allem auf kurze, informelle Unterrichtsbeobachtungen durch die Schulleitung, die in der Regel mit einfachen Checklisten erfolgten (vgl. Porter et al., 2001; Toch & Rothman, 2008). Obwohl aus der empirischen Schul- und Unterrichtsforschung bekannt ist, dass es im Hinblick auf die Wirksamkeit von Lehrkräften große Unterschiede gibt (vgl. z. B. Aaronson et al., 2007; Nye et al., 2004; Rivkin et al., 2005; Rockoff, 2004), wurde das Unterrichtsverhalten von nahezu allen Lehrkräften als zufriedenstellend beurteilt. Dieser sogenannte *Widget-Effect* (vgl. Weisberg et al., 2009) hat in den vergangenen 20 Jahren zu einer grundlegenden Reformierung der TES geführt (für einen Überblick siehe Donaldson & Papay, 2014; Papay, 2012; Steinberg & Donaldson, 2016).

Zwar setzen auch die reformierten TES bei der Unterrichtsbeurteilung vor allem auf Unterrichtsbeobachtungen durch Schulleitungen, diese erfolgen jedoch nicht länger mit einfachen Checklisten, sondern mit standardisierten Beobachtungsinstrumenten, denen ein theoretisches Unterrichtsmodell zugrunde liegt. Häufig genutzte Beobachtungsinstrumente sind der *Framework for Teaching* (FFT; vgl. Danielson, 2007) sowie das *Classroom Assessment Scoring System* (CLASS, vgl. Pianta & Hamre, 2009). Außerdem durchlaufen Schulleitungen in den meisten US-Bundesstaaten bzw. -Schuldistrikten nunmehr ein umfassendes Beobachtungstraining und sind zudem dazu verpflichtet, den Unterricht ihrer Lehrkräfte mehrmals zu beobachten. Neben den Beobachtungsdaten werden darüber hinaus in der Regel Leistungsdaten von Schülerinnen und Schülern (zumeist in Form von *Value-Added Measures* (VAMs))<sup>5</sup> und teilweise auch Unterrichtsartefakte wie

---

<sup>5</sup> Mithilfe von VAMs wird versucht zu bestimmen, wie groß der Effekt einer Lehrkraft auf die Lernzuwächse ihrer Schülerinnen und Schüler während eines Schuljahres ist. Hierfür wird das Abschneiden von Schülerinnen und Schülern in einem standardisierten Leistungstest am Ende eines Schuljahres mit dem Abschneiden verglichen, das aufgrund früherer Testergebnisse zu erwarten wäre. Zudem werden weitere lern- und leistungsrelevante Merkmale der Schülerinnen und Schüler (z. B. ihr sozio-ökonomischer Status) sowie Kompositionsmerkmale der jeweiligen Klasse als Kontrollvariablen berücksichtigt. Die Differenz zwischen dem erwarteten und tatsächlichen Abschneiden wird als *Teacher Effect* der jeweiligen Lehrkraft zugeschrieben (für eine Einführung und einen aktuellen Überblick über den Forschungsstand zu VAMs siehe Braun (2005) bzw. Koedel et al. (2015); für eine kritische Diskussion zur Nutzung von VAMs im Zusammenhang mit der Beurteilung individueller Lehrkräfte siehe z. B. Darling-Hammond et al. (2012) sowie Darling-Hammond (2015)).

Unterrichtspläne und -materialien bei der Beurteilung der Lehrkräfte berücksichtigt (für den Schuldistrikt von Cincinnati vgl. z. B. Heneman & Milanowski, 2003; Milanowski & Heneman, 2001; Taylor & Tyler, 2012; für den Schuldistrikt von Chicago vgl. z. B. Jiang et al., 2015; Sartain et al., 2011; Steinberg & Sartain, 2015).

Die TES wurden zwar mit dem Ziel reformiert, wirksame und unwirksame Lehrkräfte zuverlässig zu identifizieren und negative Beurteilungen können mit weitreichenden persönlichen Konsequenzen für die individuellen Lehrkräfte einhergehen (sogenannte *high stakes*, die sogar zu einer Entlassung der betroffenen Lehrkräfte führen können; vgl. Hill & Grossman, 2013; Papay, 2012; Steinberg & Donaldson, 2016), der Fokus in der wissenschaftlichen Diskussion liegt mittlerweile aber auf der Entwicklungsfunktion der TES. Es wird hervorgehoben, dass Lehrkräfte ein vor allem auf Unterrichtsbeobachtungen basierendes entwicklungsbezogenes Feedback erhalten sollen, auf dessen Grundlage sie ihre Unterrichtskompetenz weiterentwickeln und ihren Unterricht verbessern können (vgl. Jiang et al., 2015; Kraft & Christian, 2022; Liu et al., 2019; Papay, 2012; Rowan & Raudenbush, 2016). Dies spiegelt sich auch in den meisten Umsetzungsrichtlinien zu den TES wider, wie eine aktuelle Dokumentenanalyse zeigt: In nahezu allen US-Bundesstaaten bzw. -Schuldistrikten sollen die Unterrichtsbeurteilungen mit dem Ziel erfolgen, Lehrkräfte in ihrer professionellen Weiterentwicklung zu unterstützen (vgl. Steinberg & Donaldson, 2016). In einer repräsentativen Umfrage unter  $N = 1.825$  zufällig ausgewählten Lehrkräften berichten 86 % aller Lehrkräfte für das Schuljahr 2015/2016 zudem davon, im Rahmen der TES in ihrem Unterricht beobachtet worden zu sein und 89 % geben an, ein beobachtungsbasiertes Unterrichtsfeedback von ihrer Schulleitung erhalten zu haben (vgl. Tuma et al., 2018).

### **2.2.3 Zur Wirksamkeit beobachtungsbasierter Unterrichtsfeedbacks**

Die Wirksamkeit beobachtungsbasierter Unterrichtsfeedbacks durch Schulleitungen wurde bis dato vor allem im Zusammenhang mit den US-amerikanischen TES untersucht. Darüber hinaus lässt aber auch die Forschung zum *Instructional Leadership* sowie zum *Teacher Coaching* Rückschlüsse darauf zu, ob beobachtungsbasierte Unterrichtsfeedbacks zu einer Weiterentwicklung der Unterrichtskompetenz von Lehrkräften, einer besseren Unterrichtsqualität und besseren Leistungen von Schülerinnen und Schülern führen.

Im Hinblick auf die *Wirksamkeit von Instructional Leadership* wurde bereits die Meta-Analyse von Robinson et al. (2008) zitiert, die zum einen zeigt, dass eine unterrichtsbezogene Führung mit besseren Leistungen von Schülerinnen und Schülern einhergeht als andere Konzepte schulischen Führungshandelns und zum anderen darauf hindeutet, dass es sich bei Unterrichtsbeobachtungen und -feedbacks um eine wirksame unterrichtsbezogene Führungsaktivität handelt. Darüber hinaus lässt eine Studie von Grissom et al. (2013), in der das Führungsverhalten US-amerikanischer Schulleitungen an mehreren Tagen durch externe Raterinnen und Rater kodiert und mit den Lernzuwächsen von Schülerinnen und Schülern korreliert wurde, Rückschlüsse auf die Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback durch Schulleitungen zu. Während die insgesamt mit Aktivitäten des *Instructional Leadership* zugebrachte Zeit<sup>6</sup> in keinem Zusammenhang mit den Leistungen von Schülerinnen und Schülern stand, gingen häufige Unterrichtsbeobachtungen und -feedbacks durch Schulleitungen mit besseren Leistungen im Fach Mathematik einher. Zwischen der in Unterrichtsbeobachtungen und -feedbacks investierten Zeit und den Leseleistungen von Schülerinnen und Schülern bestand demgegenüber kein Zusammenhang (vgl. Grissom et al., 2013).

Die *Wirksamkeit der TES* wurde bislang überwiegend im Rahmen von querschnittlichen Surveystudien untersucht, in denen Lehrkräfte retrospektiv angaben, ob sie ihren Unterricht verändert bzw. verbessert haben. Die Befunde sind inkonsistent: Während die Mehrheit der befragten Lehrkräfte in einigen Studien davon berichten, ihren Unterricht aufgrund des Unterrichtsfeedbacks weiterentwickelt zu haben (vgl. Cherasaro et al., 2016; Derrington & Martinez, 2019; Jiang et al., 2019; Tuma et al., 2018), gibt die Mehrheit der Lehrkräfte in anderen Studien an, ihren Unterricht nicht weiterentwickelt zu haben (vgl. Kraft & Christian, 2022; Liu et al., 2019; Reddy et al., 2018). Dass Unterrichtsfeedbacks in einigen Fällen zu Maßnahmen der Unterrichtsentwicklung führen, in anderen jedoch nicht, deckt sich auch mit Befunden aus Interviewstudien, die mit US-amerikanischen Lehrkräften durchgeführt wurden, nachdem diese die TES durchlaufen haben (vgl. Heneman & Milanowski, 2003; Kimball 2003).

---

<sup>6</sup> Grissom et al. (2013) berichten davon, dass die Schulleitungen ihrer Stichprobe insgesamt 12.7 % ihrer Arbeitszeit mit Aktivitäten des *Instructional Leadership* zubrachten (für vergleichbare Befunde aus den USA siehe Horng et al. (2010) sowie May und Supovitz (2011)). Auch wenn für Deutschland nach meinem Kenntnisstand keine empirischen Studien zum Umfang von *Instructional Leadership* vorliegen, lassen Befunde von Brauckmann und Schwarz (2015) darauf schließen, dass Schulleitungen auch hierzulande nur wenig Zeit in unterrichtsbezogene Führungsaktivitäten investieren.

Neben Survey- und Interviewstudien, deren Aussagekraft aufgrund methodischer Einschränkungen begrenzt ist, wurde die Wirksamkeit der TES zudem in drei (quasi-)experimentellen Studien untersucht. Taylor und Tyler (2012), die die Wirksamkeit des TES im Schuldistrikt von Cincinnati evaluiert haben, berichten von positiven Effekten auf die Leistungen von Schülerinnen und Schülern im Fach Mathematik, jedoch nicht im Lesen. Steinberg und Sartain (2015), die das TES im Schuldistrikt von Chicago evaluiert haben, berichten demgegenüber von positiven Effekten auf die Leseleistung, nicht aber auf die Matheleistung. Song et al. (2021), die die Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback durch Schulleitungen in einer großangelegten Interventionsstudie in insgesamt acht US-Schuldistrikten untersucht haben, berichten von positiven Feedbackeffekten auf das von externen Beobachterinnen und Beobachtern beurteilte Unterrichtsverhalten von Lehrkräften; allerdings nur, wenn das Unterrichtsverhalten mit dem FFT beurteilt wurde. Erfolgte die Unterrichtsbeurteilung mit CLASS, zeigten sich keine Effekte. Darüber hinaus hatte das Feedback einen positiven Effekt auf die Leistungen von Schülerinnen und Schülern im Fach Mathematik, jedoch nicht im Lesen (vgl. Song et al., 2021). Zusammenfassend deuten neben den anfangs zitierten Survey- und Interviewstudien also auch die vorliegenden (quasi-)experimentellen Befunde daraufhin, dass die Unterrichtsfeedbacks, die Lehrkräfte im Rahmen der TES erhalten, nicht immer wirksam sind.

Ein weiterer Forschungsstrang, von dem auf die Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback durch Schulleitungen geschlossen werden kann, ist die Forschung zur *Wirksamkeit von Teacher Coaching*. Beim *Teacher Coaching* arbeiten Lehrkräfte über einen längeren Zeitraum mit Unterrichtsexpertinnen und -experten zusammen, um ihren Unterricht weiterzuentwickeln. Integraler Bestandteil aller Coaching-Programme sind regelmäßige, sich zyklisch wiederholende Unterrichtsbeobachtungen durch die Unterrichtsexpertinnen und -experten und hieran anschließende Beratungsgespräche, in denen die teilnehmenden Lehrkräfte unter anderem ein Feedback zu ihrem Unterricht erhalten. Vielfach werden die individuellen Beratungsgespräche zudem um weitere Professionalisierungsaktivitäten wie dem Besuch von Fortbildungen oder dem Lesen fachbezogener Literatur ergänzt (vgl. z. B. Allen et al., 2011; Fabiano et al., 2018; Kraft & Blazar, 2017). Die Wirksamkeit von *Teacher Coaching* wurde von Kraft et al. (2018) unlängst in einer Meta-Analyse untersucht, in der 60 Primärstudien mit (quasi-)experimentellem Forschungsdesign berücksichtigt wurden. Coaching wird hier „broadly as all in-service PD [professional development] programs, where coaches or peers observe

teachers' instruction and provide feedback to help them improve" (Kraft et al., 2018, S. 553) definiert. Kraft et al. (2018) berichten im Mittel von großen positiven Effekten ( $d = 0.49$ ) der Coaching-Programme auf das Unterrichtsverhalten von Lehrkräften und von kleinen positiven Effekten ( $d = 0.18$ ) auf den Lernerfolg von Schülerinnen und Schülern. Gleichzeitig zeigte sich aber auch, dass die Wirksamkeit der Coaching-Programme stark variiert: Während einige Programme keinen Effekt hatten, gingen andere sogar mit negativen Effekten einher (vgl. Kraft et al., 2018, S. 562).

Der Überblick über den Forschungsstand zeigt, dass beobachtungs-basierte Unterrichtsfeedbacks durch Schulleitungen häufig, aber nicht immer zu einer höheren Unterrichtsqualität und besseren Leistungen von Schülerinnen und Schülern führen. Vor dem Hintergrund des inkonsistenten Forschungsstands wird im folgenden Kapitel das auf Grundlage der arbeitspsychologischen Feedbackforschung entwickelte theoretische Feedbackmodell vorgestellt, das den Prozess eines beobachtungs-basierten Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte beschreibt und Bedingungen benennt, die die Feedbackwirksamkeit beeinflussen. Einige dieser Wirksamkeitsbedingungen wurden bei der Entwicklung der Feedbackintervention berücksichtigt und im Rahmen der empirischen Teilstudien untersucht.

### **2.3 Wirksamkeitsbedingungen beobachtungs-basierter Unterrichtsfeedbacks durch Schulleitungen**

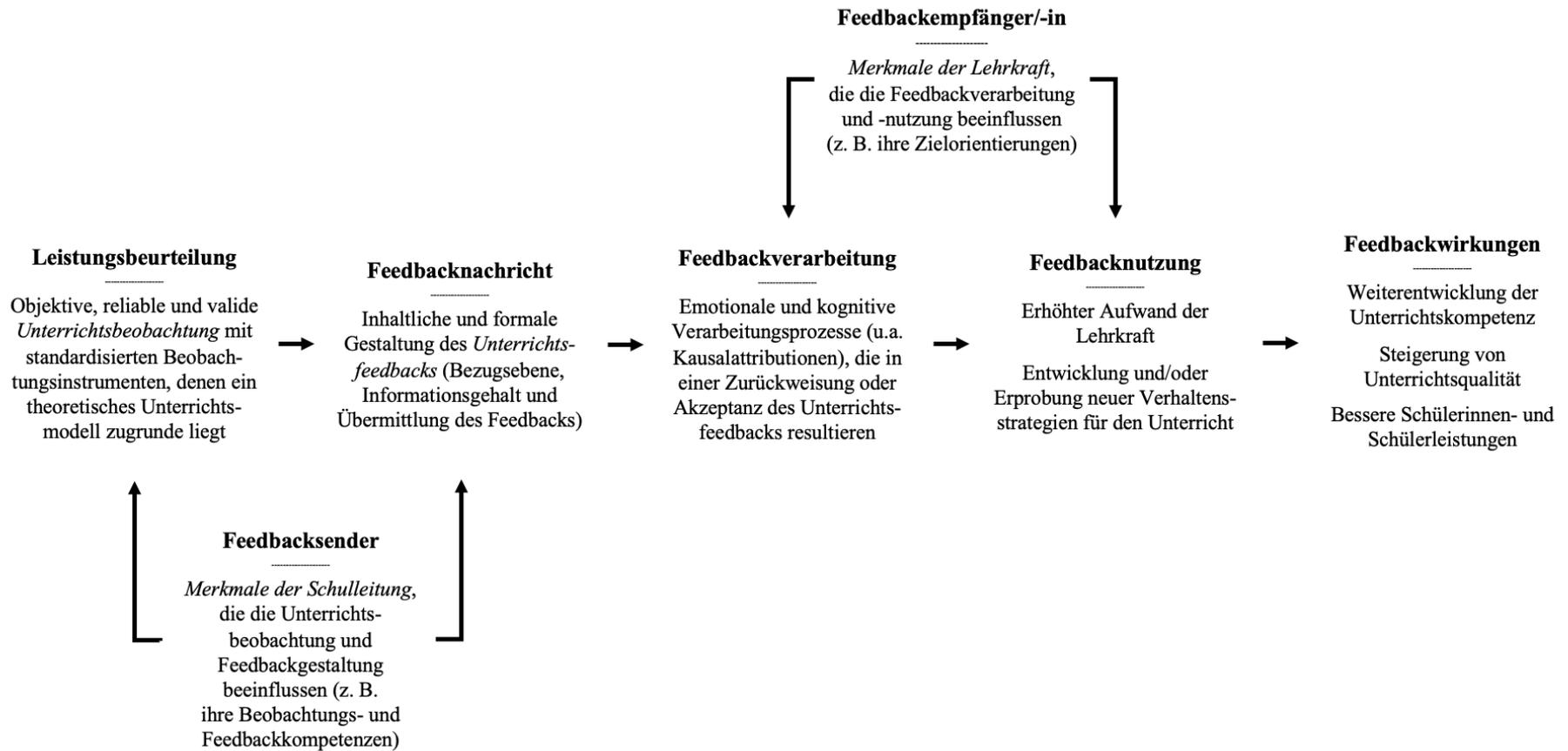
Aus der angewandten arbeitspsychologischen Feedbackforschung liegt eine Vielzahl von theoretischen Feedbackmodellen vor (vgl. Ilgen & Davis, 2000; Ilgen et al., 1979; Kluger & DeNisi, 1996; Smither & London, 2002; Smither et al., 2005a; Strijbos & Müller, 2014), die in der empirischen Schul- und Unterrichtsforschung bislang nur in Ansätzen rezipiert und noch nicht systematisch auf den Kontext eines Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte übertragen wurden (vgl. z. B. Cherasaro et al., 2016; Delvaux et al., 2013; Kraft & Christian, 2022; Tuytens & Devos, 2014, 2017). Im Folgenden werden Gemeinsamkeiten dieser Feedbackmodelle herausgearbeitet und die Aspekte hervorgehoben, die für die vorliegende Arbeit von besonderem Interesse sind.

In Anlehnung an Sender-Nachricht-Empfänger-Modelle unterscheiden alle Modelle – mal implizit (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996; London & Smither, 2002; Smither et al., 2005a), mal explizit (vgl. Ilgen et al. 1979; Strijbos & Müller, 2014) – zwischen einem Feedbacksender, der Feedbacknachricht und einer Feedbackempfängerin bzw. einem Feedbackempfänger. Der Feedbacksender sammelt die für die

Beurteilung einer Arbeitsleistung relevanten Informationen und gestaltet auf der Grundlage der Leistungsbeurteilung die Feedbacknachricht. Im Hinblick auf die Gestaltung der Feedbacknachricht kann zwischen ihrem *Inhalt*, der sich unmittelbar aus der Leistungsbeurteilung ergibt, und ihrer *Form* unterschieden werden. Sowohl die Leistungsbeurteilung als auch die formale Gestaltung der Feedbacknachricht werden von verschiedenen Merkmalen des Feedbacksenders beeinflusst (vgl. Ilgen et al., 1979; Strijbos & Müller, 2014).

Die Feedbacknachricht wird von den Feedbackempfängerinnen und -empfängern verarbeitet. Die Feedbackverarbeitung umfasst emotionale und kognitive Prozesse, die zu einer Zurückweisung oder Akzeptanz des Feedbacks führen. Vorausgesetzt das Feedback wird akzeptiert, entsteht Handlungsmotivation, die in einer Nutzung des Feedbacks und in der Folge idealerweise in einer Leistungssteigerung resultieren kann (Feedbackwirkungen). Die Feedbackverarbeitung und -nutzung können über die formale Gestaltung des Feedbacks beeinflusst werden (vgl. Kluger & DeNisi, 1996; Smither et al., 2005a). Außerdem hängen sie von verschiedenen Merkmalen der Empfängerinnen und Empfänger ab. In der arbeitspsychologischen Feedbackforschung häufig diskutiert werden unter anderem die *Big-Five*-Persönlichkeitsmerkmale und Selbstwirksamkeitserwartungen von Feedbackempfängerinnen und -empfängern, insbesondere aber ihre Zielorientierungen (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996; Smither et al., 2005a; Strijbos & Müller, 2014).

Das auf Grundlage der bisherigen Ausführungen entwickelte Feedbackmodell ist in Abb. 1 dargestellt. Die fettgedruckten Begriffe sind der arbeitspsychologischen Feedbackforschung entnommen. Die dazugehörigen Erläuterungen beziehen sich auf ein beobachtungsbasiertes Unterrichtsfeedback durch Schulleitungen an Lehrkräfte und benennen dessen Wirksamkeitsbedingungen. Sie werden in den folgenden Kapiteln näher beschrieben.



**Abb. 1:** Prozessmodell eines beobachtungs-basierten Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte

### 2.3.1 Unterrichtsbeobachtung

Am Anfang des in Abb. 1 dargestellten Feedbackmodells steht die Leistungsbeurteilung, die im Fall eines beobachtungs-basierten Unterrichtsfeedbacks durch Schulleitungen über eine oder mehrere Unterrichtsbeobachtungen erfolgt. Als zentrale Voraussetzung für systematische Unterrichtsbeobachtungen können theoretische Unterrichtsmodelle gelten, in denen Merkmale eines lernwirksamen Unterrichts beschrieben werden. Weil Schulleitungen nicht in allen Fächern über fachliche und fachdidaktische Expertise verfügen, beziehen sich diese Modelle in der Regel auf generische, d.h. fach- und jahrgangsstufenunabhängige Qualitätsmerkmale. Die Unterrichtsmodelle bilden die Grundlage für standardisierte Beobachtungsinstrumente, in denen lernwirksamer Unterricht anhand verhaltensnah formulierter Indikatoren operationalisiert wird (vgl. Heneman & Milanowski, 2001; Liu et al., 2019; Milanowski & Heneman, 2001; Papay, 2012).

Standardisierte Beobachtungsinstrumente ermöglichen es Schulleitungen zum einen, den Unterricht ihrer Lehrkräfte objektiv zu beurteilen, weil sie bei der Unterrichtsbeurteilung nicht auf ihre subjektiven Theorien über Unterrichtsqualität zurückgreifen müssen. Gleichzeitig geht mit ihnen die Erwartung einher, die Reliabilität und Validität von Unterrichtsbeobachtungen zu steigern (vgl. Heneman & Milanowski, 2003; Kimball, 2003). Während in den US-amerikanischen TES vor allem der FFT und das CLASS genutzt werden, sind in Deutschland das Modell der drei Basisdimensionen und hierauf basierende Beobachtungsinstrumente weit verbreitet (vgl. z. B. Klieme et al., 2001; Praetorius et al., 2018). Empirische Studien, in denen die Reliabilität und Validität von Unterrichtsbeobachtungen untersucht wurden, die mit den genannten Beobachtungsinstrumenten erfolgen, bestätigen die in sie gesetzten Erwartungen. Demnach können Unterrichtsbeobachtungen durch Schulleitungen bzw. durch externe Beobachterinnen und Beobachter – vorausgesetzt, diese haben zuvor ein Beobachtungstraining durchlaufen – grundsätzlich zu reliablen Urteilen über Unterrichtsqualität führen (vgl. Garrett & Steinberg, 2015; Heneman & Milanowski, 2003; Polikoff, 2015; Sartain et al., 2011). Gleichzeitig liegen aber auch Befunde vor, die darauf hindeuten, dass die Beobachtung einer einzelnen Unterrichtsstunde für eine reliable Beurteilung bei bestimmten Unterrichtsmerkmalen nicht ausreichen könnte und erst mehrere Unterrichtsbeobachtungen zu einer zufriedenstellenden Reliabilität führen (vgl. Hill et al., 2011; Kane & Staiger, 2012; Praetorius et al., 2014; van der Lans et al., 2016). Unterrichtsbeobachtungen mit standardisierten Be-

obachtungsinstrumenten sind jedoch nicht nur reliabel, sie führen darüber hinaus zu validen Urteilen über Unterrichtsqualität: In vielen Studien zeigen sich schwache bis moderate positive Zusammenhänge zwischen Unterrichtsbeobachtungen und den Lernzuwächsen von Schülerinnen und Schülern (vgl. Hill et al., 2012; Fauth et al., 2014; Garret & Steinberg, 2015; Kane & Staiger, 2012; Kane et al., 2011; Sartain et al., 2011).

Hinsichtlich der Wirksamkeitsbedingungen eines beobachtungsbasierten Unterrichtsfeedbacks lässt sich mit Blick auf die Unterrichtsbeobachtung festhalten, dass diese mit standardisierten Beobachtungsinstrumenten erfolgen sollte, denen ein theoretisches Unterrichtsmodell zugrunde liegt, in dem lernwirksamer Unterricht aus einer fach- und jahrgangsstufenübergreifenden Perspektive beschrieben wird. Weil eine einzelne Unterrichtsbeobachtung für eine reliable Unterrichtsbeurteilung nicht ausreicht, sollten Schulleitungen darüber hinaus nach Möglichkeit mehrere Unterrichtsstunden einer Lehrkraft beobachten.

### 2.3.2 Unterrichtsfeedback

Auf die Leistungsbeurteilung folgt die Feedbacknachricht (siehe Abb. 1). In Bezug auf die Feedbacknachricht kann der *Feedbackinhalt* von der *formalen Gestaltung des Feedbacks* unterschieden werden. Der Inhalt des Feedbacks ergibt sich bei einem beobachtungs-basierten Unterrichtsfeedback aus den Unterrichtsmodellen, die den jeweils eingesetzten Beobachtungsinstrumenten zugrunde liegen. Standardisierte Beobachtungsinstrumente ermöglichen damit nicht nur eine objektive, reliable und valide Unterrichtsbeurteilung, sie stecken auch den inhaltlichen Rahmen für das auf die Beobachtungen folgende Feedback ab. Auch Heneman und Milanowski (2003, S. 174) schlussfolgern dass „[s]pecific, behaviorally defined rating scales could (...) provide a better framework for the feedback from the evaluator that teachers will find useful to improve their practice“. Befunde von Englert und Sugai (1983) deuten darauf hin, dass Unterrichtsfeedbacks an Lehrkräfte, die auf Unterrichtsbeobachtungen mit standardisierten Beobachtungsinstrumenten basieren, tatsächlich wirksamer sein könnten als Feedbacks, die auf nicht standardisierten Beobachtungen basieren. Interviews mit Lehrkräften und Schulleitungen zeigen zudem, dass standardisierte Beobachtungsinstrumente im Vergleich zu einfachen Checklisten zu einer besseren Strukturierung des Unterrichtsfeedbacks, einem geteilten

Begriffsverständnis von Schulleitungen und Lehrkräften sowie einer stärkeren Fokussierung auf Unterrichtsqualität führen können (vgl. Kimball, 2003; Kraft & Gilmour, 2016; Sartain et al., 2011).

Die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks hängt jedoch nicht allein vom Feedbackinhalt ab, sondern auch von der formalen Feedbackgestaltung, die in arbeitspsychologischen Feedbacktheorien zu den zentralen Wirksamkeitsbedingungen von Feedback gezählt wird (vgl. Kluger & DeNisi, 1996; Smither et al., 2005a). Weil aus der empirischen Schul- und Unterrichtsforschung nur wenige Befunde dazu vorliegen, wie ein qualitativ hochwertiges und wirksames Feedback durch Schulleitungen gestaltet sein sollte, wurde die formale Gestaltung von Unterrichtsfeedbacks von einigen Autorinnen und Autoren dezidiert als Forschungsdesiderat benannt (vgl. Cohen & Goldhaber, 2016; Hill & Grossmann, 2013; Hunter & Springer, 2022; Kraft & Christian, 2022; Liu et al., 2019). Nachfolgend wird ausgehend von Forschungsarbeiten der arbeits- und instruktionspsychologischen Feedbackforschung erarbeitet, was ein qualitativ hochwertiges und wirksames Unterrichtsfeedback durch Schulleitungen auszeichnet. Die erarbeiteten Gestaltungsmerkmale bilden die Grundlage des im Rahmen der Dissertation entwickelten Feedbackverfahrens. In Bezug auf die formale Feedbackgestaltung wird im Folgenden zwischen der Bezugsebene des Feedbacks (Kapitel 2.3.2.1), dem Informationsgehalt des Feedbacks (Kapitel 2.3.2.2) sowie der Feedbackübermittlung (Kapitel 2.3.2.3) unterschieden (vgl. Kopp & Mandl, 2014).

### **2.3.2.1 Bezugsebene des Feedbacks**

Im Hinblick auf die Bezugsebene des Feedbacks unterscheiden Kluger und DeNisi (1996) in ihrer einflussreichen *Feedback Intervention Theory* (FIT) zwischen drei Ebenen, auf die ein Feedback ausgerichtet sein kann. Demnach kann Feedback die aufgabenbezogene Motivation (*task-motivation*), das aufgabenbezogene Lernen (*task-learning*) sowie übergeordnete Metakognitionen (*meta-task*) ansprechen. Während Feedback, das sich auf die aufgabenbezogene Motivation bezieht, nach den Annahmen der FIT zu einer erhöhten Anstrengungsbereitschaft führen kann, zieht ein Feedback, welches das aufgabenbezogene Lernen adressiert, die Entwicklung neuer Lösungsansätze bei der Aufgabebearbeitung nach sich. Feedback, das auf übergeordnete Metakognitionen ausgerichtet ist, führt nach den Annahmen der FIT demgegenüber häufig zu selbstbezogenen Kognitionen und geht mit negativen Emotionen einher, was zur Folge hat, dass den Feedbackempfängerin-

nen und -empfängern weniger kognitive Ressourcen für die eigentliche Aufgabenbearbeitung zur Verfügung stehen. Kluger und DeNisi (1996) konnten im Rahmen einer umfassenden Meta-Analyse über 131 Primärstudien mit (quasi-)experimentellem Design zeigen, dass Feedback zur aufgabenbezogenen Motivation, vor allem aber Feedback, das sich auf die Ebene des aufgabenbezogenen Lernens bezieht, positive Effekte auf Lernen und die Leistungserbringung hat. Feedback, das übergeordnete Metakognitionen adressiert, verringert demgegenüber die Feedbackwirksamkeit und führt teilweise sogar zu negativen Effekten (vgl. Kluger & DeNisi, 1996).<sup>7</sup> Für ein beobachtungs-basiertes Unterrichtsfeedback durch Schulleitungen lässt sich schlussfolgern, dass es einen hohen Aufgabenbezug aufweisen sollte, d.h. dass es sich auf das Unterrichtsverhalten von Lehrkräften beziehen sollte. Ein personenbezogenes Feedback, das auf Seiten der Lehrkräfte zu selbstbezogenen Kognitionen und negativen Emotionen führen kann, ist demgegenüber zu vermeiden. Ein hoher Unterrichtsbezug lässt sich unter anderem über den Informationsgehalt des Feedbacks herstellen.

### 2.3.2.2 Informationsgehalt des Feedbacks

Der Informationsgehalt des Feedbacks beschreibt zum einen, welche Informationen in einem Feedback enthalten sind und zum anderen den Elaborationsgrad dieser Informationen. Hattie und Timperley (2007) unterscheiden zwischen drei Informationen, über die sich ein hoher Aufgabenbezug herstellen lässt. Diese drei Informationen bezeichnen sie als FEED UP, FEED BACK und FEED FORWARD. Sie korrespondieren mit drei Fragen, die in einem Feedback adressiert werden sollten: *Where am I going?*, *How am I going?* und *Where to next?*. Die Frage *Where am I going?* bezieht sich auf die Ziele, die mit Lernen oder einer Leistungserbringung verbunden sind und den Feedbackempfängerinnen und -empfängern erläutert werden sollten. Feedback sollte darüber hinaus Informationen zur Frage *How am I going?* enthalten und aufzeigen, inwiefern ein aktueller Lern- bzw. Leistungsstand diesen Zielen entspricht. Abschließend sollte Feedback die

---

<sup>7</sup> Auch Hattie und Timperley (2007) unterscheiden Feedback, das sich auf die aufgabenbezogene Motivation (*task level*) und auf das aufgabenbezogene Lernen (*process level*) bezieht von Feedback, das übergeordnete Metakognitionen adressiert (*self level*). Zusätzlich diskutieren sie Feedback, das die bei einer Aufgabenbearbeitung involvierten selbstregulativen Prozesse anspricht (*self-regulation level*). Zwar bezieht sich letztgenanntes Feedback auch auf übergeordnete Metakognitionen, allerdings auf diejenigen, die zu einer effektiven Verhaltensregulation bei der Aufgabenbearbeitung führen können. Mit negativen Emotionen, selbstbezogenen Gedanken und einem Abzug von kognitiven Ressourcen von der Aufgabenbearbeitung ist bei dieser Feedbackart nicht zu rechnen. Dementsprechend gehen Hattie und Timperley (2007) davon aus, dass zusätzlich zum *task level*- und *process level*-Feedback auch Feedback, welches sich auf selbstregulative Prozesse bezieht, wirksam ist.

Frage *Where to next?* beantworten und den Feedbackempfängerinnen und -empfängern aufzeigen, wie die vorab definierten Ziele in Zukunft erreicht werden können.

***Where am I going?*** Für ein Unterrichtsfeedback durch Schulleitungen lässt sich aus der Feedbacktheorie von Hattie und Timperley (2007) schlussfolgern, dass Lehrkräften zunächst das der Unterrichtsbeurteilung zugrundeliegende Unterrichtsmodell erläutert werden sollte, um ihnen die Bewertungskriterien des Feedbacks transparent zu machen (FEED UP). Der hier beschriebene lernwirksame Unterricht stellt eine sachliche Bezugsnorm<sup>8</sup> dar, die die Aufmerksamkeit der Lehrkräfte auf die aufgabenbezogene Motivation bzw. das aufgabenbezogene Lernen lenken kann (vgl. Kopp & Mandl, 2014). Soziale Vergleiche mit anderen Lehrkräften sind demgegenüber zu vermeiden, weil sie mit einem hohen Personenbezug einhergehen (vgl. Shute, 2008). Darauf, dass transparente Bewertungskriterien die Wirksamkeit eines Unterrichtsfeedbacks von Schulleitungen erhöhen können, deuten auch Befunde aus der empirischen Schul- und Unterrichtsforschung hin. Demnach sind Lehrkräfte mit dem Beurteilungsprozess zufriedener (vgl. Kelly et al., 2008) und sie nehmen ein beobachtungs-basiertes Unterrichtsfeedback als fairer wahr, wenn ihnen das Unterrichtsmodell bekannt ist (vgl. Milanowski & Heneman, 2001). Außerdem ergreifen Lehrkräfte in Folge eines Unterrichtsfeedbacks häufiger Maßnahmen der professionellen Weiterentwicklung, wenn ihnen die Bewertungskriterien des Feedbacks bekannt sind (vgl. Delvaux et al., 2013).

***How am I going?*** Neben den Informationen zum Unterrichtsmodell sollte ein Unterrichtsfeedback von Schulleitungen einen Vergleich des in einer beobachteten Unterrichtsstunde gezeigten Verhaltens mit den im Modell beschriebenen lernwirksamen Verhaltensweisen enthalten (FEED BACK). Grundsätzlich kann zwischen einem bestätigenden und einem kritischen Feedback unterschieden werden. Während eine Lehrkraft bei einem bestätigenden Unterrichtsfeedback zurückgemeldet wird, einzelne oder alle der im Unterrichtsmodell beschriebenen lernwirksamen Verhaltensweisen in den beobachteten Unterrichtsstunden gezeigt zu haben, konnten diese bei einem kritischen Feedback gar

---

<sup>8</sup> In Anlehnung an Rheinberg (2006, 2008) lassen sich drei Referenzgrößen unterscheiden, auf die in einem Feedback Bezug genommen werden kann: eine sachliche, eine individuelle und eine soziale Bezugsnorm. Bei einer sachlichen Bezugsnorm wird eine erbrachte (Lern-)Leistung mit einem vorab definierten Standard verglichen. Bei der individuellen Bezugsnorm wird der aktuelle Lern- oder Leistungsstand von Feedbackempfängerinnen und -empfängern mit einem früheren Lern- oder Leistungsstand verglichen (intraindividueller Vergleich). Bei einer sozialen Bezugsnorm wiederum wird der Lern- oder Leistungsstand von Feedbackempfängerinnen und -empfängern mit dem Lern- oder Leistungsstand anderer verglichen (interindividueller Vergleich).

nicht oder kaum beobachtet werden. Befunde aus der arbeits- und instruktionspsychologischen Feedbackforschung zeigen, dass sowohl bestätigende als auch kritische Feedbacks einen positiven Effekt auf Lernen und die Leistungserbringung haben können (vgl. Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Auch eine Studie aus der empirischen Schul- und Unterrichtsforschung deutet darauf hin, dass eine Kombination aus bestätigendem und kritischem Feedback an Lehrkräfte wirksamer ist als ausschließlich bestätigendes Feedback (vgl. Wang, 2007). Schulleitungen sollten ihren Lehrkräften folglich unterrichtliche Stärken, aber auch unterrichtliche Schwächen zurückmelden.

Ein bestätigendes Feedback geht jedoch nur dann mit positiven Effekten einher, wenn es sich auf konkrete Lernerträge oder Leistungen bezieht. Auf einfaches Lob, das häufig einen hohen Personenbezug aufweist, sollte demgegenüber verzichtet werden (vgl. Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Sowohl bei einem bestätigenden als auch bei einem kritischen Feedback erhöhen zudem Erläuterungen, die es den Feedbackempfängerinnen und -empfängern ermöglichen, die Leistungsbeurteilung nachzuvollziehen, die Feedbackwirksamkeit (Shute, 2008). Auch ein Unterrichtsfeedback durch Schulleitungen sollte daher anhand konkreter Beispiele aus der beobachteten Unterrichtsstunde begründet werden. Empirische Evidenz für diese Annahme liefert eine aktuelle Studie von Kraft und Christian (2022), die zeigt, dass Lehrkräfte ihren Unterricht vor allem dann weiterentwickeln, wenn Schulleitungen ihr Unterrichtsfeedback auf zuvor beobachtete Verhaltensweisen beziehen (siehe auch Scheeler et al., 2004; Thurlings et al., 2013).

*Where to next?* Zusätzlich zu den Informationen zum Unterrichtsmodell und zur Beurteilung des beobachteten Unterrichtsverhaltens sollte ein Unterrichtsfeedback durch Schulleitungen nach der Feedbacktheorie von Hattie und Timperley (2007) Hinweise auf alternative Verhaltensstrategien enthalten (FEED FORWARD). Befunde aus der arbeits- und instruktionspsychologischen Feedbackforschung zeigen, dass Hinweise zur korrekten Aufgabenbearbeitung zu höheren Lernerträgen und besseren Leistungen führen (vgl. Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Auch aus der empirischen Schul- und Unterrichtsforschung liegen Befunde vor, die nahelegen, dass Unterrichtsfeedbacks von Schulleitungen dann besonders wirksam sind, wenn den Lehrkräften aufgezeigt wird, wie sie ihren Unterricht im Hinblick auf die identifizierten Schwächen weiterentwickeln können (vgl. Cherasaro et al., 2016; Kraft & Christian, 2022; Liu et al., 2019; Scheeler et al., 2004; Thurlings et al., 2013).

In einem engen Zusammenhang mit Hinweisen auf die korrekte Bearbeitung einer Aufgabe stehen Zielsetzungen, in denen definiert wird, *welche* Lernerträge bzw. Leistungen *wie* realisiert werden sollen. Nach der Zielsetzungstheorie von Locke und Latham (2002) sollten die im Anschluss an das Feedback gesetzten Ziele dabei möglichst konkret sein, sie sollten die Feedbackempfängerinnen und -empfänger zwar herausfordern, sie gleichzeitig aber nicht überfordern. Auch für Zielsetzungen liegt aus der arbeits- und instruktionspsychologischen Feedbackforschung Evidenz dafür vor, dass sie die Wirksamkeit von Feedback erhöhen (vgl. Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Darüber hinaus zeigt auch eine aktuelle Studie aus der empirischen Schul- und Unterrichtsforschung, dass Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte besonders wirksam sind, wenn sie um Zielsetzungen ergänzt werden (vgl. Hunter & Springer, 2022).

### 2.3.2.3 Feedbackübermittlung

Zur formalen Feedbackgestaltung kann neben der Bezugsebene und dem Informationsgehalt des Feedbacks auch die Feedbackübermittlung gezählt werden. In diesem Zusammenhang ist zum einen von Interesse, *wann* Feedback gegeben wird und zum anderen, *wie* die Feedbackgabe erfolgt, d.h. ob das Feedback mündlich oder schriftlich bzw. computerbasiert übermittelt wird. Im Hinblick auf den Zeitpunkt des Feedbacks kann zwischen einem sofortigen Feedback, bei dem die Feedbackgabe unmittelbar im Anschluss oder sogar noch während der Leistungserbringung erfolgt und einem zeitlich verzögerten Feedback unterschieden werden. Da eine sofortige Feedbackgabe im Falle eines beobachtungsbasierten Unterrichtsfeedbacks mit einer Unterbrechung des Unterrichts verbunden wäre, erhalten Lehrkräfte in der Regel ein zeitlich verzögertes Feedback. Befunde aus der empirischen Schul- und Unterrichtsforschung zeigen, dass die Effekte eines Unterrichtsfeedbacks an Lehrkräfte größer ausfallen, wenn die Feedbackgabe möglichst frühzeitig erfolgt, sodass sich Lehrkräfte noch gut an die beobachtete Unterrichtsstunde erinnern können (vgl. Cherasaro et al., 2016; Kraft & Christian, 2022; Liu et al., 2019; Scheeler et al., 2004; Thurlings et al., 2013).

Im Hinblick auf die Art der Feedbackübermittlung geht sowohl ein mündliches als auch ein schriftliches bzw. computerbasiertes Feedback mit Vor- und Nachteilen einher. Bei einem mündlichen Feedback können Feedbackempfängerinnen und -empfänger bei Verständnisschwierigkeiten Nachfragen stellen. Außerdem stehen ihnen non- und para-

verbale Hinweise der Feedbacksender zur Verfügung. Ein schriftliches bzw. computerbasiertes Feedback hat demgegenüber den Vorteil, dass es von den Empfängerinnen und Empfängern in ihrem individuellen Tempo und mehrmals gelesen werden kann (vgl. Kopp & Mandl, 2014). Befunde aus der arbeits- und instruktionspsychologischen Feedbackforschung zeigen, dass schriftliche bzw. computerbasierte Feedbacks wirksamer sind als mündliche Feedbacks (vgl. Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Gleichzeitig liegen aus der empirischen Schul- und Unterrichtsforschung aber Befunde vor, nach denen ein mündliches Unterrichtsfeedback an Lehrkräfte wirksamer sein könnte als ein schriftliches Feedback (vgl. Kraft & Christian, 2022; Liu et al., 2019). Sowohl Schulleitungen als auch Lehrkräfte berichten in Interviewstudien zudem davon, ein mündliches Feedback zu schätzen, da es einen Dialog ermöglichen und sich auch Lehrkräfte aktiv am Feedbackgespräch beteiligen könnten (vgl. Kimball, 2003; Kraft & Gilmour, 2016; Tuytens & Devos, 2017).

Zusammenfassend lässt sich mit Blick auf die formale Feedbackgestaltung festhalten, dass die Wirksamkeit eines Unterrichtsfeedbacks über einen hohen Bezug zum Unterrichtsverhalten von Lehrkräften gesteigert werden kann. Ein hoher Unterrichtsbezug lässt sich zum einen über die im Feedback enthaltenen Informationen (FEED UP, FEED BACK und FEED FORWARD) und zum anderen über den Elaborationsgrad dieser Informationen erreichen (z. B. Begründung des Feedbacks anhand konkreter Verhaltensbeispiele). Hinsichtlich der Feedbackübermittlung lässt sich schlussfolgern, dass Lehrkräfte das Unterrichtsfeedback zeitnah erhalten sollten. Außerdem könnte sich eine Kombination aus einem mündlichen und einem schriftlichen bzw. computerbasierten Feedback als vielversprechend erweisen.

### **2.3.3 Merkmale der Schulleitung**

Wie im theoretischen Feedbackmodell dargestellt (siehe Abb. 1), wird sowohl die Leistungsbeurteilung als auch die formale Gestaltung der Feedbacknachricht von Merkmalen des Feedbacksenders beeinflusst. Im Fall eines beobachtungsbasierten Unterrichtsfeedbacks durch Schulleitungen sind insbesondere Beobachtungs- und Feedbackkompetenzen von Schulleitungen von Relevanz: Schulleitungen müssen zum einen die Qualität der von Lehrkräften gestalteten unterrichtlichen Lerngelegenheiten objektiv, reliabel und valide beurteilen können und den Lehrkräften zum anderen ein Feedback geben können, „that is

personally meaningful and classroom instruction relevant“ (Ovando, 2005, S. 172). Beides kann durch ein Beobachtungs- bzw. Feedbacktraining erreicht werden.

Bei Unterrichtsbeobachtungen, die mit standardisierten Beobachtungsinstrumenten erfolgen, kann ein umfassendes Wissen über die theoretischen Unterrichtsmodelle als grundlegende Voraussetzung für eine akkurate Unterrichtsbeurteilung gelten. Aus diesem Grund werden Schulleitungen im Rahmen der reformierten US-amerikanischen TES mit den in den jeweiligen Unterrichtsmodellen beschriebenen lernwirksamen Unterrichtsmerkmalen vertraut gemacht und ihnen wird vermittelt, wie sich der Unterricht einer Lehrkraft anhand der in den Beobachtungsinstrumenten operationalisierten Verhaltensindikatoren differenziert beurteilen lässt. Darüber hinaus werden Schulleitungen häufig für Beobachtungsfehler (z. B. Konsistenz- und Erwartungseffekte) sensibilisiert, die ihre Unterrichtsbeurteilungen systematisch verzerren können und lernen Strategien kennen, mit denen sich diese minimieren lassen. Schließlich soll die Akkuratheit von Unterrichtsbeobachtungen in den Beobachtungstrainings der TES auch über sogenannte Kalibrierungsübungen gesteigert werden, bei denen Schulleitungen ihre Beurteilungen mit den Urteilen von anderen Schulleitungen oder von Unterrichtsexpertinnen und -experten abgleichen und diskutieren können (vgl. Milanowski & Heneman, 2001; Steinberg & Sartain, 2015). Die in Kapitel 2.3.1 berichteten Befunde zur Reliabilität und Validität von Unterrichtsbeobachtungen durch geschulte Schulleitungen bzw. externe Beobachterinnen und Beobachter lassen darauf schließen, dass die für eine akkurate Beurteilung von Unterrichtsqualität benötigten Kompetenzen über ein derart gestaltetes Training vermittelt werden können.

Dass Schulleitungen nicht nur ein Beobachtungs-, sondern auch ein Feedbacktraining durchlaufen sollten, lässt sich auch an der von Lehrkräften häufig bemängelten Feedbackqualität erkennen (vgl. z. B. Kimball, 2003; Kimball & Milanowski, 2009; Liu et al., 2015; Weisberg et al., 2009). Während die meisten US-amerikanischen TES ein Beobachtungstraining vorsehen, wird die Feedbackkompetenz von Schulleitungen nur selten systematisch gefördert. Nach meinem Kenntnisstand liegt nur eine Studie vor, in der ein Feedbacktraining entwickelt und evaluiert wurde, über das die Qualität und die Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback von Schulleitungen gesteigert werden sollte (vgl. Kraft & Christian, 2022). Der Fokus des Trainings lag auf „practical strategies for prioritizing time for the observation and feedback process, creating a shared vision for effective practice, helping teachers to identify actionable steps for improvement

and communicating with teachers in ways that promote positive interpersonal relationships and mutual trust“ (Kraft & Christian, 2022, S. 508). Insgesamt lässt die Beschreibung des Feedbacktrainings jedoch nur wenig Rückschlüsse auf seine konkrete Ausgestaltung zu. Die in Kapitel 2.3.2 aufgearbeiteten Befunde zu den Gestaltungsmerkmalen eines qualitativ hochwertigen Feedbacks aus der arbeits- und instruktionspsychologischen Feedbackforschung wurden bei der Entwicklung des Feedbacktrainings aber offenbar nicht berücksichtigt, was einen Erklärungsansatz für die ernüchternden Befunde der Studie darstellen könnte: Es zeigte sich, dass Lehrkräfte, die ein Unterrichtsfeedback von trainierten Schulleitungen erhielten, die Feedbackqualität nicht höher einschätzten als Lehrkräfte, die ein Unterrichtsfeedback von untrainierten Schulleitungen erhielten. Außerdem hatte das Feedbacktraining keine Effekte auf die über Unterrichtsbeobachtungen erfasste Unterrichtsqualität und die Lernzuwächse von Schülerinnen und Schülern (vgl. Kraft & Christian, 2022).

Mit Blick auf die Merkmale der Schulleitung lässt sich schlussfolgern, dass sowohl ihre Beobachtungs- als auch ihre Feedbackkompetenzen über ein Beobachtungs- bzw. Feedbacktraining gefördert werden sollten. Während sich die Beobachtungstrainings aus den US-amerikanischen TES empirisch bewährt haben, ist über die Gestaltung eines wirksamen Feedbacktrainings bislang nur wenig bekannt. Erfolgversprechend könnte es sein, Schulleitungen mit den im vorherigen Kapitel beschriebenen Merkmalen eines qualitativ hochwertigen Feedbacks vertraut zu machen.

#### **2.3.4 Feedbackverarbeitung**

Auf den Erhalt der Feedbacknachricht folgt die Verarbeitung des Feedbacks durch die Lehrkraft (siehe Abb. 1). Nach den Theorien der arbeitspsychologischen Feedbackforschung führen die hierbei involvierten emotionalen und kognitiven Prozesse zu einer Zurückweisung oder Akzeptanz des Feedbacks (vgl. Ilgen & Davis, 2000; Ilgen et al., 1979; Kluger & DeNisi, 1996; London & Smither, 2002; Smither et al., 2005a; Strijbos & Müller, 2014). Die Akzeptanz des Feedbacks stellt eine zentrale Voraussetzung dafür dar, dass die im Feedback enthaltenen Informationen auch genutzt werden. Die Feedbackverarbeitung wird dabei sowohl von der formalen Gestaltung des Unterrichtsfeedbacks als auch von verschiedenen Merkmalen der Lehrkraft beeinflusst.

Im Folgenden werden zunächst emotionale Reaktionen thematisiert, die die Feedbackverarbeitung begleiten und darüber mitentscheiden, ob Empfängerinnen und Empfänger ein Feedback akzeptieren oder zurückweisen (Kapitel 2.3.4.1). Mit Blick auf die kognitive Feedbackverarbeitung wird einerseits beschrieben, wie wechselseitige Prozesse der Diskrepanz*produktion* und Diskrepanz*reduktion* nach Erhalt eines Feedbacks Handlungsmotivation erzeugen (Kapitel 2.3.4.2). Andererseits wird auf Kausalattributionen als einem weiteren zentralen Aspekt der kognitiven Feedbackverarbeitung eingegangen (Kapitel 2.3.4.3).

#### **2.3.4.1 Emotionale Reaktionen**

Am Beginn der Feedbackverarbeitung stehen nach den Annahmen vieler arbeitspsychologischer Feedbacktheorien zunächst emotionale Reaktionen auf den Feedbackerhalt (vgl. Ilgen & Davis, 2000; London & Smither, 2002; Smither et al., 2005a). Es wird erwartet, dass Feedbackempfängerinnen und -empfänger auf ein bestätigendes Feedback mit positiven Emotionen wie z. B. Freude und Zufriedenheit reagieren, was zu einer Akzeptanz des Feedbacks führen sollte. Weil ein kritisches Feedback potenziell eine Gefahr für den Selbstwert der Empfängerinnen und Empfänger darstellt, kann es demgegenüber negative emotionale Reaktionen hervorrufen und in einer Zurückweisung des Feedbacks resultieren (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996).

Während die emotionalen Reaktionen von Lehrkräften auf den Erhalt von Unterrichtsfeedback nach meinem Kenntnisstand noch nicht untersucht wurden, deutet eine Studie aus der arbeitspsychologischen Feedbackforschung darauf hin, dass ein kritisches Feedback tatsächlich negative Emotionen nach sich zieht: Führungskräfte, die nach ihrem emotionalen Befinden nach Erhalt eines 360°-Feedbacks<sup>9</sup> gefragt wurden, berichten bei einem kritischen Feedback von Wut und Entmutigung; bei einem bestätigenden Feedback empfanden sie entgegen den Erwartungen jedoch keine positiven Emotionen (vgl. Brett & Atwater, 2001). Die Befunde einer weiteren Studie zeigen, dass negative emotionale Reaktionen von Führungskräften darüber hinaus mit einem Leistungsabfall ein Jahr nach einem 360°-Feedback einhergehen – und positive Reaktionen mit einem Leistungsanstieg (vgl. Atwater & Brett, 2005). Smither et al. (2005a, S. 48) schlussfolgern mit Blick auf

---

<sup>9</sup> Bei einem 360°-Feedback (oder *Multi-Rater-Feedback*), das als Maßnahme der individualisierten Personalentwicklung vor allem im betrieblichen Kontext Anwendung findet, erhalten Arbeitnehmerinnen und -nehmer von mehreren Personen ein Feedback zu ihren Arbeitsleistungen, z. B. von ihren Vorgesetzten, ihren Mitarbeiterinnen und Mitarbeitern, von Kolleginnen und Kollegen und/oder von Kundinnen und Kunden (für einen Überblick siehe z. B. London & Smither, 1995; Smither et al., 2005a).

diese Befunde, dass „[t]hese findings are important because they demonstrate that immediate reactions to feedback are not merely transitory mood states without relevance to subsequent behavior.“ Es kann davon ausgegangen werden, dass auch Lehrkräfte ein Feedback zu ihrem Unterricht zunächst emotional verarbeiten und dass die emotionalen Reaktionen den weiteren Verarbeitungsprozess und die Nutzung der im Feedback enthaltenen Informationen beeinflussen.

#### **2.3.4.2 Diskrepanzproduktion und Diskrepanzreduktion**

Auf die anfänglichen emotionalen Reaktionen folgt die kognitive Feedbackverarbeitung (vgl. Ilgen et al., 1979; London & Smither, 2002; Smither et al., 2005a), die unter anderem wechselseitige Prozesse der *Diskrepanzproduktion* und *Diskrepanzreduktion* umfasst. Unter Bezugnahme auf systemtheoretische Ansätze (vgl. z. B. Carver & Scheier, 1982; Miller et al., 1960) wird in der arbeitspsychologischen Feedbackforschung davon ausgegangen, dass menschliches Verhalten zielorientiert ist und auf der Grundlage von Soll-Ist-Vergleichen reguliert wird (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996; London & Smither, 2002). Demnach können Soll-Ist-Vergleiche auf Zielabweichungen aufmerksam machen, d.h. Soll-Ist-Diskrepanzen aufdecken, die in der Folge handlungsleitend werden. Auch Feedbacks enthalten Soll-Ist-Vergleiche, die Handlungsmotivation erzeugen können: Im Falle eines beobachtungsbasierten Unterrichtsfeedbacks von Schulleitungen definieren die in den Unterrichtsmodellen beschriebenen lernwirksamen Merkmale die Ziele bzw. den Soll-Standard. Den Ist-Stand stellt das beobachtete Unterrichtsverhalten der Lehrkraft dar. Macht ein kritisches Unterrichtsfeedback eine Lehrkraft auf eine Diskrepanz zwischen dem im Unterrichtsmodell definierten Soll-Standard und ihrem aktuellen Verhalten aufmerksam, sollte sie nach den Annahmen der arbeitspsychologischen Feedbacktheorien darum bemüht sein, die im Feedback enthaltene Soll-Ist-Diskrepanz zu reduzieren. Während ein kritisches Feedback also eine Soll-Ist-Diskrepanz enthält, ist bei einem bestätigenden Unterrichtsfeedback vorstellbar, dass der *Diskrepanzreduktion* eine *Diskrepanzproduktion* vorausgeht (vgl. Bandura, 1991; Bandura & Locke, 2003). Lehrkräfte könnten sich nach einem bestätigenden Feedback demnach höhere Ziele für ihren Unterricht setzen und so eine Diskrepanz zwischen ihrem aktuellen und einem angestrebten Unterrichtsverhalten erzeugen, die sie in der Folge versuchen zu verringern.

#### **2.3.4.3 Kausalattributionen**

Ob ein kritisches Feedback akzeptiert und in der Folge handlungsleitend wird, hängt nicht nur von den emotionalen Reaktionen ab. Entscheidend ist auch, auf welche Ursachen

Feedbackempfängerinnen und -empfänger die in einem kritischen Feedback enthaltene Soll-Ist-Diskrepanz zurückführen (vgl. Ilgen & Davis, 2000; Strijbos & Müller, 2014). Auch bei Ursachenzuschreibungen handelt es sich um einen Aspekt der kognitiven Feedbackverarbeitung. Nach der Attributionstheorie von Weiner (1979, 1985) lassen sich die Ursachenzuschreibungen anhand dreier Dimensionen kategorisieren: Der *Fokus* der Attribution beschreibt, ob Feedbackempfängerinnen und -empfänger die Ursachen für eine Soll-Ist-Diskrepanz auf internale oder externale Faktoren zurückführen. Die *Stabilität* der Attribution beschreibt, ob diese Ursachen als stabil oder variabel angesehen werden. Die *Kontrollierbarkeit* der Attribution definiert schließlich, ob Feedbackempfängerinnen und -empfänger der Auffassung sind, die Ursachen für die Soll-Ist-Diskrepanz beeinflussen zu können.

In arbeitspsychologischen Feedbacktheorien wird angenommen, dass kritisches Feedback vor allem dann akzeptiert und in der Folge auch handlungsleitend wird, wenn es auf internale Ursachen zurückgeführt wird, die als variabel und kontrollierbar, d.h. als veränderbar, angesehen werden (vgl. Ilgen & Davis, 2000; London & Smither, 2002; Strijbos & Müller, 2014). Lehrkräfte sollten eine Diskrepanz zwischen dem im Unterrichtsmodell definierten Soll-Standard und ihrem aktuellen Unterrichtsverhalten z. B. auf den investierten Aufwand zurückführen und der Auffassung sein, ihre Unterrichtskompetenz und damit die Qualität der von ihnen angebotenen unterrichtlichen Lerngelegenheiten über einen Mehraufwand verbessern zu können. Auch wenn eine internale Attribution eines kritischen Feedbacks grundsätzlich wünschenswert ist, können internale Attributionen hinderliche Prozesse des Selbstwertschutzes anregen (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996; Strijbos & Müller, 2014). Hiermit ist vor allem dann zu rechnen, wenn Feedbackempfängerinnen und -empfänger ein erhaltenes Feedback zwar internal attribuieren, gleichzeitig aber davon ausgehen, die Ursachen für die aufgezeigte Soll-Ist-Diskrepanz nicht verändern zu können. Lehrkräfte könnten ein kritisches Unterrichtsfeedback z. B. auf ihre berufliche Eignung zurückführen und – um ihren Selbstwert zu schützen – das Feedback zurückweisen.

Im Hinblick auf die Feedbackverarbeitung lässt sich schlussfolgern, dass sowohl ein kritisches als auch ein bestätigendes Unterrichtsfeedback zu Handlungsmotivation führen kann. Allerdings werden Unterrichtsfeedbacks nur dann handlungsleitend, wenn Lehrkräfte das Feedback auch akzeptieren. Die Akzeptanz des Feedbacks stellt mit Blick auf die Feedbackverarbeitung die zentrale Wirksamkeitsbedingung dar. Sie hängt unter

anderem von den emotionalen Reaktionen der Lehrkraft sowie ihren Kausalattributionen ab.

### 2.3.5 Feedbacknutzung

Auf die emotionale und kognitive Verarbeitung des Feedbacks folgt – vorausgesetzt das Feedback wird akzeptiert – die Feedbacknutzung (siehe Abb. 1). In den Theorien der arbeitspsychologischen Feedbackforschung wird zwischen zwei Handlungsoptionen unterschieden, auf die Feedbackempfängerinnen und -empfänger hierbei zurückgreifen können: Zum einen können sie ihren Aufwand erhöhen, zum anderen können sie neue Problemlösestrategien entwickeln und/oder erproben (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996).

Mit der ersten Handlungsoption, einem Mehraufwand, ist vor allem bei einem bestätigenden Feedback und einer hierauf folgenden Diskrepanzproduktion durch die Feedbackempfängerinnen und -empfänger zu rechnen. Sofern sich Lehrkräfte nach einem bestätigenden Unterrichtsfeedback höhere Ziele für ihren Unterricht setzen, sollten sie die erzeugte Diskrepanz zwischen ihrem aktuellen und dem nunmehr angestrebten Unterrichtsverhalten insbesondere versuchen zu reduzieren, indem sie vermehrt diejenigen Verhaltensstrategien anwenden, die sich in der Vergangenheit bewährt haben. Mit der zweiten Handlungsoption, der Entwicklung und/oder Erprobung neuer Problemlösestrategien ist demgegenüber vor allem bei einem kritischen Feedback zu rechnen. Sollte ein kritisches Unterrichtsfeedback bereits Hinweise auf alternative Verhaltensstrategien enthalten, können Lehrkräfte diese unmittelbar in ihrem Unterricht erproben. Der Mehrwert von konkreten Handlungsempfehlungen wird in vielen arbeitspsychologischen Feedbacktheorien betont (vgl. Ilgen & Davis, 2000; Kluger & DeNisi, 1996) und Hinweise auf alternative Verhaltensstrategien stellen eines der in Kapitel 2.3.2 erarbeiteten formalen Gestaltungsmerkmale eines qualitativ hochwertigen und wirksamen Unterrichtsfeedbacks dar.

Sofern das kritische Unterrichtsfeedback keine Hinweise auf alternative Verhaltensstrategien enthält oder sich die enthaltenen Verhaltensempfehlungen in der Unterrichtspraxis nicht bewähren, müssen sich Lehrkräfte selbst um Lösungsansätze bemühen. Befunde aus der empirischen Schul- und Unterrichtsforschung deuten darauf hin, dass Lehrkräfte ausgehend von einem beobachtungsbasierten Unterrichtsfeedback teilweise gezielt Aktivitäten der professionellen Weiterentwicklung ergreifen, um neue Verhaltensstrategien für ihren Unterricht kennenzulernen. Bekannt ist z. B., dass sich Lehrkräfte

aufgrund eines Unterrichtsfeedbacks um Unterstützung von erfahrenen Kolleginnen und Kollegen bemühen und bei diesen Unterrichtshospitationen durchführen (vgl. Cherasaro et al., 2016; Liu et al., 2019). Außerdem zeigen Studien, die die Feedbacknutzung durch Lehrkräfte untersuchen, dass diese den Besuch von Fort- und Weiterbildungen auf ein erhaltenes Feedback abstimmen oder gezielt fachbezogene Literatur lesen, um die Diskrepanz zwischen ihrem aktuellen und einem angestrebten Unterrichtsverhalten zu reduzieren (vgl. Cherasaro et al., 2016; Delvaux et al., 2013; Liu et al., 2019; Tuytens & Devos, 2011; 2014).

Zusammenfassend lässt sich festhalten, dass sowohl ein Mehraufwand als auch die Entwicklung und/oder Erprobung neuer Verhaltensstrategien (ggf. unter Einbezug weiterer Unterstützungsangebote) zu einer Weiterentwicklung der Unterrichtskompetenz von Lehrkräften und einer höheren Unterrichtsqualität führen kann. Wie die Feedbackverarbeitung wird auch die Nutzung der im Feedback enthaltenen Informationen von verschiedenen Merkmalen der Lehrkraft beeinflusst.

### **2.3.6 Merkmale der Lehrkraft**

Zu den in der arbeitspsychologischen Feedbackforschung häufig diskutierten und empirisch gut erforschten Merkmalen der Empfängerinnen und Empfänger, die die Feedbackverarbeitung und -nutzung beeinflussen, zählen vor allem die *Big-Five*-Persönlichkeitsmerkmale, Selbstwirksamkeitserwartungen und Zielorientierungen. Während der Forschungsstand zum Einfluss der *Big-Five*-Persönlichkeitsmerkmale, also der Offenheit, der Gewissenhaftigkeit, der Extraversion, der Verträglichkeit und des Neurotizismus, inkonsistent sind (vgl. Atwater & Brett, 2005; Bell & Arthur, 2008; Smither et al., 2005b; Walker et al., 2010), zeigen zahlreiche Studien, dass hoch ausgeprägte Selbstwirksamkeitserwartungen mit einer effektiveren Feedbackverarbeitung und -nutzung und mit größeren Effekten auf nachfolgende Arbeitsleistungen einhergehen als gering ausgeprägte Selbstwirksamkeitserwartungen (vgl. z. B. Atwater & Brett, 2005; Donovan & Hafsteinson, 2006; Heslin & Latham, 2004). Auch für Zielorientierungen liegen Forschungsbefunde vor, die zeigen, dass die Feedbackwirksamkeit in Abhängigkeit von den Zielorientierungen von Feedbackempfängerinnen und -empfängern variiert (vgl. Cianci et al., 2010; Cron et al., 2005; Donovan & Hafsteinson, 2006; VandeWalle et al., 2001). Da im Rahmen der vorliegenden Arbeit die differenziellen Effekte von Zielorientierungen auf

die Feedbackwirksamkeit untersucht werden, wird im Folgenden näher auf Zielorientierungen eingegangen. Zunächst werden die grundlegenden theoretischen Annahmen sowie empirische Befunde der Zielorientierungsforschung zusammenfassend dargestellt. Im Anschluss wird aufgezeigt, inwiefern Zielorientierungen die Verarbeitung und Nutzung eines beobachtungsbasierten Unterrichtsfeedbacks durch Lehrkräfte beeinflussen könnten.

Zielorientierungen werden als relativ stabile motivationale Merkmale definiert, die das emotionale Erleben, die Kognitionen und das Verhalten in lern- und leistungsbezogenen Situationen beeinflussen (vgl. Nitsche et al., 2011). Bei Zielorientierungen handelt es sich demnach um ein dispositionales Merkmal, das sich in Abhängigkeit von situativen Merkmalen jedoch mittel- bis langfristig verändern kann, d.h. um eine *Quasi-Trait-Variable* (vgl. DeShon & Gillespie, 2005; Kaplan & Maehr, 2007). Als ausschlaggebend für die mit Zielorientierungen assoziierten Erlebens-, Kognitions- und Verhaltensmuster werden die Ziele angesehen, die Personen in Lern- und Leistungssituationen verfolgen. Allgemein wird zwischen Lern- und Leistungszielen unterschieden, die einer Lernzielorientierung (*LZO*) bzw. einer Leistungszielorientierung zugrunde liegen. Während lernzielorientierte Personen in lern- und leistungsbezogenen Situationen das Ziel verfolgen, ihre Kompetenzen weiterzuentwickeln, verfolgen leistungszielorientierte Personen das Ziel, überlegene Kompetenzen zu demonstrieren bzw. Kompetenzdefizite zu verbergen (vgl. Ames, 1992; Dweck, 1986; Dweck & Leggett, 1988; Nicholls, 1984).

Es wird angenommen, dass einer *LZO* und einer Leistungszielorientierung unterschiedliche implizite subjektive Theorien darüber zugrunde liegen, ob Kompetenzen veränderbar sind oder nicht. Lernzielorientierte Personen gehen demnach davon aus, Kompetenzen seien veränderbar (eine sogenannte *incremental theory of intelligence*), leistungszielorientierte Personen nehmen an, Kompetenzen seien vergleichsweise stabil und ließen sich nur schwer verändern (*entity theory of intelligence*; vgl. Dweck, 1986; Dweck & Leggett, 1988). In einem engen Zusammenhang mit diesen subjektiven Theorien steht die Wahrnehmung von Aufwand. Während lernzielorientierte Personen davon ausgehen, ihre Kompetenzen über einen erhöhten Aufwand weiterentwickeln zu können und Aufwand bereits als Hinweis auf Kompetenzentwicklung interpretieren, verstehen leistungszielorientierte Personen Aufwand als Hinweis auf ein Kompetenzdefizit (vgl. Ames, 1992; Dweck, 1986; Dweck & Leggett, 1988; Nicholls, 1984).

Als ein weiteres konstituierendes Merkmal von *LZO* und Leistungszielorientierung gelten – mal implizit (vgl. Dweck, 1986; Dweck & Leggett, 1988), mal explizit (vgl.

Nicholls, 1984) – die Bezugsnormen, anhand derer Kompetenzen beurteilt werden. Während Kompetenzen bei einer *LZO* relativ zu einem aufgabenbezogenen oder individuellen Standard (sachliche bzw. individuelle Bezugsnorm) bestimmt werden, liegt der Leistungsbeurteilung bei einer Leistungszielorientierung eine soziale Bezugsnorm zugrunde. Kompetenzen werden bei der Leistungszielorientierung immer relativ zu den Kompetenzen anderer bestimmt (vgl. Elliot, 1999; Elliot & Thrash, 2001). Aus diesem Grund spricht Nicholls (1984) im Hinblick auf die *LZO* auch von *task involvement* und hinsichtlich der Leistungszielorientierung von *ego involvement*: Die sachliche bzw. individuelle Bezugsnorm bei einer *LZO* geht in lern- und leistungsbezogenen Situationen mit einem aufgabenbezogenen Fokus einher; der soziale Vergleich, den leistungszielorientierte Personen vornehmen, um ihre Kompetenzen zu bestimmen, geht demgegenüber mit einer erhöhten Aufmerksamkeit auf das Selbst einher.

Die frühe Zielorientierungsforschung hat ergeben, dass eine *LZO* mit lern- bzw. leistungsförderlichen Erlebens-, Kognitions- und Verhaltensmustern assoziiert ist. So erleben lernzielorientierte Personen in Lern- und Leistungssituationen häufig positive Emotionen, sie suchen Herausforderungen auf und weisen bei Schwierigkeiten eine hohe Widerstandsfähigkeit auf. Außerdem reagieren sie auf Probleme mit einem erhöhten Aufwand und der Entwicklung neuer Lösungsstrategien. Eine Leistungszielorientierung geht demgegenüber häufig mit lern- bzw. leistungsmindernden Erlebens-, Kognitions- und Verhaltensmustern einher. Leistungszielorientierte Personen empfinden in lern- und leistungsbezogenen Situationen oft negative Emotionen, sie meiden Herausforderungen, weisen bei Schwierigkeiten eine geringe Widerstandsfähigkeit auf und neigen zu selbstbezogenen Kognitionen. Während eine *LZO* dementsprechend in einem positiven Zusammenhang mit Lernen und Leistung steht, geht eine Leistungszielorientierung überwiegend mit negativen Folgen für Lernen und Leistung einher (für einen Überblick über den Forschungsstand der frühen Zielorientierungsforschung siehe Ames, 1992; Dweck, 1986; Dweck & Leggett, 1988; Nicholls, 1984).

Um dem Umstand Rechnung zu tragen, dass eine Leistungszielorientierung nicht immer lern- und leistungsmindernd wirkt, haben Elliot und Kolleginnen und Kollegen (vgl. Elliot, 1999; Elliot & Church, 1997; Elliot & Harackiewicz, 1996; Elliot & Thrash, 2001; siehe auch Middleton & Midgley, 1997; Skaalvik, 1997; VandeWalle, 1997) das 2-Faktorenmodell der *LZO* und Leistungszielorientierung in ein 3-Faktorenmodell weiterentwickelt, indem sie die Leistungszielorientierung in eine Annäherungs- (*ALZO*) und

eine Vermeidungsleistungszielorientierung (*VLZO*) ausdifferenziert haben. Während Personen mit einer hoch ausgeprägten *ALZO* in Lern- und Leistungssituationen das Ziel verfolgen, überlegene Kompetenzen zu demonstrieren, sind Personen mit hoch ausgeprägter *VLZO* darum bemüht, Kompetenzdefizite vor Dritten zu verbergen. Im Gegensatz zum 2-Faktorenmodell liegt dem 3-Faktorenmodell die Annahme zugrunde, dass zusätzlich zur *LZO* auch eine *ALZO* mit lern- und leistungsförderlichen Erlebens-, Kognitions- und Verhaltensmustern einhergeht (vgl. Elliot & Church, 1997; Elliot & Harackiewicz, 1996; Harackiewicz et al., 2002; Senko et al., 2011). Diese Annahme ist jedoch nicht unumstritten; einige Wissenschaftlerinnen und Wissenschaftler nehmen weiterhin an, dass sich nur eine *LZO* günstig auf Lernen und Leistung auswirkt (vgl. z. B. Kaplan & Maehr, 2007; Midgley et al., 2001).

Meta-Analysen zum 3-Faktorenmodell bestätigen die Befunde der frühen Zielorientierungsforschung, nach denen eine *LZO* lern- und leistungsförderlich ist. Darüber hinaus zeigen sie, dass eine *VLZO* mit lern- und leistungsmindernden Wirkungen einhergeht (vgl. Cellar et al., 2011; Huang, 2011, 2012; Payne et al., 2007; Wirthwein et al., 2013). Hinsichtlich der *ALZO* sind die Befunde demgegenüber inkonsistent. Annäherungsleistungszielorientierte Personen erleben in lern- und leistungsbezogenen Situationen sowohl positive Emotionen wie Zufriedenheit (vgl. Cellar et al., 2011; Huang, 2011) als auch negative Emotionen wie Angst (vgl. Huang, 2011; Payne et al., 2007). Darüber hinaus ist eine *ALZO* teilweise schwach positiv (vgl. Cellar et al., 2011) und teilweise schwach negativ (vgl. Payne et al., 2007) mit einer effektiven Verhaltensregulation korreliert. Die meisten Meta-Analysen weisen dennoch einen positiven Zusammenhang zwischen der *ALZO* und Lernen und Leistung aus, der jedoch durchweg geringer ausfällt als bei der *LZO* (vgl. Cellar et al., 2011; Huang, 2012; Wirthwein et al., 2013; für eine Ausnahme siehe Payne et al. (2007), die für die *ALZO* überwiegend Nullkorrelationen berichten).

Auch in der Forschung zu den Zielorientierungen von Lehrkräften ist das 3-Faktorenmodell die aktuell dominierende Konzeptualisierung.<sup>10</sup> Die zu den Zielorientierungen

---

<sup>10</sup> Teilweise werden die *LZO*, *ALZO* und *VLZO* um die Arbeitsvermeidung ergänzt, die bereits von Nicholls (1984) in die Diskussion eingebracht wurde. Arbeitsvermeidende Lehrkräfte verfolgen das Ziel, mit so wenig Aufwand wie möglich durch ihren Arbeitstag zu kommen (vgl. z. B. Butler, 2007; Dickhäuser et al., 2007; Janke et al., 2019; Nitsche et al., 2011). Nach Elliot (1999) handelt es sich bei der Arbeitsvermeidung jedoch nicht um eine Zielorientierung im engeren Sinne: „In essence, the work avoidance construct appears to represent the absence of an achievement goal in an achievement setting, rather than the presence of a

von Lehrkräften vorliegenden Befunde stehen dabei grundsätzlich im Einklang mit den Befunden der allgemeinen Zielorientierungsforschung: Während eine *LZO* durchweg mit lern- und leistungsförderlichen Erlebens-, Kognitions- und Verhaltensmustern einhergeht, wirkt sich eine *VLZO* auch bei Lehrkräften lern- und leistungsmindernd aus; die Befunde zur *ALZO* sind wiederum inkonsistent (vgl. Butler, 2007; Butler & Shibaz, 2008; Gorozidis & Papaioannou, 2016; Janke et al., 2019; Nitsche et al., 2011, 2013a, 2013b; Parker et al., 2013; Retelsdorf & Günther, 2011; Retelsdorf et al., 2010; Runhaar et al., 2013; Skaalvik & Skaalvik, 2013; van Daal et al., 2014; Wang et al., 2017).<sup>11</sup>

Auch wenn zu den Zielorientierungen von Lehrkräften eine Vielzahl empirischer Studien vorliegt, wurde ihre Rolle bei der Feedbackverarbeitung und -nutzung nach meinem Kenntnisstand bislang weder in der nationalen noch in der internationalen Schul- und Unterrichtsforschung untersucht. Es kann aber angenommen werden, dass Zielorientierungen beeinflussen, wie Lehrkräfte ein beobachtungsbasiertes Unterrichtsfeedback verarbeiten und ob sie die enthaltenen Informationen effektiv nutzen: Weil lernzielorientierte Lehrkräfte danach streben, ihre Unterrichtskompetenz weiterzuentwickeln und in lern- und leistungsbezogenen Situationen eine hohe *task involvement* aufweisen, ist davon auszugehen, dass sie Feedback vor allem als Lerngelegenheit wahrnehmen und akzeptieren. Mit Blick auf die Feedbackverarbeitung ist zu erwarten, dass negative Emotionen auch nach Erhalt eines kritischen Unterrichtsfeedbacks ausbleiben. Wegen der mit einer *LZO* assoziierten subjektiven Theorie, dass Kompetenzen veränderbar sind, sollten lernzielorientierte Lehrkräfte die in einem kritischen Unterrichtsfeedback enthaltene Soll-Ist-Diskrepanz vor allem auf den investierten Aufwand zurückführen, was einer günstigen Attribution auf internale, veränderbare und kontrollierbare Faktoren entspricht. Hinsichtlich der Feedbacknutzung ist zu erwarten, dass lernzielorientierte Lehrkräfte auf den Erhalt von Feedback daher mit einem erhöhten Aufwand bzw. der Entwicklung und/oder Erprobung neuer Verhaltensstrategien für ihren Unterricht reagieren. Folglich sollte eine

---

particular type of achievement goal“ (Elliot, 1999, S. 184). Neben der *LZO*, *ALZO*, *VLZO* und Arbeitsvermeidung wurden in jüngerer Zeit zudem weitere Zieltypen entwickelt: Butler (2012) diskutiert z. B. soziale Ziele, bei denen es Lehrkräften vor allem darum geht, eine gute Beziehung zu ihren Schülerinnen und Schülern aufzubauen. Mascet et al. (2017) schlagen ein 3x2-Faktorenmodell vor, in dem sie zusätzlich zur Leistungszielorientierung auch die *LZO* in eine Annäherungs- und Vermeidungskomponente ausdifferenzieren und darüber hinaus *task goals* ergänzen (inkl. Annäherungs- und Vermeidungskomponente), bei denen Lehrkräfte vorrangig das Ziel verfolgen, ihren beruflichen Anforderungen gerecht zu werden.

<sup>11</sup> Der Forschungsstand zu den Erlebens-, Kognitions- und Verhaltensmustern von lernziel-, annäherungsleistungsziel- und vermeidungsleistungszielorientierten Lehrkräften wird in den Kapiteln 3.2.4 und vor allem 5.1.3 ausführlicher dargestellt.

*LZO* die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks durch Schulleitungen begünstigen. Hierauf deuten auch Befunde aus der arbeitspsychologischen Feedbackforschung hin, die zeigen, dass Personen mit hoch ausgeprägter *LZO* nach Feedback-erhalt lern- und leistungsförderliche Erlebens-, Kognitions- und Verhaltensreaktionen zeigen und bessere Leistungen erbringen als Personen mit gering ausgeprägter *LZO* (vgl. Cianci et al., 2010; Cron et al., 2005; VandeWalle et al., 2001).

Während eine *LZO* die Feedbackwirksamkeit begünstigen sollte, lässt sich nicht eindeutig vorhersagen, wie sich eine *ALZO* auf die Feedbackwirksamkeit auswirkt. Einerseits ist davon auszugehen, dass annäherungsleistungszielorientierte Lehrkräfte ein erhaltenes Feedback vorrangig als Evaluation ihrer Kompetenz interpretieren, weil sie in Lern- und Leistungssituationen danach streben, überlegene Kompetenzen zu demonstrieren und ein hohes *ego involvement* aufweisen. Ein kritisches Feedback sollte daher zu negativen emotionalen Reaktionen führen. Die einer *ALZO* zugrundeliegende subjektive Theorie, nach der Kompetenzen nur schwer veränderbar sind, legt darüber hinaus die Vermutung nahe, dass annäherungsleistungszielorientierte Lehrkräfte eine ungünstige Attribution auf internale, stabile und nicht beeinflussbare Faktoren zeigen könnten (z. B. ihre berufliche Eignung), was zu einer Zurückweisung des Feedbacks führen könnte. Andererseits erscheint es aber auch plausibel, dass annäherungsleistungszielorientierte Lehrkräfte die in einem Feedback enthaltenen Informationen nutzen, um auch in Zukunft überlegene Kompetenzen demonstrieren zu können. Folglich könnten neben lernzielorientierten Lehrkräften auch annäherungsleistungszielorientierte Lehrkräfte mit einem erhöhten Aufwand sowie der Entwicklung und/oder Erprobung neuer Verhaltensstrategien für ihren Unterricht reagieren. Letztlich ergeben auch empirische Studien aus der arbeitspsychologischen Feedbackforschung, in denen der Einfluss der *ALZO* auf die Feedbackwirksamkeit untersucht wurde, ein uneinheitliches Bild; teilweise wirkt sich eine *ALZO* günstig auf die Feedbackwirksamkeit aus, teilweise beeinflusst diese die Feedbackwirksamkeit aber auch nicht (vgl. Cianci et al., 2010; VandeWalle et al., 2001).

Neben annäherungsleistungsorientierten Lehrkräften sollten auch vermeidungsleistungszielorientierte Lehrkräfte ein erhaltenes Feedback als Evaluation ihrer Kompetenzen wahrnehmen. Begründen lässt sich diese Annahme damit, dass sie danach streben, ihre Kompetenzdefizite vor Dritten zu verbergen und in Lern- und Leistungssituationen ein hohes *ego involvement* aufweisen. Kritisches Feedback führt daher vermutlich zu negativen Emotionen. Aufgrund der impliziten subjektiven Theorie, dass Kompetenzen nur

schwer veränderbar sind und Aufwand auf ein Kompetenzdefizit hindeutet, sollten vermeidungsleistungszielorientierte Lehrkräfte eine ungünstige Attribution auf internale Faktoren zeigen, von denen sie annehmen, sie nicht beeinflussen zu können, was in einer Zurückweisung des Feedbacks resultieren sollte. Mit Blick auf die Feedbacknutzung lässt sich schlussfolgern, dass vermeidungsleistungszielorientierte Lehrkräfte auf ein erhaltenes Feedback daher selten mit einem erhöhten Aufwand bzw. der Entwicklung und/oder Erprobung neuer Verhaltensstrategien reagieren sollten, worauf auch Befunde aus der arbeitspsychologischen Feedbackforschung hindeuten (vgl. Cron et al., 2005; VandeWalle et al., 2001).

## **2.4 Feedbackintervention, Forschungsfragen und Teilstudien**

Vor dem Hintergrund, dass beobachtungs-basierte Unterrichtsfeedbacks häufig, aber nicht immer wirksam sind, wurde in Kapitel 2.3 ein theoretisches Feedbackmodell vorgestellt, das den Prozess eines beobachtungs-basierten Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte beschreibt. Aus dem Modell wurden Bedingungen abgeleitet, von denen die Feedbackwirksamkeit abhängt. Ein Teil dieser Wirksamkeitsbedingungen wurde bei der im DFG-Projekt „Unterrichtsbeurteilungen und -feedback durch Schulleitungen“ unter Mitarbeit des Verfassers dieser Arbeit entwickelten Feedbackintervention berücksichtigt: Die Feedbackintervention bezieht sich zum einen auf die Unterrichtsbeobachtung und zum anderen auf die Gestaltung des Unterrichtsfeedbacks durch die Schulleitung. Außerdem wurde ein Beobachtungs- und Feedbacktraining entwickelt, das Schulleitungen die benötigten Kompetenzen vermitteln soll. Die einzelnen Komponenten der Intervention werden im Folgenden näher beschrieben. Anschließend werden die Forschungsfragen hergeleitet und ein Überblick über die empirischen Teilstudien gegeben.

### **2.4.1 Beschreibung der Feedbackintervention**

**Unterrichtsbeobachtung.** Mit Blick auf die Unterrichtsbeobachtung wurde gezeigt, dass standardisierte Beobachtungsinstrumente eine objektive, reliable und valide Unterrichtsbeurteilung durch Schulleitungen ermöglichen und dadurch die Wirksamkeit eines beobachtungs-basierten Unterrichtsfeedbacks begünstigen sollten. Weil eine einzelne Beobachtung für eine reliable Unterrichtsbeurteilung nicht ausreicht, wurde zudem geschlussfolgert, dass Schulleitungen mehrere Unterrichtsstunden einer Lehrkraft beobachten sollten. Die Intervention sieht daher zwei Unterrichtsbeobachtungen je Lehrkraft vor, die mit einem standardisierten Beobachtungsinstrument, dem *Beobachtungsbogen zur*

*Erfassung von Unterrichtsqualität* (BERU), erfolgen (vgl. Gärtner et al., 2021). Dem BERU liegt ein generisches Unterrichtsmodell zugrunde, das zwischen drei übergeordneten Anforderungsbereichen unterscheidet: der *Unterstützung des Wissenserwerbs* von Schülerinnen und Schülern, der *Motivierung* von Schülerinnen und Schülern sowie dem *Klassenmanagement*. Diesen drei Anforderungsbereichen sind insgesamt 14 Unterrichtsmerkmale zugeordnet, die im BERU über 14 Skalen mit jeweils vier bis acht Items operationalisiert sind. Der BERU und das zugrundeliegende Unterrichtsmodell werden in den Teilstudien 1 und 2 ausführlicher beschrieben. Die theoretische Grundlage des BERU sowie Befunde zu seiner faktoriellen Validität und Reliabilität können zudem bei Gärtner et al. (2021) nachgelesen werden.

**Feedbackverfahren.** Aufgrund der von Lehrkräften häufig bemängelten Feedbackqualität wurde neben dem BERU auch ein Feedbackverfahren entwickelt, das Schulleitungen dabei unterstützen soll, ein qualitativ hochwertiges Feedback zu gestalten. Die Entwicklung des Feedbackverfahrens erfolgte auf Grundlage der in Kapitel 2.3.2 beschriebenen Befunde der arbeits- und instruktionspsychologischen Feedbackforschung. Es wurde gezeigt, dass sich die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks über einen hohen Bezug zum Unterrichtsverhalten von Lehrkräften steigern lässt. Aus diesem Grund sieht das Verfahren vor, dass das Feedback – in Anlehnung an Hattie und Timperley (2007) – drei Informationen enthält: Zunächst soll den Lehrkräften das der Unterrichtsbeurteilung zugrunde liegende Unterrichtsmodell vorgestellt werden (FEED UP). Anschließend wird das beobachtete Verhalten den im Modell beschriebenen lernwirksamen Verhaltensweisen vergleichend gegenübergestellt (FEED BACK). Es werden unterrichtliche Stärken und unterrichtliche Schwächen thematisiert, da sowohl ein bestätigendes als auch ein kritisches Feedback wirksam sein kann. Darüber hinaus sollte das Feedback anhand von konkreten Verhaltensbeispielen aus dem Unterricht begründet werden. Im Hinblick auf das kritische Feedback sollten den Lehrkräften zudem konkrete Verhaltensempfehlungen unterbreitet werden, mit denen sie den identifizierten Schwächen begegnen können (FEED FORWARD). Abschließend sieht das Feedbackverfahren das Formulieren von Entwicklungszielen im Hinblick auf die Umsetzung der alternativen Verhaltensstrategien vor.

Hinsichtlich der Feedbackübermittlung wurde geschlussfolgert, dass Lehrkräfte das Unterrichtsfeedback zeitnah erhalten sollten. Weil sich schriftliches bzw. computerbasiertes Feedback als besonders wirksam erwiesen hat, wurde zudem ein webbasiertes Feedbackportal entwickelt, in das Schulleitungen ihre Beobachtungen übertragen können.

Im Feedbackportal werden auf Grundlage der Beobachtungsdaten lehrerspezifische Feedbackberichte generiert. Die Feedbackberichte sollen den Feedbackgesprächen zwischen der Schulleitung und den Lehrkräften zugrunde gelegt und letztgenannten im Anschluss an die Gespräche ausgehändigt werden. Auch das Feedbackverfahren inkl. des Online-Portals und der Feedbackberichte werden in den Teilstudien 1 und 2 ausführlicher beschrieben.

***Beobachtungs- und Feedbacktraining.*** Um Schulleitungen die für die Unterrichtsbeobachtung und Feedbackgabe benötigten Kompetenzen zu vermitteln, wurde zusätzlich ein Beobachtungs- und Feedbacktraining konzipiert. Weil sich die Beobachtungstrainings aus den US-amerikanischen TES empirisch bewährt haben, sieht auch die Intervention eine in Anlehnung an diese Trainings konzipierte Schulung vor: Die Schulleitungen werden mit dem BERU und dem zugrundeliegenden Unterrichtsmodell vertraut gemacht, sie lernen häufig auftretende Beobachtungsfehler und Strategien zu ihrer Vermeidung kennen. Außerdem können sie ihre Unterrichtsbeurteilungen im Rahmen von Kalibrierungsübungen mit den Urteilen anderer Schulleitungen abgleichen. Der Schwerpunkt des Feedbacktrainings liegt anders als bei Kraft und Christian (2022), deren Training sich nicht bewährt hat, auf den aus der arbeits- und instruktionspsychologischen Feedbackforschung hergeleiteten Merkmalen eines qualitativ hochwertigen Feedbacks, die dem zuvor beschriebenen Feedbackverfahren zugrunde gelegt wurden. Die Schulleitungen werden in der Anwendung des Feedbackverfahrens geschult und mit dem webbasierten Feedbackportal sowie den lehrerspezifischen Feedbackberichten vertraut gemacht, die sie bei der Feedbackgabe unterstützen sollen.

#### **2.4.2 Herleitung der Forschungsfragen**

Ob die beschriebene Feedbackintervention Schulleitungen tatsächlich dabei unterstützt, ihren Lehrkräften ein qualitativ hochwertiges und wirksames Unterrichtsfeedback zu geben, wird im Rahmen von drei Teilstudien empirisch überprüft. Folgende Forschungsfragen werden untersucht:

- *Forschungsfrage 1:* Wie schätzen Lehrkräfte die Qualität des beobachtungs-basierenden Unterrichtsfeedbacks ein? (Teilstudie 1)
- *Forschungsfrage 2:* Hat die Feedbackintervention einen Effekt auf die Unterrichts-kompetenz von Lehrkräften? (Teilstudie 1)

- *Forschungsfrage 3*: Hat die Feedbackintervention einen Effekt auf die Unterrichtsqualität? (Teilstudie 2)

Im theoretischen Feedbackmodell wurde aufgezeigt, dass die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks von Schulleitungen jedoch nicht allein von der Unterrichtsbeobachtung, der Feedbackgestaltung und den Beobachtungs- und Feedbackkompetenzen der Schulleitung abhängt, die durch die Feedbackintervention adressiert werden. Entscheidend dafür, dass ein beobachtungsbasiertes Unterrichtsfeedback zu einer Weiterentwicklung der Unterrichtskompetenz von Lehrkräften und einer besseren Unterrichtsqualität führt, ist zudem die Verarbeitung und Nutzung des Feedbacks durch die Lehrkräfte selbst. Die Feedbackverarbeitung und -nutzung wiederum werden von verschiedenen Merkmalen der Lehrkräfte, insbesondere ihren Zielorientierungen, beeinflusst. Neben den Haupteffekten der Feedbackintervention auf die Unterrichtskompetenz und die Unterrichtsqualität werden in den empirischen Studien daher zudem die differenziellen Effekte der *LZO*, *ALZO* und *VLZO* von Lehrkräften überprüft:

- *Forschungsfrage 4*: Moderieren die Zielorientierungen von Lehrkräften den Effekt der Feedbackintervention auf die Unterrichtskompetenz? (Teilstudie 1)
- *Forschungsfrage 5*: Moderieren die Zielorientierungen von Lehrkräften den Effekt der Feedbackintervention auf die Unterrichtsqualität? (Teilstudie 3)

Um die Haupteffekte der Feedbackintervention und die differenziellen Effekte der Zielorientierungen zu überprüfen, wurde im Schuljahr 2017/2018 eine Interventionsstudie mit quasi-experimentellem Design durchgeführt. Schulleitungen beurteilten den Unterricht von Lehrkräften der Interventionsgruppe mit dem BERU und gaben ihnen ausgehend von den Unterrichtsbeobachtungen ein Feedback zu ihrem Unterricht. Bei der Feedbackgabe orientierten sie sich am zuvor beschriebenen Feedbackverfahren. Außerdem durchliefen alle Schulleitungen das entwickelte Beobachtungs- und Feedbacktraining. Die Lehrkräfte der Kontrollgruppe erhielten kein Feedback. Alle an der Studie teilnehmenden Lehrkräfte haben am Anfang des Schuljahres 2017/2018 an einer Prä-Befragung teilgenommen. Die Lehrkräfte der Interventionsgruppe haben unmittelbar nach Feedbackerhalt zudem einen Implementationscheck ausgefüllt. Die Post-Befragung fand am Ende desselben Schuljahres statt. Die Unterrichtsqualität wurde über Schülerurteile erfasst. Daher haben zusätzlich zu den Lehrkräften auch die Schülerinnen und Schüler einer Klasse je Lehrkraft an der Prä- und Postbefragung teilgenommen.

Die Stichprobe umfasst  $N = 6$  Berliner Oberstufenzentren (OSZ) mit  $N = 11$  Schulleitungen (inkl. Schulleiterinnen und Schulleitern, deren Stellvertreterinnen und Stellvertretern sowie Abteilungs- und Fachleitungen),  $N = 81$  Lehrkräfte sowie  $N = 1.069$  Schülerinnen und Schüler aus  $N = 53$  Klassen. Für die *Forschungsfrage 1*, die nach den Qualitätseinschätzungen der Lehrkräfte fragt, können die Befragungsdaten von  $N = 29$  Lehrkräften genutzt werden, die den Implementationscheck ausgefüllt haben. Für die Forschungsfragen, die sich auf die Effekte des Unterrichtsfeedbacks auf die Weiterentwicklung der Unterrichtskompetenz beziehen (*Forschungsfragen 2 und 4*), können die Befragungsdaten von  $N = 81$  Lehrkräften genutzt werden (Interventionsgruppe: 35; Kontrollgruppe: 46). Für die Forschungsfragen, die sich auf die Effekte des Unterrichtsfeedbacks auf die Unterrichtsqualität beziehen (*Forschungsfragen 3 und 5*) können die Befragungsdaten von  $N = 53$  Lehrkräften und  $N = 53$  Klassen (Interventionsgruppe: 26; Kontrollgruppe: 27) mit insgesamt  $N = 1.069$  Schülerinnen und Schülern (Interventionsgruppe: 518; Kontrollgruppe: 551) genutzt werden.

### 2.4.3 Überblick über die empirischen Teilstudien

In Teilstudie 1 („Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interventionsstudie unter Berücksichtigung der Zielorientierungen von Lehrkräften“) wird zum einen untersucht, wie die Lehrkräfte der Interventionsgruppe die Qualität des beobachtungsbasierten Unterrichtsfeedbacks einschätzen (*Forschungsfrage 1*). Außerdem wird geprüft, ob die Feedbackintervention einen Effekt auf die Weiterentwicklung der Unterrichtskompetenz hat (*Forschungsfrage 2*) und ob die *LZO*, *ALZO* und *VLZO* der Lehrkräfte den Feedbackeffekt moderieren (*Forschungsfrage 4*). Im theoretischen Teil wird unter Rückgriff auf die FIT (vgl. Kluger & DeNisi, 1996) beschrieben, wie sich ein beobachtungsbasiertes Unterrichtsfeedback von Schulleitungen auf die Weiterentwicklung der Unterrichtskompetenz von Lehrkräften auswirkt. Ausgehend von den differenziellen Zusammenhängen der Zielorientierungen von Lehrkräften und ihren lernrelevanten Einstellungen und Verhaltensweisen wird zudem hergeleitet, wie die *LZO*, *ALZO* und *VLZO* die Feedbackwirksamkeit moderieren könnten. Im Methodenteil wird unter anderem die Feedbackintervention inklusive des standardisierten Beobachtungsinstruments, des entwickelten Feedbackverfahrens und des Beobachtungs- und Feedbacktrainings beschrieben. Die Qualitätseinschätzungen der Lehrkräfte werden im empirischen Teil de-

skriptiv und mithilfe eines Einstichproben-T-Test ausgewertet. Der Haupteffekt des Unterrichtsfeedbacks auf die Kompetenzentwicklung der Lehrkräfte wird mithilfe von Regressionsanalysen überprüft, bei denen diverse Kovariate, die einen Einfluss auf die Kompetenzentwicklung der Lehrkräfte haben können, als Kontrollvariablen berücksichtigt werden. Die Überprüfung der differenziellen Effekte der *LZO*, *ALZO* und *VLZO* erfolgt mithilfe moderierter Regressionsanalysen.

In Teilstudie 2 („The effect of school leaders‘ feedback on teaching quality perceived by students – results of an intervention study“) wird untersucht, ob die Feedbackintervention einen Effekt auf die von Schülerinnen und Schülern wahrgenommene Unterrichtsqualität hat (*Forschungsfrage 3*). Im theoretischen Teil wird ein Feedbackmodell vorgestellt, das große Überschneidungen mit dem in Kapitel 2.3 entwickelten Modell aufweist, dessen Beschreibung im Vergleich zu den Ausführungen im theoretischen Rahmen jedoch stark vereinfacht sind. Außerdem werden die der entwickelten Feedbackintervention zugrundeliegenden 14 Unterrichtsmerkmale aus dem Unterrichtsmodell ausführlich beschrieben. Im empirischen Teil werden die Effekte der Feedbackintervention auf die von Schülerinnen und Schülern wahrgenommene Unterrichtsqualität mithilfe von Mehrebenenregressionsanalysen überprüft, bei denen wiederum diverse Kontrollvariablen modelliert werden.

Teilstudie 3 („Improvements in teaching quality after receiving feedback: What role do teachers‘ goal orientations play for the effectiveness of feedback?“) knüpft an Teilstudie 1 und Teilstudie 2 an und setzt sich vertiefend mit den differenziellen Effekten von Zielorientierungen auseinander. Es wird untersucht, ob die *LZO*, *ALZO* und *VLZO* von Lehrkräften den Effekt des Unterrichtsfeedbacks auf die von Schülerinnen und Schülern wahrgenommene Unterrichtsqualität moderieren (*Forschungsfrage 5*). Während die differenziellen Effekte der Zielorientierungen in Teilstudie 1 vor dem Hintergrund der mit Zielorientierungen variierenden lernrelevanten Einstellungen und Verhaltensweisen von Lehrkräften begründet wird, erfolgt im theoretischen Teil der Teilstudie 3 eine Verknüpfung der FIT (vgl. Kluger & DeNisi, 1996) mit der Zielorientierungstheorie. Es wird theoretisch hergeleitet, wie die *LZO*, *ALZO* und *VLZO* von Lehrkräften die Feedbackverarbeitung und -nutzung und damit seine Wirksamkeit beeinflussen könnten. Die Überprüfung der differenziellen Effekte erfolgt mithilfe moderierter Mehrebenenregressionsanalysen.

## Literaturverzeichnis I

- Aaronson, D., Barrow, L. & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135. <https://doi.org/10.1086/508733>
- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y. & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037. <https://doi.org/10.1126/science.1207998>
- Altrichter, H. & Maag Merki, K. (2016). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulwesen* (2., überarb. und aktual. Aufl., S. 1–27). Springer VS. [https://doi.org/10.1007/978-3-531-18942-0\\_1](https://doi.org/10.1007/978-3-531-18942-0_1)
- Altrichter, H., Kemethofer, D. & George, A. C. (2019). Schulleitungen und evidenzbasierte Bildungsreform im Schulwesen. *Zeitschrift für Bildungsforschung*, 9(1), 17–35. <https://doi.org/10.1007/s35834-018-0228-5>
- Altrichter, H., Moosbrugger, R. & Zuber, J. (2016a). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulwesen* (2., überarb. und aktual. Aufl., S. 235–278). Springer VS. [https://doi.org/10.1007/978-3-531-18942-0\\_9](https://doi.org/10.1007/978-3-531-18942-0_9)
- Altrichter, H., Rürup, M. & Schuchart, C. (2016b). Schulautonomie und die Folgen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulwesen* (2., überarb. und aktual. Aufl., S. 107–149). Springer VS. [https://doi.org/10.1007/978-3-531-18942-0\\_5](https://doi.org/10.1007/978-3-531-18942-0_5)
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271. <https://doi.org/10.1037/0022-0663.84.3.261>
- Appius, S., Steger Vogt, E., Kansteiner-Schänzlin, K. & Bach-Blattner, T. (2012). Personalentwicklung an Schulen – eine Bestandsaufnahme aus Sicht deutscher und schweizerischer Schulleitungen. *Empirische Pädagogik*, 26(1), 123–141.
- Atwater, L. E. & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360° feedback. *Journal of Vocational Behavior*, 66(3), 532–548. <https://doi.org/10.1016/j.jvb.2004.05.003>

- Bach, A., Böhnke, A. & Thiel, F. (2020). Improving instructional competencies through individualized staff development and teacher collaboration in German schools. *International Journal of Educational Management*, 34(8), 1289–1302.  
<https://doi.org/10.1108/IJEM-08-2019-0294>
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50, 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Bandura, A. & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88(1), 87–99. <https://doi.org/10.1037/0021-9010.88.1.87>
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A. & Morgen, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238. <https://doi.org/10.3102/00346543061002213>
- Barfknecht, T. & von Saldern, M. (2010). Evaluation und Feedback der Lehrkräfte. In M. Demmer & M. von Saldern (Hrsg.), *Helden des Alltags. Erste Ergebnisse der Schulleitungs- und Lehrkräftebefragung (TALIS) in Deutschland (DDS – Die Deutsche Schule*, 11. Beiheft, S. 94–115).
- Bell, S. T. & Arthur, W. (2008). Feedback acceptance in developmental assessment centers: the role of feedback message, participant personality, and affective response to the feedback session. *Journal of Organizational Behavior*, 29(5), 681–703.  
<https://doi.org/10.1002/job.525>
- Böckelmann, C. & Mäder, K. (2018). Personalentwicklung im Schulbereich – Geschichte, Kontext und Kultur. In C. Böckelmann & K. Mäder (Hrsg.), *Fokus Personalentwicklung* (2., vollst. überarb. und aktual. Aufl.) (S. 3–11). Springer.  
[https://doi.org/10.1007/978-3-662-55683-2\\_1](https://doi.org/10.1007/978-3-662-55683-2_1)
- Bonsen, M. (2016). Schulleitung und Führung in der Schule. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulwesen* (2., überarb. und aktual. Aufl., S. 301–323). Springer VS. [https://doi.org/10.1007/978-3-531-18942-0\\_11](https://doi.org/10.1007/978-3-531-18942-0_11)
- Brauckmann, S. & Eder, F. (2019). Führungsforschung im Bildungsbereich: Schulleitung im Spannungsfeld erweiterter Rechte und Pflichten. *Zeitschrift für Bildungsforschung*, 9(1), 5–15. <https://doi.org/10.1007/s35834-019-00242-6>

- Brauckmann, S. & Schwarz, A. (2015). No time to manage? The trade-off between relevant tasks and actual priorities of school leaders in Germany. *International Journal of Educational Management*, 29(6), 749–765. <https://doi.org/10.1108/IJEM-10-2014-0138>
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Educational Testing Service.  
<http://files.eric.ed.gov/fulltext/ED529977.pdf>
- Brett, J. F. & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86(5), 930–942.  
<https://doi.org/10.1037//0021-9010.86.5.930>
- Butler, R. (2007). Teachers' achievement goal orientations and associations with teachers' help seeking. Examination of a novel approach to teacher motivation. *Journal of Educational Psychology*, 99(2), 241–252. <https://doi.org/10.1037/0022-0663.99.2.241>
- Butler, R. (2012). Striving to connect: Extending an achievement goal approach to teacher motivation to include relational goals for teaching. *Journal of Educational Psychology*, 104(3), 726–742. <https://doi.org/10.1037/a0028613>
- Butler, R. & Shibaz, L. (2008). Achievement goals for teaching as predictors of students' perceptions of instructional practices and students' help seeking and cheating. *Learning and Instruction*, 18(5), 453–467. <https://doi.org/10.1016/j.learninstruc.2008.06.004>
- Carver, C. S. & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality – social, clinical, and health psychology. *Psychological Bulletin*, 92(1), 111–135. <https://doi.org/10.1037/0033-2909.92.1.111>
- Cellar, D. F., Stuhlmacher, A. F., Young, S. K., Fisher, D. M., Adair, C. K., Haynes, S., Twichell, E., Arnold, K. A., Royer, K., Denning, B. L. & Riester, D. (2011). Trait goal orientation, self-regulation, and performance: A meta-analysis. *Journal of Business and Psychology*, 26(4), 467–483. <https://doi.org/10.1007/s10869-010-9201-6>
- Cherasaro, T. L., Brodersen, R. M., Reale, M. L. & Yanoski, D. C. (2016). *Teachers' responses to feedback from evaluators: What feedback characteristics matter?* (REL 2017-190). U.S. Department of Education, Institute of Education Sciences,

- National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. [https://ies.ed.gov/ncee/rel/regions/central/pdf/REL\\_2017190.pdf](https://ies.ed.gov/ncee/rel/regions/central/pdf/REL_2017190.pdf)
- Chow, A. P. Y., Wong, E. K. P., Yeung, A. S. & Mo, K. W. (2002). Teachers' perceptions of appraiser–appraisee relationships. *Journal of Personnel Evaluation in Education*, 16(2), 85–101. <https://doi.org/10.1023/A:1020096127713>
- Cianci, A. M., Schaubroeck, J. M. & McGill, G. A. (2010). Achievement goals, feedback, and task performance. *Human Performance*, 23(2), 131–154. <https://doi.org/10.1080/08959281003621687>
- Cohen, J. & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378–387. <https://doi.org/10.3102/0013189X16659442>
- Cron, W. L., Slocum, J. W., VandeWalle, D. & Fu, Q. (2005). The role of goal orientation on negative emotions and goal setting when initial performance falls short of one's performance goal. *Human Performance*, 18(1), 55–80. [https://doi.org/10.1207/s15327043hup1801\\_3](https://doi.org/10.1207/s15327043hup1801_3)
- Danielson, C. (2007). *Enhancing professional practice. A framework for teaching* (2nd ed.). Association for Supervision and Curriculum Development.
- Daniëls, E., Hondeghem, A. & Dochy, F. (2019). A review on leadership and leadership development in educational settings. *Educational Research Review*, 27, 110–125. <https://doi.org/10.1016/j.edurev.2019.02.003>
- Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132–137. <https://doi.org/10.3102/0013189X15575346>
- Darling-Hammond, L., Amrien-Beardsley, A., Haertel, E. & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8–15. <https://doi.org/10.1177/003172171209300603>
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G. & van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1–11. <https://doi.org/10.1016/j.tate.2013.06.011>
- Derrington, M. L. & Martinez, J. A. (2019). Exploring teachers' evaluation perceptions: A snapshot. *NASSP Bulletin*, 103(1), 32–50. <https://doi.org/10.1177/0192636519830770>

- DeShon, R. P. & Gillespie, J. Z. (2005). A motivated action theory account of goal orientation. *Journal of Applied Psychology*, 90(6), 1096–1127. <https://doi.org/10.1037/0021-9010.90.6.1096>
- Dickhäuser, O., Butler, R. & Tönjes, B. (2007). Das zeigt doch nur, dass ich's nicht kann. Zielorientierungen und Einstellung gegenüber Hilfe bei Lehramtsanwärtern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39(3), 120–126. <https://doi.org/10.1026/0049-8637.39.3.120>
- Donaldson, M. L. & Papay, J. P. (2015). Teacher evaluation for accountability and development. In H. F. Ladd & M. E. Goertz (Hrsg.), *Handbook of research in education finance and policy* (S. 174–173). American Educational Research Association.
- Donovan, J. & Hafsteinsson, L. (2006). The impact of goal-performance discrepancies, self-efficacy, and goal orientation on upward goal revision. *Journal of Applied Social Psychology*, 36(4), 1046–1069. <https://doi.org/10.1111/j.0021-9029.2006.00054.x>
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048. <https://doi.org/10.1037//0003-066X.41.10.1040>
- Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273. <https://doi.org/10.1037//0033-295X.95.2.256>
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169–189. [https://doi.org/10.1207/s15326985ep3403\\_3](https://doi.org/10.1207/s15326985ep3403_3)
- Elliot, A. J. & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72(1), 218–232. <https://doi.org/10.1037/0022-3514.72.1.218>
- Elliot, A. J. & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70(3), 461–475. <https://doi.org/10.1037/0022-3514.70.3.461>
- Elliot, A. J. & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review*, 13(2), 139–156. <https://doi.org/10.1023/A:1009057102306>
- Englert, C. S. & Sugai, G. (1983). Teacher training: Improving trainee performance through peer observation and observation system technology. *Teacher Education and Special Education*, 6(1), 7–17. <https://doi.org/10.1177/088840648300600102>

- Fabiano, G. A., Reddy, L. A. & Dudek, C. M. (2018). Teacher coaching supported by formative assessment for improving classroom practices. *School Psychology Quarterly*, 33(2), 293–304. <https://doi.org/10.1037/spq0000223>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Flores, M. A. (2012). The implementation of a new policy on teacher appraisal in Portugal: How do teachers experience it at school? *Educational Assessment, Evaluation and Accountability*, 24(4), 351–368. <https://doi.org/10.1007/s11092-012-9153-7>
- Flores, M. A. (2018). Teacher evaluation in Portugal: Persisting challenges and perceived effects. *Teachers and Teaching*, 24(3), 223–245. <https://doi.org/10.1080/13540602.2018.1425677>
- Gärtner, H., Thiel, F., Nachbauer, M. & Kellermann, C. (2021). Unterrichtsentwicklung durch Unterrichtsfeedback – Erste Entwicklungsschritte eines Beobachtungsboogens für Schulleitungen. *Zeitschrift für Bildungsforschung*, 11(3), 529–547. <https://doi.org/10.1007/s35834-021-00321-7>
- Garrett, R. & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores. *Educational Evaluation and Policy Analysis*, 37(2), 224–242. <https://doi.org/10.3102/0162373714537551>
- Gorozidis, G. S. & Papaioannou, A. G. (2016). Teachers' achievement goals and self-determination to engage in work tasks promoting educational innovations. *Learning and Individual Differences*, 49, 46–58. <https://doi.org/10.1016/j.lindif.2016.05.014>
- Grissom, J. A., Loeb, S. & Master, B. (2013). Effective instructional time use for school leaders. *Educational Researcher*, 42(8), 433–444. <https://doi.org/10.3102/0013189X13510020>
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329–352. <https://doi.org/10.1080/0305764032000122005>
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, 4(3), 221–239. <https://doi.org/10.1080/15700760500244793>

- Hallinger, P. & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, 32(1), 5–44. <https://doi.org/10.1177/0013161X96032001002>
- Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157–191. <https://doi.org/10.1080/0924345980090203>
- Hallinger, P. & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*, 86(2), 217–247. <https://doi.org/10.1086/461445>
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J. & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, 94(3), 638–645. <https://doi.org/10.1037//0022-0663.94.3.638>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heneman, H. G. & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(2), 173–195. <https://doi.org/10.1023/B:PEEV.0000032427.99952.02>
- Heslin, P. A. & Latham, G. P. (2004). The effect of upward feedback on managerial behavior. *Applied Psychology*, 53(1), 23–37. <https://doi.org/10.1111/j.1464-0597.2004.00159.x>
- Hill, H. C. & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2), 371–384. <https://doi.org/10.17763/haer.83.2.d11511403715u376>
- Hill, H. C., Charalambous, C. Y. & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64. <https://doi.org/10.3102/0013189X12437203>
- Hill, H. C., Kapitula, L. & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. <https://doi.org/10.3102/0002831210387916>

- Hornig, E. L., Klasik, D. & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*, 491–523.  
<https://doi.org/10.1086/653625>
- Huang, C. (2011). Achievement goals and achievement emotions: A meta-analysis. *Educational Psychology Review*, *23*(3), 359–388. <https://doi.org/10.1007/s10648-011-9155-x>
- Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(1), 48–73. <https://doi.org/10.1037/a0026223>
- Hunter, S. B. & Springer, M. G. (2022). Critical feedback characteristics, teacher human capital, and early-career teacher performance: A mixed-methods analysis. *Educational Evaluation and Policy Analysis*, *44*(3), 380–403.  
<https://doi.org/10.3102/01623737211062913>
- Ilgén, D. R. & Davis, C. A. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology: An International Review*, *49*(3), 550–565.  
<https://doi.org/10.1111/1464-0597.00031>
- Ilgén, 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>
- Janke, S., Bardach, L., Oczlon, S. & Lüftenegger, M. (2019). Enhancing feasibility when measuring teachers' motivation: A brief scale for teachers' achievement goal orientations. *Teaching and Teacher Education*, *83*, 1–11.  
<https://doi.org/10.1016/j.tate.2019.04.003>
- Jiang, J. Y., Spörte, S. E. & Lupescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH Students. *Educational Researcher*, *44*(2), 105–116.  
<https://doi.org/10.3102/0013189X15575517>
- Kane, T. J. & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill & Melinda Gates Foundation. <https://usprogram.gatesfoundation.org/-/media/dataimport/resources/pdf/2016/12/met-gathering-feedback-research-paper1.pdf>
- Kane, T. J., Taylor, E. S., Tyler, J. H. & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, *46*(3), 587–613. <https://doi.org/10.3368/jhr.46.3.587>

- Kaplan, A. & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, 19(2), 141–184.  
<https://doi.org/10.1007/s10648-006-9012-5>
- Kelly, O. K., Ang, S. Y. A., Chong, W. L. & Hu, W. S. (2008). Teacher appraisal and its outcomes in Singapore primary schools. *Journal of Educational Administration*, 46(1), 39–54. <https://doi.org/10.1108/09578230810849808>
- Kimball, S. M. (2003). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241–268.  
<https://doi.org/10.1023/A:1021787806189>
- Kimball, S. M. & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34–70.  
<https://doi.org/10.1177/0013161X08327549>
- Klieme, E. (2004). Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde. *Zeitschrift für Pädagogik*, 50(5), 625–634.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: “Aufgabenkultur” und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.), *TIMSS – Impulse für Schule und Unterricht, Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–58). Bundesministerium für Bildung und Forschung (BMBF).
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, J. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise*. Bundesministerium für Bildung und Forschung (BMBF).
- 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>
- KMK [Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Sekretariat der Kultusministerkonferenz.

[https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/Beschluesse\\_Veroeffentlichungen/Bildungsmonitoring\\_Broschuere\\_Endf.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/Beschluesse_Veroeffentlichungen/Bildungsmonitoring_Broschuere_Endf.pdf)

- KMK [Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (2016). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Sekretariat der Kultusministerkonferenz. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2015/2015\\_06\\_11-Gesamtstrategie-Bildungsmonitoring.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf)
- Koedel, C., Mihaly, K. & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47(1), 180–195.  
<https://doi.org/10.1016/j.econedurev.2015.01.006>
- Kopp, B. & Mandl, H. (2014). Aspekte der Feedbacknachricht. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 151–162). Waxmann.
- Kraft, M. A. & Blazar, D. (2017). Individualized coaching to improve teacher practice across grades and subjects: New experimental evidence. *Educational Policy*, 31(7), 1033–1068. <https://doi.org/10.1177/0895904816631099>
- Kraft, M. A. & Christian, A. (2022). Can teacher evaluation systems produce high-quality feedback? An administrator training field experiment. *American Educational Research Journal*, 59(3), 500–537. <https://doi.org/10.3102/00028312211024603>
- Kraft, M. A. & Gilmour, A. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, 52(5), 711–753.  
<https://doi.org/10.1177/0013161X16653445>
- Kraft, M. A., Blazar, D. & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Kyriakides, L., Christoforou, C. & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152.  
<https://doi.org/10.1016/j.tate.2013.07.010>
- Leithwood, K. & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, 4(3), 177–199.  
<https://doi.org/10.1080/15700760500244769>

- Leithwood, K., Harris, A. & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management*, 28(1), 27–42.  
<https://doi.org/10.1080/13632430701800060>
- 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* (Zeitschrift für Pädagogik Beiheft, Bd. 51, S. 47–70). Beltz.
- Liu, Y., Visone, J., Mongillo, M. B. & Lisi, P. (2019). What matters to teachers if evaluation is meant to help them improve? *Studies in Educational Evaluation*, 61(2), 41–54. <https://doi.org/10.1016/j.stueduc.2019.01.006>
- Locke, E. A. & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037//0003-066X.57.9.705>
- London, M. & Smither, J. W. (1995). Can multi-source feedback change self-awareness and behavior? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803–840. <https://doi.org/10.1111/j.1744-6570.1995.tb01782.x>
- London, M. & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review*, 12, 81–100. [https://doi.org/10.1016/S1053-4822\(01\)00043-2](https://doi.org/10.1016/S1053-4822(01)00043-2)
- Manzi, J., Sun, Y. & García, M. R. (Hrsg.) (2022). *Teacher evaluation around the world. Experiences, dilemmas and future challenges*. Springer.  
<https://doi.org/10.1007/978-3-031-13639-9>
- Marzano, R. J., Waters, T. & McNulty, B. A. (2005). *School leadership that works. From research to results*. Association for Supervision and Curriculum Development.
- Mascet, N., Elliot, A. J. & Cury, F. (2017). The 3 × 2 achievement goal questionnaire for teachers. *Educational Psychology*, 37(3), 346–361.  
<https://doi.org/10.1080/01443410.2015.1096324>
- May, H. & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, 47(2), 332–352.  
<https://doi.org/10.1177/0013161X10383411>
- Meetz, F. (2007). *Personalentwicklung als Element der Schulentwicklung. Bestandsaufnahme und Perspektiven*. Klinkhardt.

- Middleton, M. J. & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology*, 89(4), 710–718. <https://doi.org/10.1037/0022-0663.89.4.710>
- Midgley, C., Kaplan, A. & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93(1), 77–86. <https://doi.org/10.1037//0022-0663.93.1.77>
- Milanowski, A. T. & Heneman, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193–212. <https://doi.org/10.1023/A:1012752725765>
- Miller, G.A., Galanter, E. & Pribram, K.H. (1960). *Plans and the structure of behavior*. Holt.
- Mo, K. W., Conners, R. & McCormick, J. (1998). Teacher appraisal in Hong Kong self-managing secondary schools: Factors for effective practices. *Journal of Personnel Evaluation in Education*, 12(1), 19–42. <https://doi.org/10.1023/A:1007912532307>
- Müller, A. & Ditton, H. (2014). Feedback: Begriffe, Formen und Funktionen. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 11–28). Waxmann.
- Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H. & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256. <https://doi.org/10.1080/09243453.2014.885451>
- Neumerski, C. M. (2013). Rethinking instructional leadership, a review: What do we know about principal, teacher, and coach instructional leadership, and where should we go from here? *Educational Administration Quarterly*, 49(2), 310–347. <https://doi.org/10.1177/0013161X12456700>
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328–346. <https://doi.org/10.1037/0033-295X.91.3.328>
- Nitsche, S., Dickhäuser, O., Dresel, M. & Fasching, M. S. (2013a). Zielorientierungen von Lehrkräften als Prädiktoren lernrelevanten Verhaltens. *Zeitschrift für Pädagogische Psychologie*, 27(1-2), 95–103. <https://doi.org/10.1024/1010-0652/a000092>

- Nitsche, S., Dickhäuser, O., Fasching, M. S. & Dresel, M. (2011). Rethinking teachers' goal orientations. Conceptual and methodological enhancements. *Learning and Instruction*, 21(4), 574–586. <https://doi.org/10.1016/j.learninstruc.2010.12.001>
- Nitsche, S., Dickhäuser, O., Fasching, M. S. & Dresel, M. (2013b). Teachers' professional goal orientations: Importance for further training and sick leave. *Learning and Individual Differences*, 23, 272–278. <https://doi.org/10.1016/j.lindif.2012.07.017>
- Nye, B., Konstantopoulos, S. & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257. <https://doi.org/10.3102/01623737026003237>
- OECD [Organisation for Economic Co-operation and Development] (2009). *Creating effective teaching and learning environments: First results from TALIS*. OECD. <https://www.oecd.org/education/school/43023606.pdf>
- Ovando, M. N. (2005). Building instructional leaders' capacity to deliver constructive feedback to teachers. *Journal of Personnel Evaluation in Education*, 18(3), 171–183. <https://doi.org/10.1007/s11092-006-9018-z>
- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123–141. <https://doi.org/10.17763/haer.82.1.v40p0833345w6384>
- Parker, P. D., Martin, A. J., Colmar, S. & Liem, G. A. (2012). Teachers' workplace well-being: Exploring a process model of goal orientation, coping behavior, engagement, and burnout. *Teaching and Teacher Education*, 28(4), 503–513. <https://doi.org/10.1016/j.tate.2012.01.001>
- Payne, S. C., Youngcourt, S. S. & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92(1), 128–150. <https://doi.org/10.1037/0021-9010.92.1.128>
- 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>
- Polikoff, M. S. (2015). The stability of observational and student survey measures of teaching effectiveness. *American Journal of Education*, 121(2), 183–212. <https://doi.org/10.1086/679390>

- Porter, A. C., Youngs, P. & Odden, A. (2001). Advances in teacher assessments and their uses. In V. Richardson (Hrsg.), *Handbook of research on teaching* (S. 259–297). American Educational Research Association.
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM*, *50*(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- 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>
- Reddy, L. A., Dudek, C. M., Peters, S., Alperin, A., Kettler, R. J. & Kurz, A. (2018). Teachers' and school administrators' attitudes and beliefs of teacher evaluation: a preliminary investigation of high poverty school districts. *Educational Assessment, Evaluation and Accountability*, *30*(1), 47–70. <https://doi.org/10.1007/s11092-017-9263-3>
- Retelsdorf, J. & Günther, C. (2011). Achievement goals for teaching and teachers' reference norms: Relations with instructional practices. *Teaching and Teacher Education*, *27*(7), 1111–1119. <https://doi.org/10.1016/j.tate.2011.05.007>
- Retelsdorf, J., Butler, R., Streblov, L. & Schiefele, U. (2010). Teachers' goal orientations for teaching. Associations with instructional practices, interest in teaching, and burnout. *Learning and Instruction*, *20*(1), 30–46. <https://doi.org/10.1016/j.learninstruc.2009.01.001>
- Rheinberg, F. (2006). Bezugsnormorientierung. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 55–62). Beltz.
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 178–186). Hogrefe.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>
- Robinson, V. M. J., Lloyd, C. A. & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 635–674. <https://doi.org/10.1177/0013161X08321509>

- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252.  
<https://doi.org/10.1257/0002828041302244>
- Rowan, B. & Raudenbush, S. W. (2016). Teacher evaluation in American schools. In D. H. Gitomer & C. A. Bell (Hrsg.), *Handbook of research on teaching* (S. 1159–1216). American Educational Research Association.
- Rürup, M. & Heinrich, M. (2007). Schulen unter Zugzwang – Die Schulautonomiegesetzgebung der deutschen Länder als Rahmen der Schulentwicklung. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance* (Educational Governance, Bd. 1, S. 157–183). VS Verlag für Sozialwissenschaften.  
[https://doi.org/10.1007/978-3-531-90498-6\\_6](https://doi.org/10.1007/978-3-531-90498-6_6)
- Runhaar, P., Sanders, K. & Yang, H. (2010). Stimulating teachers' reflection and feedback asking. An interplay of self-efficacy, learning goal orientation, and transformational leadership. *Teaching and Teacher Education*, 26(5), 1154–1161.  
<https://doi.org/10.1016/j.tate.2010.02.011>
- Sartain, L., Stoelinga, S. R. & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago. Lessons learned from classroom observations, principal-teacher conference, and district implementation*. Consortium on Chicago school research.  
<https://consortium.uchicago.edu/sites/default/files/2018-10/Teacher%20Eval%20Report%20FINAL.pdf>
- Scheeler, M. C., Ruhl, K. L. & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 396–407.  
<https://doi.org/10.1177/088840640402700407>
- Schleicher, A. (2012). *Preparing teachers and developing school leaders for the 21st century. Lessons from around the world*. OECD. [https://www.oecd-ilibrary.org/preparing-teachers-and-developing-school-leaders-for-the-21st-century\\_5k9fdvsjz76g.pdf?itemId=%2Fcontent%2Fpublication%2F9789264174559-en&mimeType=pdf](https://www.oecd-ilibrary.org/preparing-teachers-and-developing-school-leaders-for-the-21st-century_5k9fdvsjz76g.pdf?itemId=%2Fcontent%2Fpublication%2F9789264174559-en&mimeType=pdf)
- Senko, C., Hulleman, C. S. & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist*, 46(1), 26–47. <https://doi.org/10.1080/00461520.2011.538646>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>

- Skaalvik, E. M. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology*, 89(1), 71–81. <https://doi.org/10.1037/0022-0663.89.1.71>
- Skaalvik, E. M. & Skaalvik, S. (2013). Teachers' perceptions of the school goal structure: Relations with teachers' goal orientations, work engagement, and job satisfaction. *International Journal of Educational Research*, 62, 199–209. <https://doi.org/10.1016/j.ijer.2013.09.004>
- Smither, J. W., London, M. & Reilly, R. R. (2005a). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, 58, 33–66. [https://doi.org/10.1111/j.1744-6570.2005.514\\_1.x](https://doi.org/10.1111/j.1744-6570.2005.514_1.x)
- Smither, J. W., London, M. & Richmond, K. R. (2005b). The relationship between leaders' personality and their reactions to and use of multisource feedback. A Longitudinal Study. *Group & Organization Management*, 30(2), 181–210. <https://doi.org/10.1177/1059601103254912>
- Song, M., Wayne, A. J., Garet, M. S., Brown, S. & Rickles, J. (2021). Impact of providing teachers and principals with performance feedback on their practice and student achievement: Evidence from a large-scale randomized experiment. *Journal of Research on Educational Effectiveness*, 14(2), 353–378. <https://doi.org/10.1080/19345747.2020.1868030>
- Steinberg, M. P. & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, 11(3), 340–359. [https://doi.org/10.1162/EDFP\\_a\\_00186](https://doi.org/10.1162/EDFP_a_00186)
- Steinberg, M. P. & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, 10(4), 535–572. [https://doi.org/10.1162/EDFP\\_a\\_00173](https://doi.org/10.1162/EDFP_a_00173)
- 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.

- Taut, S. & Sun, Y. (2014). The development and implementation of a national, standards-based, multi-method teacher performance assessment system in Chile. *Education Policy Analysis Archives*, 22(71), 1–30.  
<http://dx.doi.org/10.14507/epaa.v22n71.2014>
- Taut, S., Santelices, M. V., Araya, C. & Manzi, J. (2011). Perceived effects and uses of the national teacher evaluation system in Chilean elementary schools. *Studies in Educational Evaluation*, 37(4), 218–229.  
<https://doi.org/10.1016/j.stueduc.2011.08.002>
- Taylor, E. S. & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628–3651.  
<https://doi.org/10.1257/aer.102.7.3628>
- Thiel, F. & Schewe, C. M. (2022). Instrumente individualisierter Personalentwicklung: Mitarbeitergespräche, Zielvereinbarungen und Anreizsysteme. In F. Thiel, C. M. Schewe, B. Muslic, E.-M. Lankes, N. Maritzen & T. Riecke-Baulecke (Hrsg.), *Personalentwicklung in Schulen als Führungsaufgabe. Eine Bestandsaufnahme in den Ländern der Bundesrepublik Deutschland* (S. 169–231). Springer VS.  
[https://doi.org/10.1007/978-3-658-36925-5\\_7](https://doi.org/10.1007/978-3-658-36925-5_7)
- Thiel, F. & Thillmann, K. (2012). Interne Evaluation als Instrument der Selbststeuerung von Schulen. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung* (Educational Governance, Bd. 9, S. 35–55). Springer VS. [https://doi.org/10.1007/978-3-531-94183-7\\_2](https://doi.org/10.1007/978-3-531-94183-7_2)
- Thiel, F., Schewe, C. M. & Riecke-Baulecke, T. (2022). Dienstliche Beurteilung als multifunktionales Instrument des Personalmanagements und der Personalentwicklung in Schulen. In F. Thiel, C. M. Schewe, B. Muslic, E.-M. Lankes, N. Maritzen & T. Riecke-Baulecke (Hrsg.), *Personalentwicklung in Schulen als Führungsaufgabe. Eine Bestandsaufnahme in den Ländern der Bundesrepublik Deutschland* (S. 133–167). Springer VS. [https://doi.org/10.1007/978-3-658-36925-5\\_6](https://doi.org/10.1007/978-3-658-36925-5_6)
- Thillmann, K., Bach, A., Wurster, S. & Thiel, F. (2015a). School-based staff development in two federal states in Germany. *International Journal of Educational Management*, 29(6), 714–734. <https://doi.org/10.1108/IJEM-07-2014-0094>
- Thillmann, K., Brauckmann, S., Herrmann, C. & Thiel, F. (2015b). Praxis schulischer Personalentwicklung unter den Bedingungen der Neuen Steuerung. In H. J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungssystem. Analysen zur Mehrebenenperspektive, Steuerung und Koordination*

- (Educational Governance, Bd. 26, S. 195–228). Springer VS.  
[https://doi.org/10.1007/978-3-658-06523-2\\_8](https://doi.org/10.1007/978-3-658-06523-2_8)
- 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>
- Toch, T. & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education* (Education Sector Reports). Education Sector. <http://www.future-ed.org/wp-content/uploads/2017/02/RushToJudgment.pdf>
- Tuma, A. P., Hamilton, L. S. & Tsai, T. (2018). *A nationwide look at teacher perceptions of feedback and evaluation systems: Findings from the American Teacher Panel*. Rand Corporation. [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR2500/RR2558/RAND\\_RR2558.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR2500/RR2558/RAND_RR2558.pdf)
- Tuytens, M. & Devos, G. (2011). Stimulating professional learning through teacher evaluation: An impossible task for the school leader? *Teaching and Teacher Education*, 27(5), 891–899. <https://doi.org/10.1016/j.tate.2011.02.004>
- Tuytens, M. & Devos, G. (2014). How to activate teachers through teacher evaluation? *School Effectiveness and School Improvement*, 25(4), 509–530.  
<https://doi.org/10.1080/09243453.2013.842601>
- Tuytens, M. & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers and Teaching*, 23(1), 6–24. <https://doi.org/10.1080/13540602.2016.1203770>
- van Daal, T., Donche, V. & De Maeyer, S. (2014). The impact of personality, goal orientation and self-efficacy on participation of high school teachers in learning activities in the workplace. *Vocations and Learning*, 7(1), 21–40.  
<https://doi.org/10.1007/s12186-013-9105-5>
- van der Lans, R. M., van de Grift, W. J., van Veen, K. & Fokkens-Bruinsma, M. (2016). Once is not enough: Establishing reliability criteria for feedback and evaluation decisions based on classroom observations. *Studies in Educational Evaluation*, 50(3), 88–95. <https://doi.org/10.1016/j.stueduc.2016.08.001>
- VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement*, 57(6), 995–1015.  
<https://doi.org/10.1177/0013164497057006009>

- VandeWalle, D., Cron, W. L. & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology*, 86(4), 629–640. <https://doi.org/10.1037//0021-9010.86.4.629>
- Walker, A. G., Smither, J. W., Atwater, L. E., Dominick, P. G., Brett, J. M. & Reilly, R. R. (2010). Personality and multisource feedback improvement: A longitudinal investigation. *Journal of Behavioral and Applied Management*, 11(2), 175–204.
- Wang, W. (2007). Evaluation of 2 + 2 alternative teacher performance appraisal program in Shanxi, People's Republic of China. *Teaching and Teacher Education*, 23(7), 1012–1023. <https://doi.org/10.1016/j.tate.2006.05.005>
- Wang, H., Hall, N. C., Goetz, T. & Frenzel, A. C. (2017). Teachers' goal orientations: Effects on classroom goal structures and emotions. *The British Journal of Educational Psychology*, 87(1), 90–107. <https://doi.org/10.1111/bjep.12137>
- Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71(1), 3–25. <https://doi.org/10.1037//0022-0663.71.1.3>
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Bulletin*, 92(4), 548–573. <https://doi.org/10.1037/0033-295X.92.4.548>
- Weisberg, D., Sexton, S., Mulhern, J. & Kelling, D. (2009). *The widget effect. Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. [https://tntp.org/assets/documents/TheWidgetEffect\\_2nd\\_ed.pdf](https://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf)
- Wirthwein, L., Sparfeldt, J. R., Piquart, M., Wegerer, J. & Steinmayr, R. (2013). Achievement goals and academic achievement: A closer look at moderating factors. *Educational Research Review*, 10, 66–89. <https://doi.org/10.1016/j.edurev.2013.07.001>
- Witziers, B., Bosker, R. J. & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398–425. <https://doi.org/10.1177/0013161X03253411>

## Teilstudie 1: Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interven- tionsstudie unter Berücksichtigung der Zielorientierungen von Lehrkräften

Dies ist eine Version des Autors des Beitrags, der in der  
*Zeitschrift für Erziehungswissenschaft* veröffentlicht wurde als:

Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F. (2022). Effekte eines  
Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der  
Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interventionsstudie unter  
Berücksichtigung der Zielorientierungen von Lehrkräften.  
*Zeitschrift für Erziehungswissenschaft*, 25(5), 1267–1291.

<https://doi.org/10.1007/s11618-022-01087-y>

### **3 Teilstudie 1: Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interventionsstudie unter Berücksichtigung der Zielorientierungen von Lehrkräften**

**Zusammenfassung:** Unterrichtsfeedbacks von Schulleitungen sind häufig von geringer Qualität und daher nicht immer wirksam. Aus diesem Grund wurde eine Feedbackintervention entwickelt, die Schulleitungen dabei unterstützen soll, Lehrkräften ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben, auf dessen Grundlage diese ihre Unterrichtskompetenz weiterentwickeln können. Die Wirksamkeit der Intervention wurde in einer Studie mit quasi-experimentellem Design überprüft. Im vorliegenden Beitrag wird der Frage nachgegangen, ob Lehrkräfte, die ein Unterrichtsfeedback von ihrer Schulleitung erhielten, ihre Kompetenzentwicklung positiver einschätzen als Lehrkräfte, die kein Feedback erhielten. Weil Lehrkräfte der Interventionsgruppe ihre Kompetenzentwicklung nach dem Unterrichtsfeedback im Mittel nicht positiver einschätzten als Lehrkräfte der Kontrollgruppe, wurde zusätzlich geprüft, ob der Effekt des Unterrichtsfeedbacks mit den Zielorientierungen der Lehrkräfte interagiert. Für die Lern- und Annäherungsleistungszielorientierung zeigten sich differenzielle Effekte. Lehrkräfte mit hoch ausgeprägter Lern- und Annäherungsleistungszielorientierung, die ein Feedback erhielten, schätzten ihre Kompetenzentwicklung positiver ein als Lehrkräfte der Kontrollgruppe.

**Schlagwörter:** Unterrichtsbeobachtung, Feedback, Professionalisierung von Lehrkräften, Zielorientierungen

**Abstract:** Feedback by school leaders often is of poor quality and thus not always effective. Therefore, a feedback intervention was developed to assist school leaders in providing high-quality feedback to teachers based on which they can improve their instructional competence. The effectiveness of the intervention was tested within a quasi-experimental study. The present paper examines whether teachers who received feedback from their school leader rate their competence development more positively than teachers who did not receive feedback. After receiving feedback, teachers of the intervention group on average did not report a higher competence development than teachers of the control group. Therefore, interactions of school leaders' feedback and teachers' goal orientations were examined. With regard to teachers' learning goal orientation and performance approach goal orientation, an interaction was found. Teachers with a pronounced learning goal orientation and performance approach goal orientation who received feedback reported a higher competence development than teachers of the control group.

**Keywords:** Classroom observation, feedback, goal orientation, teachers' professional development

### 3.1 Einleitung

Die Unterrichtskompetenz von Lehrkräften nimmt entscheidenden Einfluss auf die Qualität des in Schulen angebotenen Unterrichts und damit auf den Lernerfolg von Schülerinnen und Schülern (vgl. Lipowsky 2006; Muijs et al. 2014). Ein beobachtungsbasiertes Unterrichtsfeedback von Schulleitungen wird als erfolgversprechende Maßnahme diskutiert, die zu einer Weiterentwicklung der Unterrichtskompetenz beitragen und – vermittelt über eine verbesserte Unterrichtsqualität – einen positiven Effekt auf den Lernerfolg von Schülerinnen und Schülern haben kann. Studienergebnisse zeigen jedoch, dass Unterrichtsfeedbacks nicht immer wirksam sind (vgl. Kraft et al. 2018; Steinberg und Sartain 2015; Taylor und Tyler 2012). Eine mögliche Ursache für die teilweise ausbleibenden Feedbackeffekte ist die häufig geringe Qualität eines Feedbacks von Schulleitungen (vgl. Kimball und Milanowski 2009; Kraft und Gilmour 2016; Milanowski und Heneman 2001). Aus diesem Grund wurde eine Feedbackintervention entwickelt, die Schulleitungen dabei unterstützen soll, Lehrkräften ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben, auf dessen Grundlage diese ihre Unterrichtskompetenz weiterentwickeln können. Die Wirksamkeit der Feedbackintervention wurde in einer Studie mit quasi-experimentellem Design überprüft. Im vorliegenden Beitrag wird der Frage nachgegangen, ob Lehrkräfte, die ein Unterrichtsfeedback von ihrer Schulleitung erhielten, ihre Kompetenzentwicklung positiver einschätzen als Lehrkräfte, die kein Feedback erhielten.

Vor dem Hintergrund von Angebots-Nutzungs-Modellen, die die Wirkungsweise verschiedener Lerngelegenheiten auf die Kompetenzentwicklung von Lehrkräften beschreiben (vgl. Kunter et al. 2011; siehe auch Helmke 2014), ist anzunehmen, dass die Wirksamkeit von Unterrichtsfeedback nicht allein von der Feedbackqualität abhängt, sondern auch von der Feedbacknutzung, die ihrerseits von verschiedenen kognitiven und motivationalen Merkmalen der Lehrkraft beeinflusst wird. Hinsichtlich der motivationalen Merkmale haben sich Zielorientierungen als bedeutsam für die Nutzung verschiedener Lerngelegenheiten erwiesen (vgl. z. B. Butler 2007; Janke et al. 2019; Nitsche et al. 2013a, 2013b). Daher wird zusätzlich geprüft, ob der Effekt des Unterrichtsfeedbacks auf die Kompetenzselbsteinschätzungen mit den Zielorientierungen der Lehrkräfte interagiert.

## 3.2 Theoretischer Rahmen und Forschungsbefunde

### 3.2.1 Beobachtungsbasiertes Unterrichtsfeedback und die *Feedback Intervention Theory*

Feedback kann als „bewusste Rückmeldung von Informationen an eine Person zu ihrem vorherigen Verhalten“ (Müller und Ditton 2014, S. 15) definiert werden. Ein beobachtungsbasiertes Unterrichtsfeedback stellt dementsprechend eine Rückmeldung von Informationen an eine Lehrkraft zu ihrem in einer beobachteten Unterrichtsstunde gezeigten Verhalten dar. Einem beobachtungsbasierten Unterrichtsfeedback liegt in der Regel ein theoretisches Unterrichtsmodell zugrunde, das lernwirksames Unterrichtsverhalten von Lehrkräften beschreibt. Das beschriebene Unterrichtsverhalten stellt den Soll-Standard dar, von dem ausgehend die Schulleitung das in der beobachteten Unterrichtsstunde gezeigte Verhalten einer Lehrkraft, den Ist-Stand, beurteilt. Das Unterrichtsfeedback enthält schließlich einen Vergleich des beobachteten Verhaltens mit dem im Unterrichtsmodell definierten Soll-Standard.

Die Wirkungsweise von Feedback lässt sich anhand der *Feedback Intervention Theory* (FIT) beschreiben (vgl. Kluger und DeNisi 1996). Nach der FIT ist Verhalten zielorientiert und wird über einen Soll-Ist-Vergleich reguliert. Die FIT unterscheidet vereinfacht dargestellt zwischen einer aufgaben- und einer selbstbezogenen Feedbackverarbeitung. Bei einer aufgabenbezogenen Feedbackverarbeitung zieht eine Soll-Ist-Diskrepanz eine erhöhte Anstrengungsbereitschaft nach sich (vgl. Kluger und DeNisi 1996): Macht ein kritisches Unterrichtsfeedback eine Lehrkraft auf eine Diskrepanz zwischen ihren aktuellen und den im Unterrichtsmodell beschriebenen Verhaltensweisen aufmerksam, sollte sie nach den Annahmen der FIT darum bemüht sein, diese Diskrepanz zu reduzieren. Bei einem bestätigenden Unterrichtsfeedback ist es denkbar, dass der Diskrepanzreduktion eine Diskrepanzproduktion vorausgeht (vgl. Bandura und Locke 2003), d.h. Lehrkräfte setzen sich nach einem bestätigenden Unterrichtsfeedback ggf. höhere Ziele und erzeugen so eine Diskrepanz zwischen ihrem aktuellen und einem erwünschten Unterrichtsverhalten. Führt ein Mehraufwand allein nicht zu einer Diskrepanzreduktion, werden nach den Annahmen der FIT neue Problemlösestrategien entwickelt, die zu einer Reduktion der Soll-Ist-Diskrepanz führen können: „If working harder fails, people may try to work smarter by generating a hypothesis regarding means for improved performance“ (Kluger und DeNisi 1996, S. 263). Lehrkräfte sollten demnach neue Verhaltens-

strategien für ihren Unterricht entwickeln, die – sofern sie sich bewähren – zu einer Reduktion der Diskrepanz zwischen ihren aktuellen und den erwünschten Verhaltensweisen führen. Neben einer aufgabenbezogenen Feedbackverarbeitung kann die Feedbackverarbeitung jedoch auch selbstbezogen erfolgen, was u.a. bei einem sehr kritischen Feedback zu erwarten ist (vgl. Ilgen und Davis 2000). Eine selbstbezogene Feedbackverarbeitung geht nach der FIT mit negativen Emotionen einher, hat einen Abzug von kognitiven Ressourcen von der Aufgabenbearbeitung zur Folge und kann zu einer Ablehnung des Feedbacks führen (vgl. Kluger und DeNisi 1996). Lehrkräfte könnten sich infolge eines sehr kritischen Unterrichtsfeedbacks z. B. fragen, ob sie für den Lehrerberuf ungeeignet sind und – um ihren Selbstwert zu schützen – das Unterrichtsfeedback zurückweisen.

Aus der FIT lässt sich schlussfolgern, dass Unterrichtsfeedback bei einer selbstbezogenen Feedbackverarbeitung weitgehend wirkungslos bleiben sollte. Erfolgt die Feedbackverarbeitung jedoch aufgabenbezogen, kann es zu einer Weiterentwicklung der Unterrichtskompetenz von Lehrkräften beitragen. In Anlehnung an das Kompetenzmodell der COACTIV-Studie (vgl. Baumert und Kunter 2006, 2011), wird unter Unterrichtskompetenz im vorliegenden Beitrag das für die Planung und Durchführung von Unterricht relevante professionelle Wissen und Können von Lehrkräften verstanden, das sich in ihrem Unterrichtsverhalten ausdrückt.

### **3.2.2 Forschungsstand zur Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback**

Die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks von Schulleitungen auf die Weiterentwicklung der Unterrichtskompetenz von Lehrkräften wurde nach unserem Kenntnisstand noch nicht erforscht. Es liegen aber Studien vor, die die Effekte von beobachtungsbasiertem Unterrichtsfeedback auf das Unterrichtsverhalten von Lehrkräften und den Lernerfolg von Schülerinnen und Schülern untersuchen. So zeigen Taylor und Tyler (2012) in einer quasi-experimentellen Studie, dass ein mehrmaliges Feedback von Schulleitungen und von vom Unterricht freigestellten Lehrkräften, das u.a. auf einer Unterrichtsbeobachtung basiert, einen positiven Effekt auf den Lernzuwachs von Schülerinnen und Schülern im Fach Mathematik hat. Auf den Lernzuwachs im Lesen hatte das Unterrichtsfeedback demgegenüber keinen Effekt (vgl. Taylor und Tyler 2012, S. 3648). Steinberg und Sartain (2015) berichten in einer experimentellen Studie von einem positiven Effekt eines beobachtungsbasierten Unterrichtsfeedbacks von Schulleitungen auf den Lernzuwachs im Lesen, jedoch nicht im Fach Mathematik. Weitere Hinweise auf die

Wirksamkeit von Unterrichtsfeedbacks liefert eine aktuelle Meta-Analyse zu Coaching-Programmen, in der ausschließlich experimentelle Studien berücksichtigt wurden (vgl. Kraft et al. 2018). Coaching wird hier „broadly as all in-service PD [professional development] programs, where coaches or peers observe teachers’ instruction and provide feedback to help them improve” (Kraft et al. 2018, S. 553) definiert. Im Mittel hatten die Coaching-Programme einen großen positiven Effekt auf das Unterrichtsverhalten von Lehrkräften und einen kleinen positiven Effekt auf den Lernerfolg ihrer Schülerinnen und Schüler. Allerdings variieren die Effekte stark; einige Coaching-Programme haben keinen Effekt, andere sogar negative Effekte. Weil sich hinsichtlich des Umfangs und der Dauer des Coachings keine differenziellen Effekte zeigten, vermuten die Autoren, dass die Qualität und der Fokus des Coachings entscheidend seien (vgl. Kraft et al. 2018). Diese Schlussfolgerung lässt sich auf ein Unterrichtsfeedback von Schulleitungen übertragen; auch dessen Wirksamkeit hängt von seinem Fokus und seiner Qualität ab. Interviews mit Schulleitungen und Lehrkräften deuten jedoch darauf hin, dass Unterrichtsfeedbacks von Schulleitungen häufig eine geringe Qualität aufweisen (vgl. Kimball und Milanowski 2009; Kraft und Gilmour 2016; Milanowski und Heneman 2001), was die teilweise ausbleibenden Feedbackeffekte erklären kann.

### **3.2.3 Was zeichnet ein qualitativ hochwertiges Feedback aus?**

Weil zu den Merkmalen eines qualitativ hochwertigen Unterrichtsfeedbacks vergleichsweise wenige Befunde vorliegen (für einen Überblick über den Forschungsstand siehe Scheeler et al. 2004; Thurlings et al. 2013), soll im Folgenden erneut Anschluss an die allgemeine Feedbackforschung gesucht werden, deren Theorien und Befunde der entwickelten Feedbackintervention zugrunde gelegt wurden. Nach der FIT lässt sich über die Feedbackgestaltung beeinflussen, ob die Feedbackverarbeitung aufgaben- oder selbstbezogen erfolgt (vgl. Kluger und DeNisi 1996): Während Hinweise auf die zu bearbeitende Aufgabe zu einer aufgabenbezogenen Feedbackverarbeitung führen sollten, macht ein hoher Personenbezug eine selbstbezogene Feedbackverarbeitung wahrscheinlich. Auch Hattie und Timperley (2007) postulieren, dass Feedback dann besonders wirksam ist, wenn es einen hohen Aufgabenbezug aufweist. Nach Hattie und Timperley (2007) sollte Feedback daher drei Arten von Informationen enthalten, die sie als FEED UP, FEED BACK und FEED FORWARD bezeichnen und die mit den Fragen *Where am I going?*, *How am I going?* und *Where to next?* korrespondieren. Feedback sollte Informationen zu den Zielen oder Standards enthalten (FEED UP), die die Beurteilung eines Lernertrags

oder einer Leistung erst ermöglichen. Feedback sollte darüber hinaus aufzeigen, inwiefern der aktuelle Lern- oder Leistungsstand den Zielen oder Standards entspricht (FEED BACK) und es sollte um Informationen ergänzt werden, die den Lernenden oder Leistungserbringenden dabei unterstützen, diese Ziele oder Standards zukünftig zu erreichen (FEED FORWARD).

Kluger und DeNisi (1996) und Hattie und Timperley (2007) haben im Rahmen ihrer Meta-Analysen diverse Feedbackmerkmale identifiziert, die in einem positiven Zusammenhang mit Lernen und Leistung stehen und die Annahme stützen, nach der Feedback einen hohen Aufgabenbezug aufweisen sollte. Feedback sollte demnach auf vorab definierte Ziele oder Standards bezogen sein (vgl. Hattie und Timperley 2007) oder einen Vergleich zu vorherigen Lernerträgen und Leistungen beinhalten (vgl. Kluger und DeNisi 1996). Sowohl bestätigendes als auch kritisches Feedback kann einen positiven Effekt auf Lernen und Leistung haben (vgl. Hattie und Timperley 2007; Kluger und DeNisi 1996). Bestätigendes Feedback geht allerdings nur mit positiven Effekten einher, wenn es auf konkrete Lernerträge oder Leistungen bezogen ist; einfaches Lob ist demgegenüber zu vermeiden (vgl. Hattie und Timperley 2007; Kluger und DeNisi 1996). Kritisches Feedback ist besonders wirksam, wenn es Hinweise dazu enthält, wie Lernende oder Leistungserbringende ihren Lernertrag steigern bzw. ihre Leistung verbessern können (vgl. Hattie und Timperley 2007; Kluger und DeNisi 1996). Darüber hinaus erhöhen Zielvereinbarungen, in denen konkret festgehalten wird, welche Lernerträge bzw. Leistungen in Zukunft wie realisiert werden sollen, die Wirksamkeit von Feedbackinterventionen (vgl. Kluger und DeNisi 1996). Ein weiterer Befund der Meta-Analysen von Hattie und Timperley (2007) und Kluger und DeNisi (1996) ist, dass computerbasiertes Feedback besonders wirksam ist, weil es vom Feedbackempfänger vermutlich als objektiver wahrgenommen wird.

### **3.2.4 Zielorientierungen von Lehrkräften und die Feedbacknutzung**

Die Wirksamkeit von Unterrichtsfeedback hängt jedoch nicht allein von der Feedbackqualität ab. Ein Unterrichtsfeedback stellt für Lehrkräfte eine Lerngelegenheit dar. Es führt nur dann zu einer Weiterentwicklung ihrer Unterrichtskompetenz, wenn Lehrkräfte die im Feedback enthaltenen Informationen auch nutzen: „No amount of feedback will result in professional growth if a teacher is unwilling or unable to co-construct and enact changes“ (Kraft und Gilmour 2016, S. 715). Forschungsbefunde deuten darauf hin,

dass die Nutzung verschiedener Lerngelegenheiten in einem Zusammenhang mit den Zielorientierungen von Lehrkräften steht.

Bei Zielorientierungen handelt es sich um vergleichsweise stabile motivationale Orientierungen, die Einfluss auf das Erleben und die Verhaltensregulation in Lern- und Leistungssituationen nehmen (vgl. Nitsche et al. 2011). In der frühen Zielorientierungsforschung wurde zwischen einer Lern- (*LZO*) und einer Leistungszielorientierung unterschieden (vgl. Ames 1992; Dweck 1986; Nicholls 1984). Während Personen mit hoch ausgeprägter *LZO* in Lern- und Leistungssituationen danach streben, ihre Kompetenzen weiterzuentwickeln, geht es Personen mit hoch ausgeprägter Leistungszielorientierung darum, ihre Überlegenheit zu demonstrieren. Später wurde die Leistungszielorientierung in eine Annäherungs- (*ALZO*) und eine Vermeidungsleistungszielorientierung (*VLZO*) ausdifferenziert (vgl. Elliot und Harackiewicz 1996; Harackiewicz et al. 2002). Personen mit hoch ausgeprägter *ALZO* streben danach, ihre Überlegenheit zu demonstrieren, Personen mit hoch ausgeprägter *VLZO* wollen ihre Kompetenzdefizite verbergen.

Nach unserem Kenntnisstand wurde die Bedeutung der *LZO* für die Nutzung eines Unterrichtsfeedbacks bislang nicht untersucht. Befunde zu den lernrelevanten Einstellungen und Verhaltensweisen von Lehrkräften deuten aber darauf hin, dass eine hoch ausgeprägte *LZO* die Feedbacknutzung begünstigen könnte. So haben lernzielorientierte Lehrkräfte positive Einstellungen gegenüber externer Hilfe und nehmen diese häufig in Anspruch (vgl. Butler 2007; Janke et al. 2019; Nitsche et al. 2011). Außerdem besteht zwischen einer hoch ausgeprägten *LZO* und dem Einholen von Feedback ein positiver Zusammenhang (vgl. Chughtai und Buckley 2010; Runhaar et al. 2010). Lernzielorientierte Lehrkräfte lesen häufig fachbezogene Literatur, sie haben positive Einstellungen gegenüber Fortbildungen und nehmen häufig an diesen teil (vgl. Nitsche et al. 2013a, 2013b).

Ob eine hoch ausgeprägte *ALZO* die Feedbacknutzung begünstigt, lässt sich nicht eindeutig vorhersagen. Theoretisch ist vorstellbar, dass annäherungsleistungszielorientierte Lehrkräfte ihre Unterrichtskompetenz auf der Grundlage eines Unterrichtsfeedbacks weiterentwickeln, um ihre Überlegenheit auch in Zukunft demonstrieren zu können. Die Befunde zu den lernrelevanten Einstellungen und Verhaltensweisen von annäherungsleistungszielorientierten Lehrkräften sind jedoch inkonsistent. Hinsichtlich der Einstellungen gegenüber und der Inanspruchnahme von Hilfe wurden in den zuvor zitierten Studien keine Zusammenhänge mit der *ALZO* gefunden (vgl. Butler 2007; Janke et al. 2019; Nitsche et al. 2011). Annäherungsleistungszielorientierte Lehrkräfte lesen aber häufig fachbezogene Literatur (vgl. Nitsche et al. 2013a). Zwischen einer *ALZO* und den

Einstellungen gegenüber Fortbildungen besteht kein Zusammenhang, besucht werden sie von Lehrkräften mit hoch ausgeprägter *ALZO* dennoch häufig (vgl. Nitsche et al. 2013b).

Eine hoch ausgeprägte *VLZO* sollte demgegenüber in keinem Zusammenhang mit der Feedbacknutzung stehen. Vermeidungsleistungszielorientierte Lehrkräfte haben negative Einstellungen gegenüber externer Hilfe (vgl. Butler 2007; Janke et al. 2019; Nitsche et al. 2011) und Fortbildungen, die sie auch seltener besuchen (vgl. Nitsche et al. 2013b). Außerdem besteht zwischen dem Lesen fachbezogener Literatur und einer hoch ausgeprägten *VLZO* kein Zusammenhang (vgl. Nitsche et al. 2013b).

### 3.3 Fragestellung und Hypothesen

Ein Unterrichtsfeedback von Schulleitungen stellt für Lehrkräfte eine Lerngelegenheit dar, die zu einer Weiterentwicklung ihrer Unterrichtskompetenz beitragen kann. Studienergebnisse zeigen, dass Unterrichtsfeedbacks jedoch nicht immer wirksam sind. Als ursächlich für die teilweise ausbleibenden Feedbackeffekte kann die häufig geringe Qualität von Unterrichtsfeedbacks von Schulleitungen gelten. Aus diesem Grund wurde eine Feedbackintervention entwickelt, die es Schulleitungen ermöglichen soll, Lehrkräften ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben, auf dessen Grundlage sie ihre Unterrichtskompetenz weiterentwickeln können. Für die Unterrichtsbeobachtungen wurde ein standardisiertes Beobachtungsinstrument entwickelt. Für die Feedbackgespräche wurde – ausgehend von den Theorien und Befunden der allgemeinen Feedbackforschung – ein Feedbackverfahren entworfen, das zu einer aufgabenbezogenen Feedbackverarbeitung führen soll. Die Schulleitungen wurden in einem Beobachtungs- und Feedbacktraining in der Nutzung des Beobachtungsinstruments und der Anwendung des Feedbackverfahrens geschult. Die Effekte der Feedbackintervention auf die von Lehrkräften eingeschätzte Weiterentwicklung ihrer Unterrichtskompetenz wurde in einer Studie mit quasi-experimentellem Design überprüft.

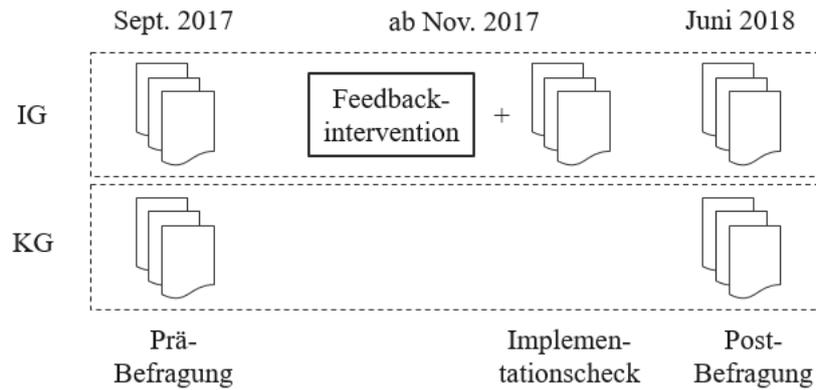
Im vorliegenden Beitrag wird zunächst untersucht, wie Lehrkräfte die Qualität des Unterrichtsfeedbacks einschätzen (*Forschungsfrage 1*). Das Hauptaugenmerk gilt jedoch dem Effekt des Unterrichtsfeedbacks auf die selbst eingeschätzte Kompetenzentwicklung der Lehrkräfte. Es wird überprüft, ob Lehrkräfte, die ein Unterrichtsfeedback von ihrer Schulleitung erhielten, ihre Kompetenzentwicklung positiver einschätzen als Lehrkräfte, die kein Feedback bekamen (*Forschungsfrage 2*). Die Wirksamkeit von Unterrichtsfeedback hängt jedoch nicht allein von der Feedbackqualität ab. Unter Rückgriff auf Ange-

bots-Nutzungs-Modelle lässt sich die Kompetenzentwicklung als Interaktion des Feedbacks und verschiedener individueller Merkmale der Lehrkräfte verstehen, von denen die Feedbacknutzung abhängt (vgl. Kunter et al. 2011). Zielorientierungen spielen dabei vermutlich eine besondere Rolle. Daher wird zusätzlich untersucht, ob der Effekt des Unterrichtsfeedbacks auf die Kompetenzselbsteinschätzung mit den Zielorientierungen von Lehrkräften interagiert (*Forschungsfrage 3*). Vor dem Hintergrund des dargestellten Forschungsstands wird erwartet, dass Lehrkräfte mit hoch ausgeprägter *LZO* ihre Kompetenzentwicklung nach dem Unterrichtsfeedback positiver einschätzen als Lehrkräfte mit hoch ausgeprägter *LZO*, die kein Feedback erhielten. Hinsichtlich der *ALZO* wird aufgrund des inkonsistenten Forschungsstands keine gerichtete Hypothese formuliert. Darüber hinaus wird erwartet, dass der Effekt des Unterrichtsfeedbacks nicht mit der *VLZO* der Lehrkräfte interagiert. Auch im Hinblick auf die *VLZO* wird eine ungerichtete Hypothese getestet.

## **3.4 Methodik**

### **3.4.1 Design**

Alle teilnehmenden Lehrkräfte haben im September 2017 an einer Prä-Befragung teilgenommen (siehe Abb. 1). Das Beobachtungs- und Feedbacktraining fand im September und Oktober 2017 statt und wurde über einen Kurzfragebogen evaluiert. Während Lehrkräfte der Interventionsgruppe (IG) im Winter 2017/2018 ein Unterrichtsfeedback von ihrer Schulleitung erhielten, wurde den Lehrkräften der Kontrollgruppe (KG) kein Feedback gegeben. Die Post-Befragung der Lehrkräfte fand im Juni 2018 statt und wurde wie die Prä-Befragung von geschulten Testleiterinnen und Testleitern administriert. Darüber hinaus liegen die Ergebnisse eines Implementationschecks vor, den die Lehrkräfte der IG im Anschluss an den Erhalt des Unterrichtsfeedbacks ausfüllten und der Autorengruppe postalisch zuschickten. Die Anonymität der Lehrkräfte war jederzeit gewährleistet, sowohl gegenüber der Autorengruppe als auch gegenüber den Schulleitungen.



**Abb. 1:** Studiendesign

### 3.4.2 Stichprobe

Die Stichprobe umfasst  $N = 6$  Berliner Oberstufenzentren (OSZ)<sup>12</sup> mit  $N = 11$  Schulleitungen (inkl. stellvertretenden Schulleitungen, Abteilungs- und Fachleitungen) und  $N = 81$  Lehrkräften.  $N = 35$  Lehrkräfte haben ein Unterrichtsfeedback erhalten (IG),  $N = 46$  Lehrkräfte haben kein Unterrichtsfeedback erhalten (KG). 55 % der teilnehmenden Lehrkräfte sind weiblich. 27.5 % der Lehrkräfte waren zum Zeitpunkt der Datenerhebung 30 bis 39 Jahre alt, 21.3 % waren 40 bis 49 Jahre alt, 45 % waren 50 bis 59 Jahre alt und 6.3 % waren 60 Jahre alt oder älter. Die Lehrkräfte der IG und KG unterscheiden sich hinsichtlich ihres Geschlechts und Alters nicht signifikant voneinander.

Die Schulen nahmen freiwillig an der Studie teil. Die Schulleitungen aller Berliner OSZ wurden auf einer turnusmäßigen Schulleitungskonferenz über die Studie informiert und konnten im Anschluss über eine Teilnahme entscheiden. Eine randomisierte Zuteilung der Lehrkräfte zur IG und KG war nicht möglich. Die Entscheidung, welche Lehrkräfte an der Studie teilnehmen und welcher Gruppe sie angehören würden, wurde schulintern getroffen. Es ist davon auszugehen, dass an einigen Schulen die Lehrkräfte über ihre Teilnahme und Gruppenzugehörigkeit entschieden, während diese Entscheidungen an anderen Schulen von der Schulleitung getroffen wurden. An jeder Schule gibt es Lehrkräfte der IG und KG, die sich weitgehend gleichmäßig auf die Gruppen verteilen.

Weil eine Randomisierung nicht möglich war, könnten sich die Lehrkräfte der IG und KG systematisch voneinander unterscheiden. In Fällen, in denen die Lehrkräfte über ihre Teilnahme und Gruppenzugehörigkeit entschieden, ist anzunehmen, dass sich vor

<sup>12</sup> An Berliner OSZ werden berufsvorbereitende und -qualifizierende sowie studienberechtigende Bildungsgänge angeboten, die zu verschiedenen Schulabschlüssen führen können (z. B. (erweiterte) Berufsbildungsreife, mittlerer Schulabschluss, Fach- sowie allgemeine Hochschulreife).

allem Lehrkräfte für eine Teilnahme in der IG entschieden, die ohnehin eine hohe Bereitschaft zeigen, sich professionell weiterzuentwickeln. In Fällen, in denen diese Entscheidungen von der Schulleitung getroffen wurden, erscheint es denkbar, dass z. B. Lehrkräfte überproportional häufig für eine Teilnahme in der IG ausgewählt wurden, die kein Lehramtsstudium abgeschlossen haben und/oder über wenig Berufserfahrung verfügen. Möglichen Unterschieden zwischen der IG und KG wurde Rechnung getragen, indem weitere Lernaktivitäten und diverse Lehrkräfte Merkmale erfasst wurden. Diese Variablen werden in den späteren Analysen als Kontrollvariablen berücksichtigt, um den Feedbackeffekt zu isolieren.

### **3.4.3 Beschreibung der Feedbackintervention**

Bei der Feedbackintervention handelt es sich um ein einmaliges Unterrichtsfeedback von Schulleitungen an Lehrkräfte. Schulleitungen beurteilten den Unterricht der Lehrkräfte der IG mit einem standardisierten Beobachtungsinstrument und gaben ihnen ausgehend von der Unterrichtsbeobachtung ein Feedback zu ihrem Unterricht. Bei der Feedbackgabe orientierten sie sich an einem Feedbackverfahren, das auf Grundlage der Theorien und Befunde der allgemeinen Feedbackforschung entworfen wurde und eine aufgabenbezogene Feedbackverarbeitung evozieren soll. Alle Schulleitungen wurden in einem Beobachtungs- und Feedbacktraining in der Nutzung des Beobachtungsinstruments und der Anwendung des Feedbackverfahrens geschult.

#### **3.4.3.1 Unterrichtsbeobachtung**

Die Schulleitungen beobachteten je Lehrkraft zwei 90-minütige Unterrichtseinheiten und beurteilten das Unterrichtsverhalten der Lehrkräfte mit einem von der Autorengruppe entwickelten Beobachtungsinstrument (vgl. Gärtner et al. 2021). Das Beobachtungsinstrument soll vorrangig im Rahmen von Unterrichtsevaluationen durch Schulleitungen eingesetzt werden, die mit dem Ziel eines entwicklungsbezogenen Unterrichtsfeedbacks erfolgen. Daher liegt der Fokus des Beobachtungsinstruments auf dem Unterrichtsverhalten von Lehrkräften bzw. auf beeinflussbaren Verhaltensweisen von Schülerinnen und Schülern.

Das Beobachtungsinstrument basiert auf einem theoretischen Unterrichtsmodell, das an das Modell der drei Basisdimensionen anknüpft (vgl. Klieme et al. 2001; Praetorius et al. 2018). Weil Schulleitungen nicht in allen Fächern über fachliche und fachdi-

daktische Expertise verfügen, liegt der Schwerpunkt des Unterrichtsmodells auf der fachunabhängigen Gestaltung von Lehr-Lern-Prozessen. Auf der Grundlage von lehr-lernpsychologischen Theorien der Informationsverarbeitung, Motivationstheorien und Theorien der sozialen Interaktion im Klassenzimmer wurden drei übergeordnete Anforderungsbereiche des Unterrichtens identifiziert: die Unterstützung von Schülerinnen und Schülern beim Wissenserwerb, die Motivierung von Schülerinnen und Schülern sowie das Klassenmanagement. Die theoretische Grundlage des Anforderungsbereichs *Unterstützung des Wissenserwerbs* bildet das Lehr-Lern-Prozessmodell von Klauer und Leutner (2012), das zwischen vier Komponenten eines kognitiven Lernprozesses unterscheidet, die in die Unterrichtsmerkmale (1) Erklärung und Präsentation, (2) Unterstützung von Verstehensprozessen, (3) Speicherung durch Üben sowie (4) Anwendung und Transfer überführt wurden. Die theoretische Grundlage des Anforderungsbereichs *Motivierung* bildet die Erwartungs-mal-Wert-Theorie der Motivation nach Eccles und Wigfield (2002; siehe auch Rheinberg 2008), aus der die Unterrichtsmerkmale (5) Weckung von Interesse, (6) Förderung von Lernfreude, (7) Verdeutlichung des Nutzens sowie (8) Förderung der Selbstwirksamkeit hergeleitet wurden. Ausgehend von Theorien der sozialen Interaktion im Klassenzimmer (vgl. Doyle 2006; Ophardt und Thiel 2013) wurden hinsichtlich des Anforderungsbereichs *Klassenmanagement* die Unterrichtsmerkmale (9) Lerndienliches Unterrichtsklima, (10) Monitoring, (11) Steuerung des Unterrichtsflusses, (12) Gruppenaktivierung sowie (13) Umgang mit Störungen abgeleitet. Weil sich Schülerinnen und Schüler hinsichtlich ihrer kognitiven und motivationalen Lernvoraussetzungen unterscheiden, wird (14) Individualisierung als zu den Anforderungsbereichen *Unterstützung des Wissenserwerbs* und *Motivierung* querliegendes Unterrichtsmerkmal verstanden.

Die 14 Unterrichtsmerkmale wurden im Beobachtungsinstrument über 14 Skalen mit vier bis acht Items operationalisiert, die von den Schulleitungen auf einer vierstufigen Likert-Skala eingeschätzt wurden.<sup>13</sup> Die theoretische Grundlage des Beobachtungsinstruments und Befunde zu seiner faktoriellen Validität und Reliabilität werden in einer weiteren Publikation der Autorengruppe ausführlicher beschrieben (vgl. Gärtner et al. 2021).

### 3.4.3.2 Feedbackverfahren

Die Schulleitungen machten sich während der Unterrichtsbeobachtungen Notizen zum Unterrichtsverhalten der Lehrkräfte. Weil sich computerbasiertes Feedback als besonders

---

<sup>13</sup> Eine tabellarische Übersicht über die 14 Unterrichtsmerkmale und Beispielitems finden sich im Online-Supplement 1.

wirksam erwiesen hat, füllten die Schulleitungen das Beobachtungsinstrument im Anschluss an die Beobachtungen in einem webbasierten Feedbackportal aus, das auf der Grundlage der Beobachtungsdaten lehrerspezifische Feedbackberichte generierte. Die Feedbackberichte enthielten eine Beschreibung des Unterrichtsmodells, ein überblicksartiges Feedback sowie bestätigendes und kritisches Feedback. Ein Algorithmus entschied in einem ersten Schritt über die Auswahl der Unterrichtsmerkmale: Die drei Unterrichtsmerkmale mit den höchsten Skalenmittelwerten wurden in den Feedbackberichten als bestätigendes Feedback aufgeführt, die drei Unterrichtsmerkmale mit den niedrigsten Mittelwerten als kritisches Feedback. In einem zweiten Schritt wurde das bestätigende Feedback um Beispiele gelungener Unterrichtspraxis ergänzt, die je Unterrichtsmerkmal aus den Items hergeleitet wurden, die die höchsten Ausprägungen innerhalb der entsprechenden Skalen aufwiesen. Das kritische Unterrichtsfeedback wurde um Hinweise auf alternative Verhaltensstrategien ergänzt, die je Unterrichtsmerkmal aus den Items hergeleitet wurden, die die geringsten Ausprägungen innerhalb der jeweiligen Skalen aufwiesen.

Die lehrerspezifischen Feedbackberichte bildeten die Grundlage für ein mündliches Unterrichtsfeedback. In den Feedbackgesprächen, die eine durchschnittliche Dauer von  $M = 44$  Minuten ( $SD = 13$  Minuten) hatten, wurde den Lehrkräften zu Beginn erläutert, dass das Ziel des Unterrichtsfeedbacks in der Weiterentwicklung ihres Unterrichts liegt und nicht der Kontrolle durch die Schulleitung dient. Anschließend erhielten die Lehrkräfte in Anlehnung an Hattie und Timperley (2007) drei Arten von Informationen: Zunächst wurde den Lehrkräften das dem Feedback zugrundeliegende Unterrichtsmodell vorgestellt, das den Soll-Standard der Unterrichtsbeurteilung definiert (FEED UP). Dadurch wurde den Lehrkräften verdeutlicht, dass dem Feedback eine sachliche Bezugsnorm zugrunde liegt, wodurch der „Fokus auf die Aufgabenstellung und die dafür nötige Anstrengung gelenkt [wird] und weniger auf das Selbst“ (Kopp und Mandl 2014, S. 151). Anschließend erhielten die Lehrkräfte Informationen zu ihrem beobachteten Verhalten (Ist-Stand) im Vergleich zum Soll-Standard (FEED BACK). Die Schulleitungen gaben den Lehrkräften bestätigendes und kritisches Unterrichtsfeedback zu den in den Feedbackberichten aufgeführten Unterrichtsmerkmalen, da beide Feedbackarten einen positiven Effekt auf Lernen und Leistung haben können. Das bestätigende Unterrichtsfeedback wurde anhand der Beispiele gelungener Unterrichtspraxis aus den Feedbackberichten erläutert, das kritische Unterrichtsfeedback um die Hinweise auf alternative Verhaltensstrategien ergänzt (FEED FORWARD). Abschließend formulierten die Schulleitungen und

die Lehrkräfte Entwicklungsziele im Hinblick auf die Umsetzung der alternativen Verhaltensstrategien. Weil ein schriftliches Feedback gegenüber einem mündlichen Feedback den Vorteil hat, dass es von der Lehrkraft mehrmals und zu einem späteren Zeitpunkt erneut gelesen werden kann (vgl. Kopp und Mandl 2014), händigten die Schulleitungen den Lehrkräften die Feedbackberichte im Anschluss an die Feedbackgespräche aus.<sup>14</sup>

### 3.4.3.3 Beobachtungs- und Feedbacktraining

Alle Schulleitungen durchliefen ein Beobachtungs- und Feedbacktraining im Umfang von 18 Stunden. Im Beobachtungstraining wurde den Schulleitungen ein grundlegendes Verständnis des dem Beobachtungsinstrument zugrundeliegenden Unterrichtsmodells vermittelt. Außerdem wurden häufig auftretende Beobachtungsfehler und Strategien zu ihrer Vermeidung vorgestellt. Die Nutzung des Beobachtungsinstruments wurde anhand von videobasierten Beobachtungsübungen trainiert. Im Feedbacktraining wurde den Schulleitungen erläutert, wie das Feedbackverfahren inklusive des webbasierten Feedbackportals und der lehrerspezifischen Feedbackberichte zu nutzen ist. Die Anwendung des Feedbackverfahrens wurde im Rahmen von Rollenspielen eingeübt.

### 3.4.4 Variablen

Die Feedbackqualität wurde mit einer an den Kontext eines Unterrichtsfeedbacks von Schulleitungen angepassten Skala der *Feedback Environment Scale* (vgl. Steelman et al. 2004) erfasst (5 Items; Beispielitem: „Meine Kollegin/mein Kollege hat mir nützliches Feedback gegeben“; Cronbachs  $\alpha = .82$ ). Eine Erfassung der Kompetenzentwicklung über einen Wissens- oder Performanztest konnte in der vorliegenden Studie nicht realisiert werden, weil aktuell kein geeignetes Testverfahren vorliegt und die zeitlichen Ressourcen eine Testentwicklung und -validierung nicht zuließen. Die Weiterentwicklung der Unterrichtskompetenz wurde stattdessen über eine retrospektive Selbsteinschätzung erfasst. Während eine retrospektive Selbsteinschätzung eine direkte Erfassung der Kompetenzentwicklung erlaubt, kann bei einer indirekten Messung der Kompetenzentwicklung, d.h. bei einer Selbsteinschätzung des Kompetenzstandes vor und nach der Feedbackintervention und einer anschließenden Differenzwertbildung, nicht ausgeschlossen werden, dass es zu einem *Response Shift* kommt, also zu einem neuen Bewertungsstandard bei der Beurteilung von Kompetenz oder einem veränderten Begriffsverständnis von Kompetenz

---

<sup>14</sup> Ein exemplarischer Ausschnitt eines Feedbackberichts findet sich im Online-Supplement 2.

(vgl. Piwowar und Thiel 2014; Schwartz und Sprangers 1999), was zu einer Über- oder Unterschätzung des Feedbackeffekts führen kann. Die Lehrkräfte wurden daher darum gebeten, ihre Kompetenzentwicklung für das Schuljahr 2017/2018 rückblickend zu beurteilen (4 Items; Beispielitem: "Ich habe viel dazugelernt, was die Verbesserung meines Unterrichts betrifft"; Cronbachs  $\alpha = .91$ ).

Neben dem Unterrichtsfeedback der Schulleitung können weitere Faktoren die selbst eingeschätzte Kompetenzentwicklung der Lehrkräfte beeinflussen. Zuvorderst sind dies weitere Lernaktivitäten sowie verschiedene Merkmale der Lehrkräfte. Weil eine Randomisierung nicht möglich war und es hinsichtlich der Lernaktivitäten und Lehrkräfte Merkmale systematische Unterschiede zwischen den Lehrkräften der IG und KG geben könnte, wurden diese als Kovariaten erfasst und in den späteren Analysen als Kontrollvariablen berücksichtigt. Im Hinblick auf die Lernaktivitäten wurde die Anzahl der im Schuljahr 2017/2018 besuchten Fort- und Weiterbildungen erfragt. Darüber hinaus wurde das Lesen fachbezogener Literatur mit einem in Anlehnung an Kwakman (2003) entwickelten Instrument erhoben (4 Items; Beispielitem: „Um auf dem aktuellen Stand zu bleiben, lese ich intensiv pädagogische Fachzeitschriften“; Cronbachs  $\alpha = .84$ ). Außerdem wurde erfragt, ob Lehrkräfte im Rahmen einer kollegialen Hospitation ein Feedback erhalten oder sich ein Schülerfeedback eingeholt haben. Hinsichtlich der Lehrkräfte Merkmale wurde der Abschluss eines Lehramtsstudiums und die Berufserfahrung erhoben.

Die Zielorientierungen wurden mit einem von Nitsche et al. (2011) entwickelten Instrument in der Prä-Befragung erfasst, das aus erhebungsökonomischen Gründen gekürzt wurde. Während die *LZO* mit der originalen Skala erfasst wurde, wurde auf eine nach Adressaten differenzierte Erfassung der *ALZO* und *VLZO* verzichtet. Die Kürzungen stehen im Einklang mit einer Kurzversion des Instruments von Nitsche et al. (2011), die von Janke et al. (2019) entwickelt und validiert wurde. Ein Beispielitem der *LZO* (9 Items; Cronbachs  $\alpha = .90$ ) lautet: „In meinem Unterricht strebe ich danach, mit kritischen Unterrichtssituationen besser umgehen zu können“. Ein Beispielitem der *ALZO* (3 Items; Cronbachs  $\alpha = .91$ ) lautet: „In meinem Unterricht strebe ich danach, zu zeigen, dass ich kritische Unterrichtssituationen besser bewältigen kann als andere Lehrkräfte“. Ein Beispielitem der *VLZO* (3 Items; Cronbachs  $\alpha = .83$ ) lautet: „In meinem Unterricht strebe ich danach, nicht zu zeigen, wenn mir die Unterrichtsanforderungen schwerere fallen als anderen Lehrkräften“.

### 3.4.5 Analysen

Die Qualitätseinschätzungen der Lehrkräfte werden deskriptiv ausgewertet. Ein Einstichproben-T-Test wird genutzt, um zu prüfen, ob der Mittelwert der Skala *Qualität des Feedbacks* signifikant vom theoretischen Skalenmittel abweicht. Die Effekte des Unterrichtsfeedbacks auf die selbst eingeschätzte Kompetenzentwicklung werden mithilfe von Regressionsanalysen überprüft. Da von jeder Schule mehrere Lehrkräfte an der Studie teilgenommen haben, liegt eine hierarchische Datenstruktur vor (Lehrkräfte geclustert in Schulen), der mit einer Mehrebenenregressionsanalyse Rechnung getragen werden könnte (vgl. Raudenbush und Bryk 2002). Die im vorliegenden Fall relevanten Random-Intercept-Modelle setzen jedoch mindestens  $N = 30$  Cluster voraus, um zu nicht verzerrten Parameterschätzungen zu kommen (vgl. McNeish und Stapleton 2016). Weil diese Voraussetzung bei  $N = 6$  Schulen nicht erfüllt ist, werden die Effekte des Unterrichtsfeedbacks auf die Kompetenzselbsteinschätzung der Lehrkräfte mithilfe multipler Regressionsanalysen überprüft. Der Modellaufbau erfolgt schrittweise; mit jedem Schritt wird die Modellierung restriktiver. Zunächst wird das Unterrichtsfeedback als einziger Prädiktor modelliert. Im Anschluss werden weitere Lernaktivitäten eingeführt und abschließend die Lehrkräftemerkmale ergänzt. Nicht signifikante Prädiktoren ( $p \geq .10$ ) werden aus den Analysen ausgeschlossen. Weil sich standardisierte Regressionskoeffizienten für dichotome Variablen inhaltlich nicht sinnvoll interpretieren lassen und sie bei einer 0/1-Kodierung von der Größe der Gruppe abhängen, die mit 1 kodiert ist, wurde eine Teilstandardisierung vorgenommen: Während die metrischen Variablen z-standardisiert wurden, ist von einer z-Standardisierung der dichotomen Variablen abgesehen worden (vgl. Hayes 2013; Urban und Mayerl 2018). Die Interaktionsterme zur Berechnung der differenziellen Effekte der Zielorientierungen wurden aus den z-standardisierten Moderatorvariablen und der dichotomen Feedbackvariable gebildet. Da signifikante Moderationseffekte nur Auskunft darüber geben, ob die Zielorientierungen mit der Feedbackvariable interagieren, sich anhand der Moderationseffekte aber nicht beurteilen lässt, ob der Feedbackeffekt bei einer bestimmten Ausprägung der Zielorientierungen das Niveau statistischer Signifikanz erreicht, werden alle signifikanten Moderationseffekte mit sogenannten *Simple Slopes*-Tests überprüft, mit denen sich der Feedbackeffekt bei verschiedenen Ausprägungen der Zielorientierungen ( $-1 SD$ ,  $M$  und  $+1 SD$ ) schätzen lässt (vgl. Hayes 2013). Die Regressionsanalysen wurden mit Mplus 8 durchgeführt (vgl. Muthén und Muthén 2017). Weil nicht alle Variablen eine Normalverteilung aufweisen, wurde ein robustes

*Maximum Likelihood Schätzverfahren* genutzt. Fehlende Werte wurden mit der *Full Information Likelihood Methode* behandelt (vgl. Lüdtke et al. 2007).

## 3.5 Ergebnisse

### 3.5.1 Evaluations des Beobachtungs- und Feedbacktrainings und Implementationscheck

Bevor die Ergebnisse zu den Forschungsfragen präsentiert werden, sollen die Evaluationsergebnisse zum Beobachtungs- und Feedbacktraining der Schulleitungen sowie die Ergebnisse des Implementationschecks vorgestellt werden. Das Beobachtungs- und Feedbacktraining wurde über selbst entwickelte Einzelitems evaluiert, denen die Schulleitungen auf einer 6-stufigen Likert-Skala zustimmen konnten. Hinsichtlich des Beobachtungstrainings gaben die Schulleitungen überwiegend an, sie hätten ihr Wissen über Unterrichtsbeobachtungen erweitert ( $M = 5.08$ ,  $SD = 0.95$ ) und ihre Beobachtungskompetenz weiterentwickelt ( $M = 5.08$ ,  $SD = 0.86$ ). Auch das Feedbacktraining hat in der Wahrnehmung der Schulleitungen zu einer Erweiterung ihres Wissens über wirksames Feedback ( $M = 5.29$ ,  $SD = 1.07$ ) und einer Weiterentwicklung ihrer Feedbackkompetenz geführt ( $M = 5.07$ ,  $SD = 0.62$ ).

Anhand des Implementationschecks lässt sich prüfen, ob die Schulleitungen das Unterrichtsfeedback in Übereinstimmung mit dem entwickelten Feedbackverfahren gestaltet haben. Der Implementationscheck liegt für  $N = 29$  Lehrkräfte vor (82.9 % aller Lehrkräfte der IG). Auch die Umsetzung des Feedbackverfahrens wurde über selbst entwickelte Einzelitems mit einer 6-stufigen Likert-Skala erfasst. Die Lehrkräfte der IG gaben überwiegend an, das Unterrichtsfeedback habe aus ihrer Sicht der Weiterentwicklung ihres Unterrichts ( $M = 5.17$ ,  $SD = 1.14$ ) und nicht der Kontrolle durch die Schulleitung ( $M = 1.52$ ,  $SD = 0.91$ ) gegolten. Neben einer hohen Zielklarheit berichten die Lehrkräfte davon, dass ihnen das theoretische Unterrichtsmodell vorgestellt wurde ( $M = 5.55$ ,  $SD = 0.78$ ). Das bestätigende Unterrichtsfeedback wurde anhand von Beispielen gelungener Unterrichtspraxis erläutert ( $M = 5.55$ ,  $SD = 0.63$ ). Auch der Aussage, das kritische Feedback habe konkrete Hinweise auf alternative Verhaltensstrategien für die Weiterentwicklung ihres Unterrichts enthalten, stimmten die Lehrkräfte überwiegend zu ( $M = 4.97$ ,  $SD = 1.18$ ). Die Zielvereinbarungen wurden akzeptiert, als realistisch eingeschätzt und als konkret wahrgenommen ( $M = 5.58$ ,  $SD = 0.55$ ).

Zusammenfassend lässt die Kurzevaluation des Beobachtungs- und Feedbacktrainings darauf schließen, dass die Schulleitungen ihr Wissen und ihre Kompetenzen über Unterrichtsbeobachtungen und -feedbacks erweitert haben. Die Daten des Implementationschecks zeigen, dass die Schulleitungen das Unterrichtsfeedback in Übereinstimmung mit dem entwickelten Feedbackverfahren gestaltet haben.

### 3.5.2 Qualität des Unterrichtsfeedbacks

Die erste Forschungsfrage bezog sich auf die Feedbackqualität. Der Mittelwert der Skala *Qualität des Feedbacks* fällt mit  $M = 5.33$  ( $SD = 0.71$ ) deutlich höher aus als das theoretische Mittel von  $\mu = 3.50$ . Der Mittelwertunterschied ist statistisch signifikant ( $t(27) = 14.39$ ,  $p < .01$ ) und praktisch bedeutsam ( $d = 2.6$ ) (vgl. Cohen 1988). Weiterführende Analysen zeigen zudem, dass die Qualitätseinschätzungen der Lehrkräfte mit den meisten der Gestaltungsmerkmale des Feedbacks korrespondieren: Je stärker die Lehrkräfte mit dem Unterrichtsfeedback das Ziel der Weiterentwicklung ihres Unterrichts assoziierten ( $r_s = .51$ ,  $p < .01$ ), je besser das bestätigende Unterrichtsfeedback anhand von Beispielen gelungener Unterrichtspraxis erläutert wurde ( $r_s = .44$ ,  $p < .05$ ), je konkreter die alternativen Verhaltensstrategien im Hinblick auf das kritische Feedback waren ( $r_s = .71$ ,  $p < .01$ ) und je mehr Akzeptanz die Zielvereinbarungen erfuhren ( $r_s = .62$ ,  $p < .01$ ), desto höher bewerteten die Lehrkräfte die Feedbackqualität.

### 3.5.3 Effekte des Unterrichtsfeedbacks auf die selbst eingeschätzte Kompetenzentwicklung

#### 3.5.3.1 Deskriptive Analysen

In der zweiten Forschungsfrage wurde nach dem Haupteffekt des Unterrichtsfeedbacks auf die selbst eingeschätzte Kompetenzentwicklung der Lehrkräfte gefragt. Weil sich die Lehrkräfte der IG und KG systematisch voneinander unterscheiden könnten, werden zunächst die Mittelwerte der Kovariaten zwischen den Gruppen verglichen (siehe Tab. 1). Das Signifikanzniveau der Gruppenunterschiede wurde bei metrischen Variablen mithilfe von T-Tests für unabhängige Stichproben oder dem Mann-Whitney-U-Test überprüft und bei dichotomen Variablen mithilfe des  $X^2$ -4-Felder-Tests. Um die Größe der Gruppenunterschiede zu bestimmen, wurde für metrische Variablen Cohens  $d$  und für dichotome Variablen Cohens  $h$  berechnet (vgl. Cohen 1988).

Im Hinblick auf die Lernaktivitäten gibt es Unterschiede zwischen den Lehrkräften der IG und KG. Während sich die Gruppen hinsichtlich der besuchten Fort- und Weiterbildungen und dem Lesen fachbezogener Literatur kaum voneinander unterscheiden, erhielten Lehrkräfte der KG signifikant häufiger ein kollegiales Feedback ( $h = -.47, p < .10$ ) und häufiger ein Schülerfeedback ( $h = -.24, p > .10$ ) als Lehrkräfte der IG. Auch im Hinblick auf die Lehrkräfte Merkmale gibt es Unterschiede zwischen der IG und KG. Während der Anteil an Lehrkräften mit abgeschlossenem Lehramtsstudium und die Berufserfahrung in den Gruppen vergleichbar sind, haben Lehrkräfte der IG im Mittel eine signifikant höher ausgeprägte *LZO* ( $d = .48, p < .05$ ) und eine höher ausgeprägte *ALZO* ( $d = .20, p > .10$ ) als Lehrkräfte der KG. Hinsichtlich der *VLZO* unterscheiden sich die Gruppen kaum voneinander.

**Tab. 1: Gruppenvergleich IG und KG**  
(Mittelwerte (*M*), Standardabweichungen (*SD*) sowie Gruppenunterschiede (*p* und *d/h*))

	Gesamt		IG		KG		Gruppenunterschiede	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>d/h</i>
<b>Lernaktivitäten</b>								
Besuchte Fort- und Weiterbildungen	3.36	1.88	3.33	1.56	3.39	2.17	.80	-.03
Lesen fachbezogener Literatur	3.06	1.03	3.10	1.06	3.03	1.02	.80	.07
Kollegiales Feedback erhalten <sup>1</sup>	0.26	-	0.15	-	0.35	-	.06	-.47
Schülerfeedback erhalten <sup>1</sup>	0.30	-	0.24	-	0.35	-	.32	-.24
<b>Lehrkraftmerkmale</b>								
Lehramtsstudium abgeschlossen <sup>1</sup>	0.75	-	0.74	-	0.76	-	.84	-.02
Berufserfahrung (in Jahren)	13.87	9.47	13.91	9.13	13.84	9.81	.97	.01
<i>LZO</i>	4.93	0.88	5.16	0.66	4.76	0.99	.05	.48
<i>ALZO</i>	1.75	0.94	1.86	1.02	1.67	0.87	.39	.20
<i>VLZO</i>	1.90	0.99	1.88	1.02	1.92	0.98	.83	-.04

Anmerkungen: <sup>1</sup> Dichotome Variable (0 „nein“, 1 „ja“); *LZO* = Lernzielorientierung; *ALZO* = Annäherungsleistungszielorientierung; *VLZO* = Vermeidungsleistungszielorientierung

### 3.5.3.2 Haupteffekte auf die Kompetenzentwicklung

Die Haupteffekte des Unterrichtsfeedbacks können den ersten drei Spalten von Tab. 2 entnommen werden.<sup>15</sup> Für dichotome Prädiktoren werden teilstandardisierte und für metrische Prädiktoren standardisierte Regressionskoeffizienten angegeben. Zunächst wurde der Effekt des Unterrichtsfeedbacks in einer einfachen Regressionsanalyse überprüft (Modell 1). Im bivariaten Modell hat das Unterrichtsfeedback keinen Effekt auf die selbst eingeschätzte Kompetenzentwicklung der Lehrkräfte ( $\beta = .32, p > .10$ ).

Im Modell 2 wurden weitere Lernaktivitäten bei der Modellbildung berücksichtigt. Von den erhobenen Lernaktivitäten hat nur das Lesen fachbezogener Literatur einen signifikanten Effekt auf die Kompetenzselbsteinschätzung. Die besuchten Fort- und Weiterbildungen, das kollegiale Feedback und das Schülerfeedback wurden bei der weiteren Modellbildung daher nicht berücksichtigt. Auch unter Kontrolle des Lesens fachbezogener Literatur hat das Unterrichtsfeedback keinen Effekt auf die selbst eingeschätzte Kompetenzentwicklung der Lehrkräfte ( $\beta = .30, p > .10$ ).

<sup>15</sup> Eine Korrelationstabelle der bivariaten Zusammenhänge zwischen der Kompetenzselbsteinschätzung und den Kovariaten findet sich im Online-Supplement 3.

**Tab. 2:** Ergebnisse der Regressionsanalysen (teilstandardisierte und standardisierte Regressionskoeffizienten ( $\beta$ ) und Standardfehler (SE))

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5	Modell 6
	$\beta$ (SE)					
<b>Feedbackintervention</b>						
Feedbackintervention erhalten <sup>1</sup>	.32 (.24)	.30 (.22)	.09 (.19)	.11 (.22)	.15 (.21)	.09 (.19)
<b>Lernaktivitäten</b>						
Lesen fachbezogener Literatur		.36** (.13)	.23* (.11)		.24+ (.13)	.21+ (.11)
<b>Lehrkraftmerkmale</b>						
Lehramtsstudium <sup>1</sup>			-.54* (.26)			-.50* (.25)
Berufserfahrung (in Jahren)			-.30** (.11)			-.28* (.11)
<i>LZO</i>			.18* (.09)	.13 (.11)	.09 (.10)	.11 (.10)
<i>ALZO</i>			.21* (.09)	-.16 (.23)	-.17 (.22)	.02 (.22)
<i>VLZO</i>				.12 (.18)	.17 (.17)	.06 (.18)
<b>Interaktionsterme</b>						
<i>LZO</i> * Unterrichtsfeedback				.40+ (.21)	.36+ (.20)	.26 (.17)
<i>ALZO</i> * Unterrichtsfeedback				.62* (.26)	.56* (.27)	.24 (.25)
<i>VLZO</i> * Unterrichtsfeedback				-.31 (.22)	-.28 (.22)	-.03 (.22)
<b>Erklärte Varianz</b>						
R <sup>2</sup>	.03	.15+	.43**	.28+	.29*	.39**

Anmerkungen: +  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$

<sup>1</sup> Dichotome Variable (0 „nein“, 1 „ja“); *LZO* = Lernzielorientierung; *ALZO* = Annäherungsleistungszielorientierung; *VLZO* = Vermeidungsleistungszielorientierung

Im Modell 3 wurden abschließend die Lehrkräftemerkmale als Kontrollvariablen in die Analysen aufgenommen. Die *VLZO* wurde nicht modelliert, weil sie keinen Effekt auf die Kompetenzselbsteinschätzung hat. Auch im Modell 3 erreicht der Feedbackeffekt nicht das Niveau statistischer Signifikanz ( $\beta = .09$ ,  $p > .10$ ). Zusammenfassend lässt sich festhalten, dass das Unterrichtsfeedback in keinem der Regressionsmodelle einen Effekt auf die selbst eingeschätzte Kompetenzentwicklung hat. Lehrkräfte, die ein Feedback zu ihrem Unterricht erhielten, schätzten ihre Kompetenzentwicklung nicht positiver ein als Lehrkräfte, die kein Feedback bekamen.

### 3.5.3.3 Differenzielle Effekte der Zielorientierungen

In der dritten Forschungsfrage wurde gefragt, ob der Feedbackeffekt mit den Zielorientierungen der Lehrkräfte interagiert. Die differenziellen Effekte der Zielorientierungen können den letzten drei Spalten von Tab. 2 entnommen werden. Für dichotome Prä-

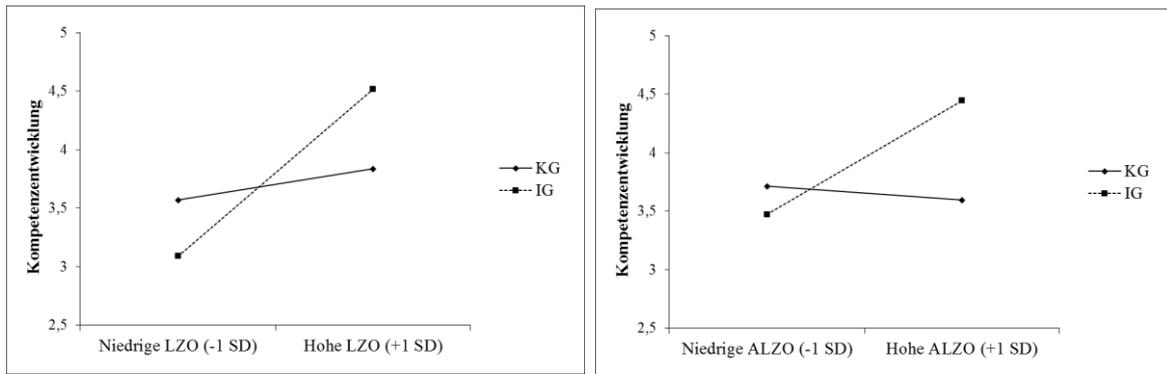
diktoren und die Interaktionsterme werden teilstandardisierte und für metrische Prädiktoren standardisierte Regressionskoeffizienten angegeben. Modell 4 umfasst das Unterrichtsfeedback, die Moderatorvariablen sowie die Interaktionsterme. Für die *LZO* ( $\beta = .40, p < .10$ ) und die *ALZO* ( $\beta = .62, p < .05$ ) konnte ein signifikant positiver Moderationseffekt und für die *VLZO* kein Moderationseffekt gefunden werden ( $\beta = -.31, p > .10$ ).

Im Modell 5 wurde das Lesen fachbezogener Literatur ergänzt. Auch unter Kontrolle des Lesens fachbezogener Literatur moderieren die *LZO* ( $\beta = .36, p < .10$ ) und die *ALZO* ( $\beta = .56, p < .05$ ) den Effekt des Unterrichtsfeedbacks. Der Moderationseffekt der *VLZO* erreicht weiterhin nicht das Niveau statistischer Signifikanz ( $\beta = -.28, p > .10$ ). Modell 6 ist am restriktivsten und enthält zusätzlich die Lehrkräfte Merkmale abgeschlossenes Lehramtsstudium und Berufserfahrung. Im Modell 6 konnte weder für die *LZO* ( $\beta = .26, p > .10$ ) noch die *ALZO* ( $\beta = .24, p > .10$ ) oder die *VLZO* ( $\beta = -.03, p > .10$ ) ein Moderationseffekt gefunden werden.

Werden die Effekte des Lehramtsstudiums und der Berufserfahrung nicht kontrolliert, interagiert sowohl die *LZO* als auch die *ALZO* signifikant positiv mit dem Feedbackeffekt. Um zu überprüfen, bei welchen Ausprägungen der *LZO* und *ALZO* der Feedbackeffekt das Niveau statistischer Signifikanz erreicht, wurden Simple-Slopes-Tests durchgeführt. Für die *LZO* zeigte sich bei einer um +1 *SD* über dem Mittelwert liegenden *LZO* ein signifikant positiver Feedbackeffekt ( $\beta = .68, p < .05$ ). Auch für die *ALZO* konnte bei einer um +1 *SD* über dem Mittelwert liegenden *ALZO* ein signifikant positiver Feedbackeffekt gefunden werden ( $\beta = .81, p < .05$ ). Die differenziellen Effekte der *LZO* und *ALZO* werden in Abb. 2 veranschaulicht.<sup>16</sup>

---

<sup>16</sup> Für die grafische Veranschaulichung wurde ein von Dawson (2014) entwickeltes Online-Tool genutzt (<http://www.jeremydawson.co.uk/slopes.htm>).



**Abb. 2:** Differenzielle Effekte der Lern- (LZO) und Annäherungsleistungszielorientierung (ALZO) auf die selbst eingeschätzte Weiterentwicklung der Unterrichtskompetenz

### 3.6 Diskussion

Im vorliegenden Beitrag wurde eine Feedbackintervention vorgestellt, die Schulleitungen dabei unterstützen soll, Lehrkräften ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben, auf dessen Grundlage diese ihre Unterrichtskompetenz weiterentwickeln können. Die Qualität des Unterrichtsfeedbacks wurde von den Lehrkräften als hoch eingeschätzt. Trotz der hohen Qualitätseinschätzungen hatte das Feedback jedoch keinen Haupteffekt auf die Kompetenzselbsteinschätzungen der Lehrkräfte. Dieser Befund steht im Widerspruch zu den Ergebnissen einer weiteren Studie der Autorengruppe, in der der Effekt der Feedbackintervention auf die von Schülerinnen und Schülern wahrgenommene Unterrichtsqualität untersucht wurde (vgl. Kellermann et al. eingereicht): Schülerinnen und Schüler, die von Lehrkräften der IG unterrichtet wurden, schätzten die Unterrichtsqualität nach dem Unterrichtsfeedback bezogen auf 3 der 14 Unterrichtsmerkmale des theoretischen Unterrichtsmodells signifikant besser ein als Schülerinnen und Schüler von Lehrkräften der KG. Die widersprüchlichen Befunde der beiden Studien sollen später erneut aufgegriffen werden.

Auch wenn das Unterrichtsfeedback keinen Haupteffekt auf die selbst eingeschätzte Kompetenzentwicklung der Lehrkräfte hatte, zeigten sich für die LZO und die ALZO differenzielle Effekte, für die VLZO nicht. Lehrkräfte mit hoch ausgeprägter LZO und hoch ausgeprägter ALZO, die ein Unterrichtsfeedback erhielten, schätzten ihre Kompetenzentwicklung positiver ein als lern- oder annäherungsleistungszielorientierte Lehrkräfte, die kein Feedback bekamen. Der positive Moderationseffekt der LZO und der ausbleibende Moderationseffekt der VLZO sind hypothesenkonform. Im Hinblick auf die ALZO wurde aufgrund der inkonsistenten Befunde zu den lernrelevanten Einstellungen und Verhal-

tensweisen von annäherungsleistungszielorientierten Lehrkräften keine Hypothese formuliert. Eine Erklärung für den positiven Moderationseffekt bietet die Operationalisierung der *ALZO*. Die inkonsistenten Befunde, die auch aus anderen Bereichen bekannt sind, können häufig auf die Konzeptualisierung der *ALZO* zurückgeführt werden; teilweise liegt der Fokus auf dem Streben, Kompetenz zu demonstrieren, teilweise auf dem Streben, Überlegenheit zu demonstrieren (vgl. Senko et al. 2011). Letztgenannte Konzeptualisierung steht häufig in einem positiven Zusammenhang mit Lernen und Leistung (vgl. Hulleman et al. 2010; Senko und Dawson 2017) und liegt der Operationalisierung im vorliegenden Beitrag zugrunde. Es wird vermutet, dass das Ziel, andere zu übertreffen, mit einem hohen Leistungsstandard und einer hohen Leistungsmotivation einhergeht (vgl. Senko et al. 2011). Insofern erscheint es plausibel, dass Lehrkräfte mit hoch ausgeprägter *ALZO* ihre Unterrichtskompetenz auf Grundlage des Unterrichtsfeedbacks weiterentwickeln, um auch in Zukunft ihre Überlegenheit demonstrieren zu können. Vor dem Hintergrund der Zusammenhänge von Zielorientierungen mit den lernrelevanten Einstellungen und Verhaltensweisen von Lehrkräften ist vorstellbar, dass annäherungsleistungs- ebenso wie lernzielorientierte Lehrkräfte das Unterrichtsfeedback mit weiteren Lernaktivitäten verknüpfen haben; so könnten sie z. B. gezielt fachbezogene Literatur gelesen oder Fort- und Weiterbildungen besucht haben (vgl. Nitsche et al. 2013a, 2013b), um neue Verhaltensstrategien für ihren Unterricht kennenzulernen.

Einschränkend ist zu erwähnen, dass die Moderationseffekte der *LZO* und der *ALZO* nur dann das Niveau statistischer Signifikanz erreichen, wenn die Variablen abgeschlossenes Lehramtsstudium und Berufserfahrung nicht kontrolliert werden. Dies kann vorrangig auf die hohe geteilte Varianz der beiden Variablen mit der Kompetenzselbsteinschätzung zurückgeführt werden. Es ist plausibel, dass Lehrkräfte, die über wenig Berufserfahrung verfügen und kein Lehramtsstudium abgeschlossen haben, ihre Kompetenzentwicklung positiver einschätzen als Lehrkräfte mit viel Berufserfahrung sowie Lehrkräfte mit einem abgeschlossenen Lehramtsstudium. Offen bleibt zudem, worauf die Feedbackeffekte bei den Lehrkräften mit hoch ausgeprägter *LZO* und *ALZO* zurückzuführen sind. Welchen Anteil das standardisierte Beobachtungsinstrument, das Feedbackverfahren und die lehrerspezifischen Feedbackberichte an den Feedbackeffekten haben, lässt sich anhand der vorliegenden Daten nicht beurteilen. Auch über die Wirksamkeit des Beobachtungs- und Feedbacktrainings liegen nur wenige Informationen vor. Die Kurzevaluation deutet zwar darauf hin, dass die Schulleitungen ihr Wissen und ihre Kompetenz über Unterrichtsbeobachtungen und -feedbacks erweitert haben, inwiefern die

Feedbackeffekte bei lern- und annäherungsleistungszielorientierten Lehrkräften auf das Training der Schulleitungen zurückzuführen sind, lässt sich jedoch nicht beurteilen.

Darüber hinaus weist die vorliegende Studie weitere Limitationen auf. Eine Limitation betrifft die nicht randomisierte Zuordnung der Lehrkräfte zur IG und KG, die eine kausale Interpretation erschwert. Durch die Berücksichtigung diverser Kontrollvariablen konnte der Feedbackeffekt nach unserer Einschätzung jedoch weitgehend isoliert werden. Eine weitere Limitation betrifft die Operationalisierung der Weiterentwicklung der Unterrichtskompetenz. Durch eine retrospektive Selbsteinschätzung der Kompetenzentwicklung kann es zwar nicht zu einem *Response Shift* kommen (vgl. Piwowar und Thiel 2014; Schwartz und Spranger 1999), im Vergleich zu einem Wissens- oder Performanztest kann eine retrospektive Kompetenzselbsteinschätzung jedoch mit Validitätsproblemen wie Verzerrungen einhergehen. Beispielsweise ist vorstellbar, dass Lehrkräfte der IG ihre Kompetenzentwicklung rückblickend allein schon deshalb positiver einschätzten, weil sie ein Unterrichtsfeedback erhalten haben, was in besonderer Weise auf Lehrkräfte mit hoch ausgeprägter *LZO* und *ALZO* zutreffen könnte. Inwiefern retrospektive Selbsteinschätzungen der Kompetenzentwicklung tatsächliche Veränderungen im professionellen Wissen und Können von Lehrkräften abbilden, wurde nach unserem Kenntnisstand noch nicht erforscht. Es liegen aber Studien vor, die nahelegen, dass Selbsteinschätzungen von Lehrkräften im Allgemeinen und Kompetenzselbsteinschätzungen im Besonderen positive Zusammenhänge mit extern erhobenen Kriterien aufweisen. So zeigt die Unterrichtsforschung, dass die von Lehrkräften selbsteingeschätzte Unterrichtsqualität überwiegend moderate Zusammenhänge mit den Qualitätseinschätzungen von Schülerinnen und Schülern aufweist (vgl. Fauth et al. 2014; Kunter und Baumert 2006; Kunter et al. 2008; Wagner et al. 2016). Darüber hinaus liegen Befunde vor, die zeigen, dass die Kompetenzselbsteinschätzung von angehenden Lehrkräften zwar schwach mit ihrem Abschneiden in Wissens- bzw. Performanztests zusammenhängt (vgl. König et al. 2012; Seifert et al. 2018), dass dieser Zusammenhang aber größer ausfällt, nachdem Lehramtsstudierende eigene Unterrichtserfahrungen sammeln konnten (vgl. Mertens und Gräsel 2018). Auch wenn es vereinzelt zu Verzerrungen bei der selbsteingeschätzten Kompetenzentwicklung kommen kann, gehen wir vor dem Hintergrund der zuvor zitierten Befunde davon aus, dass die in der vorliegenden Studie befragten Lehrkräfte, die überwiegend über eine langjährige Unterrichtserfahrung verfügen, die Weiterentwicklung ihrer Unterrichtskompetenz auch rückblickend differenziert beurteilen können.

Eine weitere Limitation betrifft die Passung der Feedbackintervention und der selbst eingeschätzten Kompetenzentwicklung. Während sich das Unterrichtsfeedback auf einzelne Unterrichtsmerkmale des theoretischen Unterrichtsmodells bezog, wurde die selbst eingeschätzte Kompetenzentwicklung global erfasst. Dass Schülerinnen und Schüler die Unterrichtsqualität von Lehrkräften der IG im Hinblick auf einzelne Merkmale des theoretischen Unterrichtsmodells signifikant besser einschätzten als Schülerinnen und Schüler von Lehrkräften der KG (vgl. Kellermann et al. eingereicht), kann als Hinweis darauf interpretiert werden, dass eine nach den 14 Unterrichtsmerkmalen differenzierte Erfassung der Kompetenzentwicklung die Wahrscheinlichkeit, einen Haupteffekt zu finden, erhöhen könnte. Denkbar ist auch, dass erst ein mehrmaliges und ggf. regelmäßiges Unterrichtsfeedback der Schulleitung, das in Übereinstimmung mit der entwickelten Feedbackintervention gegeben wird, zu einer Weiterentwicklung der Unterrichtskompetenz von Lehrkräften führt. Der ausbleibende Haupteffekt könnte darüber hinaus auf die kleine Stichprobe von  $N = 81$  Lehrkräften und eine damit einhergehende geringe Teststärke zurückzuführen sein.

Zusammenfassend kann festgehalten werden, dass die Feedbackintervention Schulleitungen dabei unterstützt, ein qualitativ hochwertiges Unterrichtsfeedback zu gestalten. Dass das Unterrichtsfeedback nur bei Lehrkräften mit hoch ausgeprägter *LZO* und *ALZO* mit einer Weiterentwicklung ihrer selbst eingeschätzten Unterrichtskompetenz einhergeht, zeigt aber auch, dass individuelle Merkmale von Lehrkräften für die Wirksamkeit von Unterrichtsfeedbacks von Relevanz sind. Dies sollte bei der Entwicklung von Feedbackinterventionen in Zukunft berücksichtigt werden. Insbesondere gilt es zu klären, wie ein Unterrichtsfeedback zu gestalten ist, von dem auch Lehrkräfte mit niedrig ausgeprägter *LZO* und *ALZO* profitieren. Hierbei könnte eine Verzahnung des Unterrichtsfeedbacks mit weiteren Formaten der professionellen Weiterentwicklung eine entscheidende Rolle spielen.

### 3.7 Literatur

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271. <https://doi.org/10.1037/0022-0663.84.3.261>
- Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88(1), 87–99. <https://doi.org/10.1037/0021-9010.88.1.87>

- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, & W. Blum (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Butler, R. (2007). Teachers' achievement goal orientations and associations with teachers' help seeking. Examination of a novel approach to teacher motivation. *Journal of Educational Psychology*, 99(2), 241–252. <https://doi.org/10.1037/0022-0663.99.2.241>
- Chughtai, A. A., & Buckley, F. (2010). Assessing the effects of organizational identification on inrole job performance and learning behaviour. *Personnel Review*, 39(2), 242–258. <https://doi.org/10.1108/00483481011017444>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsday, NJ: Lawrence Erlbaum.
- Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, 29(1), 1–19. <https://doi.org/10.1007/s10869-013-9308-7>
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson, & C. S. Weinstein (Hrsg.), *Handbook of classroom management* (S. 97–126). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048. <https://doi.org/10.1037//0003-066X.41.10.1040>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, 53, 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70(3), 461–475. <https://doi.org/10.1037/0022-3514.70.3.461>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>

- Gärtner, H., Thiel, F., Nachbauer, M., & Kellermann, C. (2021). Unterrichtsentwicklung durch Unterrichtsfeedback – Erste Entwicklungsschritte eines Beobachtungsboogens für Schulleitungen. *Zeitschrift für Bildungsforschung*, 11(3), 529–547. <https://doi.org/10.1007/s35834-021-00321-7>
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, 94(3), 638–645. <https://doi.org/10.1037//0022-0663.94.3.638>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: The Guilford Press.
- Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität – Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Friedrich Verlag GmbH.
- Hulleman, C. S., Schragar, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, 136(3), 422–449. <https://doi.org/10.1037/a0018947>
- Ilgén, D. R., & Davis, C. A. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology: An International Review*, 49(3), 550–565. <https://doi.org/10.1111/1464-0597.00031>
- Janke, S., Bardach, L., Oczlon, S., & Lüftenegger, M. (2019). Enhancing feasibility when measuring teachers' motivation: A brief scale for teachers' achievement goal orientations. *Teaching and Teacher Education*, 83, 1–11. <https://doi.org/10.1016/j.tate.2019.04.003>
- Kellermann, C., Nachbauer, M., Gärtner, H., & Thiel, F. (eingereicht). The effect of school leaders' feedback on teaching quality perceived by students – Results of an intervention study.
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34–70. <https://doi.org/10.1177/0013161X08327549>
- Klauer, K. J., & Leutner, D. (2012). *Lehren und Lernen. Einführung in die Instruktionspsychologie* (2., überarbeitete Auflage). Weinheim: Beltz.

- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: “Aufgabenkultur” und Unterrichtsgestaltung. In BMBF (Hrsg.), *TIMSS – Impulse für Schule und Unterricht, Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–58). Bonn: BMBF.
- 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>
- König, J., Kaiser, G., & Felbrich, A. (2012). Spiegelt sich pädagogisches Wissen in den Kompetenzselbsteinschätzungen angehender Lehrkräfte? Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerausbildung. *Zeitschrift für Pädagogik*, *58*(4), 476–491.
- Kopp, B., & Mandl, H. (2014). Aspekte der Feedbacknachricht. In H. Ditton, & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 151–162). Münster: Waxmann.
- Kraft, M. A., & Gilmour, A. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, *52*(5), 711–753.  
<https://doi.org/10.1177/0013161X16653445>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251. <https://doi.org/10.1007/s10984-006-9015-7>
- Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, & W. Blum (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55–69). Münster: Waxmann.
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, *18*(5), 468–482.  
<https://doi.org/10.1016/j.learninstruc.2008.06.008>

- Kwakman, K. (2003). Factors affecting teachers' participation in professional learning activities. *Teaching and Teacher Education*, 19(2), 149–170.  
[https://doi.org/10.1016/S0742-051X\(02\)00101-4](https://doi.org/10.1016/S0742-051X(02)00101-4)
- 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* (Zeitschrift für Pädagogik Beiheft, Bd. 51, S. 47–70). Weinheim: Beltz.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28(2), 295–314. <https://doi.org/10.1007/s10648-014-9287-x>
- Mertens, S., & Gräsel, C. (2018). Entwicklungsbereiche bildungswissenschaftlicher Kompetenzen von Lehramtsstudierenden im Praxissemester. *Zeitschrift für Erziehungswissenschaft*, 21(6), 1109–1133. <https://doi.org/10.1007/s11618-018-0825-z>
- Milanowski, A. T., & Heneman, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193–212.  
<https://doi.org/10.1023/A:1012752725765>
- Müller, A., & Ditton, H. (2014). Feedback: Begriffe, Formen und Funktionen. In H. Ditton, & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 11–28). Münster: Waxmann.
- Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256.  
<https://doi.org/10.1080/09243453.2014.885451>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus. Statistical analysis with latent variables. User's guide*. Los Angeles, CA: Muthén & Muthén.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328–346.  
<https://doi.org/10.1037/0033-295X.91.3.328>

- Nitsche, S., Dickhäuser, O., Dresel, M., & Fasching, M. S. (2013a). Zielorientierungen von Lehrkräften als Prädiktoren lernrelevanten Verhaltens. *Zeitschrift für Pädagogische Psychologie*, 27(1-2), 95–103. <https://doi.org/10.1024/1010-0652/a000092>
- Nitsche, S., Dickhäuser, O., Fasching, M. S., & Dresel, M. (2011). Rethinking teachers' goal orientations. Conceptual and methodological enhancements. *Learning and Instruction*, 21(4), 574–586. <https://doi.org/10.1016/j.learninstruc.2010.12.001>
- Nitsche, S., Dickhäuser, O., Fasching, M. S., & Dresel, M. (2013b). Teachers' professional goal orientations: Importance for further training and sick leave. *Learning and Individual Differences*, 23, 272–278. <https://doi.org/10.1016/j.lindif.2012.07.017>
- Ophardt, D., & Thiel, F. (2013). *Klassenmanagement. Ein Handbuch für Studium und Praxis*. Stuttgart: Kohlhammer.
- Piwovar, V., & Thiel, F. (2014). Evaluating response shift in training evaluation: comparing the retrospective pretest with an adapted measurement invariance approach in a classroom management training program. *Evaluation Review*, 38(5), 420–448. <https://doi.org/10.1177/0193841X14546932>
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Rheinberg, F. (2008). *Motivation*. Stuttgart: Kohlhammer.
- Runhaar, P., Sanders, K., & Yang, H. (2010). Stimulating teachers' reflection and feedback asking. An interplay of self-efficacy, learning goal orientation, and transformational leadership. *Teaching and Teacher Education*, 26(5), 1154–1161. <https://doi.org/10.1016/j.tate.2010.02.011>
- Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 396–407. <https://doi.org/10.1177/088840640402700407>
- Schwartz, C. E., & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*, 48, 1531–1548. [https://doi.org/10.1016/S0277-9536\(99\)00047-7](https://doi.org/10.1016/S0277-9536(99)00047-7)

- Seifert, A., Schaper, N., & König, J. (2018). Bildungswissenschaftliches Wissen und Kompetenzeinschätzungen von Studierenden im Praxissemester: Veränderungen und Zusammenhänge. In J. König, M. Rothland, & N. Schaper (Hrsg.), *Learning to Practice, Learning to Reflect? Ergebnisse aus der Längsschnittstudie LtP zur Nutzung und Wirkung des Praxissemesters in der Lehrerbildung* (S. 325–347). Wiesbaden: Springer VS.
- Senko, C., & Dawson, B. (2017). Performance-approach goal effects depend on how they are defined: Meta-analytic evidence from multiple educational outcomes. *Journal of Educational Psychology, 109*(4), 574–598.  
<https://doi.org/10.1037/edu0000160>
- Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist, 46*(1), 26–47. <https://doi.org/10.1080/00461520.2011.538646>
- Stelman, 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>
- Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy, 10*(4), 535–572.  
[https://doi.org/10.1162/EDFP\\_a\\_00173](https://doi.org/10.1162/EDFP_a_00173)
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628–3651.  
<https://doi.org/10.1257/aer.102.7.3628>
- 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>
- Urban, D., & Mayerl, J. (2018). *Angewandte Regressionsanalyse: Theorie, Technik und Praxis*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-01915-0>
- 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>

# 4

## Teilstudie 2:

### The effect of school leaders' feedback on teaching quality perceived by students – results of an intervention study

Dies ist eine Version des Autors des Beitrags, der bei  
*School Effectiveness and School Improvement* veröffentlicht wurde als:

Kellermann, C., Nachbauer, M., Gaertner, H. & Thiel, F. (2023). The effect of school leaders' feedback on teaching quality perceived by students – results of an intervention study. *School Effectiveness and School Improvement*, 34(1), 24–42.

<https://doi.org/10.1080/09243453.2022.2075018>

#### **4 Teilstudie 2: The effect of school leaders' feedback on teaching quality perceived by students – results of an intervention study**

**Abstract:** Feedback from school leaders often is of low quality and not always effective. Because school leaders spend only limited time on instructional leadership activities, an intervention was developed to assist them in efficiently providing high-quality feedback to teachers. The effectiveness of the intervention was evaluated within a quasi-experimental study.  $N = 11$  school leaders were trained in conducting classroom observations and providing feedback. School leaders observed teachers of the intervention group ( $N = 26$ ) and provided feedback with regard to 14 teaching characteristics.  $N = 27$  teachers served as a control group. It is examined whether the intervention leads to improvements in teaching quality perceived by students. Multilevel regression analyses were conducted. For three of the 14 teaching characteristics, intervention group students ( $N = 518$ ) rated changes in teaching quality significantly better than control group students ( $N = 551$ ). The results provide first evidence for the effectiveness of the intervention.

**Keywords:** Teacher evaluation, feedback, school improvement, instructional leadership, intervention study

# 5

## Teilstudie 3:

Improvements in teaching quality after receiving feedback: What role do teachers' goal orientations play for the effectiveness of feedback?

Dies ist eine Version des Autors des Beitrags,  
der zur Veröffentlichung eingereicht wurde als:

Kellermann, C., Nachbauer, M., Gaertner, H. & Thiel, F. (eingereicht). Improvements in teaching quality after receiving feedback: What role do teachers' goal orientations play for the effectiveness of feedback?

## **5 Teilstudie 3: Improvements in teaching quality after receiving feedback: What role do teachers' goal orientations play for the effectiveness of feedback?**

**Abstract:** Feedback on teaching constitutes a learning opportunity for teachers. However, it is not always effective. While some teachers improve their teaching after receiving feedback, others do not. Research shows that teachers' attitudes towards and use of various learning opportunities other than feedback vary as a function of the goals they pursue. Therefore, it can be assumed that the effectiveness of feedback also depends on teachers' goal orientations. Based on survey data from  $N = 53$  teachers and  $N = 1.069$  students, it is examined whether teachers' goal orientations moderate the effectiveness of feedback on changes in teaching quality as perceived by their students. Results indicate that a pronounced learning goal orientation increases the effectiveness of feedback. While a performance avoidance goal orientation does not moderate the effectiveness of feedback, results are inconclusive with respect to a performance approach goal orientation. Feedback valence is discussed as a potential explanation.

**Keywords:** Feedback, Goal orientation, Teachers' professional learning, Teaching quality

## 5.1 Introduction

According to Butler (2007, p. 242), schools not only are an “achievement arena” for students, but also for teachers. Teachers too strive to succeed in their job. However, depending on the goals they pursue, they may differ in the way they define success. In general, there are two types of goals individuals pursue in achievement-related settings, learning goals and performance goals. These goals are assumed to be associated with different patterns of affect, cognition, and behavior. For example, teachers’ attitudes towards and their use of various learning opportunities vary as a function of the goals they pursue (e.g., Butler, 2007; Nitsche et al., 2013a, 2013b; Runhaar et al., 2010).

Feedback on teaching can be considered a learning opportunity for teachers, too. Research, however, indicates that feedback to teachers is not always effective (Kraft et al., 2018; Song et al., 2021; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). In view of the associations between the goals teachers pursue and their professional learning, it is reasonable to assume that the effectiveness of feedback also varies with teachers’ goal orientations. By linking goal orientation theory with Kluger and DeNisi’s (1996) feedback intervention theory (FIT), we elaborate on the role teachers’ goal orientations play for the effectiveness of feedback. Using data from a quasi-experimental study, we then empirically test whether the effectiveness of feedback is moderated by teachers’ goal orientations.

The next section provides a brief introduction to goal orientation theory. Afterwards, feedback processing according to FIT is described. Goal orientation theory and FIT then are combined in an attempt to predict how teachers’ goal orientations may influence the processing of feedback and in turn its effectiveness. The intervention study is described in the methods section. Thereafter, results are presented and discussed.

### 5.1.1 Goal orientation theory

Goal orientation theory assumes that the goals individuals pursue determine their affect, cognition, and behavior in achievement-related settings. In general, a distinction is made between learning goals and performance goals, which form the basis of a learning goal orientation (LGO) and a performance goal orientation (PGO), respectively. While individuals with a LGO strive to *develop* their competence, individuals with a PGO strive to *demonstrate* superior and *hide* inferior competence (Ames, 1992; Dweck, 1986; Dweck & Leggett, 1988; Nicholls, 1984).

One key distinction between learning and performance goals relates to the evaluation of competence: While competence in case of a learning goal is determined in relation to a task-based or an individual standard, in case of a performance goal it is determined in relation to a social standard (Elliot, 1999). That is, competence is assessed relative to the competence of others and therefore is accompanied by an increased attention to the self. It is for this reason that Nicholls (1984) refers to LGO as *task involvement* and to PGO as *ego involvement*.

Early goal orientation research shows that learning goals are associated with adaptive response patterns, whereas performance goals predominantly correspond with maladaptive response patterns: Learning-oriented individuals experience positive affect in achievement-related settings (e.g., joy), they prefer challenging tasks, see problems as cues to increase their effort and/or to develop new problem-solving strategies. Performance-oriented individuals, on the other hand, often experience negative affect (e.g., anxiety), tend to avoid challenge, often exhibit low persistence in the face of obstacles and interpret failure as an indication of low ability (for an overview see Ames, 1992; Dweck, 1986; Dweck & Leggett, 1988; Nicholls, 1984). Therefore, it is assumed that LGO is beneficial for achievement and PGO in most cases is not.

To account for findings that PGO does not always correspond with maladaptive response patterns, Elliot and associates (Elliot, 1999; Elliot & Church, 1997; Elliot & Harackiewicz, 1996) further refined the 2-factor model by distinguishing between a performance approach goal orientation (PAPGO) and a performance avoidance goal orientation (PAVGO). While individuals with a PAPGO strive to demonstrate superior competence, individuals with a PAVGO strive to hide inferior competence. With regard to the 3-factor model, it is assumed that not only LGO but also PAPGO can be beneficial for achievement (Harackiewicz et al., 2002; Senko et al., 2011). However, there is some debate about the beneficial effects of PAPGO (Kaplan & Maehr, 2007; Midgley et al., 2001). In addition, recent meta-analyses do not provide a clear picture of whether PAPGO is associated with adaptive response patterns and benefits for achievement (Cellar et al., 2011; Huang, 2011, 2012; Payne et al., 2007; Hulleman et al., 2010; Wirthwein et al., 2013).

### **5.1.2 Feedback processing according to feedback intervention theory**

Feedback can be defined as “actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance” (Kluger & DeNisi, 1996, p.

255). Accordingly, feedback on teaching refers to an external agent, for example a school leader or coach, providing information regarding some aspect(s) of a teacher's performance. Classroom observations are considered being especially useful for gathering information on teachers' performances (Cohen & Goldhaber, 2016; Goldring et al., 2015). Therefore, feedback on teaching often is based on classroom observations. For classroom observations, a theoretical model of teaching quality is needed. It describes teaching behaviors that have proven to be effective for student learning and defines the standards for the performance evaluation. At the same time, the model provides a framework for the feedback (Heneman & Milanowski, 2003), which contains a comparison between a teacher's observed performance and the standards defined.

Although observation-based feedback is seen as a promising measure for improving the quality of teaching and promoting student achievement (Hill & Grossmann, 2013; Papay, 2012; Kraft & Gilmour, 2016), research indicates it is not effective per se: After receiving feedback, teachers often, but not always, improve their teaching and students of these teachers often, but not always, show better achievement (Kraft et al., 2018; Song et al., 2021; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). These findings are in line with results on the effectiveness of feedback in general. For example, Kluger and DeNisi (1996) in a meta-analysis show that although feedback interventions have a positive effect on performance on average, feedback often has no effect and, in some cases, even decreases performance.

To explain these inconsistent findings, Kluger and DeNisi (1996) developed FIT. According to FIT, feedback processing, simplified, may focus either on the task or the person executing the task. If feedback processing is task-oriented, a discrepancy between a teacher's observed performance and the standards defined in the model of teaching quality can result in an increased effort: "Working harder is the output of the motivational process, and it is accomplished by activating programs or scripts for action that are available from past experience" (Kluger & DeNisi, 1996, p. 263). Hence, after receiving negative feedback, teachers should strive to reduce the discrepancy between their performance and the standards defined. They may do so by applying teaching strategies that have proven to be successful in the past.

However, not only negative feedback can result in an increased effort. In case of positive feedback, discrepancy *reduction* may be preceded by discrepancy *production* (Bandura, 1991). That is, by setting higher goals after receiving positive feedback, teachers may produce a discrepancy between their actual performance and an altered standard

of good teaching, which they subsequently strive to reduce by increasing their effort. If additional effort alone is deemed insufficient, feedback recipients in case of both positive and negative feedback may develop new problem-solving strategies for discrepancy reduction: “If working harder fails, people may try to work smarter by generating a hypothesis regarding means for improved performance” (Kluger & DeNisi, 1996, p. 263). Thus, teachers should develop new teaching strategies that can—given they prove to be successful in practice—reduce a discrepancy between their actual and a desired performance.

Feedback processing, however, can also be self-oriented. Self-oriented feedback processing is to be expected, amongst others, in the case of extremely negative feedback (Ilgen & Davis, 2000). According to FIT, self-oriented feedback processing often is associated with negative affect, leads to a depletion of cognitive resources for task performance, and may result in a rejection of feedback (Kluger & DeNisi, 1996). For instance, after receiving negative feedback, teachers might wonder whether they are unsuited for the teaching profession and—to protect their self—reject the feedback. However, if feedback is related to an important self-goal, self-oriented feedback processing may as well lead to an increased effort expenditure and/or the development of new problem-solving strategies (Kluger & DeNisi, 1996). In the latter case, teachers might utilize the information on a performance-standard discrepancy to align with their self-goal of being a good teacher. Even then, however, task performance could be impaired by initial negative emotions and a depletion of cognitive resources.

From FIT, it can be concluded that task-oriented feedback processing can result in improvements in teaching quality. Self-oriented feedback processing, on the other hand, provides an explanation for feedback sometimes being ineffective. Whether feedback processing is task-oriented or self-oriented not only depends on feedback characteristics, for example its valence, but also on personal characteristics of the feedback recipient (Kluger & DeNisi, 1996). Amongst others, goal orientations are assumed to influence the processing of feedback (Ilgen & Davis, 2000; VandeWalle et al., 2001). By linking goal orientation theory with FIT, in the next section, we elaborate on how teachers’ goal orientations may influence the processing of feedback and its effectiveness.

### **5.1.3 Teachers’ goal orientations and feedback processing**

Because teachers with a pronounced LGO strive to develop their teaching competence (Butler, 2007; Nitsche et al., 2011; Retelsdorf et al., 2010) and exhibit high levels of *task*

*involvement* in achievement-related settings, they should perceive information on a performance-standard comparison primarily as cues to improve their teaching and focus on the task when processing feedback. Therefore, it can be assumed that learning-oriented teachers use feedback to reduce a discrepancy between their actual and a desired performance. They may do so by increasing their effort and, if necessary, by developing new teaching strategies potentially resulting in an improved teaching quality.

To our knowledge, the influence of teachers' goal orientations on the effectiveness of feedback has not been investigated to date. However, there are studies showing that teachers with a pronounced LGO have positive attitudes towards and tend to use various learning opportunities other than feedback. For example, learning-oriented teachers perceive external help as beneficial and frequently reach out for it (Butler, 2007; Janke et al., 2019; Nitsche et al., 2011). In addition, they often seek feedback (Chughtai & Buckley, 2010; Runhaar et al., 2010), frequently read professional literature, and have positive attitudes towards further training, which they also attend regularly (Nitsche et al., 2013a, 2013b). Participation is intrinsically motivated and learning-oriented teachers are willing to implement the training content in the future (Gorozidis & Papaioannou, 2016). Furthermore, they tend to reflect their teaching (Runhaar et al., 2010) and often experiment with new teaching strategies (van Daal et al., 2014).

Teachers with a pronounced PAPGO, on the other hand, strive to demonstrate superior teaching competence (Butler, 2007; Nitsche et al., 2011; Retelsdorf et al., 2010). Since PAPGO is associated with an increased attention to the self (*ego involvement*), it can be assumed that approach-oriented teachers perceive information on a performance-standard comparison primarily as an evaluation of their teaching competence, which in case of negative feedback is likely to result in feedback rejection. However, if high-PAPGO teachers consider the feedback being useful for demonstrating superior competence in the future, they might as well respond with an increased effort and/or the development of new teaching strategies. That is, they utilize the information on a performance-standard discrepancy to align with their self-goal of being a better teacher than their colleagues are. While feedback in the case of rejection is ineffective, it could result in an improved teaching quality in the latter case.

Results on performance approach-oriented teachers' attitudes towards and their use of other learning opportunities than feedback are inconclusive. There are no associations between teachers' PAPGO and their perception of help seeking (Butler, 2007; Janke et

al., 2019; Nitsche et al., 2011). However, performance approach-oriented teachers frequently read professional literature (Nitsche et al., 2013a). Although there are no associations with their attitudes towards further training, high-PAPGO teachers nonetheless attend training workshops regularly (Nitsche et al., 2013b). Participation, however, often is extrinsically motivated and performance approach-oriented teachers are not willing to implement the training content in the future (Gorozidis & Papaioannou, 2016). Furthermore, there are no associations between PAPGO and the extent to which teachers experiment with new teaching methods (van Daal et al., 2014).

In contrast to teachers with a pronounced PAPGO, teachers with a pronounced PAVGO do not strive to demonstrate superior teaching competence, but to avoid the demonstration of inferior competence (Butler, 2007; Nitsche et al., 2011; Retelsdorf et al., 2010). As with PAPGO, PAVGO is associated with an increased attention to the self (*ego involvement*). Therefore, performance avoidance-oriented teachers also should perceive information on a performance-standard comparison as an evaluation of their competence, probably resulting in feedback rejection. Hence, high-PAVGO teachers are unlikely to improve their teaching after receiving feedback.

In general, this assumption is supported by findings on high-PAVGO teachers' attitudes towards and their use of other learning opportunities. Performance avoidance-oriented teachers have negative attitudes towards help seeking (Butler, 2007; Janke et al., 2019; Nitsche et al., 2011) and further training, which they rarely participate in (Nitsche et al., 2013b). If they attend training workshops, they do so because they are extrinsically motivated, which is not related to the willingness to implement the training content in the future (Gorozidis & Papaioannou, 2016).

#### **5.1.4 Research question and hypotheses**

Section 1.3 elaborated on how teachers' goal orientations might influence feedback processing and its effectiveness. In the remainder of the article, the question of whether teachers' goal orientations moderate the effectiveness of feedback is addressed empirically. For this purpose, data from a quasi-experimental study is used. School leaders conducted classroom observations and provided feedback to one group of teachers, while another group of teachers served as a control group. Effects of the intervention were tested based on students' ratings of teaching quality. In another study by the authors, the main

effects of the intervention are reported (Kellermann et al., 2022b). The present study examines the interactions of teachers' LGO, PAPGO, and PAVGO and school leaders' feedback in predicting changes in teaching quality.

The following hypotheses are tested: In view of learning-oriented teachers' *task involvement* in achievement-related settings as well as findings on their attitudes towards and use of various learning opportunities other than feedback, it is expected that a LGO moderates the effectiveness of feedback. Specifically, we test the hypotheses that high-LGO teachers show greater improvements in teaching quality after receiving feedback than low-LGO teachers (*Hypothesis 1*). A clear prediction of whether teachers' PAPGO moderates the effectiveness of feedback cannot be established. The theorized role of PAPGO in feedback processing as well as performance approach-oriented teachers' competence-related attitudes and activities are inconclusive. Therefore, we test no specific hypothesis in relation to teachers' PAPGO (*Hypothesis 2*). Performance avoidance-oriented teachers' *ego involvement* suggests that teachers' PAVGO does not moderate the effectiveness of feedback. This assumption is supported by findings on performance avoidance-oriented teachers' attitudes towards and their use of other learning opportunities. Therefore, no interaction is expected (*Hypothesis 3*).

## 5.2 Method

### 5.2.1 Study design

The intervention study was conducted in the school year 2017/2018. Teachers of the intervention group (IG) received feedback from their school leaders in the middle of the school year, teachers of the control group (CG) did not receive feedback. Teachers as well as students of one of each teachers' classes were surveyed using paper-pencil questionnaires in the beginning ( $t_1$ ) and end ( $t_2$ ) of the school year.

### 5.2.2 Sample

The sample includes  $N = 53$  teachers as well as  $N = 1.069$  students from six vocational schools in Berlin, Germany.<sup>17</sup> IG consists of  $N = 26$  teachers and classes, respectively, with  $N = 518$  students. CG consists of  $N = 27$  teachers/classes and  $N = 551$  students.

---

<sup>17</sup> After completing general education, students in Germany may attend vocational schools where they are prepared for a specific occupation. Vocational schools encompass different educational programs some of which are combined with on-the-job training.

55.8% of teachers are female. 30.8% were 30 to 39 years old, 21.2% were 40 to 49 years old, and 44.2% were 50 to 59 years old. In addition, 3.8% were 60 years of age or older. With regard to students, 39.6% were in 11th grade, 52.1% were in 12th grade, and 8.3% were in 13th grade. That is, all students were at least 16 years of age.

A random assignment of teachers and classes to IG and CG could not be ensured. The decision as to which teachers and classes would participate was made within schools. It can be assumed that in some schools, principals decided on participation and assigned teachers and classes to IG and CG, while in other schools, teachers decided. However, there are IG and CG teachers/classes from every school, distributed evenly between the two groups. We return to non-randomization and how we dealt with it in section 2.4.3.

### 5.2.3 Intervention

School leaders conducted two classroom observations and completed a standardized observation instrument. The instrument is based on a theoretical model of teaching quality that differentiates between 14 teaching characteristics that relate to three teaching requirements: supporting students' knowledge acquisition, motivating students, and classroom management (Gaertner et al., 2021). In relation to supporting knowledge acquisition, the model encompasses the teaching characteristics (1) *explanation and presentation*, (2) *supporting comprehension*, (3) *practice and retention* as well as (4) *application and transfer*. The teaching requirement motivating students comprises four teaching characteristics: (5) *arousing interest*, (6) *promoting joy*, (7) *clarifying utility* and (8) *promoting self-efficacy*. In relation to classroom management, the model includes the teaching characteristics (9) *supportive learning climate*, (10) *monitoring*, (11) *flow of the lesson*, (12) *group activation* as well as (13) *dealing with disruptions*. Furthermore, (14) *individualization* is conceptualized as an additional teaching characteristic primarily relevant for both supporting students' knowledge acquisition and motivating students. The 14 teaching characteristics were operationalized using 14 scales each comprising four to eight items, which school leaders rated on a four-point Likert scale. The theoretical model, the observation instrument, and its psychometric properties are described in more detail by Gaertner et al. (2021).

After classroom observations, school leaders transmitted the observational data to a web-based feedback portal that generated individual feedback reports for each teacher. The feedback reports comprised a description of the model of teaching quality as well as positive and negative feedback (the three teaching characteristics with the highest and

lowest mean values, respectively). The negative feedback was complemented with specific suggestions for improvement. Based on the individual feedback reports, school leaders provided verbal feedback to teachers following specific guidelines. According to these guidelines, school leaders first outlined the model of teaching quality that defines the standards of the performance evaluation. School leaders then provided positive as well as negative feedback in relation to the teaching characteristics from the feedback reports. In this regard, it is important to note, that not every teacher received feedback on all 14 teaching characteristics. In the end, school leaders suggested how teachers could improve their teaching. The feedback conferences on average lasted  $M = 44$  min. ( $SD = 13$  min.) The feedback intervention and its theoretical and empirical foundations are described in more detail in another study by the authors (Kellermann et al., 2022b).

## 5.2.4 Variables

### 5.2.4.1 Teaching quality

The moderating effect of teachers' goal orientations on the effectiveness of feedback was tested based on students' ratings of teaching quality. Teaching quality was captured with the student questionnaires at  $t_1$  and  $t_2$ . The student questionnaires correspond to the model of teaching quality described in section 2.3. Each of the 14 teaching characteristics was measured with a four-point Likert scale. Construct validity of the student questionnaires as well as reliabilities in terms of internal consistency and intra-class-correlation proved to be satisfactory.<sup>18</sup>

### 5.2.4.2 Teachers' goal orientations

LGO, PAPGO, and PAVGO were assessed with an instrument by Nitsche et al. (2011) in the teacher questionnaire at  $t_1$ . To capture LGO, the original six-point Likert scale comprising nine items ( $\Omega = .82$ ; "In my vocation, I aspire to constantly deal better with critical class situations.") was used. The scales capturing PAPGO and PAVGO were shortened for reasons of survey efficiency. While Nitsche et al. (2011) assessed teachers' striving to demonstrate superior competence as well as their striving to hide inferior competence in relation to various addressees (students, colleagues, superiors as well as themselves),

---

<sup>18</sup> More information on the psychometric properties of the student questionnaires can be found in another publication by the authors (Kellermann et al., 2022b). Item examples and descriptive statistics are provided in Appendix A.

in the present study, PAPGO and PAVGO were measured in relation to no specific addressee. The modifications are in line with a brief version of the instrument by Janke et al. (2019). PAPGO ( $\Omega = .91$ ; “In my vocation, I aspire to show that I deal better with critical lessons than other teachers.”) and PAVGO ( $\Omega = .81$ ; “In my vocation, I aspire to not show when I have more troubles to meet the job demands than other teachers.”) both were measured with six-point Likert scales each comprising three items.

### 5.2.4.3 Covariates

Given the non-random assignment of teachers and classes to IG and CG, further professional development activities teachers engaged in during the intervention year as well as teacher, class, and student characteristics likely to influence changes in teaching quality were captured as covariates. These factors are considered as control variables in order to isolate the intervention effect.

With regard to *teachers' professional development activities*, the teacher questionnaire at  $t_2$  captured whether teachers received feedback by peer teachers and/or their students. In addition, teachers were asked to indicate how many training workshops they attended in the intervention year and how often they read professional literature. The latter was captured with a six-point Likert scale comprising four items ( $\Omega = .82$ ; “I am reading pedagogical literature in my leisure time.”). Regarding *teacher characteristics*, teachers' work experience was captured at  $t_1$ . Furthermore, teachers were asked to indicate whether they have a professional teaching qualification. In relation to *class characteristics*, teachers at  $t_2$  described the class in terms of size and class composition (proportion of girls, proportion of low-SES students, proportion of migrant students, and number of students with special educational needs). Regarding *student characteristics*, subject grade was captured as an indicator for achievement level in the student questionnaire at  $t_2$ .

### 5.2.5 Analyses

The interactions of teachers' LGO, PAPGO, and PAVGO and school leaders' feedback were tested by means of multilevel regression analysis (Hox, 2010; Raudenbush & Bryk, 2002). An intercept-as-outcome model was conducted for each of the 14 teaching characteristics with the respective teaching characteristic at  $t_2$  as the outcome, school leaders' feedback, teachers' goal orientations, and the interaction terms as predictors. In addition, the aforementioned covariates were considered as control variables. The models were two-level models with students at L1 and teachers/classes at L2.

While all metric predictors were centered around their grand mean, dichotomous predictors remained uncentered. The interaction terms were built from the centered LGO, PAPGO, and PAVGO, respectively, and the uncentered feedback variable. Accordingly, the interactions between teachers' goal orientations and school leaders' feedback represent an interaction at L2 rather than a cross-level interaction typically examined in multilevel regression analysis. However, the principles of moderation analyses as they apply in the context of multiple regression analyses (Hayes, 2013; Jaccard & Turrisi, 2003) can be generalized to multilevel regression analyses with interactions at L2 (Bauer & Curran, 2005; Dawson, 2014; Preacher et al., 2006). The analyses were conducted using Mplus 8 (Muthén & Muthén, 1998-2017). As some variables were not normally distributed, a robust maximum likelihood estimator (MLR) was used. Missing values were treated with the full-information maximum likelihood estimator (FIML).

Since interaction effects only indicate if teachers' goal orientations moderate the effectiveness of feedback, additional inferential tests are needed to ascertain at what values of LGO, PAPGO, and PAVGO the feedback effect reaches the level of statistical significance. On the one hand, tests of simple slopes were conducted to estimate the feedback effect at  $+1 SD$ ,  $M$ , and  $-1 SD$  of teachers' goal orientations. On the other hand, the Johnson-Neyman (J-N) technique was applied (Bauer & Curran, 2005). While tests of simple slopes estimate the feedback effect at rather arbitrarily selected values of the moderator, the J-N technique identifies *regions of significance*, that is, ranges of teachers' goal orientations at which the feedback has a significant effect. Test of simple slopes and the J-N technique were carried out using an online tool developed by Preacher et al. (2006) (<https://www.quantpsy.org/interact/hlm2.htm>).

## 5.3 Results

### 5.3.1 Descriptive results

Because random assignments of teachers and classes to IG and CG could not be ensured, particular attention was paid to initial group differences. Differences between IG and CG were assessed based on Cohen's  $d$  (metric variables) and Cohen's  $h$  (dichotomous variables) (Cohen, 1988).

As can be seen in table 1, regarding *teachers' professional development activities*, CG teachers considerably more often received feedback from peer teachers ( $h = -0.61$ ) and students ( $h = -0.28$ ) than IG teachers. In relation to the number of training workshops

attended and the amount of professional literature read, there are no meaningful group differences. In relation to *teacher characteristics*, IG teachers have more work experience ( $d = 0.30$ ) than CG teachers, but less often a professional teaching qualification ( $h = -0.23$ ). Moreover, IG teachers have a more pronounced LGO ( $d = 0.53$ ) and PAPGO ( $d = 0.37$ ), but there are no differences in relation to PAVGO. The same is true for *class characteristics* and *student characteristics*, except that low-SES students are more common in CG classes than in IG classes ( $d = -0.29$ ). In addition, there are no meaningful differences between IG and CG classes in relation to students' perceptions of teaching quality at  $t_1$  (see Appendix A).

**Table 1: Differences between IG and CG**  
(means ( $M$ ), standard deviations ( $SD$ ) and Cohen's  $d/h$ )

	Total		IG		CG		Cohens $d/h^2$
	$M$	$SD$	$M$	$SD$	$M$	$SD$	
<b><i>Professional development activities</i></b>							
Feedback by peer teachers received <sup>1</sup>	0.25	-	0.12	-	0.37	-	-0.61
Feedback by students received <sup>1</sup>	0.31	-	0.24	-	0.37	-	-0.28
Number of training workshops	3.00	2.12	2.84	1.84	3.15	2.38	-0.15
Professional literature read	3.15	1.03	3.21	1.01	3.10	1.06	0.11
<b><i>Teacher characteristics</i></b>							
Work experience (in years)	13.36	8.19	14.64	8.63	12.17	7.74	0.30
Professional teaching qualification <sup>1</sup>	0.80	-	0.76	-	0.85	-	-0.23
LGO	4.89	0.89	5.12	0.64	4.66	1.04	0.53
PAPGO	1.85	0.95	2.03	1.07	1.68	0.80	0.37
PAVGO	1.90	0.96	1.91	1.06	1.88	0.88	0.03
<b><i>Class characteristics</i></b>							
Class size	18.49	5.17	18.29	5.57	18.67	4.88	-0.05
Proportion of girls	0.47	0.39	0.48	0.39	0.46	0.40	0.05
Proportion of low-SES students	0.24	0.27	0.20	0.19	0.28	0.33	-0.29
Proportion of migrant students	0.32	0.22	0.32	0.23	0.32	0.22	0.00
Number of students with SEN	2.22	4.17	2.00	3.42	2.42	4.85	-0.10
<b><i>Student characteristics</i></b>							
Subject grade	2.56	1.05	2.57	1.05	2.55	1.05	0.02

Notes: <sup>1</sup> dichotomous variable (0 "no", 1 "yes"); <sup>2</sup> Cohen's  $d$  for metric variables and Cohen's  $h$  for dichotomous variables; LGO = learning goal orientation, PAPGO = performance approach goal orientation, PAVGO = performance avoidance goal orientation; SEN = special educational needs

### 5.3.2 Results of multilevel regression analyses

Results of multilevel regression analyses are presented in table 2. All coefficients are unstandardized regression coefficients. Empty cells (-) indicate that the corresponding

covariate did not have a significant effect on the respective teaching characteristic ( $p < .10$ ) and was excluded from the final model for reasons of parsimony. Of particular interest for the research question under study are the coefficients reported in section Interactions. A significant regression coefficient indicates that the respective goal orientation and school leaders' feedback interact in predicting changes in teaching quality.

In relation to teachers' LGO, a positive interaction was found for the teaching characteristics *clarifying utility* ( $\gamma = .16, p = .09$ ) and *supportive learning climate* ( $\gamma = .16, p = .01$ ). These results support *Hypothesis 1*, which stated that high-LGO teachers show greater improvements in teaching quality after receiving feedback than low-LGO teachers. However, for 12 teaching characteristics, teachers' LGO did not moderate the effectiveness of feedback.

With regard to teachers' PAPGO, results are inconsistent. For the teaching characteristic *supporting comprehension*, a positive interaction was identified ( $\gamma = .21, p = .07$ ), suggesting that feedback effects are greater for high-PAPGO teachers than for low-PAPGO teachers. In contrast, with regard to the teaching characteristics *flow of the lesson* ( $\gamma = -.35, p < .01$ ), *group activation* ( $\gamma = -.25, p = .01$ ), and *dealing with disruptions* ( $\gamma = -.22, p = .02$ ) negative interactions suggest that the feedback effect is smaller for high-PAPGO teachers than for low-PAPGO teachers. In summary, these findings provide no clear picture of how a teacher's PAPGO moderates the effectiveness of feedback (*Hypothesis 2*). In addition, for 10 teaching characteristics, no interaction could be identified.

Regarding teachers' PAVGO, contrary to our expectations (*Hypotheses 3*), one interaction could be detected. In relation to the teaching characteristic *flow of the lesson*, in addition to the negative interaction of teachers' PAPGO and school leaders' feedback, a positive interaction was identified for teachers' PAVGO ( $\gamma = .18, p = .09$ ). That is, high-PAVGO teachers show greater improvements in teaching quality after receiving feedback than low-PAVGO teachers. However, for 13 teaching characteristics, teachers' PAVGO did not moderate the effectiveness of feedback.

To shed further light on the moderation, each significant interaction effect was tested using tests of simple slopes and the J-N technique. Prior to this, for reasons of parsimony, all non-significant interaction terms and the corresponding moderator ( $p < .10$ ) were excluded from the respective model. Regarding the teaching characteristic *flow of the lesson*, after excluding teachers' LGO, only teachers' PAPGO interacted significantly with school leaders' feedback, teachers' PAVGO did not. Consequently, we tested only the moderating effect of teachers' PAPGO.

Kapitel 5 | Teilstudie 3

**Table 2: Results of multilevel regression analyses (unstandardized coefficients ( $\gamma$ ) and standard errors (S.E.) in parantheses)**

	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	TC11	TC12	TC13	TC14
<b>Student level predictors</b>														
L1 Teaching characteristic t <sub>1</sub>	.35* (.04)	.29* (.04)	.40* (.04)	.34* (.05)	.48* (.04)	.44* (.05)	.34* (.05)	.44* (.04)	.46* (.03)	.39* (.04)	.39* (.04)	.33* (.05)	.33* (.04)	.44 (.06)
L1 Subject grade	-.10* (.02)	-	-	-.07* (.02)	-.08* (.03)	-.08* (.02)	-.07* (.02)	-.08* (.03)	-	-.04 <sup>+</sup> (.03)	-	-	-.06* (.02)	-
<b>Class level predictors</b>														
L2 Teaching characteristic t <sub>1</sub>	.91* (.12)	.83* (.21)	1.05* (.17)	.92* (.11)	.90* (.11)	.98* (.11)	.88* (.09)	1.20* (.13)	.42* (.07)	1.29* (.24)	.93* (.09)	1.05* (.17)	.74* (.12)	.91* (.15)
L2 Subject grade	.16* (.07)	-	-	-	-	-	-	.17* (.07)	-	-	.13* (.06)	-	-	-
<b>Professional development activities</b>														
Feedback by school leaders received (intervention) <sup>1</sup>	.00 (.08)	.17* (.07)	-.03 (.08)	.04 (.07)	.08 (.11)	.01 (.09)	.02 (.08)	.11 (.08)	.17* (.05)	.05 (.08)	-.12* (.06)	.12 <sup>+</sup> (.07)	-.01 (.06)	.30* (.11)
Feedback by peer teachers received <sup>1</sup>	-	.23* (.07)	-	.19* (.06)	.23* (.09)	-	-	.21* (.07)	-	-	-	.19* (.09)	-	.29* (.10)
Feedback by students received <sup>1</sup>	-	-	-	-	-	-	-	-	.17* (.06)	-	-	-	-	-
Number of training workshops	.03* (.02)	-	-	-	-	.04* (.02)	-	.04* (.01)	-	-	-	-	-	-
Reading professional literature	-	-	-	-	-	-	-	-	-	-	.08* (.03)	-	-	-
<b>Teacher characteristics</b>														
Professional experience	-	-	.01* (.00)	.01* (.00)	.01* (.00)	-	.01* (.00)	-	-	-	-	-	.01* (.00)	-
Professional teaching qualification <sup>1</sup>	-	-	-	-.19* (.06)	-	-	-	-.16 <sup>+</sup> (.09)	-	-	-	-	-	-
LGO	.01 (.05)	-.02 (.04)	.00 (.05)	-.04 (.03)	.02 (.04)	-.02 (.05)	-.06 (.04)	-.03 (.04)	-.03 (.03)	-.05 (.04)	.05 (.03)	-.01 (.03)	.06 <sup>+</sup> (.03)	-.04 (.05)
PAPGO	.16 (.13)	-.18 <sup>+</sup> (.10)	-.05 (.11)	.05 (.10)	.07 (.14)	.12 (.12)	.11 (.13)	.11 (.13)	-.16* (.06)	.03 (.09)	.15 (.11)	.19* (.07)	.10 (.07)	-.18 (.12)
PAVGO	-.15 (.12)	.10 (.09)	.06 (.09)	.01 (.09)	-.05 (.13)	.01 (.10)	-.09 (.12)	-.02 (.13)	.16* (.04)	-.01 (.08)	-.12 (.10)	-.13 <sup>+</sup> (.07)	-.09 (.07)	.21* (.08)
<b>Class characteristics</b>														
Class size	-	-	-	-	-	-	-	-	-.02* (.01)	-	-	-	-	-
Proportion of girls	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Proportion of low-SES students	-	-	-	-	-	-	-	-	-	-	-	-	-.25 <sup>+</sup> (.13)	-
Proportion of migrant students	-	.49* (.16)	-	.36* (.14)	-	-	.47* (.16)	.39* (.13)	.36* (.16)	-	-	-	-	-
Number of students with SEN	-	-.02* (.01)	-	-.02* (.00)	-.02* (.01)	-	-.02* (.01)	-.02* (.01)	-.03* (.01)	-	-	-	-	-.02* (.01)
<b>Interactions</b>														
Feedback intervention * LGO	.00 (.11)	-.02 (.07)	.07 (.11)	.06 (.07)	.04 (.12)	.21 (.13)	.16 <sup>+</sup> (.09)	.03 (.08)	.16* (.06)	.12 (.08)	.05 (.06)	.08 (.08)	-.02 (.09)	.07 (.10)
Feedback intervention * PAPGO	-.20 (.15)	.21 <sup>+</sup> (.11)	-.05 (.11)	-.06 (.11)	-.09 (.17)	-.19 (.15)	-.12 (.14)	-.07 (.15)	.03 (.08)	-.06 (.13)	-.35* (.12)	-.25* (.10)	-.22* (.09)	.16 (.15)
Feedback intervention * PAVGO	.18 (.13)	-.06 (.11)	-.06 (.11)	-.01 (.10)	.06 (.16)	.02 (.13)	.10 (.13)	.02 (.14)	-.03 (.06)	.09 (.10)	.18 <sup>+</sup> (.11)	.12 (.09)	.07 (.08)	-.16 (.11)
<b>Explained variance</b>														
Explained variance at student level	16 %	10 %	15 %	14 %	26 %	23 %	15 %	25 %	21 %	18 %	16 %	11 %	16 %	19 %
Explained variance at class level	71 %	70 %	70 %	92 %	70 %	78 %	83 %	92 %	99 %	86 %	91 %	86 %	72 %	68 %

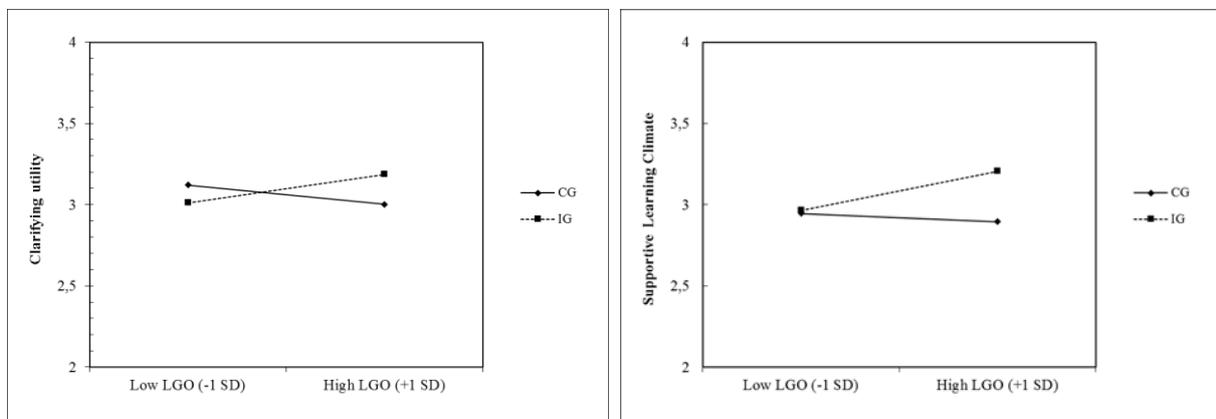
Notes: <sup>+</sup>p < .10, \*p < .05

<sup>1</sup> dichotomous variable (0 “no”, 1 “yes”); LGO = learning goal orientation, PAPGO = performance approach goal orientation; PAVGO = performance avoidance goal orientation; SEN = special educational needs; TC1 = Explanation and presentation; TC2 = Supporting comprehension; TC3 = Practice and retention; TC4 = Application and transfer; TC5 = Arousing interest; TC6 = Promoting joy; TC7 = Clarifying utility; TC8 = Promoting self-efficacy; TC9 = Supportive learning climate; TC10 = Monitoring; TC11 = Flow of the lesson; TC12 = Group activation; TC13 = Dealing with disruptions; TC14 = Individualisation.

### 5.3.3 Testing the interactions

With regard to teachers LGO, for *clarifying utility* tests of simple slopes yielded a positive feedback effect ( $\gamma = .19, p = .04$ ) for high-LGO teachers (+1 *SD*) and no effect for teachers with average (*M*) and low LGO (-1 *SD*). The J-N technique reveals that the effect of school leaders' feedback is significant at any value of teachers' LGO higher than 0.50 above the mean. That is, only teachers with a rather high LGO improved their teaching after receiving feedback. The interaction of teachers' LGO and school leaders' feedback alone explains an additional 5.1 % of variance.

Regarding *supportive learning climate*, an average LGO (*M*) as well as a high LGO (+1 *SD*) are associated with improvements in teaching quality ( $\gamma = .17, p < .01$  and  $\gamma = .31, p < .01$ , respectively); a low LGO (-1 *SD*) is not. In fact, starting from a LGO of 0.38 below the mean, school leaders' feedback yielded a positive effect on changes in teaching quality. That is, feedback was effective even for teachers with a LGO just below average and ineffective only for teachers with a rather low LGO. The interaction alone explains 4.5 % of variance. To facilitate interpretation, the interactions of teachers' LGO and school leaders' feedback are illustrated in figure 1.<sup>19</sup>



**Figure 1:** Interaction of teachers' learning goal orientation (LGO) and school leaders' feedback in predicting changes in teaching quality

As shown before, results are inconclusive with respect to teachers' PAPGO. Regarding *supporting comprehension*, a positive interaction was detected. Test of simple

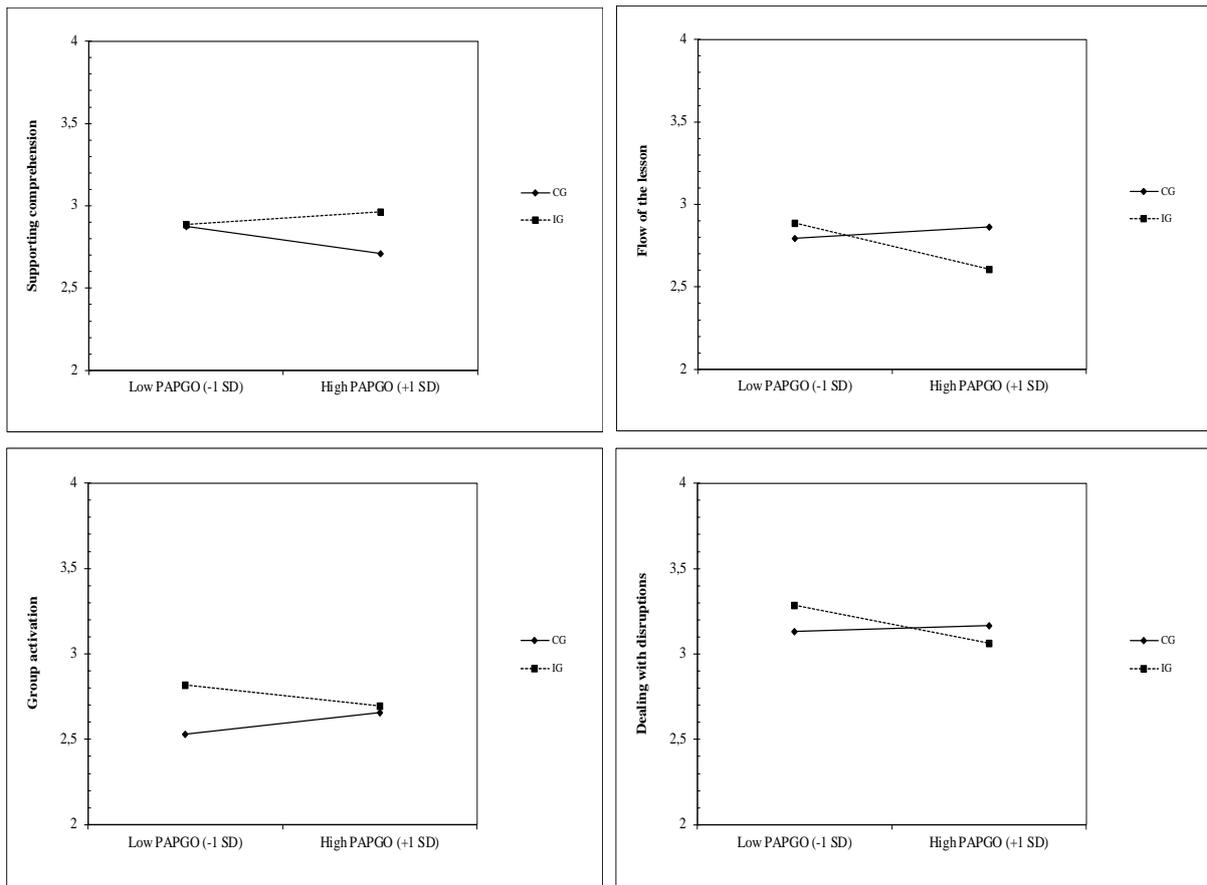
<sup>19</sup> For illustration, an online tool developed by Dawson (2014) was used (<http://www.jeremydawson.co.uk/slopes.htm>).

slopes reveal that an average PAPGO ( $M$ ) as well as a high PAPGO ( $+1 SD$ ) are associated with improvements in teaching quality ( $\gamma = .13, p = .06$  and  $\gamma = .25, p = .01$ , respectively); a low PAPGO ( $-1 SD$ ) is not. The J-N technique reveals that starting from a PAPGO of 0.13 below the mean, IG teachers significantly improved their teaching. That is, school leaders' feedback even yielded an effect for teachers with a PAPGO just below average and was ineffective only for teachers with a low PAPGO. The interaction contributes 6.8 % to the variance explained.

In contrast, when it comes to *flow of the lesson*, the interaction of teachers' PAPGO and school leaders' feedback was negative. Tests of simple slopes yielded no effect for teachers with average ( $M$ ) and low PAPGO ( $-1 SD$ ) and a negative feedback effect ( $\gamma = -.26, p = .01$ ) for high-PAPGO teachers ( $+1 SD$ ). Starting from a PAPGO of 0.10 above the mean, the feedback had a negative effect on changes in teaching quality. That is, IG teachers with a PAPGO above average show a lower teaching quality after receiving feedback than CG teachers. The variance explained by the interaction alone amounts to 5.5 %.

Regarding *group activation*, an average ( $M$ ) as well as a low PAPGO ( $-1 SD$ ) are associated with improvements in teaching quality ( $\gamma = .17, p = .02$  and  $\gamma = .29, p < .01$ , respectively); a high PAPGO ( $+1 SD$ ) is not. The J-N technique indicates that the effect of school leaders' feedback on improvements in teaching quality is significant at any value of teachers' PAPGO lower than 0.36 above the mean. That is, school leaders' feedback yielded no effect for teachers with a rather high PAPGO but was effective for teachers with a PAPGO just above average and below. The interaction term alone explains an additional 4.6 % of variance.

With regard to *dealing with disruptions*, tests of simple slopes yielded a positive feedback effect ( $\gamma = .15, p = .08$ ) for teachers with low PAPGO ( $+1 SD$ ) and no effect for teachers with average ( $M$ ) and high PAPGO ( $+1 SD$ ). Starting from a PAPGO of 0.74 below the mean, IG teachers significantly improved their teaching after receiving feedback. That is, the feedback only yielded an effect for teachers with a rather low PAPGO. The interaction term accounts for 4.0 % of variance explained. The interaction effects of teachers' PAPGO and school leaders' feedback are illustrated in figure 2.



**Figure 2:** Interaction of teachers' performance approach goal orientation (PAPGO) and school leaders' feedback in predicting changes in teaching quality

## 5.4 Discussion

### 5.4.1 Summary of results

Feedback on teaching is considered a promising measure for improving the quality of teaching. However, feedback to teachers is not always effective. By linking goal orientation theory with Kluger and DeNisi's (1996) FIT, we elaborated on how teachers' goal orientations may influence the processing and usage of feedback. Using data from an intervention study, we thereafter empirically tested whether teachers' goal orientations moderate the effectiveness of feedback. Interactions with school leaders' feedback were found for teachers' LGO and PAPGO.

With regard to LGO, it was hypothesized that high-LGO teachers show greater improvements in teaching quality after receiving feedback than low-LGO teachers. However, for most of the teaching characteristics examined, teachers' LGO and school leaders' feedback did not interact in predicting changes in teaching quality. The assumed positive interaction was found for only two teaching characteristics. In relation to *clarifying*

*utility* as well as *supportive learning climate*, the feedback was more effective for high-LGO teachers than for low-LGO teachers. More specifically, teaching quality improved among teachers with an average to high LGO, while feedback was ineffective among low-LGO teachers.

In relation to PAPGO, no specific hypothesis was tested. Again, for most of the teaching characteristics examined, no moderation was found. With regard to four teaching characteristics, however, teachers' PAPGO and school leaders' feedback interactively predicted changes in teaching quality. Negative interactions in terms of *flow of the lesson*, *group activation*, and *dealing with disruptions* show that feedback was less effective for high-PAPGO teachers than for low-PAPGO teachers. More specifically, teachers with a low to average PAPGO improved their teaching after receiving feedback, while high-PAPGO teachers did not. In one instance, high-PAPGO teachers even exhibited a lower teaching quality after receiving feedback, suggesting that feedback may in some cases have a negative impact among teachers with a pronounced PAPGO. While these results suggest that feedback is less effective for high-PAPGO teachers, the positive interaction relating to *supporting comprehension* indicates that high-PAPGO teachers may also show greater improvements in teaching quality after receiving feedback than low-PAPGO teachers.

The next section takes a closer look at the moderating effects of teachers' LGO and, in particular, the contradictory findings on teachers' PAPGO. For this purpose, we draw on the observational data from the web-based feedback portal, which formed the basis of the teacher-specific feedback reports that guided school leaders in providing feedback. As described in section 2.3, the feedback reports contained positive and negative feedback in relation to the three teaching characteristics with the highest and lowest mean values, respectively. That is, not every teacher received feedback on all teaching characteristics. Therefore, feedback effects and in turn interaction effects were most likely to occur for those teaching characteristics on which many IG teachers received feedback. In addition, feedback may have been predominately positive with respect to some teaching characteristics, while it was predominately negative with respect to other teaching characteristics.<sup>20</sup>

---

<sup>20</sup> Although we have access to the observational data, due to data protection regulations, it is not possible to assign the feedback reports to individual teachers. For the entire sample, however, it is possible to estimate how many IG teachers received feedback on each teaching characteristic and whether the feedback was predominately positive or negative.

#### 5.4.2 Teachers' goal orientations and the valence of feedback

In relation to LGO, the observational data show that both *clarifying utility* and *supportive learning climate* were frequently addressed by school leaders. While IG teachers often received both positive and negative feedback with regard to *clarifying utility*, the feedback predominantly was positive in terms of *supportive learning climate*.

In accordance with our theoretical assumptions, high-LGO teachers because of their high level of *task involvement* in achievement-related settings may have perceived information on a performance-standard comparison as cues to improve their teaching and focused on the task when processing feedback. Therefore, high-LGO teachers who received negative feedback on the teaching characteristic *clarifying utility* could have strived to reduce the performance-standard discrepancy by increasing their effort and/or by developing new strategies on how to emphasize the usefulness of the content to their students. At the same time, high-LGO teachers who received positive feedback may have set higher goals for themselves, resulting in a discrepancy between their actual and a desired performance that they subsequently sought to reduce. The same could be true for the teaching characteristic *supportive learning climate*. According to the observational data, the learning climate in IG teachers' classes predominantly was characterized by mutual respect and the willingness to help one another. However, IG teachers with a pronounced LGO might have strived for an even better learning climate after receiving positive feedback. The positive interactions of teachers' LGO and school leaders' feedback in predicting changes in teaching quality generally are in line with findings of another study by the authors. Here too, high-LGO teachers reported a higher competence development after receiving feedback than high-LGO teachers who did not receive feedback (Kellermann et al., 2022a).

With regard to PAPGO, the observational data show that the teaching characteristics *supporting comprehension*, *flow of the lesson*, and *dealing with disruptions* all were frequently addressed by school leaders. However, with regard to *group activation*, not many IG teachers received feedback, raising the question of whether improvements in teaching quality among teachers with an average and low PAPGO can be attributed to the feedback. Therefore, the moderating effect of teachers' PAPGO in relation to *group activation* will not be discussed any further.

While IG teachers often received negative feedback with regard to *supporting comprehension*, the feedback predominantly was positive in terms of *flow of the lesson* and

*dealing with disruptions*. From a theoretical point of view, we assumed that high-PAPGO teachers because of their high level of *ego involvement* in achievement-related settings would perceive information on a performance-standard comparison primarily as an evaluation of their teaching competence and therefore would focus on the self when processing feedback. Self-oriented feedback processing, according to FIT, often is associated with negative affect, a depletion of cognitive resources for task performance and is likely to result in a rejection of feedback (Kluger & DeNisi, 1996). However, self-oriented feedback processing primarily is to be expected in case of negative feedback. Because IG teachers in relation to *dealing with disruptions* and *flow of the lesson* predominantly received positive feedback, self-oriented feedback processing is insufficient to explain the negative interaction of teachers' PAPGO and school leaders' feedback. Instead, it seems plausible that high-PAPGO teachers felt they had accomplished their goal of demonstrating superior competence when receiving positive feedback and, as a result, did not set higher goals for themselves. Furthermore, the negative feedback effect in relation to *flow of the lesson* indicates that after goal accomplishment, high-PAPGO teachers could even dedicate less attention to the respective teaching characteristic resulting in a decline of teaching quality.

With regard to *supporting comprehension*, IG teachers often received negative feedback. Although negative feedback potentially poses a threat to high-PAPGO teachers' self-esteem and feedback rejections seems likely, high-PAPGO teachers apparently used the information on the performance-standard discrepancy to align with their self-goal of being a better teacher than their colleagues are. They might have strived to reduce the performance-standard discrepancy by increasing their effort and/or by developing new strategies on how to stimulate students' in-depth understanding. The positive interaction of teachers' PAPGO and school leaders' feedback regarding *supporting comprehension* is in line with findings from the previously quoted study by the authors. Not only high-LGO teachers, but also high-PAPGO teachers reported a higher competence development after receiving feedback (Kellermann et al, 2022a).

In summary, our findings suggest that a pronounced LGO is beneficial for the effectiveness of feedback; whether feedback is positive or negative seems to be irrelevant. With regard to teachers' PAPGO, however, the above discussion suggests that valence might play a crucial role. After receiving positive feedback, low-PAPGO teachers showed greater improvements in teaching quality than high-PAPGO teachers. In contrast, after receiving negative feedback, high-PAPGO teachers showed greater improvements than

low-PAPGO teachers. However, as with teachers' LGO, for many of the teaching characteristics examined, no interaction of teachers' PAPGO and school leaders' feedback could be identified. This aspect and other limitations will be addressed in the next section.

### 5.4.3 Limitations and directions for future research

One limitation relates to the sample size of  $N = 53$  teachers/classes at L2. The small sample is associated with low statistical power, which is especially true for testing interaction effects (Aguinis, 1995; Shieh, 2009). Given a larger sample, more interactions might have reached the level of statistical significance. Therefore, more studies, ideally with larger samples than the one at hand, are needed to gain further insights into the role teachers' goal orientations play for the effectiveness of feedback.

Another limitation relates to the nature of feedback. There is no information on the focus of feedback received by individual teachers. Although IG teachers did not receive feedback on every teaching characteristic, interactions of teachers' goal orientations and school leaders' feedback were tested for the total sample, which again impedes the detection of interaction effects. In addition, there is no teacher-specific information on the valence of feedback. Whether the moderating effect of teachers' PAPGO varies as a function of feedback valence needs to be verified in the future.

A third limitation concerns the processing of feedback and the assumed causal mechanisms. By linking goal orientation theory with FIT, we offered an explanation on how teachers' goal orientations might influence feedback processing and in turn its effectiveness. However, we do not have information on teachers' affect and cognitions after receiving feedback, nor do we know what measures teachers implemented to improve their teaching. Future studies need to capture teachers' affective and cognitive responses and scrutinize on the actions they undertake to resolve a performance-standard discrepancy.

### 5.4.4 Conclusion and practical implications

In conclusion, our results provide preliminary evidence for the moderating effect of teachers' goal orientations on the effectiveness of feedback. Regardless of the valence of feedback, LGO appears to be beneficial for the effectiveness of feedback. In addition, PAPGO could also be associated with larger feedback effects—at least when feedback is negative. In case of positive feedback, in contrast, our results indicate that PAPGO might attenuate the effectiveness of feedback.

Although teachers' goal orientations are considered as relatively stable motivational orientations, the underlying achievement goals do vary over time, which is especially true for learning goals (Praetorius et al., 2014). Hence, it can be concluded that achievement goals as well as the corresponding patterns of affect, cognition, and behavior can be modified. Since a high LGO not only is associated with improvements in teaching quality after receiving feedback, but also with other beneficial competence-related attitudes and activities, it seems worthwhile to identify factors that foster a learning goal among teachers. Although research on this issue is scarce, there are studies indicating that various factors of the school environment are associated with a teacher's LGO. For example, Cho and Shim (2013) found that a teacher's perception of the school's goal structure is related to the teacher's goal orientation. While a perceived performance goal structure is related to a teacher's PGO, a perceived learning goal structure is related to a teacher's LGO. The authors conclude that by creating a school culture that values learning and competence development, a LGO among teachers could be promoted (see also Skaalvik & Skaalvik, 2013).

Furthermore, a teacher's perceived need satisfaction in terms of autonomy, competence, and relatedness in the working environment has been found to be positively related to a teacher's LGO (Janke et al., 2015). In this regard, Dickhäuser et al. (2021) show that school leaders who, for example, provide teachers with opportunities to develop and enact their own approaches to instruction promote a LGO among teachers, presumably because they satisfy a teacher's need for autonomy. Of particular interest in the context of the present study is the finding that schools that value and rely on feedback procedures also foster a LGO among teachers (Dickhäuser et al., 2021). Thus, it can be concluded that not only is a LGO beneficial for the effectiveness of feedback, but feedback—when provided regularly—can also promote a LGO among teachers.

## 5.5 References

- Aguinis, H. (1995). Statistical power with moderated multiple regression in management research. *Journal of Management*, *21*(6), 1141–1158. <https://doi.org/10.1177/014920639502100607>
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*(3), 261–271. <https://doi.org/10.1037/0022-0663.84.3.261>

- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Butler, R. (2007). Teachers' achievement goal orientations and associations with teachers' help seeking. Examination of a novel approach to teacher motivation. *Journal of Educational Psychology*, 99(2), 241–252. <https://doi.org/10.1037/0022-0663.99.2.241>
- Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, 40(3), 373–400. [https://doi.org/10.1207/s15327906mbr4003\\_5](https://doi.org/10.1207/s15327906mbr4003_5)
- Cellar, D. F., Stuhlmacher, A. F., Young, S. K., Fisher, D. M., Adair, C. K., Haynes, S., Twichell, E., Arnold, K. A., Royer, K., Denning, B. L., & Riester, D. (2011). Trait goal orientation, self-regulation, and performance: A meta-analysis. *Journal of Business and Psychology*, 26(4), 467–483. <https://doi.org/10.1007/s10869-010-9201-6>
- Chughtai, A. A., & Buckley, F. (2010). Assessing the effects of organizational identification on inrole job performance and learning behaviour. *Personnel Review*, 39(2), 242–258. <https://doi.org/10.1108/00483481011017444>
- Cho, Y., & Shim, S. S. (2013). Predicting teachers' achievement goals for teaching: The role of perceived school goal structure and teachers' sense of efficacy. *Teaching and Teacher Education*, 32, 12–21. <https://doi.org/10.1016/j.tate.2012.12.003>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.
- Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378–387. <https://doi.org/10.3102/0013189X16659442>
- Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, 29(1), 1–19. <https://doi.org/10.1007/s10869-013-9308-7>
- Dickhäuser, O., Janke, S., Daumiller, M., & Dresel, M. (2021). Motivational school climate and teachers' achievement goal orientations: A hierarchical approach. *The British Journal of Educational Psychology*, 91(1), 391–408. <https://doi.org/10.1111/bjep.12370>

- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048. <https://doi.org/10.1037//0003-066X.41.10.1040>
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273. <https://doi.org/10.1037//0033-295X.95.2.256>
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169–189. [https://doi.org/10.1207/s15326985ep3403\\_3](https://doi.org/10.1207/s15326985ep3403_3)
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72(1), 218–232. <https://doi.org/10.1037/0022-3514.72.1.218>
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70(3), 461–475. <https://doi.org/10.1037/0022-3514.70.3.461>
- Gaertner, H., Thiel, F., Nachbauer, M., & Kellermann, C. (2021). Unterrichtsentwicklung durch Unterrichtsfeedback – Erste Entwicklungsschritte eines Beobachtungsbogens für Schulleitungen. [Instructional development through feedback – First development steps of an observation sheet for school management]. *Zeitschrift für Bildungsforschung*, 89(11), 529–547. <https://doi.org/10.1007/s35834-021-00321-7>
- Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96–104. <https://doi.org/10.3102/0013189X15575031>
- Gorozidis, G. S., & Papaioannou, A. G. (2016). Teachers' achievement goals and self-determination to engage in work tasks promoting educational innovations. *Learning and Individual Differences*, 49, 46–58. <https://doi.org/10.1016/j.lindif.2016.05.014>
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, 94(3), 638–645. <https://doi.org/10.1037//0022-0663.94.3.638>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. The Guilford Press.

- Heneman, H. G., & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(2), 173–195.  
<https://doi.org/10.1023/B:PEEV.0000032427.99952.02>
- Hill, H. C. & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2), 371–384. <https://doi.org/10.17763/haer.83.2.d11511403715u376>
- Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). Routledge.
- Huang, C. (2011). Achievement goals and achievement emotions: A meta-analysis. *Educational Psychology Review*, 23(3), 359–388. <https://doi.org/10.1007/s10648-011-9155-x>
- Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(1), 48–73. <https://doi.org/10.1037/a0026223>
- Hulleman, C. S., Schrage, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, 136(3), 422–449. <https://doi.org/10.1037/a0018947>
- Ilgén, D. R., & Davis, C. A. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology: An International Review*, 49(3), 550–565. <https://doi.org/10.1111/1464-0597.00031>
- Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression*. Sage Publications. <https://doi.org/10.4135/9781412984522>
- Janke, S., Bardach, L., Oczlon, S., & Lüftenegger, M. (2019). Enhancing feasibility when measuring teachers' motivation: A brief scale for teachers' achievement goal orientations. *Teaching and Teacher Education*, 83, 1–11. <https://doi.org/10.1016/j.tate.2019.04.003>
- Janke, S., Nitsche, S., & Dickhäuser, O. (2015). The role of perceived need satisfaction at work for teachers' work-related learning goal orientation. *Teaching and Teacher Education*, 47, 184–194. <https://doi.org/10.1016/j.tate.2015.01.009>
- Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, 19(2), 141–184. <https://doi.org/10.1007/s10648-006-9012-5>

- Kellermann, C., Nachbauer, M., Gaertner, H., & Thiel, F. (2022). Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interventionsstudie unter Berücksichtigung der Zielorientierungen von Lehrkräften [Effects of school leaders' feedback on teaching on teachers' self-rated instructional competence development – Results of an intervention study considering teachers' goal orientations]. *Zeitschrift für Erziehungswissenschaft*, 25(5), 1267–1291.  
<https://doi.org/10.1007/s11618-022-01087-y>
- Kellermann, C., Nachbauer, M., Gaertner, H., & Thiel, F. (2022). The effect of school leaders' feedback on teaching quality perceived by students – Results of an intervention study. *School Effectiveness and School Improvement*. Advance online publication. <https://doi.org/10.1080/09243453.2022.2075018>
- 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>
- Kraft, M. A., & Gilmour, A. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, 52(5), 711–753.  
<https://doi.org/10.1177/0013161X16653445>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93(1), 77–86. <https://doi.org/10.1037//0022-0663.93.1.77>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus. Statistical analysis with latent variables. User's guide*.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328–346.  
<https://doi.org/10.1037/0033-295X.91.3.328>

- Nitsche, S., Dickhäuser, O., Dresel, M., & Fasching, M. S. (2013a). Zielorientierungen von Lehrkräften als Prädiktoren lernrelevanten Verhaltens [Teachers goal orientations as predictors of vocational learning behavior]. *Zeitschrift für Pädagogische Psychologie*, 27(1-2), 95–103. <https://doi.org/10.1024/1010-0652/a000092>
- Nitsche, S., Dickhäuser, O., Fasching, M. S., & Dresel, M. (2011). Rethinking teachers' goal orientations. Conceptual and methodological enhancements. *Learning and Instruction*, 21(4), 574–586. <https://doi.org/10.1016/j.learninstruc.2010.12.001>
- Nitsche, S., Dickhäuser, O., Fasching, M. S., & Dresel, M. (2013b). Teachers' professional goal orientations: Importance for further training and sick leave. *Learning and Individual Differences*, 23, 272–278. <https://doi.org/10.1016/j.lindif.2012.07.017>
- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123–141. <https://doi.org/10.17763/haer.82.1.v40p0833345w6384>
- Praetorius, A.-K., Nitsche, S., Janke, S., Dickhäuser, O., Drexler, K., Fasching, M., Dresel, M. (2014). Here today, gone tomorrow? Revisiting the stability of teachers' achievement goals. *Contemporary Educational Psychology*, 39(4), 379–387. <https://doi.org/10.1016/j.cedpsych.2014.10.002>
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modelling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31(3), 437–448. <https://doi.org/10.3102/10769986031004437>
- Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92(1), 128–150. <https://doi.org/10.1037/0021-9010.92.1.128>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Sage Publications.
- Retelsdorf, J., Butler, R., Streblow, L., & Schiefele, U. (2010). Teachers' goal orientations for teaching. Associations with instructional practices, interest in teaching, and burnout. *Learning and Instruction*, 20(1), 30–46. <https://doi.org/10.1016/j.learninstruc.2009.01.001>

- Runhaar, P., Sanders, K., & Yang, H. (2010). Stimulating teachers' reflection and feedback asking. An interplay of self-efficacy, learning goal orientation, and transformational leadership. *Teaching and Teacher Education*, *26*(5), 1154–1161.  
<https://doi.org/10.1016/j.tate.2010.02.011>
- Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist*, *46*(1), 26–47. <https://doi.org/10.1080/00461520.2011.538646>
- Shieh, G. (2010). Sample size determination for confidence intervals of interaction effects in moderated multiple regression with continuous predictor and moderator variables. *Behavior research methods*, *42*(3), 824–835.  
<https://doi.org/10.1177/1094428108320370>
- Skaalvik, E. M., & Skaalvik, S. (2013). Teachers' perceptions of the school goal structure: Relations with teachers' goal orientations, work engagement, and job satisfaction. *International Journal of Educational Research*, *62*, 199–209.  
<https://doi.org/10.1016/j.ijer.2013.09.004>
- Song, M., Wayne, A. J., Garet, M. S., Brown, S., & Rickles, J. (2021). Impact of providing teachers and principals with performance feedback on their practice and student achievement: Evidence from a large-scale randomized experiment. *Journal of Research on Educational Effectiveness*, *14*(2), 353–378.  
<https://doi.org/10.1080/19345747.2020.1868030>
- Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, *10*(4), 535–572.  
[https://doi.org/10.1162/EDFP\\_a\\_00173](https://doi.org/10.1162/EDFP_a_00173)
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*(7), 3628–3651.  
<https://doi.org/10.1257/aer.102.7.3628>
- van Daal, T., Donche, V., & De Maeyer, S. (2014). The impact of personality, goal orientation and self-efficacy on participation of high school teachers in learning activities in the workplace. *Vocations and Learning*, *7*(1), 21–40.  
<https://doi.org/10.1007/s12186-013-9105-5>
- VandeWalle, D., Cron, W. L., & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology*, *86*(4), 629–640.  
<https://doi.org/10.1037//0021-9010.86.4.629>

Wirthwein, L., Sparfeldt, J. R., Piquart, M., Wegerer, J., & Steinmayr, R. (2013). Achievement goals and academic achievement: A closer look at moderating factors. *Educational Research Review*, *10*, 66–89. <https://doi.org/10.1016/j.edurev.2013.07.001>

# 6

## Gesamtdiskussion

## 6 Gesamtdiskussion

Seit Einführung des neuen Steuerungsparadigmas im deutschen Schulsystem müssen Schulleitungen zusätzlich zu operativen Managementaufgaben vermehrt strategische Führungsaufgaben im Hinblick auf die innerschulische Qualitätssicherung und -entwicklung übernehmen (vgl. Altrichter et al., 2019; Bensen, 2016; Brauckmann & Eder, 2019; Brauckmann & Schwarz, 2015). Im Fokus des Schulleitungshandelns steht dabei vor allem die Qualität der von Lehrkräften gestalteten unterrichtlichen Lerngelegenheiten, die für den Lernerfolg von Schülerinnen und Schülern ausschlaggebend sind (vgl. z. B. Kyriakides et al., 2013; Lipowsky, 2006; Muijs et al., 2014). Unterrichtsbeobachtungen und -feedbacks gehören zu den Führungsaktivitäten, über die Schulleitungen die Unterrichtsqualität an ihren Schulen positiv beeinflussen können. In Deutschland sind beobachtungsbasierte Unterrichtsfeedbacks durch Schulleitungen jedoch kaum verbreitet (vgl. Bach et al., 2020; Barfknecht & von Saldern, 2010; Schleicher, 2012; Thillmann et al., 2015a, 2015b). Darüber hinaus zeigen Studienbefunde, dass Unterrichtsfeedbacks nicht immer zu einer besseren Unterrichtsqualität bzw. einer Steigerung des Lernerfolgs von Schülerinnen und Schülern führen (vgl. z. B. Kraft et al., 2018; Song et al., 2021; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). Aus diesen Gründen wurde im DFG-Projekt „Unterrichtsbeurteilungen und -feedback durch Schulleitungen“ unter Mitarbeit des Verfassers dieser Arbeit eine Feedbackintervention entwickelt, die es Schulleitungen ermöglichen soll, ihren Lehrkräften ein wirksames Feedback zu ihrem Unterricht zu geben.

Die vorliegende Dissertation beschäftigt sich vor diesem Hintergrund mit den Gelingensbedingungen von beobachtungsbasiertem Unterrichtsfeedback durch Schulleitungen. In Kapitel 2.3 wurde ein theoretisches Feedbackmodell entwickelt, das den Feedbackprozess beschreibt und Bedingungen für seine Wirksamkeit benennt. Ausgehend von diesem Feedbackmodell wurde aufgezeigt, dass die Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback unter anderem von der Objektivität, Reliabilität und Validität der Unterrichtsbeurteilung, der formalen Feedbackgestaltung sowie den für die Unterrichtsbeobachtung und Feedbackgabe benötigten Kompetenzen der Schulleitung abhängt. Alle drei Gelingensbedingungen wurden bei der Entwicklung der Feedbackintervention berücksichtigt. In den Teilstudien 1 und 2 wurde die Wirksamkeit der Intervention empi-

risch überprüft. Es wurde untersucht, ob das Unterrichtsfeedback der Schulleitungen einen Effekt auf die Kompetenzentwicklung von Lehrkräften hat und zu einer besseren Unterrichtsqualität führt.

Wie im theoretischen Feedbackmodell herausgearbeitet, hängt die Feedbackwirksamkeit jedoch nicht allein von der Objektivität, Reliabilität und Validität der Unterrichtsbeurteilung, der formalen Feedbackgestaltung sowie den Beobachtungs- und Feedbackkompetenzen der Schulleitung ab, sondern auch von der Verarbeitung und Nutzung des Feedbacks durch die Lehrkräfte selbst. Die Feedbackverarbeitung und -nutzung wiederum wird von verschiedenen Merkmalen der Lehrkraft beeinflusst. Der Fokus dieser Arbeit liegt auf Zielorientierungen, von denen bekannt ist, dass sie differenzielle Zusammenhänge mit lernrelevanten Einstellungen und Verhaltensweisen von Lehrkräften aufweisen (vgl. z. B. Butler, 2007; Janke et al., 2019; Nitsche et al., 2013a, 2013b). Daher kann angenommen werden, dass Zielorientierungen auch die Verarbeitung und Nutzung und in der Folge die Wirksamkeit von Feedback beeinflussen. In Kapitel 2.3.6 wurde theoretisch hergeleitet, wie die Feedbackverarbeitung und -nutzung in Abhängigkeit von der *LZO*, *ALZO* und *VLZO* von Lehrkräften variieren könnte. In den Teilstudien 1 und 3 wurden die differenziellen Effekte der Zielorientierungen auf die Kompetenzentwicklung der Lehrkräfte und die Unterrichtsqualität empirisch überprüft.

Nachfolgend werden zunächst die Ergebnisse der empirischen Studien zusammenfassend dargestellt und diskutiert (Kapitel 6.1). Anschließend werden Limitationen der vorliegenden Arbeit benannt (Kapitel 6.2) und Implikationen für die schulische Praxis aufgezeigt (Kapitel 6.3). Im Ausblick (Kapitel 6.4) wird auf den weiteren Forschungsbedarf verwiesen, bevor in Kapitel 6.5 ein abschließendes Fazit gezogen wird.

## **6.1 Zusammenfassung und Diskussion der Ergebnisse**

Die Ergebnisse der Teilstudien wurden in den jeweiligen Abschnitten bereits im Einzelnen diskutiert. Aus diesem Grund liegt der Fokus im Folgenden auf einer integrativen Betrachtung der zentralen Befunde, die entlang der in Kapitel 2.4.2 formulierten Forschungsfragen erfolgt.

### *Forschungsfrage 1: Qualitätseinschätzungen der Lehrkräfte*

Weil Lehrkräfte die Qualität von Unterrichtsfeedbacks durch Schulleitungen häufig bemängeln (vgl. Kimball, 2003; Kimball & Milanowski, 2009; Kraft & Gilmour, 2016; Liu

et al., 2019; Weisberg et al., 2009), lag ein Schwerpunkt der Dissertation auf der Entwicklung eines Feedbackverfahrens, das Schulleitungen dabei unterstützen soll, ihren Lehrkräften ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben. Die Entwicklung des Feedbackverfahrens erfolgte auf Grundlage der in Kapitel 2.3.2 beschriebenen formalen Gestaltungsmerkmale, die ausgehend von Theorien und Befunden der arbeits- und instruktionspsychologischen Feedbackforschung erarbeitet wurden.

Die erste Forschungsfrage bezog sich auf die Qualität des beobachtungsbasierten Unterrichtsfeedbacks. Es zeigte sich, dass Lehrkräfte die Feedbackqualität überwiegend als sehr hoch bewerten und die Qualitätseinschätzungen signifikant und inhaltlich bedeutsam vom theoretischen Mittel abweichen. Weiterführende Analysen ergaben, dass einzelne der über einen Implementationscheck erfassten formalen Gestaltungsmerkmale des Feedbacks zudem positive Zusammenhänge mit den Qualitätseinschätzungen der Lehrkräfte aufweisen. Als bedeutsam für die Qualitätseinschätzungen erwiesen sich insbesondere die im Feedback enthaltenen Verhaltensempfehlungen sowie Zielvereinbarungen zu deren Umsetzung, auf die sich die Schulleitungen und Lehrkräfte am Ende der Feedbackgespräche einigten. Diese Befunde können als erster Hinweis darauf interpretiert werden, dass die in der arbeits- und instruktionspsychologischen Feedbackforschung identifizierten Gestaltungsmerkmale, die im Feedbackverfahren umgesetzt wurden, auch für ein beobachtungsbasiertes Unterrichtsfeedback von Schulleitungen an Lehrkräfte von Relevanz sind.

Ein Ziel des Promotionsvorhabens wurde somit erreicht: Das entwickelte Feedbackverfahren unterstützt Schulleitungen dabei, Lehrkräften ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben. Eine positive Bewertung der Feedbackqualität durch die Lehrkräfte – hierauf deuten Befunde aus der empirischen Schul- und Unterrichtsforschung hin (vgl. Cherasaro et al., 2016; Delvaux et al., 2013; Kraft & Christian, 2022; Tuytens & Devos, 2011, 2014) – stellt eine grundlegende Voraussetzung für die Nutzung der im Feedback enthaltenen Informationen und dessen Wirksamkeit dar. Inwiefern die formalen Gestaltungsmerkmale die Qualitätseinschätzungen der Lehrkräfte beeinflussen und ob die hohe Feedbackqualität ihrerseits die Wirksamkeit des beobachtungsbasierten Unterrichtsfeedbacks begünstigt, konnte aus methodischen Gründen zwar nicht untersucht werden (siehe auch den Abschnitt *Modellierung von Wirkmechanismen* in Kapitel 6.2), ob das beobachtungsbasierte Unterrichtsfeedback aber grundsätzlich wirksam war, war Gegenstand der Forschungsfragen 2 und 3.

*Forschungsfragen 2 und 3: Haupteffekte der Feedbackintervention auf die Unterrichtskompetenz und Unterrichtsqualität*

Die Befunde zu den Haupteffekten der Feedbackintervention sind inkonsistent. Während das Unterrichtsfeedback keinen Effekt auf die von Lehrkräften selbst eingeschätzte Weiterentwicklung ihrer Unterrichtskompetenz hatte, führte es bei drei der 14 untersuchten Unterrichtsmerkmale zu einer besseren Unterrichtsqualität. Schülerinnen und Schüler von Lehrkräften der Interventionsgruppe schätzten die Veränderung der Unterrichtsqualität im Hinblick auf die Unterrichtsmerkmale *Lerndienliches Unterrichtsklima*, *Gruppenaktivierung* und *Individualisierung* nach dem Feedback signifikant besser ein als Schülerinnen und Schüler in der Kontrollgruppe.

Dieses Befundmuster ist erklärungsbedürftig – nicht nur, weil es keine eindeutigen Rückschlüsse auf die Wirksamkeit der Intervention zulässt, sondern auch, weil die Weiterentwicklung der Unterrichtskompetenz und ein daraus resultierendes verändertes Unterrichtsverhalten eine Voraussetzung für die Verbesserung von Unterricht darstellt. Hierauf lassen auch Befunde schließen, nach denen zwischen dem professionellen Wissen und Können von Lehrkräften und der Qualität der von ihnen angebotenen Lerngelegenheiten grundsätzlich ein positiver Zusammenhang besteht (vgl. Baumert et al., 2010; Blömeke et al., 2022; König & Pflanzl, 2016; König et al., 2021; Kunter et al., 2013; Voss et al., 2014). Nachfolgend sollen zwei potenzielle Erklärungen für die widersprüchlichen Befunde thematisiert werden. Zum einen wird unter Rückgriff auf Studienbefunde zur Validität von (retrospektiven) Kompetenzselbsteinschätzungen durch Lehrkräfte und zur Validität von Schülerinnen- und Schülerurteilen zur Unterrichtsqualität diskutiert, bei welchem der beiden Kriterien es sich um das verlässlichere Maß zur Beurteilung der Interventionswirksamkeit handelt. Zum anderen wird die Passung zwischen den Inhalten des Unterrichtsfeedbacks auf der einen und den beiden abhängigen Variablen auf der anderen Seite thematisiert.

Hinsichtlich der Validität von Kompetenzselbsteinschätzungen durch Lehrkräfte ist der Forschungsstand uneindeutig. Wie in der Diskussion von Teilstudie 1 (siehe Kapitel 3.6) aufgezeigt, berichten mehrere Studien zwar von nur schwach positiven Zusammenhängen zwischen den Kompetenzselbsteinschätzungen von angehenden Lehrkräften und ihrem Abschneiden in Wissens- bzw. Performanztests (vgl. König et al., 2012; Kunter & Klusmann, 2010; Mertens & Gräsel, 2018; Seifert et al., 2018), diese Zusammenhänge fallen in zwei Studien jedoch größer aus, wenn angehende Lehrkräfte über Unterrichts-

erfahrungen verfügen (vgl. Mertens & Gräsel, 2018; Seifert et al., 2018). Vor diesem Hintergrund kann geschlussfolgert werden, dass die Lehrkräfte aus der Interventionsstudie, die überwiegend über eine langjährige Berufserfahrung verfügen, ihren Kompetenzstand valide beurteilen können. Inwiefern diese Schlussfolgerung jedoch auch auf eine retrospektive Selbsteinschätzung der eigenen Kompetenzentwicklung zutrifft, ist noch weitgehend ungeklärt.

Die Validität von retrospektiven Kompetenzselbsteinschätzungen durch Lehrkräfte wurde nach meinem Kenntnisstand bislang nur in einer Studie untersucht (vgl. Copur-Gencturk & Thacker, 2021), in der die selbst eingeschätzte Kompetenzentwicklung von Lehrkräften nach Teilnahme an einer Fortbildung mit ihrem Wissenszuwachs in einem standardisierten Test verglichen wurde. Obwohl die beiden Erhebungsinstrumente parallelisiert wurden und sich auf dieselben Wissensdomänen bezogen, korrelierten die selbst eingeschätzte Kompetenzentwicklung und die Testergebnisse der Lehrkräfte nicht miteinander (vgl. Copur-Gencturk & Thacker, 2021). Dieser Befund kann als erster Hinweis darauf interpretiert werden, dass es Lehrkräften – obwohl Sie ihren Kompetenzstand valide beurteilen können – schwerfällt, ihre Kompetenzentwicklung rückblickend valide zu beurteilen. Hierauf lassen auch die Befunde einer Meta-Analyse aus der Arbeitspsychologie schließen (vgl. Sitzmann et al., 2010), in der 166 Primärstudien berücksichtigt wurden, die die Wirksamkeit verschiedener universitärer Lernangebote sowie beruflicher Fortbildungsmaßnahmen evaluiert haben. Die Autorinnen und Autoren berichten im Mittel zwar von moderaten positiven Zusammenhängen zwischen dem selbst eingeschätzten Wissensstand von Studierenden bzw. Arbeitnehmerinnen und -nehmern und ihrem Abschneiden in Wissenstests, zwischen der selbst eingeschätzten Wissensentwicklung und objektiven Maßen des Wissenszuwachses bestand jedoch kein Zusammenhang (vgl. Sitzmann et al., 2010).

Zusammenfassend kann geschlussfolgert werden, dass es grundsätzlich Hinweise auf die Validität von Kompetenzselbsteinschätzungen durch Lehrkräfte gibt. Gleichzeitig deuten die zuletzt zitierten Studien jedoch darauf hin, dass die von Lehrkräften selbst eingeschätzte Weiterentwicklung ihrer Unterrichtskompetenz ein nur unzureichendes Maß für tatsächliche Veränderungen ihres professionellen Wissens und Könnens darstellen könnte. Wie in der Diskussion von Teilstudie 1 bereits angesprochen, erscheint es z. B. denkbar, dass die Kompetenzselbsteinschätzungen von selbstwertdienlichen Verzerrungen oder Verzerrungen nach sozialer Erwünschtheit betroffen sind. Eine Verzer-

rung der Schülerinnen- und Schülerurteile zur Unterrichtsqualität ist demgegenüber unwahrscheinlich, weil diesen zum einen nicht bekannt war, ob ihre Lehrkräfte ein Unterrichtsfeedback erhalten haben oder nicht. Zum anderen zeigen Studienbefunde, dass die Einschätzungen von Schülerinnen und Schülern zur Unterrichtsqualität grundsätzlich schwache bis moderate Zusammenhänge mit den Urteilen externer Beobachterinnen und Beobachter aufweisen (vgl. de Jong & Westerhof, 2001; Fauth et al., 2014a) und prädiktiv für ihren Lernerfolg sind (vgl. de Jong & Westerhof, 2001; Fauth et al., 2014a, 2014b; Kunter & Baumert, 2006; Wagner et al., 2016). Dementsprechend werden Schülerinnen- und Schülerurteile in der empirischen Schul- und Unterrichtsforschung als valides Maß zur Erfassung von Unterrichtsqualität angesehen (vgl. Göllner et al., 2016).

Neben einer möglichen Verzerrung der retrospektiven Kompetenzselbsteinschätzung kommt als zweiter Erklärungsansatz für die widersprüchlichen Befunde die Passung zwischen den Feedbackinhalten und den abhängigen Variablen in Frage: Während die Weiterentwicklung der Unterrichtskompetenz mit Items wie „Ich habe meine Kompetenz als Lehrkraft erweitert.“ oder „Ich habe viel dazugelernt, was die Verbesserung meines Unterrichts betrifft.“ global gemessen wurde, bezogen sich die Schülerinnen- und Schülerurteile zur Unterrichtsqualität explizit auf die im Feedback thematisierten 14 Merkmale eines lernwirksamen Unterrichts. Die globale Erfassung der Kompetenzentwicklung könnte einen Nachweis der Interventionswirksamkeit daher erschwert haben. Bei einer größeren Kongruenz mit den Inhalten des Unterrichtsfeedbacks wäre es ggf. möglich gewesen, neben Haupteffekten auf die Unterrichtsqualität auch Haupteffekte auf die selbst eingeschätzte Kompetenzentwicklung festzustellen.

In der Zusammenschau kann geschlussfolgert werden, dass die Urteile der Schülerinnen und Schüler zur Unterrichtsqualität aufgrund ihrer konvergenten und prädiktiven Validität ein verlässlicheres Maß für die Beurteilung der Interventionswirksamkeit darstellen als die retrospektiven Kompetenzselbsteinschätzungen der Lehrkräfte. Zusätzlich erschwerte die globale Erfassung der Kompetenzentwicklung das Auffinden eines signifikanten Feedbackeffekts. Wie erwähnt, führte die Feedbackintervention jedoch nur bei drei Unterrichtsmerkmalen zu einer besseren Unterrichtsqualität. Folglich stellt sich auch die Frage, warum sich der Unterricht nur im Hinblick auf bestimmte Unterrichtsmerkmale verbesserte, die Intervention bei anderen Merkmalen hingegen wirkungslos blieb.

Die plausibelste Erklärung hierfür wurde bereits in der Diskussion zu Teilstudie 2 angeführt (siehe Kapitel 4.6): Weil die Lehrkräfte der Interventionsgruppe nur zu ausgewählten Unterrichtsmerkmalen ein Feedback erhalten haben, waren Feedbackeffekte vor

allem bei Merkmalen zu erwarten, zu denen viele Lehrkräfte ein Feedback bekamen. Dieser Erklärungsansatz steht im Einklang damit, dass die Intervention einen Effekt auf die Unterrichtsmerkmale *Lernförderliches Unterrichtsklima* und *Individualisierung* hatte, die von allen 14 Unterrichtsmerkmalen in den Feedbackgesprächen am häufigsten thematisiert wurden. Der Feedbackeffekt beim Unterrichtsmerkmal *Gruppenaktivierung* lässt sich mit der Schwerpunktsetzung des Feedbacks jedoch nicht überzeugend erklären. Zur *Gruppenaktivierung* haben nur wenige Lehrkräfte ein Feedback erhalten, was die Frage aufwirft, ob die Verbesserung der Unterrichtsqualität bei diesem Unterrichtsmerkmal kausal auf die Intervention zurückgeführt werden kann. Zu berücksichtigen ist in diesem Zusammenhang jedoch, dass die Feedbackinhalte nur auf Grundlage der Beobachtungsdaten aus dem Feedbackportal bestimmt werden konnten, von denen auf die Unterrichtsmerkmale aus den Feedbackberichten und damit auf die inhaltliche Schwerpunktsetzung des Feedbacks geschlossen wurde. Darüber, auf welche Merkmale sich das Feedback tatsächlich bezog, liegen keine Informationen vor. Letztlich wurde den Schulleitungen bei der inhaltlichen Gestaltung des Feedbacks ein gewisser Handlungsspielraum gelassen, sodass sie auch *Gruppenaktivierung* thematisiert haben könnten, obwohl sie in den Feedbackberichten selten aufgeführt wurde.<sup>21</sup>

Mit Blick auf die *Forschungsfragen 2 und 3* lässt sich abschließend bilanzieren, dass die Feedbackintervention wirksam war. Auch wenn Effekte auf die Kompetenzentwicklung der Lehrkräfte aus den genannten Gründen nicht nachgewiesen werden konnten, hatte die Intervention bei einzelnen Unterrichtsmerkmalen einen positiven Effekt auf die Unterrichtsqualität. Worauf die Effekte im Einzelnen zurückzuführen sind, lässt sich anhand der vorliegenden Daten zwar nicht beurteilen, die Interventionswirksamkeit verweist grundsätzlich aber darauf, dass die im theoretischen Feedbackmodell aus Kapitel 2.3 identifizierten Gelingensbedingungen erfolgreich umgesetzt wurden: So lassen die positiven Effekte darauf schließen, dass die mit dem BERU durchgeführten Beobachtungen von zwei Unterrichtsstunden je Lehrkraft zu einer objektiven, reliablen und validen

---

<sup>21</sup> Grundsätzlich zeigten sich aber auch bei weiteren Unterrichtsmerkmalen, die in den Feedbackberichten vieler Lehrkräfte aufgeführt waren, inhaltlich bedeutsame Gruppenunterschiede, was darauf schließen lässt, dass sich die Schulleitungen bei der Feedbackgabe überwiegend an den Feedbackberichten orientiert haben. Beispielsweise wurden auch die Unterrichtsmerkmale *Unterstützung von Verstehensprozessen*, *Förderung der Selbstwirksamkeit* und *Anwendung und Transfer* häufig in den Feedbackberichten thematisiert und die Lehrkräfte der Interventionsgruppe wiesen nach dem Feedback eine höhere Unterrichtsqualität auf als Lehrkräfte der Kontrollgruppe.

Beurteilung von Unterrichtsqualität geführt haben, wozu auch das Beobachtungstraining der Schulleitungen beigetragen haben dürfte.<sup>22</sup> Außerdem können die Feedbackeffekte und die hohen Qualitätseinschätzungen der Lehrkräfte als Hinweis darauf interpretiert werden, dass sich neben dem BERU und dem Beobachtungstraining auch das im Rahmen dieser Arbeit entwickelte Feedbackverfahren sowie das hiermit korrespondierende Feedbacktraining bewährt haben.

*Forschungsfragen 4 und 5: Differenzielle Effekte der Zielorientierungen auf die Unterrichtskompetenz und Unterrichtsqualität*

Weil die Feedbackwirksamkeit auch von Merkmalen der Lehrkräfte abhängt, wurden neben den Haupteffekten der Feedbackintervention auch die differenziellen Effekte von Zielorientierungen untersucht, von denen angenommen werden kann, dass sie die Feedbackverarbeitung und -nutzung durch Lehrkräfte beeinflussen. Während für die VLZO keine differenziellen Effekte gefunden wurden, moderierten sowohl die LZO als auch die ALZO die Feedbackeffekte auf die Kompetenzentwicklung der Lehrkräfte und die Veränderung der Unterrichtsqualität.

Konkret zeigte sich, dass sich eine LZO günstig auf die Feedbackwirksamkeit auswirkt. Lernzielorientierte Lehrkräfte schätzten ihre Kompetenzentwicklung nach dem Feedback positiver ein als Lehrkräfte mit gering ausgeprägter LZO. Während das Feedback bei erstgenannten einen positiven Effekt auf ihre selbst eingeschätzte Kompetenzentwicklung hatte, blieb es bei letztgenannten wirkungslos. Einschränkend ist in diesem Zusammenhang jedoch zu erwähnen, dass Verzerrungen bei der Kompetenzselbsteinschätzung – wie zuvor erwähnt – nicht ausgeschlossen werden können. So könnten lernzielorientierte Lehrkräfte ihre Kompetenzentwicklung beispielsweise allein deshalb positiver eingeschätzt haben, weil sie ein Feedback zu ihrem Unterricht erhalten haben.

---

<sup>22</sup> Auch wenn die Interventionswirksamkeit diesen Schluss nahelegt, steht eine empirische Prüfung der Reliabilität und Validität von Unterrichtsbeobachtungen, die mit dem BERU erfolgen, noch weitgehend aus. In einer ersten Validierungsstudie (vgl. Gärtner et al., 2021) konnte die theoretisch angenommene Faktorenstruktur zwar empirisch abgebildet werden und es liegen erste Hinweise auf zeitliche Stabilität der Unterrichtsbeurteilungen vor (Test-Retest-Reliabilität), weitere wichtige Validierungsschritte müssen aber noch erfolgen. Ob z. B. mehre Beobachterinnen und Beobachter zu vergleichbaren Urteilen kommen, wenn beide dieselbe Unterrichtssequenz mit dem BERU beurteilen, ist nicht bekannt (Inter-Rater-Reliabilität). Außerdem liegen keine Informationen darüber vor, inwiefern eine Beurteilung mit dem BERU z. B. mit den Einschätzungen von Schülerinnen und Schülern oder mit Beurteilungen, die mit anderen Beobachtungsinstrumenten erfolgen, korrespondiert (konvergente Validität). Auch darüber, ob eine positive Unterrichtsbeurteilung mit Lernzuwächsen von Schülerinnen und Schülern einhergeht, liegen keine Informationen vor (Kriteriumsvalidität).

Grundsätzlich steht der positive Moderationseffekt auf die Kompetenzentwicklung aber im Einklang mit den differenziellen Effekten der *LZO* auf die Verbesserung der Unterrichtsqualität. Es zeigten sich zwar nur bei zwei Unterrichtsmerkmalen signifikante Moderationseffekte; diese korrespondieren aber mit den differenziellen Effekten auf die Kompetenzentwicklung. Lehrkräfte mit hoch ausgeprägter *LZO* wiesen im Hinblick auf die Unterrichtsmerkmale *Verdeutlichung des Nutzens* und *Lerndienliches Unterrichtsklima* nach dem Feedback eine bessere Unterrichtsqualität auf als Lehrkräfte mit gering ausgeprägter *LZO*. De facto führte das Feedback bei diesen beiden Unterrichtsmerkmalen nur bei lernzielorientierten Lehrkräften zu einer besseren Unterrichtsqualität und blieb bei allen anderen Lehrkräften wirkungslos.<sup>23</sup> Die differenziellen Effekte der *LZO* können somit als Hinweis darauf interpretiert werden, dass die theoretischen Überlegungen zum Einfluss der *LZO* auf die Feedbackverarbeitung und -nutzung aus Kapitel 2.3.6 zutreffend sein könnten. Empirisch prüfen lässt sich dies jedoch nicht, da die involvierten emotionalen und kognitiven Verarbeitungsprozesse und die Feedbacknutzung im Rahmen der Interventionsstudie nicht vollständig erfasst werden konnten (siehe auch den Abschnitt *Modellierung von Wirkmechanismen* in Kapitel 6.2).

Während eine hoch ausgeprägte *LZO* die Feedbackwirksamkeit begünstigte, sind die Befunde zu den differenziellen Effekten der *ALZO* inkonsistent. Einerseits schätzten Lehrkräfte mit hoch ausgeprägter *ALZO* ihre Kompetenzentwicklung nach dem Feedback positiver ein als Lehrkräfte mit niedrig ausgeprägter *ALZO* – wobei Verzerrungen auch bei annäherungsleistungszielorientierten Lehrkräften nicht ausgeschlossen werden können. Andererseits konnte mit Blick auf die Verbesserung der Unterrichtsqualität nur bei einem Unterrichtsmerkmal ein positiver Moderationseffekt gefunden werden: Bei der *Unterstützung von Verstehensprozessen* führte das Feedback bei Lehrkräften mit hoch ausgeprägter *ALZO* zu einer besseren Unterrichtsqualität und blieb bei Lehrkräften mit gering ausgeprägter *ALZO* wirkungslos. Bei drei Unterrichtsmerkmalen wies die *ALZO* demgegenüber einen negativen Interaktionszusammenhang mit dem Unterrichtsfeedback auf. Hinsichtlich der Unterrichtsmerkmale *Steuerung des Unterrichtsflusses*, *Gruppenaktivierung* und *Umgang mit Störungen* profitierten Lehrkräfte mit niedrig ausgeprägter

---

<sup>23</sup> Wie bei den Haupteffekten ist auch bei den differenziellen Effekten zu berücksichtigen, dass Moderationseffekte vor allem bei Unterrichtsmerkmalen zu erwarten waren, zu denen viele Lehrkräfte ein Feedback erhalten haben. Sowohl *Verdeutlichung des Nutzens* als auch *Lerndienliches Unterrichtsklima* wurden – darauf lassen die Beobachtungsdaten aus dem Feedbackportal schließen – von den Schulleitungen häufig in den Feedbackgesprächen thematisiert.

*ALZO* stärker von der Feedbackintervention als Lehrkräfte mit hoch ausgeprägter *ALZO*.<sup>24</sup>

Die inkonsistenten Befunde lassen damit – ebenso wie die theoretischen Überlegungen zu ihrem Einfluss auf die Feedbackverarbeitung und -nutzung aus Kapitel 2.3.6 – keine eindeutigen Rückschlüsse darauf zu, wie sich eine *ALZO* auf die Feedbackwirksamkeit auswirkt. Damit knüpfen sie an eine Kontroverse aus der allgemeinen Zielorientierungsforschung an, in der die Frage, ob eine *ALZO* lern- und leistungsförderlich ist oder nicht, intensiv diskutiert wird. Während einige Autorinnen und Autoren annehmen, nur eine *LZO* gehe mit lern- und leistungsförderlichen Erlebens-, Kognitions- und Verhaltensmustern einher (die sogenannte *mastery goal perspective*; vgl. z. B. Kaplan & Maehr, 2007; Midgley et al., 2001), gehen andere Autorinnen und Autoren davon aus, dass sich neben der *LZO* auch eine *ALZO* günstig auf Lernen und Leistung auswirken kann (die sogenannte *multiple goals perspective*; vgl. z. B. Harackiewicz et al., 2002). Auch in mehreren Meta-Analysen finden sich für annäherungsleistungszielorientierte Personen sowohl lern- und leistungsförderliche als auch lern- und leistungsmindernde Erlebens-, Kognitions- und Verhaltensmuster (vgl. Cellar et al., 2011; Huang, 2011, 2012; Payne et al., 2007; Wirthwein et al., 2013).

Zwei Meta-Analysen, die die beschriebene Kontroverse explizit adressieren, zeigen jedoch, dass die inkonsistenten Befunde auf unterschiedliche Konzeptualisierungen der *ALZO* zurückgeführt werden können (vgl. Hulleman et al., 2010; Senko & Dawson, 2017): Während der Fokus teilweise auf der Demonstration von Kompetenz liegt, wird die *ALZO* teilweise als Streben nach der Demonstration von *überlegenen* Kompetenzen definiert. Für die erstgenannte Konzeptualisierung zeigten sich überwiegend lern- und leistungsmindernde Erlebens-, Kognitions-, und Verhaltensmuster. Die letztgenannte Konzeptualisierung ging demgegenüber überwiegend mit lern- und leistungsförderlichen Erlebens-, Kognitions- und Verhaltensmustern einher und wies positive Zusammenhänge mit Lernen und Leistung auf (vgl. Hulleman et al., 2010; Senko & Dawson, 2017). Begründet werden diese Befunde unter anderem mit einem hohen Leistungsstandard und

---

<sup>24</sup> Wie im Zusammenhang mit den Haupteffekten auf die Unterrichtsqualität angemerkt, haben nur wenige Lehrkräfte ein Feedback zum Unterrichtsmerkmal *Gruppenaktivierung* erhalten. Die Unterrichtsmerkmale *Unterstützung von Verstehensprozessen*, *Steuerung des Unterrichtsflusses* und *Umgang mit Störungen* wurden demgegenüber häufig in den Feedbackgesprächen thematisiert. Weil der Moderationseffekt bei *Gruppenaktivierung* nicht kausal auf die Intervention zurückzuführen sein könnte, wird er im Folgenden nicht weiter interpretiert.

einer damit einhergehenden hohen Leistungsmotivation von annäherungsleistungszielorientierten Personen, die danach streben, anderen überlegen sein zu wollen (vgl. Senko et al., 2011).

Auch in der vorliegenden Arbeit wurde die *ALZO* als Streben definiert, *überlegene* Kompetenzen demonstrieren zu wollen. Sie wurde über Items wie „In meinem Beruf strebe ich danach, zu beweisen, dass ich mehr weiß als andere Lehrkräfte.“ oder „In meinem Beruf strebe ich danach, zu zeigen, dass ich besser unterrichte als andere Lehrkräfte.“ operationalisiert (vgl. Nitsche et al., 2011; Janke et al., 2019). Die positiven Moderationseffekte der *ALZO* auf die Kompetenzentwicklung und die Unterrichtsqualität bei der *Unterstützung von Verstehensprozessen* können daher mit einem hohen Leistungsstandard und einer hohen Leistungsmotivation von annäherungsleistungszielorientierten Lehrkräften erklärt werden. Demnach könnten Lehrkräfte mit hoch ausgeprägter *ALZO* das Feedback genutzt haben, um auch in Zukunft *überlegene* Kompetenzen demonstrieren zu können. Die negativen Moderationseffekte der *ALZO* lassen sich mit ihrer Konzeptualisierung jedoch nicht überzeugend erklären; im Gegenteil scheinen sie den bisherigen Überlegungen zu widersprechen. Warum haben Lehrkräfte, die danach streben, *überlegene* Kompetenz zu demonstrieren und entsprechend einen hohen Leistungsstandard verfolgen und eine hohe Leistungsmotivation aufweisen, bei den beiden Unterrichtsmerkmalen *Steuerung des Unterrichtsflusses* und *Umgang mit Störungen* in einem geringeren Maße vom Unterrichtsfeedback profitiert als Lehrkräfte, die dieses Streben nicht antreibt? Wie in der Diskussion zu Teilstudie 3 (siehe Kapitel 5.4.2) erläutert, stellt die Feedbackvalenz einen möglichen Erklärungsansatz dar. Weil das Feedbackverfahren sowohl ein bestätigendes als auch ein kritisches Feedback zu ausgewählten Unterrichtsmerkmalen vorsah, wurden einige Unterrichtsmerkmale überwiegend als bestätigendes Feedback in den Feedbackgesprächen thematisiert und andere Unterrichtsmerkmale überwiegend als kritisches Feedback.

Die Beobachtungsdaten aus dem Feedbackportal deuten darauf hin, dass viele Lehrkräfte im Hinblick auf das Unterrichtsmerkmal *Unterstützung von Verstehensprozessen*, bei dem sich ein positiver Moderationseffekt zeigte, ein kritisches Feedback erhalten haben. Zu den Unterrichtsmerkmalen *Steuerung des Unterrichtsflusses* und *Umgang mit Störungen*, bei denen sich negative Moderationseffekte zeigten, haben viele Lehrkräfte demgegenüber ein bestätigendes Feedback erhalten. Vor diesem Hintergrund kann vermutet werden, dass Lehrkräfte mit einer hoch ausgeprägten *ALZO* ihren Unterricht nur im Falle eines kritischen Unterrichtsfeedbacks weiterentwickeln könnten, weil sie ihrem

Streben, *überlegene* Kompetenzen zu demonstrieren, nicht gerecht geworden sind. Annäherungsleistungszielorientierte Lehrkräfte, die ein bestätigendes Feedback erhalten, könnten demgegenüber keine Notwendigkeit sehen, ihren Unterricht weiterzuentwickeln, weil sie das bestätigende Feedback bereits als Hinweis darauf interpretieren, anderen Lehrkräften überlegen zu sein. Empirisch prüfen lässt sich der Einfluss der Feedbackvalenz auf die Feedbackverarbeitung und -nutzung durch annäherungsleistungszielorientierte Lehrkräfte jedoch nicht, weil hierzu – wie zuvor erwähnt – keine Informationen vorliegen.

In der Zusammenschau lässt sich mit Blick auf die *Forschungsfragen 4 und 5* bilanzieren, dass eine *LZO* die Feedbackwirksamkeit begünstigt. Die Befunde zu den differenziellen Effekten der *ALZO* sind demgegenüber inkonsistent. Einerseits liegen Hinweise darauf vor, dass auch eine hoch ausgeprägte *ALZO* mit größeren Feedbackeffekten einhergeht, andererseits deuten die negativen Moderationseffekte darauf hin, dass annäherungsleistungszielorientierte Lehrkräfte in einem geringeren Maße von einem Unterrichtsfeedback profitieren könnten als andere Lehrkräfte. Grundsätzlich zeigen die differenziellen Effekte der Zielorientierungen aber, dass beobachtungsbasierte Unterrichtsfeedbacks nicht bei allen Lehrkräften gleichermaßen wirksam sind, was die Annahmen aus dem theoretischen Feedbackmodell bestätigt. Hieraus lässt sich zum einen folgern, dass die differenziellen Effekte von Zielorientierungen sowie weiterer Lehrkraftmerkmale auch in zukünftigen Studien untersucht werden sollten. Zum anderen lässt sich konstatieren, dass Lehrkraftmerkmale bei der Feedbackgabe in Zukunft stärker berücksichtigt werden sollten. Beide Aspekte werden im Zusammenhang mit den praktischen Implikationen (Kapitel 6.3) bzw. im Ausblick (Kapitel 6.4) erneut aufgegriffen, vorher wird jedoch auf die Limitationen der vorliegenden Arbeit eingegangen.

## 6.2 Limitationen

Die Dissertation und die durchgeführte Interventionsstudie weisen verschiedene inhaltliche und methodische Limitationen auf, die bei der Interpretation der Befunde zu berücksichtigen sind. Einige dieser Einschränkungen wurden in den Teilstudien bereits angesprochen, andere klangen in der Gesamtdiskussion der Befunde an. Im Folgenden sollen sie zusammenfassend diskutiert werden. Zunächst werden Einschränkungen aufgeführt, die sich aus dem Studiendesign sowie der Stichprobe ergeben. Anschließend werden die

im theoretischen Feedbackmodell beschriebenen Wirkmechanismen aufgegriffen, die aus methodischen bzw. erhebungsökonomischen Gründen nicht untersucht werden konnten.

**Studiendesign.** Im Hinblick auf das Studiendesign ist zum einen zu konstatieren, dass die interne Validität der Interventionsstudie aufgrund ihres quasi-experimentellen Designs und der damit einhergehenden nicht-randomisierten Zuweisung der Lehrkräfte und Schülerinnen und Schüler zur Interventions- und Kontrollgruppe eingeschränkt ist (vgl. Döring & Bortz, 2016). Die Entscheidung, welche Lehrkräfte und welche Schülerinnen und Schüler an der Interventionsstudie teilnehmen und welcher Gruppe sie angehören würden, wurde schulintern getroffen. Aus diesem Grund kann nicht ausgeschlossen werden, dass sich die Interventions- und Kontrollgruppe im Hinblick auf potenzielle Störvariablen systematisch voneinander unterscheiden. Durch die Erfassung diverser Kovariaten und ihrer statistischen Kontrolle in den Wirksamkeitsanalysen wurde ein Teil dieser Störvariablen zwar konstant gehalten, um den Feedbackeffekt zu isolieren. Fraglich ist jedoch, ob alle personenbezogenen Störvariablen auf Seiten der Lehrkräfte und der Schülerinnen und Schüler berücksichtigt wurden. Zukünftige Studien, die die Wirksamkeit dieser oder vergleichbarer Interventionen untersuchen, sollten daher nach Möglichkeit ein klassisches experimentelles Design mit Randomisierung umsetzen, welches auch in der pädagogischen Interventionsforschung aufgrund seiner hohen internen Validität als Königsweg gilt, um Ursache-Wirkungs-Annahmen zu überprüfen (vgl. Leutner, 2010; Köller, 2020).

Eine weitere Limitation, die im Zusammenhang mit dem Studiendesign zu diskutieren ist, betrifft die Messung der Unterrichtsqualität zu nur zwei Messzeitpunkten. Die Unterrichtsqualität wurde am Anfang des Schuljahres 2017/2018 (Prä-Befragung) und am Ende desselben Schuljahres (Post-Befragung) erfasst. Eine Follow-Up-Befragung am Ende des Schuljahres 2018/2019 wurde zwar durchgeführt, die erhobenen Daten konnten aufgrund einer großen Stichprobenmortalität im zweiten Studienjahr jedoch nicht für weitere Wirksamkeitsanalysen genutzt werden. Dabei wären Daten aus einer Follow-Up-Befragung aus mindestens zwei Gründen von Interesse: Zum einen ließe sich prüfen, ob das Unterrichtsfeedback auch zu einer langfristigen Verbesserung der Unterrichtsqualität führt. Zum anderen deuten Studienbefunde darauf hin, dass beobachtungsbasierte Unterrichtsfeedbacks teilweise erst im Jahr nach der Intervention wirksam werden (vgl. Allen et al., 2011; Taylor & Tyler, 2012), sodass sich bei einzelnen Unterrichtsmerkmalen ggf. erst zu einem späteren Zeitpunkt eine Verbesserung eingestellt hätte.

**Stichprobe.** Hinsichtlich der Stichprobe ist zum einen die *Schulstichprobe* kritisch zu diskutieren. Bei allen Schulen der Interventionsstudie handelt es sich um Berliner OSZ, die im Vergleich zu allgemeinbildenden Schulen eine Reihe von Besonderheiten aufweisen (z. B. wechseln sich in der klassischen dualen Berufsausbildung schulische Phasen mit Praxisphasen in den Ausbildungsbetrieben ab und das Fächerprinzip wird teilweise zugunsten des sogenannten Lernfeldunterrichts aufgeben; vgl. Seifried & Weyland, 2021). Vor diesem Hintergrund kann angenommen werden, dass sich auch die Arbeitsbedingungen von Lehrkräften und Schulleitungen an Berufsschulen von den Arbeitsbedingungen an allgemeinbildenden Schulen unterscheiden, was die Frage aufwirft, inwiefern die Befunde zur Interventionswirksamkeit auf letztgenannte übertragbar sind. Replikationsstudien sollten die Wirksamkeit der Feedbackintervention aus diesem Grund nach Möglichkeit an einer Stichprobe aus dem allgemeinbildenden Schulsystem überprüfen.

Das Sampling der Lehrkräfte und der Schülerinnen und Schüler wurde im Abschnitt *Studiendesign* bereits angesprochen und soll an dieser Stelle nicht erneut thematisiert werden. Stattdessen liegt der Fokus auf der *Schulleitungsstichprobe*, bei der es sich ebenfalls um eine Gelegenheitsstichprobe handelt, welcher eine Selbstselektion zugrunde liegt (vgl. Döring & Bortz, 2016). Zwar wurden die Schulleitungen aller Berliner OSZ auf einer turnusmäßig stattfindenden Schulleitungskonferenz über die Interventionsstudie informiert, sie konnten im Anschluss jedoch selbst über ihre Teilnahme entscheiden. Daher ist zu vermuten, dass vor allem Schulleitungen teilgenommen haben, die Unterrichtsbeobachtungen und -feedbacks grundsätzlich aufgeschlossen gegenüberstehen. Diese Positivselektion könnte die Wirksamkeit der Feedbackintervention begünstigt haben. Inwiefern die Intervention auch dann wirksam ist, wenn sie von Schulleitungen implementiert wird, die einer unterrichtsbezogenen Führung im Allgemeinen und beobachtungsbasierendem Unterrichtsfeedback im Besonderen weniger aufgeschlossen gegenüberstehen, ist eine offene Frage.

Eine weitere Limitation ergibt sich aus der *Stichprobengröße*, die mit Einschränkungen im Hinblick auf die statistischen Analysen verbunden ist. Dies betrifft zum einen die Schulstichprobe, die es nicht erlaubt hat, der hierarchischen Datenstruktur vollumfänglich Rechnung zu tragen. Zwar wurde in den Teilstudien 2 und 3 berücksichtigt, dass Schülerinnen und Schüler in Lehrkräften bzw. Klassen genestet sind, dass aber auch Lehrkräfte in Schulen genestet sind, konnte bei nur  $N = 6$  teilnehmenden Schulen in keiner der drei Teilstudien modelliert werden (vgl. Bryk & Raudenbush, 1988). Die geringe Stichprobengröße betrifft aber nicht nur die Schulstichprobe, sondern auch die Stichproben der

Lehrkräfte sowie der Schülerinnen und Schüler, die zu klein waren, um latente Modellierung in Form von komplexen (Mehrebenen-)Strukturgleichungsmodellen durchzuführen (vgl. Kleinke et al., 2017). Stattdessen wurde ausschließlich mit manifesten Skalenmittelwerten gerechnet, bei denen zufällig auftretende Messfehler in den beobachteten Variablen nicht berücksichtigt werden (vgl. Geiser, 2010). Mit der geringen Stichprobengröße geht darüber hinaus eine geringe Teststärke einher, die das Auffinden statistisch signifikanter Haupteffekte erschwert hat. Bei einer größeren Stichprobe hätten ggf. weitere inhaltlich bedeutsame Gruppenunterschiede, z. B. bei den Unterrichtsmerkmalen *Unterstützung von Verstehensprozessen* und *Förderung der Selbstwirksamkeit*, das Niveau statistischer Signifikanz erreicht. Außerdem hätte eine höhere statistische Power zu mehr signifikanten Moderationseffekten der *LZO* und *ALZO* führen können.

**Modellierung von Wirkmechanismen.** Mit Blick auf die theoretisch angenommenen Wirkmechanismen aus dem Feedbackmodell ist zu konstatieren, dass nicht alle Wirkungspfade empirisch geprüft werden konnten. Dies betrifft zunächst die Frage, welche formalen Gestaltungsmerkmale des Unterrichtsfeedbacks zu einer höheren Qualitätseinschätzung durch die Lehrkräfte geführt haben und ob die Feedbackqualität ihrerseits den Feedbackeffekt mediiert. Der beschriebene Wirkungspfad könnte prinzipiell mithilfe von (Mehrebenen-)Pfadanalysen bzw. (Mehrebenen-)Strukturgleichungsmodellen untersucht werden (vgl. Kleinke et al., 2017), in denen die Feedbackmerkmale als unabhängige Variablen, die Qualitätseinschätzung als Mediatorvariable und die Kompetenzentwicklung der Lehrkräfte bzw. die Veränderung der Unterrichtsqualität als abhängige Variablen modelliert werden. Mit den vorliegenden Daten war dies jedoch nicht möglich, weil nur die Lehrkräfte der Interventionsgruppe ein Feedback erhalten haben und folglich nur auf die Teilstichprobe dieser Lehrkräfte zurückgegriffen werden konnte, welche für die beschriebenen Analysen deutlich zu klein ist. Beispiele dafür, wie sich die Effekte der Feedbackmerkmale bzw. der Mediationseffekt der Feedbackqualität überprüfen ließe, finden sich in der empirischen Schul- und Unterrichtsforschung z. B. bei Cherasaro et al. (2016), Delvaux et al. (2013) sowie Tuytens und Devos (2011, 2014).

Neben den Effekten der Feedbackmerkmale und dem Mediationseffekt der Feedbackqualität ließen sich auch die in Kapitel 2.3.4 beschriebenen emotionalen und kognitiven Verarbeitungsprozesse, die in einer Zurückweisung oder Akzeptanz des Unterrichtsfeedback resultieren können, nicht empirisch abbilden. Die emotionalen Reaktionen auf den Feedbackerhalt, die wechselseitigen Prozesse der Diskrepanzproduktion und Dis-

krepanzreduktion und die Kausalattributionen der Lehrkräfte konnten aus erhebungs-ökonomischen Gründen nicht erfasst werden. Daher war es auch nicht möglich, zu prüfen, inwiefern die beschriebenen Verarbeitungsprozesse – wie theoretisch angenommen – mit den Zielorientierungen von Lehrkräften variieren und ob sich die differenziellen Effekte der *LZO* und *ALZO* hiermit tatsächlich erklären lassen. Daten zu den emotionalen und kognitiven Verarbeitungsprozessen hätten ggf. auch dazu beitragen können, die inkonsistenten Befunde zur *ALZO* zu erklären und die Rolle, die die Feedbackvalenz in diesem Zusammenhang spielt, weiter zu beleuchten. Unabhängig davon, dass die Verarbeitungsprozesse in der Interventionsstudie nicht erfasst wurden, wäre die Teilstichprobe der Lehrkräfte, die ein Unterrichtsfeedback erhalten haben, ohnehin zu klein gewesen. Derart komplexe Wirkzusammenhänge lassen sich am ehesten mit einem längsschnittlichen Studiendesign untersuchen, in dem alle Probandinnen und Probanden ein Feedback erhalten. Beispiele, in denen das Zusammenspiel von Verarbeitungsprozessen und Zielorientierungen untersucht wird, finden sich in der arbeitspsychologischen Feedbackforschung z. B. bei Brett und Atwater (2001), Cron et al. (2005) sowie VandeWalle et al. (2001).

### **6.3 Praktische Implikationen**

Trotz der beschriebenen Limitationen ist die vorliegende Arbeit von großer Relevanz für die schulische Praxis. Obwohl Schulleitungen seit Einführung des neuen Steuerungsparadigmas vermehrt strategische Führungsaufgaben im Hinblick auf die innerschulische Qualitätssicherung und -entwicklung übernehmen müssen, werden sie bislang kaum auf diese neuen Anforderungsbereiche vorbereitet (vgl. Klein & Tulowitzki, 2020; Tulowitzki et al., 2019). Vor diesem Hintergrund sind auch die Befunde des *Schulleitungsmonitor*, der sich auf die Befragungsdaten von  $N = 1.323$  freiwillig teilnehmenden Schulleitungen aus neun Bundesländern stützt (vgl. Schwanenberg et al., 2018), wenig überraschend: Sowohl im Bereich der Personalentwicklung als auch im Hinblick auf eine unterrichtsbezogene Führung sahen die befragten Schulleitungen einen vergleichsweise hohen Unterstützungsbedarf. Konkret zeigte sich, dass sich über 50 % der Schulleitungen Unterstützung bei der Durchführung von Unterrichtsbeobachtungen wünschen. Über 60 % wollen ihre Lehrkräfte zudem häufiger zum Nachdenken über den eigenen Unterricht anregen (vgl. Schwanenberg et al., 2018), wozu ein beobachtungsbasiertes Unterrichtsfeedback beitragen könnte. Die entwickelte Feedbackintervention stellt ein Qualifizierungsangebot dar, das diesen Unterstützungsbedarf direkt adressiert. Der BERU, das

Feedbackverfahren und das Beobachtungs- und Feedbacktraining ermöglichen es Schulleitungen, den Unterricht ihrer Lehrkräfte objektiv, reliabel und valide zu beurteilen und ihnen ausgehend von den Beobachtungen ein qualitativ hochwertiges Feedback zu ihrem Unterricht zu geben. Damit können sie einer der zentralen unterrichtsbezogenen Führungsaktivitäten nachkommen und die Qualität der unterrichtlichen Lerngelegenheiten an ihren Schulen über ein beobachtungsbasiertes Unterrichtsfeedback positiv beeinflussen.

Die Wirksamkeit des Unterrichtsfeedbacks ließe sich zudem steigern, wenn die Befunde zu den differenziellen Effekten der Zielorientierungen bei der formalen Feedbackgestaltung berücksichtigt würden. Aus den positiven Auswirkungen einer *LZO* ließe sich beispielsweise folgern, dass Schulleitungen zu Beginn der Feedbackgespräche – noch stärker als im Feedbackverfahren ohnehin vorgesehen – betonen, dass das Unterrichtsfeedback der professionellen Weiterentwicklung der Lehrkräfte dient und nicht ihrer Kontrolle. Hierdurch ließe sich ggf. dem hohen *ego involvement* begegnen, das annäherungs- und vor allem vermeidungsleistungszielorientierte Lehrkräfte in lern- und leistungsbezogenen Situationen aufweisen (vgl. Nicholls, 1984), sodass diese das Unterrichtsfeedback nicht als Evaluation ihrer Kompetenz wahrnehmen, sondern als Lerngelegenheit. Außerdem könnten Schulleitungen verdeutlichen, dass es bei entsprechender Anstrengungsbereitschaft möglich ist, neue Verhaltensweisen für den eigenen Unterricht zu erlernen und diesen zu verbessern. Hiermit ließe sich unter Umständen den impliziten subjektiven Theorien von annäherungs- und vermeidungsleistungszielorientierten Lehrkräften entgegenwirken, nach denen ihre Unterrichtskompetenz vergleichsweise stabil ist und sich auch über einen Mehraufwand nicht verändern lasse (vgl. Dweck, 1986; Dweck & Leggett, 1988).

Die vorliegende Arbeit ist jedoch nicht nur für das deutsche Schulsystem von praktischer Relevanz, sondern auch für die US-amerikanischen TES. Weil mit diesen ursprünglich das Ziel verfolgt wurde, wirksame und unwirksame Lehrkräfte zuverlässig zu identifizieren und negative Beurteilungen teilweise auch heute noch mit weitreichenden persönlichen Konsequenzen für individuelle Lehrkräfte einhergehen können, lag der Fokus der wissenschaftlichen Begleitforschung bislang vor allem auf der objektiven, reliablen und validen Erfassung von Unterrichtsqualität (vgl. z. B. Holtzapple, 2003; Jacob & Lefgren, 2008; Kimball et al., 2004; Milanowski, 2004). Wie in Kapitel 2.2.2 beschrieben, wird zuletzt jedoch vermehrt die Entwicklungsfunktion der TES betont; Lehrkräfte sollen ihren Unterricht auf Grundlage eines beobachtungsbasierten Unterrichtsfeedbacks weiterentwickeln. Aus diesem Grund fordern viele Autorinnen und Autoren in jüngerer

Zeit, das Augenmerk auf das Unterrichtsfeedback der Schulleitungen zu richten und zu erforschen, wie sich die Feedbackqualität – auch über ein Feedbacktraining für Schulleitungen – steigern ließe (vgl. Cohen & Goldhaber, 2016; Hill & Grossmann, 2013; Hunter & Springer, 2022; Kraft & Christian, 2022; Papay, 2012; Liu et al., 2019). Der in der vorliegenden Arbeit gewählte Ansatz, die Theorien und Befunde der arbeits- und instruktionspsychologischen Feedbackforschung auf ein beobachtungsbasiertes Unterrichtsfeedback durch Schulleitungen zu übertragen und das Feedback über Gesprächsleitfäden und lehrerspezifische Feedbackberichte stärker zu standardisieren, könnte sich auch im Kontext der US-amerikanischen TES als nützlich erweisen.

Darüber hinaus könnte die Feedbackintervention auch im Rahmen von kollegialen Hospitationen Anwendung finden. Würden sich Fachkolleginnen und -kollegen gegenseitig im Unterricht beobachten und sich wechselseitig ein Feedback geben, hätte dies den Vorteil, dass sie die im theoretischen Unterrichtsmodell beschriebenen generischen Merkmale eines lernwirksamen Unterrichts um fachspezifische Aspekte ergänzen könnten. Der Mehrwert einer fachspezifischen Konkretisierung generischer Qualitätsmerkmale wird in der deutschsprachigen Schul- und Unterrichtsforschung in jüngerer Zeit intensiv diskutiert (vgl. z. B. Lindmeier & Heinze, 2020; Praetorius & Gräsel, 2021; Praetorius et al., 2020). Darüber hinaus zeigen Forschungsbefunde aus dem internationalen Forschungsraum, dass gegenseitige Unterrichtsbeobachtungen und wechselseitige Feedbacks durch Kolleginnen und Kollegen zu einer besseren Unterrichtsqualität sowie einer Steigerung des Lernerfolgs von Schülerinnen und Schülern führen können (vgl. Burgess et al., 2021; Papay et al., 2020; Murphy et al. 2021). Die Studie von Burgess et al. (2021) verweist dabei darauf, dass kollegiale Hospitation nicht allein über das Feedback wirken, das Lehrkräfte zu ihrem eigenen Unterricht erhalten, sondern auch über die Beobachtungen, die sie bei ihren Kolleginnen und Kollegen durchführen. Insofern könnten Lehrkräfte nicht nur von dem qualitativ hochwertigen Unterrichtsfeedback ihrer Kolleginnen und Kollegen profitieren, das das Feedbackverfahren ermöglicht, sondern auch von der kriteriengeleiteten Unterrichtsbeobachtung mit dem BERU, die sie selbst bei ihren Kolleginnen und Kollegen durchführen.

## **6.4 Ausblick**

Bevor die Feedbackintervention jedoch in der schulischen Praxis implementiert wird, sollte ihre Wirksamkeit idealerweise in weiteren Studien empirisch geprüft werden. Wie in Kapitel 6.2 erläutert, sollte nach Möglichkeit ein klassisches experimentelles Design

mit Randomisierung umgesetzt werden, um eine hohe interne Validität der Befunde zu gewährleisten. Darüber hinaus sollte sich die Schulstichprobe aus allgemeinbildenden Schulen zusammensetzen. Neben einer Prä- und Post-Befragung sollte auch eine Follow-Up-Befragung durchgeführt werden, mit der sich langfristige Effekte der Intervention überprüfen ließen. Außerdem könnte die Interventionswirksamkeit anhand des Lernerfolgs von Schülerinnen und Schülern evaluiert werden. Sollte das Feedback der Schulleitungen nicht nur zu einer besseren Unterrichtsqualität führen, sondern auch zu Lernzuwächsen bei Schülerinnen und Schülern, wäre gleichzeitig ein Nachweis über die Kriteriumsvalidität des BERU erbracht (siehe die Ausführungen aus *Fußnote 25*).

Zusätzlich zu den Haupteffekten der Feedbackintervention sollten auch die differenziellen Effekte der Zielorientierungen erneut untersucht werden. In diesem Zuge ist zum einen zu prüfen, ob sich die positiven Moderationseffekte der *LZO* replizieren lassen. Zum anderen könnten die Befunde weiterer Studien zur Klärung der Frage beitragen, wie sich eine *ALZO* auf die Feedbackwirksamkeit auswirkt und welche Rolle die emotionale und kognitive Feedbackverarbeitung sowie die Feedbackvalenz in diesem Zusammenhang spielen. Neben den differenziellen Effekten von Zielorientierungen sollten zudem die Moderationseffekte weiterer Lehrkraftmerkmale untersucht werden. Anknüpfend an die arbeitspsychologische Feedbackforschung würden sich z. B. Selbstwirksamkeitserwartungen von Lehrkräften sowie ihre *Big-Five*-Persönlichkeitsmerkmale anbieten. Bei beiden handelt es sich um für die erfolgreiche Ausübung des Lehrberufs relevante Lehrkraftmerkmale, wie aktuelle Meta-Analysen und Literaturreviews aus der empirischen Schul- und Unterrichtsforschung zeigen (vgl. Bardach et al., 2022; Cramer & Binder, 2015; Kim et al., 2019; Klassen & Tze, 2014; Zee & Koomen, 2016). Insofern erscheint es plausibel, dass die Wirksamkeit eines beobachtungsbasierten Unterrichtsfeedbacks nicht nur von den Zielorientierungen von Lehrkräften beeinflusst wird, sondern auch von ihren Selbstwirksamkeitserwartungen und *Big-Five*-Persönlichkeitsmerkmalen.

Wenn sich die Feedbackintervention in Replikationsstudien als wirksam und die gefundenen Effekte über unterschiedliche Stichproben und Settings hinweg als robust erweisen, wäre in einem nächsten Schritt zu prüfen, wie eine erfolgreiche Implementation in die schulische Praxis gelingen kann (vgl. Gräsel, 2010; Gräsel & Parchmann, 2004; Hasselhorn et al., 2014). Dabei ist zu berücksichtigen, dass von einer „linearen“ Übernahme der Feedbackintervention nicht auszugehen ist; stattdessen ist zu erwarten, dass Schulleitungen und Lehrkräfte Anpassungen an die spezifischen Gegebenheiten an ihren

Schulen vornehmen (vgl. Petermann, 2014). Mit der Frage, wie erfolgreiche Implementierungen von evidenten Interventionen gelingen können und welche Anpassungen sich förderlich bzw. hemmend auf ihre Wirksamkeit auswirken, beschäftigt sich die in der deutschsprachigen Schul- und Unterrichtsforschung bislang nur in Ansätzen etablierte Implementationsforschung (vgl. Hasselhorn et al., 2014; Schrader et al., 2020).

Diesem Forschungsstrang ist auch eine bereits bewilligte Folgestudie zuzurechnen, die das Ziel verfolgt, die Implementation der Feedbackintervention in die schulische Praxis zu begleiten und mithilfe eines *Multi-Method*-Ansatzes zu erforschen.<sup>25</sup> Hierbei sollen unter anderem regelmäßige Reflexionsgespräche mit den an der Implementationsstudie beteiligten Schulleitungen geführt und die vorgenommenen Anpassungen dokumentiert werden. Im Anschluss soll empirisch geprüft werden, wie sich diese Anpassungen auf die Interventionswirksamkeit auswirken. Die Befunde dieser Studie können zum einen wichtige Erkenntnisse darüber liefern, welche Modifikationen angezeigt sind, um die „Alltagstauglichkeit“ der Feedbackintervention zu erhöhen, ohne dass diese ihre Wirksamkeit einbüßt. Darüber hinaus können ihre Befunde genutzt werden, um das theoretische Feedbackmodell aus Kapitel 2.3 um weitere Gelingensbedingungen zu ergänzen, die bisher nicht konzeptualisiert wurden. So deuten Befunde aus der Implementationsforschung z. B. darauf hin, dass eine erfolgreiche Implementation von pädagogischen Innovationen von diversen Merkmalen der Einzelschule wie ihren Kooperationsstrukturen beeinflusst wird (zusammenfassend vgl. Gräsel, 2010; Schrader et al., 2020). Auch in der Forschung zu den US-amerikanischen TES wird diskutiert, dass die Kooperation innerhalb eines Lehrkräftekollegiums die Wirksamkeit beobachtungsbasierter Unterrichtsfeedbacks durch Schulleitungen begünstigen könnte (vgl. Darling-Hammond et al., 2012; Jiang et al., 2015; siehe auch Tuytens & Devos, 2017). Sollte sich dies in der Implementationsstudie bestätigen, könnte das theoretische Feedbackmodell entsprechend um dieses Schulmerkmal erweitert werden.

---

<sup>25</sup> Das Folgeprojekt „Implementation einer nachhaltigen Feedbackkultur auf Basis regelmäßiger Unterrichtsbeobachtungen und Feedback durch Schulleitungen“ soll unter der Leitung von Prof. Dr. Holger Gärtner und Prof. Dr. Felicitas Thiel im Frühjahr 2024 starten. Gefördert wird das Forschungsvorhaben durch das Bundesministerium für Bildung und Forschung (BMBF).

## 6.5 Fazit

In der vorliegenden Arbeit wurden der Prozess und die Gelingenbedingungen eines beobachtungsbasierten Unterrichtsfeedbacks von Schulleitungen an Lehrkräfte erstmals anhand eines umfassenden theoretischen Feedbackmodells beschrieben. Hierfür wurden zentrale Theorien sowie empirische Befunde der arbeits- und instruktionspsychologischen Feedbackforschung aufgearbeitet und für die Schul- und Unterrichtsforschung nutzbar gemacht. Ausgehend von diesem Feedbackmodell wurde eine Feedbackintervention entwickelt, deren Wirksamkeit in einer quasi-experimentellen Studie nachgewiesen werden konnte. Darüber hinaus wurde erstmals gezeigt, dass die Wirksamkeit von beobachtungsbasiertem Unterrichtsfeedback auch von den Zielorientierungen von Lehrkräften abhängt.

Auch wenn die durchgeführte Interventionsstudie einige Limitationen aufweist und Replikations- sowie Implementationsstudien noch ausstehen, stellt die Feedbackintervention Schulleitungen empirisch geprüftes Handlungswissen zur Verfügung, mit dem sie die Unterrichtsqualität an ihren Schulen verbessern können. Dies ist auch in Anbetracht der Herausforderungen, vor denen das deutsche Schulsystem aktuell steht, von besonderer Relevanz: Weil hierzulande vermehrt Lehrkräfte ohne universitäre Lehramtsausbildung unterrichten (vgl. Driesner & Arndt, 2020; Klemm, 2019) und gleichzeitig ein großer Anteil von Schülerinnen und Schülern erhebliche Lernrückstände aufweist (vgl. Reiss et al., 2019; Schwippert et al., 2020; Stanat et al., 2019), die zuletzt unter anderem aufgrund der pandemiebedingten Schulschließungen sogar zugenommen haben (vgl. Betthäuser et al., 2023, Stanat et al., 2022), erscheinen wirksame Maßnahmen zur Steigerung der Unterrichtsqualität wichtiger denn je.

## Literaturverzeichnis II

- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y. & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037. <https://doi.org/10.1126/science.1207998>
- Altrichter, H., Kemethofer, D. & George, A. C. (2019). Schulleitungen und evidenzbasierte Bildungsreform im Schulwesen. *Zeitschrift für Bildungsforschung*, 9(1), 17–35. <https://doi.org/10.1007/s35834-018-0228-5>
- Bach, A., Böhnke, A. & Thiel, F. (2020). Improving instructional competencies through individualized staff development and teacher collaboration in German schools. *International Journal of Educational Management*, 34(8), 1289–1302. <https://doi.org/10.1108/IJEM-08-2019-0294>
- Bardach, L., Klassen, R. M. & Perry, N. E. (2022). Teachers' psychological characteristics: Do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative Review. *Educational Psychology Review*, 34(1), 259–300. <https://doi.org/10.1007/s10648-021-09614-9>
- Barfknecht, T. & von Saldern, M. (2010). Evaluation und Feedback der Lehrkräfte. In M. Demmer & M. von Saldern (Hrsg.), *Helden des Alltags. Erste Ergebnisse der Schulleitungs- und Lehrkräftebefragung (TALIS) in Deutschland (DDS – Die Deutsche Schule*, 11. Beiheft, S. 94–115).
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Kraus, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Bethhäuser, B.A., Bach-Mortensen, A.M. & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*. Advance online publication. <https://doi.org/10.1038/s41562-022-01506-4>
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G. & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, Article 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>

- Bonsen, M. (2016). Schulleitung und Führung in der Schule. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulwesen* (2., überarb. und aktual. Aufl., S. 301–323). Springer VS. [https://doi.org/10.1007/978-3-531-18942-0\\_11](https://doi.org/10.1007/978-3-531-18942-0_11)
- Brauckmann, S. & Eder, F. (2019). Führungsforschung im Bildungsbereich: Schulleitung im Spannungsfeld erweiterter Rechte und Pflichten. *Zeitschrift für Bildungsforschung*, 9(1), 5–15. <https://doi.org/10.1007/s35834-019-00242-6>
- Brauckmann, S. & Schwarz, A. (2015). No time to manage? The trade-off between relevant tasks and actual priorities of school leaders in Germany. *International Journal of Educational Management*, 29(6), 749–765. <https://doi.org/10.1108/IJEM-10-2014-0138>
- Brett, J. F. & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86(5), 930–942. <https://doi.org/10.1037//0021-9010.86.5.930>
- Bryk, A. S. & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education*, 97(1), 65–108. <https://doi.org/10.1086/443913>
- Butler, R. (2007). Teachers' achievement goal orientations and associations with teachers' help seeking. Examination of a novel approach to teacher motivation. *Journal of Educational Psychology*, 99(2), 241–252. <https://doi.org/10.1037/0022-0663.99.2.241>
- Burgess, S., Rawal, S. & Taylor, E. S. (2021). Teacher peer observation and student test scores: Evidence from a field experiment in English secondary schools. *Journal of Labor Economics*, 39(4), 1155–1186. <https://doi.org/10.1086/712997>
- Cellar, D. F., Stuhlmacher, A. F., Young, S. K., Fisher, D. M., Adair, C. K., Haynes, S., Twichell, E., Arnold, K. A., Royer, K., Denning, B. L. & Riester, D. (2011). Trait goal orientation, self-regulation, and performance: A meta-analysis. *Journal of Business and Psychology*, 26(4), 467–483. <https://doi.org/10.1007/s10869-010-9201-6>
- Cherasaro, T. L., Brodersen, R. M., Reale, M. L. & Yanoski, D. C. (2016). *Teachers' responses to feedback from evaluators: What feedback characteristics matter?* (REL 2017-190). U.S. Department of Education, Institute of Education Sciences,

- National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. [https://ies.ed.gov/ncee/rel/regions/central/pdf/REL\\_2017190.pdf](https://ies.ed.gov/ncee/rel/regions/central/pdf/REL_2017190.pdf)
- Cohen, J. & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378–387. <https://doi.org/10.3102/0013189X16659442>
- Copur-Gencturk, Y. & Thacker, I. (2021). A comparison of perceived and observed learning from professional development: Relationships among self-reports, direct assessments, and teacher characteristics. *Journal of Teacher Education*, 72(2), 138–151. <https://doi.org/10.1177/0022487119899101>
- Cramer, C. & Binder, K. (2015). Zusammenhänge von Persönlichkeitsmerkmalen und Beanspruchungserleben im Lehramt. Ein internationales systematisches Review. *Zeitschrift für Erziehungswissenschaft*, 18(1), 101–123. <https://doi.org/10.1007/s11618-014-0605-3>
- Cron, W. L., Slocum, J. W., VandeWalle, D. & Fu, Q. (2005). The role of goal orientation on negative emotions and goal setting when initial performance falls short of one's performance goal. *Human Performance*, 18(1), 55–80. [https://doi.org/10.1207/s15327043hup1801\\_3](https://doi.org/10.1207/s15327043hup1801_3)
- Darling-Hammond, L., Amrien-Beardsley, A., Haertel, E. & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8–15. <https://doi.org/10.1177/003172171209300603>
- De Jong, R. & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, 4(1), 51–84. <https://doi.org/10.1023/A:1011402608575>
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G. & van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1–11. <https://doi.org/10.1016/j.tate.2013.06.011>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollst. überarb., aktual. und erw. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Driesner, I. & Arndt, M. (2020). Die Qualifizierung von Quer- und Seiteneinsteiger\*innen. Konzepte und Lerngelegenheiten im bundesweiten Überblick. *DDS – Die Deutsche Schule*, 112(4), 414–427. <https://doi.org/10.31244/dds.2020.04.05>

- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048. <https://doi.org/10.1037//0003-066X.41.10.1040>
- Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273. <https://doi.org/10.1037//0033-295X.95.2.256>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014a). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014b). 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>
- Gärtner, H., Thiel, F., Nachbauer, M. & Kellermann, C. (2021). Unterrichtsentwicklung durch Unterrichtsfeedback – Erste Entwicklungsschritte eines Beobachtungsboogens für Schulleitungen. *Zeitschrift für Bildungsforschung*, 11(3), 529–547. <https://doi.org/10.1007/s35834-021-00321-7>
- Geiser, C. (2010). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92042-9>
- 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 (BMBF) (Hrsg.), *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments* (Bildungsforschung, Bd. 44, S. 63–82). Bundesministerium für Bildung und Forschung (BMBF).
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13(1), 7–20. <https://doi.org/10.1007/s11618-010-0109-8>
- Gräsel, C. & Parchmann, I. (2004). Implementationsforschung – oder: der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft*, 32(3), 238–256.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J. & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, 94(3), 638–645. <https://doi.org/10.1037//0022-0663.94.3.638>

- Hasselhorn, M., Köller, O., Maaz, K. & Zimmer, K. (2014). Implementation wirksamer Handlungskonzepte im Bildungsbereich als Forschungsaufgabe. *Psychologische Rundschau*, 65(3), 140–149. <https://doi.org/10.1026/0033-3042/a000216>
- Hill, H. C. & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2), 371–384. <https://doi.org/10.17763/haer.83.2.d11511403715u376>
- Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(3), 207–219. <https://doi.org/10.1007/s11092-005-2980-z>
- Huang, C. (2011). Achievement goals and achievement emotions: A meta-analysis. *Educational Psychology Review*, 23(3), 359–388. <https://doi.org/10.1007/s10648-011-9155-x>
- Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(1), 48–73. <https://doi.org/10.1037/a0026223>
- Hulleman, C. S., Schrage, S. M., Bodmann, S. M. & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, 136(3), 422–449. <https://doi.org/10.1037/a0018947>
- Hunter, S. B. & Springer, M. G. (2022). Critical feedback characteristics, teacher human capital, and early-career teacher performance: A mixed-methods analysis. *Educational Evaluation and Policy Analysis*, 44(3), 380–403. <https://doi.org/10.3102/01623737211062913>
- Jacob, B. A. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101–136. <https://doi.org/10.1086/522974>
- Janke, S., Bardach, L., Oczlon, S. & Lüftenegger, M. (2019). Enhancing feasibility when measuring teachers' motivation: A brief scale for teachers' achievement goal orientations. *Teaching and Teacher Education*, 83, 1–11. <https://doi.org/10.1016/j.tate.2019.04.003>
- Jiang, J. Y., Spote, S. E. & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH Students. *Educational Researcher*, 44(2), 105–116. <https://doi.org/10.3102/0013189X15575517>

- Kaplan, A. & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, 19(2), 141–184.  
<https://doi.org/10.1007/s10648-006-9012-5>
- Kim, L., Jörg, V. & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review*, 31(1), 163–195. <https://doi.org/10.1007/s10648-018-9458-2>
- Kimball, S. M. (2003). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241–268.  
<https://doi.org/10.1023/A:1021787806189>
- Kimball, S. M. & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34–70.  
<https://doi.org/10.1177/0013161X08327549>
- Kimball, S. M., White, B., Milanowski, A. T. & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54–78.  
[https://doi.org/10.1207/s15327930pje7904\\_4](https://doi.org/10.1207/s15327930pje7904_4)
- Klassen, R. M. & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59–76.  
<https://doi.org/10.1016/j.edurev.2014.06.001>
- Klein, E. D. & Tulowitzki, P. (2020). Die Fortbildung von Schulleiter\*innen in Forschung und Praxis – ein Systematisierungsversuch. *DDS – Die Deutsche Schule*, 112(3), 257–276. <https://doi.org/10.31244/dds.2020.03.02>
- Kleinke, K., Schlüter, E. & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung*. De Gruyter. <https://doi.org/10.1524/9783486714807>
- Klemm, K. (2019). *Seiten-und Quereinsteiger\_innen an Schulen in den 16 Bundesländern*. Friedrich-Ebert-Stiftung. <http://library.fes.de/pdf-files/studienfoerderung/15305.pdf>
- Köller, O. (2020). Evaluation pädagogisch-psychologischer Maßnahmen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (3., vollst. überarb. und aktual. Aufl., S. 335–347). Springer. [https://doi.org/10.1007/978-3-662-61403-7\\_14](https://doi.org/10.1007/978-3-662-61403-7_14)

- König, J. & Pflanzl, B. (2016). Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality. *European Journal of Teacher Education*, 39(4), 419–436.  
<https://doi.org/10.1080/02619768.2016.1214128>
- König, J., Kaiser, G. & Felbrich, A. (2012). Spiegelt sich pädagogisches Wissen in den Kompetenzselbsteinschätzungen angehender Lehrkräfte? Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerausbildung. *Zeitschrift für Pädagogik*, 58(4), 476–491.
- König, J., Blömeke, S., Jentsch, A., Schlesinger, L., Musekamp, F. & Kaiser, G. (2021). The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educational Studies in Mathematics*, 107(1), 189–212. <https://doi.org/10.1007/s10649-020-10021-0>
- Kraft, M. A. & Christian, A. (2022). Can teacher evaluation systems produce high-quality feedback? An administrator training field experiment. *American Educational Research Journal*, 59(3), 500–537. <https://doi.org/10.3102/00028312211024603>
- Kraft, M. A. & Gilmour, A. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, 52(5), 711–753.  
<https://doi.org/10.1177/0013161X16653445>
- Kraft, M. A., Blazar, D. & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 231–251. <https://doi.org/10.1007/s10984-006-9015-7>
- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. *Unterrichtswissenschaft*, 38(1), 68–86.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers. Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820.  
<https://doi.org/10.1037/a0032583>
- Kyriakides, L., Christoforou, C. & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective

- teaching. *Teaching and Teacher Education*, 36, 143–152.  
<https://doi.org/10.1016/j.tate.2013.07.010>
- Leutner, D. (2010). Perspektiven pädagogischer Interventionsforschung. In T. Hascher & B. Schmitz (Hrsg.), *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen* (S. 63–72). Juventa.
- Lindmeier, A. & Heinze, A. (2020). Die fachdidaktische Perspektive in der Unterrichtsqualitätsforschung: (bisher) ignoriert, implizit enthalten oder nicht relevant? In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierung* (Zeitschrift für Pädagogik Beiheft, Bd. 66, S. 255–268). Beltz.
- 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* (Zeitschrift für Pädagogik Beiheft, Bd. 51, S. 47–70). Beltz.
- Liu, Y., Visone, J., Mongillo, M. B. & Lisi, P. (2019). What matters to teachers if evaluation is meant to help them improve? *Studies in Educational Evaluation*, 61(2), 41–54. <https://doi.org/10.1016/j.stueduc.2019.01.006>
- Mertens, S. & Gräsel, C. (2018). Entwicklungsbereiche bildungswissenschaftlicher Kompetenzen von Lehramtsstudierenden im Praxissemester. *Zeitschrift für Erziehungswissenschaft*, 21(6), 1109–1133. <https://doi.org/10.1007/s11618-018-0825-z>
- Midgley, C., Kaplan, A. & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93(1), 77–86. <https://doi.org/10.1037//0022-0663.93.1.77>
- Milanowski, A. T. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33–53. [https://doi.org/10.1207/s15327930pje7904\\_3](https://doi.org/10.1207/s15327930pje7904_3)
- Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H. & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256.  
<https://doi.org/10.1080/09243453.2014.885451>
- Murphy, R., Weinhardt, F. & Wyness, G. (2021). Who teaches the teachers? A RCT of peer-to-peer observation and feedback in 181 schools. *Economics of Education Review*, 82, Article 102091. <https://doi.org/10.1016/j.econedurev.2021.102091>

- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328–346. <https://doi.org/10.1037/0033-295X.91.3.328>
- Nitsche, S., Dickhäuser, O., Dresel, M. & Fasching, M. S. (2013a). Zielorientierungen von Lehrkräften als Prädiktoren lernrelevanten Verhaltens. *Zeitschrift für Pädagogische Psychologie*, 27(1-2), 95–103. <https://doi.org/10.1024/1010-0652/a000092>
- Nitsche, S., Dickhäuser, O., Fasching, M. S. & Dresel, M. (2011). Rethinking teachers' goal orientations. Conceptual and methodological enhancements. *Learning and Instruction*, 21(4), 574–586. <https://doi.org/10.1016/j.learninstruc.2010.12.001>
- Nitsche, S., Dickhäuser, O., Fasching, M. S. & Dresel, M. (2013b). Teachers' professional goal orientations: Importance for further training and sick leave. *Learning and Individual Differences*, 23, 272–278. <https://doi.org/10.1016/j.lindif.2012.07.017>
- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123–141. <https://doi.org/10.17763/haer.82.1.v40p0833345w6384>
- Papay, J. P., Taylor, E. S., Tyler, J. H. & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, 12(1), 359–388. <https://doi.org/10.1257/pol.20170709>
- Payne, S. C., Youngcourt, S. S. & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92(1), 128–150. <https://doi.org/10.1037/0021-9010.92.1.128>
- Petermann, F. (2014). Implementationsforschung: Grundbegriffe und Konzepte. *Psychologische Rundschau*, 65(3), 122–128. <https://doi.org/10.1026/0033-3042/a000214>
- 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., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48(3), 409–446. <https://doi.org/10.1007/s42010-020-00082-8>

- Reiss, K., Weis, M., Klieme, E. & Köller, O. (Hrsg.) (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Waxmann.
- Schleicher, A. (2012). *Preparing teachers and developing school leaders for the 21st century. Lessons from around the world*. OECD. [https://www.oecd-ilibrary.org/preparing-teachers-and-developing-school-leaders-for-the-21st-century\\_5k9fdvsjz76g.pdf?itemId=%2Fcontent%2Fpublication%2F9789264174559-en&mimeType=pdf](https://www.oecd-ilibrary.org/preparing-teachers-and-developing-school-leaders-for-the-21st-century_5k9fdvsjz76g.pdf?itemId=%2Fcontent%2Fpublication%2F9789264174559-en&mimeType=pdf)
- Schrader, J., Hasselhorn, M., Hetfleisch, P. & Goeze, A. (2020). Stichwortbeitrag Implementationsforschung: Wie Wissenschaft zu Verbesserungen im Bildungssystem beitragen kann. *Zeitschrift für Erziehungswissenschaft*, 23(1), 9–59. <https://doi.org/10.1007/s11618-020-00927-z>
- Schwanenberg, J., Klein, E. D. & Walpuski, M. (2018). Wie erfolgreich fühlen sich Schulleitungen und welche Unterstützungsbedürfnisse haben sie? Ergebnisse aus dem Projekt Schulleitungsmonitor. *SHIP Working Paper Reihe, No. 03*. Universität Duisburg Essen.
- Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. & Wendt, H. (Hrsg.) (2020). *Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann.
- Seifert, A., Schaper, N. & König, J. (2018). Bildungswissenschaftliches Wissen und Kompetenzeinschätzungen von Studierenden im Praxissemester: Veränderungen und Zusammenhänge. In J. König, M. Rothland & N. Schaper (Hrsg.), *Learning to Practice, Learning to Reflect? Ergebnisse aus der Längsschnittstudie LtP zur Nutzung und Wirkung des Praxissemesters in der Lehrerbildung* (S. 325–347). Springer VS. [https://doi.org/10.1007/978-3-658-19536-6\\_12](https://doi.org/10.1007/978-3-658-19536-6_12)
- Seifried, J. & Weyland, U. (2021). Schulische Formen der Berufsbildung. In T. Ha-scher, T.-S. Idel & W. Helsper (Hrsg.), *Handbuch Schulforschung* (S. 1–18). Springer VS. [https://doi.org/10.1007/978-3-658-24734-8\\_17-1](https://doi.org/10.1007/978-3-658-24734-8_17-1)
- Senko, C. & Dawson, B. (2017). Performance-approach goal effects depend on how they are defined: Meta-analytic evidence from multiple educational outcomes. *Journal of Educational Psychology*, 109(4), 574–598. <https://doi.org/10.1037/edu0000160>

- Senko, C., Hulleman, C. S. & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist*, 46(1), 26–47. <https://doi.org/10.1080/00461520.2011.538646>
- Sitzmann, T., Ely, K., Brown, K. G. & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, 9(2), 169–191. <https://doi.org/10.5465/amle.9.2.zqr169>
- Song, M., Wayne, A. J., Garet, M. S., Brown, S. & Rickles, J. (2021). Impact of providing teachers and principals with performance feedback on their practice and student achievement: Evidence from a large-scale randomized experiment. *Journal of Research on Educational Effectiveness*, 14(2), 353–378. <https://doi.org/10.1080/19345747.2020.1868030>
- Stanat, P., Schipolowski, S., Mahler, N., Weirich, S. & Henschel, S. (Hrsg.) (2019). *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich*. Waxmann.
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S. & Henschel, S. (Hrsg.) (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Waxmann.
- Steinberg, M. P. & Sartin, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, 10(4), 535–572. [https://doi.org/10.1162/EDFP\\_a\\_00173](https://doi.org/10.1162/EDFP_a_00173)
- Taylor, E. S. & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628–3651. <https://doi.org/10.1257/aer.102.7.3628>
- Thillmann, K., Bach, A., Wurster, S. & Thiel, F. (2015a). School-based staff development in two federal states in Germany. *International Journal of Educational Management*, 29(6), 714–734. <https://doi.org/10.1108/IJEM-07-2014-0094>
- Thillmann, K., Brauckmann, S., Herrmann, C. & Thiel, F. (2015b). Praxis schulischer Personalentwicklung unter den Bedingungen der Neuen Steuerung. In H. J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungssystem. Analysen zur Mehrebenenperspektive, Steuerung und Koordination* (Educational Governance, Bd. 26, S. 195–228). Springer VS. [https://doi.org/10.1007/978-3-658-06523-2\\_8](https://doi.org/10.1007/978-3-658-06523-2_8)

- Tulowitzki, P., Hinzen, I. & Roller, M. (2019). Die Qualifizierung von Schulleiter\*innen in Deutschland – ein bundesweiter Überblick. *DDS – Die Deutsche Schule*, 111(2), 149–169. <https://doi.org/10.31244/dds.2019.02.04>
- Tuytens, M. & Devos, G. (2011). Stimulating professional learning through teacher evaluation: An impossible task for the school leader? *Teaching and Teacher Education*, 27(5), 891–899. <https://doi.org/10.1016/j.tate.2011.02.004>
- Tuytens, M. & Devos, G. (2014). How to activate teachers through teacher evaluation? *School Effectiveness and School Improvement*, 25(4), 509–530. <https://doi.org/10.1080/09243453.2013.842601>
- Tuytens, M. & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers and Teaching*, 23(1), 6–24. <https://doi.org/10.1080/13540602.2016.1203770>
- VandeWalle, D., Cron, W. L. & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology*, 86(4), 629–640. <https://doi.org/10.1037//0021-9010.86.4.629>
- Voss, T., Kunter, M., Seiz, J., Hoehne, V. & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik*, 60(2), 184–201.
- 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>
- Weisberg, D., Sexton, S., Mulhern, J. & Kelling, D. (2009). *The widget effect. Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. [https://tntp.org/assets/documents/TheWidgetEffect\\_2nd\\_ed.pdf](https://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf)
- Wirthwein, L., Sparfeldt, J. R., Pinquart, M., Wegerer, J. & Steinmayr, R. (2013). Achievement goals and academic achievement: A closer look at moderating factors. *Educational Research Review*, 10, 66–89. <https://doi.org/10.1016/j.edurev.2013.07.001>
- Zee, M. & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/0034654315626801>

## Eigenanteil an Studien mit Co-Autorenschaft

Die nachfolgende Tabelle gibt einen Überblick über den Eigenanteil an den in Co-Autorenschaft entstandenen empirischen Teilstudien der Dissertation.

Autorinnen und Autoren	Titel	Status	Eigenanteil
Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F	Effekte eines Unterrichts- feedbacks von Schullei- tungen auf die selbst ein- geschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergeb- nisse einer Interventions- studie unter Berücksichti- gung der Zielorientie- rungen von Lehrkräften	veröffentlicht	Federführung bei der Konzeption; Aufbereitung der Theorie und des empi- rischen Forschungsstands; statistische Analysen; Schriftlegung der ersten Version und der Revisio- nen
Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F	The effect of school lead- ers' feedback on teaching quality perceived by stu- dents – results of an inter- vention study	veröffentlicht	Federführung bei der Konzeption; Aufbereitung der Theorie und des empi- rischen Forschungsstands; statistische Analysen und Schriftlegung der ersten Version (zusammen mit Max Nachbauer); Reana- lysen im Zuge des Re- viewverfahrens und Schriftlegung der Revisio- nen
Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F	Improvements in teaching quality after receiving feedback: What role do teachers' goal orientations play for the effectiveness of feedback?	eingereicht	Federführung bei der Konzeption; Aufbereitung der Theorie und des empi- rischen Forschungsstands; statistische Analysen; Schriftlegung des Manu- skripts

## **Selbstständigkeitserklärung**

Hiermit erkläre ich, die vorliegende Dissertation selbstständig verfasst und ohne unerlaubte Hilfe angefertigt habe. Alle Hilfsmittel, die verwendet wurden, habe ich angegeben. Die Dissertation ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin, den 06.02.2023

---

Christopher Kellermann

## Publikationsliste

### *Zeitschriftenartikel mit Peer Review*

- Gärtner, H., Thiel, F., Nachbauer, M. & Kellermann, C. (2022). Entwicklung eines theoriebasierten Schülerfragebogens für Unterrichtsevaluation. *Journal for Educational Research Online*, 14(1), 147–173. <https://doi.org/10.31244/jero.2022.01.07>
- Kellermann, C., Nachbauer, M., Gaertner, H. & Thiel, F. (2022). The effect of school leaders' feedback on teaching quality perceived by students – results of an intervention study. *School Effectiveness and School Improvement*. Advance online publication. <https://doi.org/10.1080/09243453.2022.2075018>
- Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F. (2022). Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die selbst eingeschätzte Entwicklung der Unterrichtskompetenz von Lehrkräften – Ergebnisse einer Interventionsstudie unter Berücksichtigung der Zielorientierungen von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 25(5), 1267–1291. <https://doi.org/10.1007/s11618-022-01087-y>
- Gärtner, H., Thiel, F., Nachbauer, M. & Kellermann, C. (2021). Unterrichtsentwicklung durch Unterrichtsfeedback – Erste Entwicklungsschritte eines Beobachtungsbogens für Schulleitungen. *Zeitschrift für Bildungsforschung*, 11(3), 529–547. <https://doi.org/10.1007/s35834-021-00321-7>

### *Sonstige Beiträge*

- Kellermann, C. & Gärtner, H. (2022). *Feed up, feed back und feed forward! Durch Unterrichtsbeobachtungen und Feedback an Lehrkräfte können Schulleitungen zu besserem Unterricht beitragen*. Online-Magazin Schulmanagement. <https://www.campus-schulmanagement.de/magazin/feed-up-feed-back-feed-forward>

### *Vorträge und Posterpräsentationen*

- Kellermann, C. Nachbauer, M., Gärtner, H. & Thiel, F. (2022). *Die Effekte eines Unterrichtsfeedbacks von Schulleitungen auf die von Schülerinnen und Schülern wahrgenommene Unterrichtsqualität. Ergebnisse einer Interventionsstudie*. Vortrag auf der 10. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF) in Bamberg, Deutschland

- Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F. (2019). *Does feedback by school leaders affect teaching quality?* Vortrag auf der 18. Tagung der European Association for Research on Learning and Instruction (EARLI) in Aachen, Deutschland.
- Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F. (2019). *Führt Unterrichtsfeedback von Schulleitungen zu einer professionellen Weiterentwicklung von Lehrpersonen?* Vortrag auf der 7. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF) in Köln, Deutschland.
- Nachbauer, M., Kellermann, C., Gärtner, H. & Thiel, F. (2018). *Entwicklung und psychometrische Überprüfung eines Beobachtungsinstruments für Unterrichtsqualität.* Vortrag auf der 83. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF) in Lüneburg, Deutschland.
- Kellermann, C., Nachbauer, M., Gärtner, H. & Thiel, F. (2018). *Unterrichtsfeedback von Schulleitungen – Erste Befunde zur Wirksamkeit einer Feedbackintervention.* Vortrag auf der 83. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF) in Lüneburg, Deutschland.
- Nachbauer, M., Gärtner, H. Kellermann, C. & Thiel, F. (2018) *Dimensionen der Unterrichtsqualität. Unterrichtsbeobachtungen durch schulische Führungskräfte.* Vortrag auf der 6. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF) in Basel, Schweiz.