



KLEINE KIELER
BEITRÄGE ZU

Künstlicher Intelligenz

BAND 1

KLEINE KIELER BEITRÄGE ZU

Künstlicher Intelligenz

Pädagogische und fachdidaktische Perspektiven
auf technologische Entwicklungen

BAND 1

Herausgeben von

Kirsten Schindler, Nicolaus Wilder, Virtuelles Kompetenzzentrum
für Künstliche Intelligenz und Wissenschaftliches Arbeiten (VK:KIWA)

Lisanne Pitzen

Peer- vs. KI-Feedback

Wie Schüler*innen der Sekundarstufe I
Textfeedback zu einer Zusammenfassung
rezipieren

KLEINE KIELER BEITRÄGE ZU KÜNSTLICHER INTELLIGENZ | BAND 1

eISSN: 3053-4488

HERAUSGEGEBEN VON

Prof.in Dr. Kirsten Schindler , Bergische Universität Wuppertal

Dr. Nicolaus Wilder , Christian-Albrechts-Universität zu Kiel

Virtuelles Kompetenzzentrum für Künstliche Intelligenz und

Wissenschaftliches Arbeiten

<https://www.vkkiwa.de/>

AUTORIN

Lisanne Pitzen ist als wissenschaftliche Hilfskraft am Lehrstuhl von Prof.in Dr. Kirsten Schindler in der Didaktik der deutschen Sprache an der Bergischen Universität Wuppertal tätig. Der vorliegende Beitrag wurde von ihr als Masterthesis verfasst. Im Oktober 2025 beginnt Lisanne Pitzen ihre Promotion. In ihrem Promotionsvorhaben erforscht sie die Rezeption und Umsetzung von KI-basiertem Feedback durch Schüler*innen.

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliographische Daten sind über <https://dnb.d-nb.de> abrufbar.

Open Access



Das Werk ist unter der Creative-Commons-Lizenz Namensnennung-4.0 International veröffentlicht.

Den Vertragstext finden Sie unter: <https://creativecommons.org/licenses/by/4.0/>.

Bitte beachten Sie, dass einzelne, entsprechend gekennzeichnete Teile des Werks von der genannten Lizenz ausgenommen sein bzw. anderen urheberrechtlichen Bedingungen unterliegen können.

Das Werk ist auf dem Open-Access-Publikationsserver MACAU der Universitätsbibliothek Kiel (<https://macau.uni-kiel.de>) frei verfügbar: <https://doi.org/10.38071/2025-01281-7>.

Kiel, 2025

Selbstverlag des Virtuellen Kompetenzzentrums für Künstliche Intelligenz und Wissenschaftliches Arbeiten (VK:KIWA) im Hosting-Service der UB Kiel.

Titelbild: Erstellt mit Recraft (<https://www.recraft.ai/>).

Vorwort der Herausgebenden

Mit den *Kleinen Kieler Beiträgen zu Künstlicher Intelligenz* legen wir den Grundstein für eine neue Open-Access-Reihe, die auf zweierlei zielt: wissenschaftlich fundierte und gesellschaftlich relevante Beiträge aktuell veröffentlichen. Künstliche Intelligenz ist längst kein Nischenthema mehr, sondern durchdringt in rasanter Geschwindigkeit Bildungsinstitutionen, öffentliche Debatten und unseren Alltag. Gerade in Schule und Hochschule stellen sich dabei drängende Fragen: Wie verändert KI das Lehren und Lernen? Welche Chancen eröffnet sie für individuelle Förderung? Welche Risiken birgt sie für Bildungsgerechtigkeit? Zwischen unkritischer Euphorie und reflexhafter Ablehnung entsteht so ein oftmals ideologiegetriebenes Spannungsfeld, in dem differenzierte wissenschaftliche Stimmen dringender sind denn je.

Genau hier setzt unsere Reihe an. Wir möchten ein Forum eröffnen, in dem empirisch ausgerichtete Studien ebenso Platz finden wie reflektierte theoretische Beiträge. Unsere Leitidee ist es, wissenschaftliche Erkenntnisse niedrigschwellig, transparent und schnell zugänglich zu machen. Letzteres unterscheidet uns von Sammelbänden oder Zeitschriften, deren langwierige Produktionszyklen häufig im Widerspruch zur Dynamik des hier zur Diskussion stehenden Feldes stehen. Indem wir auf digitale Open-Access-Publikationen setzen, wollen wir zugleich einen Beitrag zur Demokratisierung von Wissen leisten: Forschungsergebnisse sollen frei und unmittelbar zugänglich sein.

Ein besonderes Anliegen ist uns dabei die Öffnung für den wissenschaftlichen Nachwuchs. Exzellente Abschlussarbeiten sind mitunter Schätze, die mit hohem Aufwand erstellt dennoch in Prüfungsakten verschwinden, anstatt im wissenschaftlichen Diskurs weiterzuwirken. In unserer Reihe sollen sie sichtbar werden – als gleichwertiger Bestandteil fachlicher Auseinandersetzung. Denn gerade junge Forscher*innen bringen eigene Fragen, kreative Zugänge und Offenheit mit. Diese neugierige Offenheit, gepaart mit einer Orientierung an den Kriterien für gute wissenschaftliche Praxis, ist der Anspruch der Kleinen Kieler Beiträge.

Dass unsere Reihe mit einer Masterarbeit eröffnet wird, unterstreicht diesen Anspruch auf ideale Weise. Im folgenden, jedem Band dieser Reihe vorangestellten, offenen Editorial Review, wollen wir transparent machen, welche Aspekte der Arbeit uns dazu bewogen haben, sie in die Reihe mit aufzunehmen, und sie kritisch im Diskurs verorten.

Lisanne Pitzen hat sich in der vorliegenden Arbeit mit einem Thema beschäftigt, das von großer fachdidaktischer und unterrichtspraktischer Relevanz ist: der Frage, wie lernförderliche Arrangements (zum Schreiben von Texten) mit Hilfe von KI gestaltet werden können. Der Fokus auf Feedback ist dabei von besonderer Bedeutung, denn formatives Feedback (auf Schüler*innen-texte) ist im Kontext von Schule (also einer großen Gruppe von Lernenden) durch die Lehrkraft allein kaum zu realisieren, ist aber für gelingende Lernprozesse von großer Bedeutung. KI kann hier eine enorme Hilfestellung (und Arbeitserleichterung) sein, wie es auch bereits einige Gründer*innen entsprechender KI-Tools in der Bewerbung ihrer Technik beschreiben (vgl. u. a. Hendrik Haverkamp für FelloFish). Das gilt allerdings nur, wenn das Feedback angemessen ist, also fachlich richtig, adressatenorientiert und idealerweise adaptiv umgesetzt wird. Auf den letzten Aspekt weisen u. a. die Arbeiten von Sara Rezat hin, die z. Zt. ein KI-Tool mit der Hilfe von Schüler-texten zum argumentativen Schreiben trainiert.

In der Masterarbeit wird entsprechend KI-Feedback in Einsatz gebracht, wobei sich Lisanne Pitzen vor allem auf die Wirkung des Feedbacks bei den Schreibenden interessiert, und mit einer anderen Form von Feedback vergleicht, der Rückmeldung, die Peers geben. Die Ergebnisse können daher nicht nur Hinweise auf die Qualität und insbesondere die Einschätzung des Feedbacks durch Schüler*innen liefern, sondern machen auch Überlegungen anknüpfbar, die sich zum Feedback von Peers stellen, wie sie kürzlich in der Dissertation von Michaela Mörs (allerdings für ältere Lernende) diskutiert wurden. Die Arbeit ist empirisch ausgerichtet und setzt eine kleine Studie von je drei Unterrichtsstunden in zwei Klassen am Gymnasium an, Schüler*innen einer sechsten und einer siebten Klasse.

Zentral sind dabei folgende Fragen, denen Lisanne Pitzen nachgeht: „Wie unterscheiden sich die Feedbackarten hinsichtlich ihrer Affektwirkung, der Akzeptanz durch die Schüler*innen, der wahrgenommenen Fairness und

Nützlichkeit und der Identifikation von Fehlerstellen und Verbesserungsvorschlägen“, „Welches Feedback wird von den Schüler*innen bevorzugt“, „Welche Komponenten der einzelnen Feedbacks finden verstärkt Beachtung bei den Schüler*innen?“ und „Ändert sich ihre Meinung, wenn sie über die Quelle des Feedbacks informiert werden?“.

Insgesamt nehmen 21 Schüler*innen aus Klasse 6 und 26 Schüler*innen aus Klasse 7 an der Erprobung teil: Alle schreiben einen Sachtext, geben ein Peer-Feedback und beurteilen die beiden Feedbacks zu ihrem Text (Peer-Feedback und FelloFish-Feedback) und füllen dazu einen Fragebogen aus, der am Ende auch die Auflösung gibt, welches Feedback von der KI und welches von einem Peer verfasst wurde. Interessant ist einerseits, dass Peer-Feedback wird (insbesondere in Klasse 7) sehr positiv eingeschätzt, andererseits bleibt die Einstellung (in jedem Fall) bestehen, wenn die Feedbackform aufgelöst wird, es verschieben sich aber die Gründe, die für die Entscheidung herangezogen werden. Tiefere Analysen zeigen, dass der insgesamt eher negative Eindruck bezüglich des KI-Feedbacks in Klasse 7 vor allem darin begründet ist, dass die KI eine fehlerhafte Einschätzung vornimmt. Diese Fehler überstrahlen auch alle anderen (positiven) Aspekte; so wird das KI-Feedback insgesamt durchaus als detailreicher eingeschätzt und auch eher in der Lage, Verbesserungsvorschläge, Tipps und/oder Beispiele zu geben. Ein Schluss, der sich aus diesen Befunden auch ableiten lässt, ist, dass Peer-Feedback, selbst wenn es nicht völlig überzeugt, doch durchaus zielführend sein kann – eine angesichts knapper Ressourcen durchaus wichtige Schlussfolgerung. Zudem können die Feedbackgebenden beim Peer-Feedback auch selbst etwas lernen, beim KI-Feedback fehlt diese Lernerfahrung.

Das erscheint auch für die weitere Entwicklung entsprechender Tools eine wichtige Erkenntnis zu sein. Das Vertrauen in die Tools ist entscheidend – auch für die Akzeptanz von Feedback. Dieses Vertrauen ist durchaus sensibel und lässt sich vergleichsweise schnell erschüttern.

Unabhängig von diesem inhaltlichen Ertrag lässt sich die Qualität der Arbeit durchaus so zusammenfassen: Lisanne Pitzens Masterarbeit überzeugt vollständig. Das Thema und die Fragestellung sind neu und innovativ, die fachliche Grundlegung ihrer Arbeit ist souverän, sprachlich durchgehend präzise und sehr klug aufbereitet und zusammengefasst. Bemerkenswert

ist die durchgehende Konzentration und Fokussierung auf den Gegenstand (Feedbackrezeption). Die empirische Umsetzung ist methodisch reflektiert und entspricht durchgehend empirischen Standards.

Wir hätten uns keinen besseren Auftakt für die *Kleinen Kieler Beiträge zu Künstlicher Intelligenz* wünschen können. Lisanne Pitzens Arbeit zeigt exemplarisch, was wir erreichen möchten: wissenschaftliche Strenge, gesellschaftliche Relevanz und die Öffnung des Diskurses für neue, zukunftsweisende Perspektiven.

Kirsten Schindler & Nicolaus Wilder

Inhalt

Vorwort der Herausgebenden	4
1 Einleitung	9
2 Theoretischer Rahmen: Feedback und (Deutsch-)Unterricht	12
2.1 Feedback: Eine Definition	12
2.2 Feedback im schulischen Schreibprozess	17
2.2.1 Charakteristika schulischen Schreibens	18
2.2.2 Notwendigkeit formativen Feedbacks	23
2.3 Formen von Feedback	26
2.3.1 Feedback durch die Peers	28
2.3.2 Feedback durch die Künstliche Intelligenz	34
2.4 Vom Feedback zu den Lernenden: Feedbackrezeption	46
2.5 Zusammenfassung und Einblick in den aktuellen Forschungsstand	52
3 Rezeption von KI- und Peerfeedback im Deutschunterricht der Klasse 6 und 7	57
3.1 FelloFish als Tool für KI-basiertes Feedback	57
3.1.1 Funktionsweise und Aufbau	59
3.1.2 Einsatzmöglichkeiten im Schreibunterricht	60
3.2 Forschungsfragen und Hypothesen	63
3.3 Methodisches Vorgehen	65
3.3.1 Beschreibung der Stichprobe	65
3.3.2 Zeitlicher Ablauf	66
3.3.3 Material und Messinstrumente	69
3.4 Ergebnisse	79
3.4.1 Klasse 7	79
3.4.2 Klasse 6	96
3.5 Diskussion	111
3.6 Limitationen	119
4 Fazit	121
5 Literaturverzeichnis	123
6 Anhang	132

1 Einleitung

Spätestens seit Hattie und Timperley (2007) gilt Feedback als einflussreicher Faktor, der den Lernprozess von Schüler*innen beeinflussen, und sie so im Kompetenzerwerb unterstützen kann (vgl. Hattie & Timperley, 2007, S. 102). Auch im Deutschunterricht kann Feedback dabei als wirksames Tool auftreten und sowohl im Rahmen der Textproduktion als auch -rezeption eingesetzt werden. Die hier zu erwerbende literale Kompetenz ist dabei hoch komplex, fachübergreifend relevant und kann auch über den unmittelbaren Schulkontext hinaus wirken. Sie ist »Grundvoraussetzung für Bildung und gesellschaftliche Teilhabe« (Stanat et al., 2019, S. 5). Umso wichtiger ist es, Schüler*innen im Erwerb von Lese- und Schreibfertigkeiten nicht nur anzuleiten, sondern sie in ihrem Prozess zu begleiten.

Gerade im Schreibunterricht der Schüler*innen kann Feedback entscheidend sein, so steht doch beispielsweise bereits am Ende der Sekundarstufe I das Ziel der selbstständigen Gestaltung des eigenen Schreibprozesses (vgl. KMK, 2022, S. 21). Damit dieses Ziel erreicht werden kann, ist es essentiell, Schüler*innen nicht erst am Ende ihres Lernweges über ihre Leistung zu informieren, sondern ihnen kontinuierlich Rückmeldungen – in anderen Worten: formatives Feedback – zu geben, um den individuellen Lernfortschritt zu fördern (vgl. Shute, 2008).

Das Feedback sollte dabei nicht nur möglichst zielgerichtet und präzise formuliert sein, sondern auch »bei den Lernenden eine aktive Informationsverarbeitung in Gang setzen« (Hattie, 2015, S. 211). Die Schüler*innen nehmen somit als Rezipierende des Feedbacks eine Schlüsselposition ein. Feedback wird nicht unmittelbar und unverändert von den Gebenden zu den Empfangenden übertragen, sondern von den Schüler*innen durch persönliche Wahrnehmungen und Interpretationen gefiltert. In anderen Worten: »Wichtiger als das Feedback, das ich gebe, ist das Feedback, das der Lernende erhält.« (Zierer, 2019, S. 7). Dies bedeutet, dass für die Umsetzung des Feedbacks die Rezeption der Lernenden entscheidend ist: Ist der*die Schüler*in mit dem Feedback zufrieden? Erachtet er*sie das Feedback als hilfreich? Und welche Emotionen löst das Feedback bei ihm*ihr aus?

Ein wie hier dargestelltes formatives Feedback lern- und schreibprozessbegleitend zu geben, kann für Lehrkräfte einen hohen Zeitaufwand bedeuten. Dieser wächst mit zunehmender Klassengröße oder Grad an Textkomplexität weiter (vgl. u. a. Lipnevich et al., 2016, S. 175; Banihashem et al., 2024, S. 2). Neben einem Selbstfeedback kann in Lehr-/Lernkontexten daher auch zu Peerfeedback – einem Feedback unter den Lernenden – gegriffen werden. Richert (200, S. 49) beschreibt dies als oftmals »regelmäßige Praxis im Unterricht«.

Mit dem Einzug von Künstlicher Intelligenz (KI) in (schulische) Lehr-Lern-Kontexte gewinnt auch KI-basiertes Feedback zunehmend an Bedeutung. Es verspricht auch abseits des Einflussbereichs der Lehrkraft passgenau und ohne örtliche oder zeitliche Beschränkungen zu fördern und stellt damit eine potenzielle Alternative oder Ergänzung zum etablierten Peerfeedback dar.

Im Rahmen dieser Masterarbeit sollen diese beiden Feedbackquellen im Hinblick auf die Rezeption durch Schüler*innen der Sekundarstufe I (genauer: einer sechsten und einer siebten Klasse eines Gymnasiums in Nordrhein-Westfalen) vergleichend untersucht werden. Die durchgeführte Studie zielt dabei einerseits darauf ab, den bestehenden Kenntnisrahmen zu KI-gestütztem Feedback systematisch zu erweitern. Andererseits bietet sie empirisch gestützte Einblicke in die Rezeption von Schüler*innen in Feedbackprozessen, die insbesondere für die Weiterentwicklung theoretischer Konzepte in der Grundlagenforschung relevant sind.

Im Zentrum der empirischen Erhebung steht dabei die Fragestellung, ob und wie sich die Rezeption von Peerfeedback und KI-gestütztem Feedback bei Schüler*innen der Sekundarstufe I unterscheidet, wenn letztere die Quelle des Feedbacks nicht kennen. Genauer sollen dabei die Bereiche wahrgenommene Fairness, Nützlichkeit, Akzeptanz und Affektwirkung fokussiert werden. Darüber hinaus liegt ein weiterer Untersuchungsschwerpunkt auf den Komponenten der jeweiligen Feedbacks, die Schüler*innen besonders beachten bzw. durch diese hervorgehoben werden, sowie der Wahrnehmung der Schüler*innen der Auflösung im Hinblick auf die Feedbackquellen.

Um während der Erhebung neben dem Peerfeedback auch ein möglichst schulalltagsnahes KI-Feedback zu gewährleisten, wurde das DSGVO-konforme Tool *FelloFish* (FelloFish GmbH, 2025, <https://www.fellofish.com>) ausgewählt. Als Messinstrument wurden Items des *Feedback Perception Questionnaire* von Strijbos et al. (2010) übernommen und inhaltlich durch offene Fragen ergänzt. Die Auswertung dieser erfolgte mittels inhaltlich strukturierender qualitativer Inhaltsanalyse nach Kuckartz (hier: Kuckartz/Rädiker, 2022).

Im Folgenden wird in einem ersten Schritt der theoretische Rahmen der Arbeit dargelegt. Nach einer Annäherung an den Begriff *Feedback*, wird dieser in den Kontext des Schreibunterrichts gesetzt. Hierbei sollen sowohl Einblicke in die Charakteristika schulischen Schreibens gegeben als auch auf die Notwendigkeit von formativem Feedback in ebendiesem Prozess eingegangen werden.

In einem nächsten Schritt werden Peer- und KI-Feedback als Formen von Feedback in ihren Potenzialen, Limitationen und Rahmenbedingungen vorgestellt und mit aktuellen Forschungsbefunden verknüpft. Im Anschluss wird der Fokus auf die Feedbackrezeption gelegt, die neben den Feedbackformen das Kernstück dieser Untersuchung abbildet. Nach einer abschließenden Zusammenfassung der theoretischen Fundierung, wird die durchgeführte Studie thematisiert. Zuerst wird dabei in *FelloFish* als KI-Tool eingeführt. Im Anschluss erfolgt eine Vorstellung und Einbettung der Forschungsfragen und Hypothesen sowie des methodischen Vorgehens. Nach einer Darlegung der Ergebnisse nach Klassen erfolgt eine Diskussion und das Aufzeigen möglicher Limitationen der Studie. Final wird ein Fazit und ein Ausblick auf zukünftige Forschungsmöglichkeiten gegeben.

2 Theoretischer Rahmen: Feedback und (Deutsch-)Unterricht

Die Fähigkeit, mit Texten umgehen zu können, geht einher mit der Möglichkeit zur gesellschaftlichen Teilhabe: Die Schule – und insbesondere der Deutschunterricht – nimmt in ihrem Anspruch der Entwicklung produktiver und rezeptiver Textkompetenz so eine essentielle Schlüsselposition ein (vgl. Rezat, 2022, S. 178f.). Es gilt, die Entwicklung dieser Kompetenzen im Unterricht aktiv auch durch Feedback zu unterstützen, welches Jäger (2019, S. 11) als »Ausgangspunkt jeder Förderung im schulischen Kontext« bezeichnet. Verankerung findet Feedback in diversen lerntheoretischen Grundlagen. So wird es im Behaviorismus zumeist als Verstärker, im Kognitivismus als Information und im Konstruktivismus als Angebot, das eigenaktives Verarbeiten benötigt, verstanden (vgl. Haag & Götz, 2019, S. 14).

Effektives Feedback wird dabei zu einem komplexen Prozess. Es benötigt sowohl auf Seite der Sendenden als auch auf der Seite der Empfangenden eine Vielzahl an Kompetenzen, die sowohl das Erstellen als auch das Verarbeiten von Feedback im konkreten Lehr-Lernsetting und darüber hinaus beinhalten (Wrede et al., 2023, S. 137). Um diesen Feedbackprozess im Deutschunterricht besser erfassen zu können, wird im Folgenden auf Basis theoretischer Modellierungen eine allgemeine Definition von Feedback angestrebt. Im Anschluss wird der Fokus auf den Schreib- bzw. Deutschunterricht und den Schreibprozess als Schlüsselstelle in diesem gerichtet. Nach einer exemplarischen Darlegung einzelner Facetten schulischen Schreibens wird die Notwendigkeit formativen Feedbacks im Deutschunterricht adressiert. Im Anschluss wird der Blick auf das aktive Rezipieren der Lernenden im Feedbackprozess gerichtet.

2.1 Feedback: Eine Definition

Folgt man den Ausführungen von Haag und Götz (2019, S. 14), ist Feedback im Lehr-/Lernsetting »eine bewusste, auf Daten basierende Rückmeldung an eine Person bzw. Personengruppe zu ihrem vorherigen Verhalten«. Die

Begriffe *Feedback* und *Rückmeldung* werden dabei synonym verwendet¹ und inkludieren sowohl kurze Ja-/Nein-Antworten als auch ausführlichere oder elaboriertere Formen des Feedbacks.

Im Allgemeinen wird zwischen summativem und formativem Feedback unterschieden. Summative Feedbacks sind dabei am Ende eines Prozesses situiert: Sie zeigen auf, wo die Lernenden am Ende einer Unterrichtseinheit im Hinblick auf einzelne Lernziele stehen, und begegnen Lernenden oft in Form von Klassenarbeiten oder Klausuren (vgl. Wendt, 2023, S. 40). Formatives Feedback ist auf der anderen Seite nicht am Ende des Lernprozesses, sondern inmitten von diesem positioniert und kann ihn daher beeinflussen: Formatives Feedback dient der Unterstützung des Lernprozesses während des Lernens: »Es gibt den Lernenden Informationen darüber, wo sie im Lernprozess stehen und hilft ihnen, ihr Wissen und ihre Fähigkeiten zu verbessern.« (Wendt, 2023, S. 39; vgl. auch Philipp, 2023, S. 10). Feedback ist dabei eng an Diagnostik und Evaluation geknüpft. Erst durch die Kommunikation mittels Feedback, werden Ergebnisse aus der Diagnostik und Evaluation für die Lernenden konkret zugänglich (vgl. Jäger, 2019, S. 11).

Der Einsatz von Feedback in konkreten Unterrichtszusammenhängen findet Verankerung in einer Vielzahl an Modellen und Theorien. Zwei davon sollen im Folgenden exemplarisch vorgestellt werden, um ein besseres Verständnis des Feedbackbegriffs zu ermöglichen.

1996 entwickeln Kluger und DeNisi die *Feedback Intervention Theory* (FIT), eine Theorie, die Feedbackinterventionen als Auslöser für Verhaltensänderungen basierend auf einer Verlagerung des Aufmerksamkeitsfokus versteht. Kluger und DeNisi nehmen an, dass Verhalten durch einen Vergleich der erhaltenen Rückmeldung bzw. des erhaltenen Feedbacks mit den

1 Demgegenüber zeigen Ditton und Müller (2014) in ihren Ausführungen Unterschiede zwischen Feedback und Rückmeldung auf. So geschieht für sie ein Feedback beispielsweise »eher verhaltensnah (proximal)« und »eher bezogen auf [ein] Individuum«, eine Rückmeldung dagegen »eher verhaltensfern (distal)« und »eher bezogen auf Systeme oder Systemeinheiten« (Ditton & Müller, 2014, S. 14). Die Autoren ergänzen hierbei jedoch auch, dass beide Begriffe auf einem gemeinsamen Kontinuum zu verorten sind und dabei deutliche Schnittstellen aufweisen (vgl. Ditton & Müller, 2014, S. 15). Im Folgenden sollen die Begriffe *Feedback* und *Rückmeldung* synonym verwendet werden.

angestrebten Zielen und Standards reguliert wird. Wird eine Diskrepanz zwischen dem rückgemeldeten eigenen Leistungsstand und dem des Standards oder Ziels festgestellt, kann es zu einer von vier Bewältigungsmöglichkeiten kommen: Das eigene Verhalten kann geändert werden, der Standard oder das Ziel verändert werden, der Standard oder das Ziel aufgegeben oder das Feedback abgelehnt werden (vgl. Kluger & DeNisi, 1996, S. 260). Da die Aufmerksamkeitskapazität der Feedback-Rezipient*innen begrenzt ist, wird nur auf die Diskrepanzen zwischen der eigenen Leistung und der Norm bzw. dem Ziel reagiert, die die Aufmerksamkeit der Rezipient*innen auf sich ziehen.

Im Rahmen der FIT sind Ziele und Standards hierarchisch organisiert: Während Meta-Aufgabenprozesse, die das Selbst betreffen, weiter oben platziert sind, liegen Aufgabenmotivationsprozesse, die die Fokusaufgabe betreffen, mittig und Aufgabenlernprozesse, die die Aufgabendetails der Fokusaufgabe betreffen, unten (vgl. Kluger & DeNisi, 1996, S. 261–263). Der Aufmerksamkeitsfokus ruht meistens auf der mittleren Ebene. Durch Feedbackinterventionen kann der Ort der Aufmerksamkeit verändert werden, sich also auch auf Meta-Aufgabenprozesse oder Aufgabenlernprozesse richten (vgl. Kluger & DeNisi, 1996, S. 262). Folgt man dem Modell, ist Feedback dann besonders förderlich, wenn die Intervention mittig bis hierarchieniedriger auf Ebene der Aufgabenmotivationsprozesse oder Aufgabenlernprozesse angesetzt wird (vgl. Kluger & DeNisi, 1996, S. 267).²

Auch wenn die hier angeführte Theorie gerade für den Bildungsbereich kritisch eingeordnet werden muss – so führen beispielsweise Lipnevich et al. (2016, S. 171) u. a. an, dass sowohl der Kontext des Feedbacks als auch die Eigenschaften der Feedbackkämpfenden hier nicht berücksichtigt werden –, kann die FIT als Ausgangspunkt in der Annäherung an den schulischen Feedbackprozess betrachtet werden.

Ein Ansatz, der dem Anspruch der schulischen Passung nachgeht, lässt sich bei Hattie und Timperley (2007) bzw. Hattie (2015) finden (vgl. auch

2 Lenkt die Feedbackintervention die Aufmerksamkeit auf die Meta-Aufgabenebene, also die selbstbezogenen Prozesse, kann sich dies negativ auf die Leistung auswirken, da es kognitive Ressourcen bindet und von der eigentlichen Aufgabe ablenkt (vgl. Kluger & DeNisi, 1996, S. 266).

Lipnevich et al., 2016, S. 171f.): Laut den Autor*innen gehört Feedback zu den einflussreichsten Faktoren, die im Lernprozess von Schüler*innen wirken (vgl. Hattie & Timperley, 2007, S. 102). Unter Feedback verstehen sie hier »eine Information [...], die von einem Akteur (z. B. Lehrperson, Peer, Buch, Eltern oder die eigene Erfahrung) über Aspekte der eigenen Leistung oder des eigenen Verstehens gegeben wird.« (Hattie, 2015, S. 206, vgl. auch Hattie & Timperley, 2007, S. 81).

Feedback ist eine Folge von oder eine Reaktion auf eine vorherige Leistung, es ist »das, was als Zweites passiert« (Hattie, 2015, S. 211). Dabei ist das Grundziel, Schüler*innen (aber auch Lehrer*innen) zu zeigen, wo diese im aktuellen Lernprozess stehen (vgl. Hattie & Timperley, 2007, S. 89) und die Diskrepanz zwischen Soll (dem Lernziel) und Ist (dem aktuellen Lernstand) zu reduzieren (vgl. Hattie & Timperley, 2007, S. 86).

Damit Feedback besonders effektiv wirken kann, sollte es

klar, zweckgerichtet, sinnvoll und mit dem Vorwissen der Lernenden kompatibel sein. Außerdem muss es logische Verbindungen bieten. Es muss bei den Lernenden eine aktive Informationsverarbeitung in Gang setzen, über eine geringe Aufgabenkomplexität verfügen, sich auf spezifische und klare Ziele beziehen und es darf keine ernsthafte Bedrohung der Person auf der Ebene des Selbst darstellen. (Hattie, 2015, S. 211)

Feedback geben erfordert Fingerspitzengefühl: Es muss für die Lernenden verständlich sein und ihnen die Möglichkeit geben, an ihr Vorwissen anzuknüpfen, sie denken zu lassen, ohne ihnen eine vorgefertigte Antwort zu geben. Es ist gebunden an den Unterrichtskontext und der Zeitpunkt seiner Verwendung sollte mit Bedacht gewählt werden. So ist ein sehr frühes Einsetzen komplexen Feedbacks zum Beispiel in der Erarbeitungsphase oft wenig hilfreich, da die behandelten Themen erst noch vertieft werden müssen (vgl. Hattie, 2015, S. 211).

Charakteristisch für das Feedbackverständnis nach Hattie und Timperley (2007) sind die von ihnen formulierten Fragen, die ein »gutes« Feedback bedingen und eine starke Verknüpfung aufweisen. Hier wird unterschieden zwischen:

- Feed Up: *Wohin gehe ich?* bzw. *Wohin gehst du?*
- Feed Back: *Wie komme ich voran?* bzw. *Wie kommst du voran?*
- Feed Forward: *Wohin geht es danach?* bzw. *Wohin geht es als Nächstes?*

(Hattie, 2015, S. 209–210, vgl. auch Hattie & Timperley, 2007, S. 86)

Das *Feed Up* beinhaltet Ziele oder Erfolgskriterien, die Lernenden durch eine möglichst konkrete und klare Formulierung eine Orientierung geben können (vgl. Hattie & Timperley, 2007, S. 88f.). Auch die anderen beiden Fragen dürfen in ihren Antworten nicht zu vage für Schüler*innen sein. Auf die Frage des *Feed Forwards* sollte den Lernenden so kein einfaches »weiter« als Antwort, sondern möglichst konkrete Informationen an die Hand gegeben werden (vgl. Hattie & Timperley, 2007, S. 90).

Folgt man dem Feedbackmodell von Hattie (2015, S. 209) weiter, so wirkt jede dieser Feedback-Fragen auf vier verschiedenen Ebenen. Es gibt aufgabenbezogenes Feedback, welches sich auf das Lerner*innenprodukt bezieht, und prozessbezogenes Feedback, welches den Informationsverarbeitungs- oder Lernprozess fokussiert. Hinzu treten das Feedback zur Selbstregulation, in dem Elemente wie Selbstüberwachung oder -regulationsfähigkeiten situiert sind, die das selbstständige Auseinandersetzen mit der Aufgabe fördern, und das persönliche Feedback. Dieses zielt z. B. in Form von Lob auf die Lernenden als Person ab und wird daher als weniger nützlich verstanden (vgl. Hattie, 2015, S. 210; Hattie & Timperley, 2007, S. 90).³

Auf welcher Ebene das Feedback ansetzt, hängt in erster Linie von der konkreten Lernumgebung ab: Auch, wenn ein Feedback als besonders effektiv aufgefasst wird, wenn es Schüler*innen von der Aufgabe zum Prozess und vom Prozess zur Selbstregulation führt (Hattie & Timperley, 2007, S. 91), muss in erster Linie auf ein lerner*innen- und kontextabhängiges Timing geachtet werden. So geschieht Lernen, folgt man Hattie im Interview mit Klaus Zierer, »selten linear« (Zierer, 2019, S. 9).

3 Für eine Begriffsdefinition dieser Ebenen im Kontext des Textfeedbacks, vgl. Graham (2018, S. 150).

Die Feedbackebenen müssen dabei nicht klar voneinander getrennt stehen, sondern treten auch in Kombination mit einander auf. Wisniewski et al. (2019) zeigen in ihrer Metanalyse dabei, dass ein *high-information feedback* (ein Feedback, welches Informationen über Aufgabe und Prozess, manchmal in Kombination mit Selbstregulation, enthält) besonders effektiv wirken kann: »Feedback is more effective the more information it contains. Simple forms of reinforcement and punishment have low effects, while high-information feedback is most effective." (Wisniewski et al., 2019, S. 12). Mittels dieses *high-information feedbacks* können Schüler*innen nicht nur die eigenen Fehler erkennen, sondern auch deren Ursprung verstehen und für Folgeaufgaben beachten (vgl. Wisniewski et al., 2019, S. 12).

Wird es richtig eingesetzt, kann Feedback somit als effektives Tool zwischen den Akteur*innen im Bildungskontext angewendet werden. Dabei ist Feedback für alle Beteiligten ein komplexer Vorgang: Es handelt sich nicht (nur) um eine reine Informationsweitergabe von der einen zur anderen Instanz, sondern um ein vielschichtiges Agieren im Lehr-Lernkontext, das von den beteiligten Akteur*innen und der äußeren Aufgaben- oder Schulfachumgebung beeinflusst wird. Während also der Grundgedanke des Feedbacks, d. h. die hier skizzierten Wirkungsmechanismen, fächerübergreifend verstanden werden kann, kann sich der konkrete Einsatz des Feedbacks, z. B. der inhaltliche Bezug oder der Umfang, von Fach zu Fach unterscheiden.

2.2 Feedback im schulischen Schreibprozess

In der Schule wird Feedback ein Teil des Lernprozesses, Unterricht und Feedback werden miteinander verwoben. Dies bedeutet, dass nicht nur der Unterricht das Feedback lenkt, sondern das Feedback auch den Unterricht beeinflussen kann (vgl. Hattie, 2015, S. 207). Hierbei muss das Feedback zusätzlich zur Thematisierung der Diskrepanz zwischen Lernstand und Lernziel (vgl. Kapitel 2.1) auch Informationen bereitstellen, die eine direkte Anknüpfung an die Aufgabe oder den Lernprozess der Schüler*innen bedeuten (vgl. Hattie, 2015, S. 207).

Der Umgang mit Texten stellt im schulischen Kontext eine besondere Herausforderung für Schüler*innen dar. Der Deutschunterricht bildet dabei den Ausgangspunkt von Textrezeption und -produktion. Der Erwerb dieser Kompetenzen geht jedoch über die Fachbindung hinaus. So schreibt Rezat (2022, S. 193f.), dass »Texte [...] in allen Fächern gelesen und geschrieben [werden] und die damit verbundenen (bildungs-)sprachlichen Textkompetenzen [...] zentral für das Lernen in allen Fächern [sind]«. Aber auch darüber hinaus ist und bleibt der Umgang mit und die gezielte Nutzung von Texten unverzichtbar und stellt eine »Grundvoraussetzung für Bildung und gesellschaftliche Teilhabe« (Stanat et al., 2019, S. 5) dar.

Im Deutschunterricht treffen Schüler*innen auf eine Vielzahl an textuellen Anforderungen, die sie bewältigen müssen. So begegnen ihnen hier sowohl prototypische bzw. traditionellere Texte, die ein enges Textverständnis aufweisen, wie z. B. Sachtexte oder Argumentationen, aber auch multimediale Texte, die ein weites Textverständnis bedingen, wie PowerPoint-Präsentationen (vgl. Rezat, 2022, S. 185). Diese Komplexität und Vielfalt an Anforderungen benötigen im schulischen Alltag eine passende (formative) Begleitung durch Feedback (vgl. Jansen et al., 2024, S. 81; Graham, 2018, S. 145). Im Folgenden werden in einem ersten Schritt die Grundzüge schulischen Schreibens aufgegriffen, bevor im Anschluss die Notwendigkeit von Feedback – gerade im Überarbeitungsprozess – schulischen Schreibens aufgegriffen wird.

2.2.1 Charakteristika schulischen Schreibens

Im Alltag geht die Produktion eines Textes mit der Intention zur Kommunikation einher: »Texte schreiben bedeutet, ein kommunikatives Ziel mithilfe eines Textes zu erreichen, also beispielsweise *jemanden einladen* oder *jemanden von etwas überzeugen*« (Stanat et al., 2019, S. 7). Das Formulieren eines Textes ist dabei hoch komplex und wird von Gredel et al. (2024, S. 385) als »Schaltstelle« zwischen Gedanken und Sprache verstanden:

Schreibende müssen [...] auf Ebene des Mediums, der Orthographie, der Lexik, der Syntax, des Textmusters und der Leserorientierung deklaratives, Problemlöse-, prozedurales und metakognitives Wissen ausbilden, um über Schreibkompetenz zu verfügen. (Gredel et al., 2024, S. 385)

Schreibkompetenz als solche setzt sich aus den Elementen *Vertexten* und *Verschriften* zusammen. Unter *Vertexten* werden hierarchiehöhere Prozesse gefasst. Es geht nicht nur um das Verstehen auf inhaltlicher Ebene, sondern auch um Vorgänge der Steuerung und Anpassung. So werden hier zusätzlich zur Frage der Textsorte auch Fragen nach Funktion, Intention und Adressat*in geklärt (vgl. Sturm, 2022, S. 216f.). Das *Verschriften* ist hierarchieniedriger angesetzt. Hier werden die Gedanken und Ideen in Sprache verwandelt und »orthografisch wie auch grammatisch möglichst korrekt verschriftet« (Sturm, 2022, S. 217). Schreibkompetenz meint somit »die Fähigkeit zur Produktion von Texten« (Becker-Mrotzek & Böttcher, 2015, S. 49) und inkludiert dabei die Befähigung, »sprachliche Äußerungen so zu konzipieren, dass sie aus sich heraus verständlich sind, also über Raum und Zeit hinweg zerdehnt werden können« (Becker-Mrotzek & Böttcher, 2015, S. 49).

Das Entstehen eines Textes bedarf eines vielschichtigen Prozesses. Baron (2023) verknüpft das Schreiben mit dem Begriff »journey«. Schreiben wird hier zu einer Reise, die zwar zu einem finalen Endprodukt – der »destination« – führt, in ihrer Prozesshaftigkeit aber keineswegs vernachlässigt werden sollte (Baron, 2023). Auch bei Philipp (2015) nehmen Schreibprozesse eine Sonderstellung ein. Er bezeichnet sie als »Herz des Schreibens und damit der Schreibkompetenz« (Philipp, 2015, S. 10).

Bei dem Schreibprozess handelt es sich um einen zeitintensiven Prozess. Dies liegt unter anderem daran, dass vor dem Schreiben der Gedanke steht (s. oben *Vertexten* und *Verschriften*), aber auch daran, dass die mechanische Komponente des Niederschreibens zeitlich kostspieliger als das reine Sprechen ist (vgl. Becker-Mrotzek & Böttcher, 2015, S. 19). Ideen oder Vorstellungen müssen in sprachliche Ausdrücke übertragen und verschriftlicht werden. Der bereits erstellte Textentwurf beeinflusst dabei laufend die weiteren Schritte, was zu einer fortwährenden Veränderung der textuellen Basis und des Schreibprozesses führt (vgl. Becker-Mrotzek & Böttcher, 2015, S. 22).

Philipp (2015, S. 10) bezeichnet die Vorgänge Planen, Verschriften und Revidieren als »die drei zentralen Teilprozesse des umfassenden Schreibprozesses«. Dabei knüpft er u. a. an das Schreibprozessmodell von Flower und

Hayes (1980) bzw. Hayes (2012) an. Während ersteres die Prozesse noch explizit erwähnt (hier als *Planning*, *Modelling* und *Writing*, vgl. Flower & Hayes, 1980, S. 11), tauchen sie bei Hayes (2012) nicht mehr so distinkt und explizit nebeneinander auf (vgl. Hayes, 2012, S. 371). Auch bei Phillip (2015, S. 13) stehen im Schreibprozess Planen, Verschriften und Revidieren nicht trennscharf nebeneinander, sondern überschneiden sich und wirken aufeinander ein.

Aufschlüsse über die Verankerung des Schreibprozesses im schulischen Kontext geben u. a. Blicke in die Bildungsstandards der Kultusministerkonferenz für den ersten Schulabschluss im Fach Deutsch. Hier heißt es:

Sie [die Schüler*innen] gestalten ihren Schreibprozess weitgehend selbstständig und bauen ihre Fähigkeiten im Umgang mit den Anforderungen des Schreibens im Hinblick auf Texteigenschaften, allgemeine Texteigenschaften, auf Adressaten, Textfunktion, Schreibanlass und Textsorte, auf Formen kooperativen Schreibens, digitale Schreibwerkzeuge und Schreibstrategien weiter aus. (KMK, 2022, S. 21)

In den Bildungsstandards für den Mittleren Schulabschluss werden diese Kompetenzen dann ausgeweitet, sodass von Schüler*innen ein komplett selbstständiges Agieren im Schreibprozess erwartet wird (vgl. KMK, 2022, S. 21).

Die Anforderungen an das Schreiben verändern sich über Jahrgangsstufen hinweg. Während zu Beginn der Sekundarstufe I eher von Grundlagen im Bereich des Verschriftens ausgegangen wird, wird in den Folgejahren an diese Kenntnisse angeknüpft. Auch »zeigen sich zwischen den Fächern bezogen auf die Anforderungen wie auch die fachspezifischen Textsorten Unterschiede, die für den Schreiberwerb nicht unerheblich sind« (Sturm, 2022, S. 217) und so die Schüler*innen im Verstehen von Fremdtexten, aber auch im Verfassen eigener Textprodukte prägen.

Ortner (2006) grenzt elaboriertes (akademisches) Schreiben von spontanem (schulischem) Schreiben ab. Beim Spontanschreiben handelt es sich um eher kürzere Texte, die in Alltagssprache geschrieben werden und dabei in Alltagsthemen verankert sind. Zur Informationsbeschaffung werden meist eher keine weiteren Quellen herangezogen. Vielmehr zählt das aktuell

vorhandene, stabil verankerte Alltagswissen, das für eine eher pauschale Themenbehandlung verwendet wird. Hier erfolgt eine Textproduktion von Satz zu Satz, bei welcher der Text »am Stück« – in ein, zwei, drei Stunden (Ortner, 2006, S. 77) entsteht. Meist wird linear geschrieben, ohne eine weitreichende Vorausplanung. Der schnelle Vollzug ist dabei – laut Ortner – typisch in schulischen Schreibsituationen (vgl. Ortner, 2006, S. 79–82).

Hier scheint der Prozess beim Schreiben eine eher geringere Beachtung zu erhalten. Der Fokus liegt auf der Formulierung des Textes und weniger auf der Planung und der Überarbeitung. Dabei ist gerade eine ganzheitliche Be- trachtung des Schreibprozesses auch für Schüler*innen hilfreich. So wird mit fortschreitender Jahrgangsstufe das Schreiben zunehmend komplexer. Man denke beispielsweise an das materialgestützte Schreiben, das in seiner wissenschaftspropädeutischen Verankerung auch den Umgang mit Fremdquellen schulen soll (Decker et al., 2022). Oder aber, man schaue über den schulischen Tellerrand hinaus und denke an akademisches oder berufliches Schreiben.

Um auf diese Schritte vorzubereiten, scheint es ratsam, die Prozesshaftigkeit des Schreibens auch in jüngeren Jahrgangsstufen nicht aus den Augen zu verlieren und sie aktiv in das Blickfeld der Schüler*innen zu setzen. Dies kann beispielsweise bedeuten, die Entwicklung eines Schreibplans und damit die Phase der Planung in den Vordergrund zu stellen. Laut Becker-Mrotzek und Böttcher (2015, S. 21) kann dieser Plan selbst entwickelt oder aus bestehenden Vorlagen übernommen werden. Bei der Planung sind dann sowohl die eigene Zielsetzung als auch materielle Gegebenheiten entscheidend.

Auch ist es jedoch möglich, die Phase der Überarbeitung – die nicht zwingend nur am Ende des Schreibprozesses, sondern in Wechselwirkung mit den anderen Prozessen steht (vgl. Philipp, 2015, S. 10) – in den Fokus zu setzen. Das Überarbeiten muss dabei erst gelernt werden, es wird von Schreibenden nicht sofort beherrscht, »denn es erfordert eine Distanz zum eigenen Text« (Gredel et al., 2024, S. 387).

Jantzen (2003) definiert Überarbeitungskompetenz als Zusammenspiel folgender Komponenten: die Überarbeitungshaltung (diese wird gebildet aus dem Wissen, dass und auf welche Art ein Text bearbeitet werden kann, in

Kombination mit dem Willen, dies auch zu tun), die Beurteilungsfähigkeit (die Fähigkeit, den Text zum Beispiel auf bestimmte Zielvorgaben hin zu überprüfen), die Diagnosefähigkeit (die Fähigkeit, die Textteile zu erkennen, die überarbeitet werden sollten; hier werden sowohl Richtigkeitsnormen als auch Angemessenheitsnormen berücksichtigt) und die Alternativenkompetenz (die Fähigkeit, Lösungen für die zuvor identifizierten Probleme zu finden, und gleichzeitig mit den durch die Alternativen hervorgerufenen Änderungen umzugehen) (vgl. Jantzen, 2003, S. 114–116).

Hayes (2012) bezeichnet das Überarbeiten, bzw. Revidieren, als eine spezialisierte Schreibtätigkeit und betont den integrativen Charakter dieser:

Revising written text is also best thought of as a specialized writing activity. Revising is typically initiated by the detection of a problem in an existing text. It involves planning a solution to the problem (in written form or not), translating that solution into language, and transcribing that language into new text to replace the old text. In this view, revision, like planning, is seen not as a separate writing process parallel to the other writing processes [...] but rather as a special application of the writing model. (Hayes, 2012, S. 376)

Für Hayes gelten also (ähnlich Jantzen, 2003), die Identifizierung eines Problems, das Finden einer Lösung und schließlich das Übersetzen dieser Lösung in Sprache und das Verschriften als zentrale Bausteine der Überarbeitungskompetenz. Auch Philipp (2015) greift diesen prozesshaften Charakter des Überarbeitens auf und konkretisiert ihn in einem Drei-Schritt: das Problem erkennen, über Änderungen Entscheidungen treffen und schließlich »die Ausführung der Veränderungen (das Überarbeiten)« (Philipp, 2015, S. 28).

Wie an diesen Stellen deutlich wird, bedeutet Überarbeitung mehr als das »bloße« Umschreiben eines Textes. Vielmehr greifen hier einzelne Komponenten vorbereitend ineinander, die das finale Ändern des Textes erst möglich machen. So können auch Schüler*innen ihre Texte erst dann verändern, wenn sie zusätzlich zu einer Überarbeitungshaltung auch Problemstellen erkennen und mit diesen adäquat umgehen können. Im konkreten Unterrichtskontext bedeutet dies, sich von der Idee des Schüler*innentextes als »fehlerfreie Fassung« (Jantzen, 2003, S. 112) wegzubewegen und ihn

stattdessen in erster Linie als Entwurf wahrzunehmen, der überarbeitet werden kann, und diese Überarbeitung auch ins Zentrum aktiver Bemühungen zu setzen (vgl. Becker-Mrotzek & Böttcher, 2015, S. 135).

Gerade im Anfangsstadium, aber auch darüber hinaus, kann es Schreibenden schwerfallen, Problemstellen nicht nur zu identifizieren, sondern auch passende Lösungen zu finden. Hier kann Feedback von außen greifen. Philipp (2015, S. 30) spricht dabei von »externe[n] Leser[*innen] als Ressource [...], die quasi ›unbelastet‹ den Text sichten und einem Schreiber Rückmeldungen zu Problemen geben können« und auch Jantzen empfiehlt, Distanz zum eigenen Text aufzubauen oder Unterstützung von außerhalb zu suchen (vgl. Jantzen, 2003, S. 115f.). Daher wird im Folgenden die Notwendigkeit formativen Feedbacks im Deutschunterricht thematisiert.

2.2.2 Notwendigkeit formativen Feedbacks

Feedback ist ein wirksames Tool im Schreibunterricht. Im Hinblick auf die Textqualität von Schüler*innen stellen Graham et al. (2011, S. 17) eine durchschnittliche Effektgröße von 0,77 fest – dies entspricht einem mittleren bis starken Effekt.⁴

Im Rahmen seines Textes *Instructional Feedback in Writing* formuliert Graham (2018, S. 163) Empfehlungen für Feedback beim Schreiben. Laut ihm könne u. a. eine zu große Menge an Feedback sowohl für die Lehrkraft als auch die Lernenden hinderlich sein. Zudem sollte Feedback möglichst konstruktiv sein, die erbrachten Leistungen würdigen und dabei nicht als einmaliges Ereignis auftreten, sondern Schüler*innen in ihrem Arbeiten begleiten. Graham ergänzt seine Empfehlungen um weitere von Underwood und Tregidgo (2006). Hiernach sollte detailliertes Feedback nicht mit dem Geben einer Gesamtnote einhergehen, dafür jedoch möglichst spezifisch sein und je nach Schwerpunkt auf Oberflächen- oder Inhaltsebene ansetzen (vgl. Underwood & Tregidgo 2006, S. 90f.). Auch das bewusste

4 Im Rahmen der Überwachung des Schreibprozesses stellen die Autor*innen einen ebenfalls positiven und signifikanten, aber eher mildernden Effekt mit einer durchschnittlichen Effektstärke von 0,24 fest.

Inkludieren von Schüler*innen in Überlegungen hinsichtlich des Feedbackprozesses wird hier empfohlen (vgl. Underwood & Tregidgo 2006, S. 93).

Formative Diagnostik ist im Deutsch- bzw. Schreibunterricht essentiell:

Damit Lehrkräfte Schreibkompetenzen gezielt fördern können, ist eine Diagnostik erforderlich, die es erlaubt, individuelle Stärken und Schwächen der Schülerinnen und Schüler zu identifizieren. Sie sollte die Entwicklung der Schreibkompetenzen über die Zeit formativ, also begleitend zum Lehr-Lern-Prozess, erfassen – von der Grundschule bis zur 10. Klasse. Eine Beschränkung auf das Testen der Rechtschreibfähigkeiten ist dabei nicht ausreichend, um Schreibkompetenz zu erfassen. Anzustreben sind vielmehr Verfahren, die alle Aspekte von Schreibkompetenz in den Blick nehmen: Rechtschreibung, Schreibflüssigkeit und Textproduktion. (Stanat et al., 2019, S. 7)

Wer formativ begleitet, kann auch im Prozess selbst fördern – und nicht nur danach. An die Notwendigkeit formativer Diagnostik schließt sich auch die Notwendigkeit formativen Feedbacks. Dieses

dient der Unterstützung des Lernprozesses während des Lernens. Es gibt den Lernenden Informationen darüber, wo sie im Lernprozess stehen und hilft ihnen, ihr Wissen und ihre Fähigkeiten zu verbessern. Es kann von den Lehrenden, Lernenden, Peers oder externen Quellen kommen und in Form von Fragen, Anweisungen oder Leistungsbeobachtung gegeben werden. Formatives Feedback als Teil des formativen Assessments versucht, den Schüler*innen Rückmeldung zu Lernzielen, dem individuellen Lernstand in Bezug zu den jeweiligen Lernzielen sowie zu Strategien zu deren Erreichung zu geben. (Wendt, 2023, S. 39f.)

Prägend für den Begriff des formativen Feedbacks ist die Arbeit von Shute (2008). In ihrem Beitrag *Focus on Formative Feedback* formuliert sie Handlungsanweisungen für Lehrkräfte, mit denen das Geben von formativem Feedback gelingen kann. Hierzu gehört, dass das ausführliche Feedback in verarbeitbaren Häppchen kommuniziert, es spezifisch, klar und für Lerner*innen verständlich formuliert und dabei möglichst objektiv konstruiert werden soll. Zusätzlich sollte das Feedback nicht zu früh positioniert sein –

d. h., es sollte erst dann erfolgen, wenn die Lernenden bereits selbst tätig geworden sind (vgl. Shute, 2008, S. 177f.).

Durch formatives Feedback sollen Lernende nicht nur in der Diskrepanz zwischen Ist- und Soll-Zustand unterstützt, sondern auch auf kognitiver Ebene entlastet werden (vgl. Shute, 2008, S. 157). Zu vermeiden sind laut Shute dafür Feedbacks, die normative Vergleiche anstellen, das Selbstwertgefühl der Lernenden angreifen oder sie entmutigen, rein mündlich erfolgen und die Lernenden unterbrechen, wenn diese gerade aktiv in der Auseinandersetzung mit einer Aufgabe sind. Es sollte auch vermieden werden, schrittweise Hinweise zu geben, die schließlich mit der richtigen Antwort enden, oder sich ausschließlich auf die schriftliche Übermittlung in Textform zu verlassen (vgl. Shute, 2008, S. 178f.).

Im Kontext des Deutschunterrichts finden sich bei der Rückgabe einer Klassenarbeit oder Klausur häufig Erwartungshorizonte. Diese geben den Schüler*innen kriterienbasierte Auskünfte über deren Leistungen. Auch wenn diese Art des Feedbacks durchaus den Ansprüchen der klaren, spezifischen und leicht verarbeitbaren Darstellung genügen kann, wirkt sie eher retrospektiv und mindert dabei die Verarbeitung für den aktuellen Lernprozess. Auch Sturm (2022, S. 236) schreibt:

Kriterienbasiertes Beurteilen von Texten hat sich im Schreibunterricht weitgehend durchgesetzt, vor allem aber nach Abschluss eines Textproduktionsprozesses und damit im Sinne des summativen Beurteilens. Damit Schülerinnen und Schüler von Rückmeldungen jedoch profitieren, diese insbesondere auch auf ihre Texte umsetzen können, müssen Rückmeldungen formativ erfolgen.

Der Text wird in der summativen Bewertung nicht als Entwurf, sondern als Endprodukt wahrgenommen. Rückmeldungen dieser Form haben »eher Informations- und Beurteilungsfunktion und stehen somit der Weiterentwicklung und Förderung sprachlicher Fähigkeiten im Weg« (Jünger & Escher, 2023, S. 66). Schüler*innen werden dabei in ihrer Rolle als aktive Feedbackrezipient*innen gehemmt, sie sehen die Rückmeldung zwar, müssen sich mit dieser aber nicht unmittelbar auseinandersetzen (vgl. Jünger & Escher, 2023, S. 66).

Der Prozess, den Schüler*innen bei der Überarbeitung des eigenen Textes durchmachen, ist jedoch zentral. Jantzen (2003, S. 122) beschreibt diesen auch als »Beobachten – Verstehen – Lernen«. Zur prozessorientierten Förderung sollte das Überarbeiten daher nicht abseits des Lernsettings stehen, sondern in dieses integriert sein (vgl. auch Sturm, 2022, S. 217).

Formatives Feedback greift dabei nicht nur auf Aufgaben- oder Prozessebene, sondern kann auch darüber hinausgehen und im Kontext des Schreibunterrichts entscheidend für die Förderung selbstregulierten Schreibens sein (vgl. Philipp, 2023, S. 16). Letzteres definiert Philipp (2023, S. 14) als

die eigenständige Planung, Überwachung, Ausführung, Aufrechterhaltung, etwaige Anpassung, Beurteilung und Reflexion des eigenen Schreibens sowohl der Textprodukte als auch der Schreibprozesse.

Anders ausgedrückt können Schüler*innen durch diese Art von Begleitung in der eigenen Entwicklung des*r selbstständigen und selbsttätigen Schreibenden unterstützt werden. Damit kann die auf die individuellen Lernenden zugeschnittene Praxis des formativen Feedbacks im Deutsch- und Schreibunterricht zur Ausbildung von Kompetenzen beitragen, die sowohl das Fach- als auch das Schulspezifische übersteigen.

2.3 Formen von Feedback

Im Kontext des Schulunterrichts ist nach traditionellem Verständnis die Lehrkraft Dialog- und somit auch Feedbackpartner*in für die Lernenden (vgl. Jantzen, 2003, S. 123). Gerade in Fächern wie denen des Deutschunterrichtes, wo Langtexte produziert werden, die eines komplexen Feedbacks bedürfen, kann dies Lehrkräfte jedoch vor große Herausforderungen stellen. Beispielsweise müssen sie aufgrund der hohen Klassengröße ein individuelles und passgenaues Feedback einer Vielzahl von Schüler*innen zur Verfügung stellen. Hierbei handelt es sich um einen sowohl mental als auch

zeitlich kostspieligen Prozess (vgl. Banihashem et al., 2024, S. 2; Jansen et al., 2024, S. 81; Lipnevich et al., 2016, S. 175).⁵

Doch, auch wenn der Lehrkraft-Schüler*innen-Dialog zum Geben und Rezipieren von Feedback im (Deutsch-)Unterricht traditionell verankert ist, stellt er nicht die einzige mögliche Option dar. Folgt man Haag und Götz (2019, S. 14), so

ist ein Feedbackaustausch sinnvoll zwischen unterschiedlichen Konstellationen von Lehrkräften, Referendarinnen und Referendaren, Schulleitung, Schülerinnen und Schülern sowie Eltern. Auch ein Feedback an sich selbst ist inkludiert.

Und auch Wendt (2023, S. 39) betont, dass formatives Feedback »von den Lehrenden, Lernenden, Peers oder externen Quellen kommen« kann (vgl. auch Kapitel 2.2.2). Dies bedeutet, dass Feedback zwischen allen Akteur*innen des schulischen Alltags möglich ist und so neben dem Feedback durch die Lehrkraft auch Peers als Fremdquelle⁶ für Feedback im Unterricht eingesetzt werden können.

Peerfeedback stellt dabei jedoch nicht (mehr) die einzige Möglichkeit dar, Schüler*innen im konkreten Unterrichtssetting ein externes Feedback zu geben und gleichzeitig die Lehrkraft zu entlasten. So hat im Zuge der aktuellen Entwicklungen auch KI Einzug in den Klassenraum gefunden. Lehnens (2023, S. 29) spricht hierbei von einem »Wandel« in der Feedbackkultur: »KI-basierte Programme wie ChatGPT können künftig zu kompetenten Feedbackpartner:innen bei der gemeinsamen Arbeit am Text gemacht werden« (Lehnens, 2023, S. 29). Und auch Fleckenstein et al. (2023, S. 5) schreiben, dass automatisiertes formatives Assessment besonders lernförderlich im

-
- 5 Der hohe Korrekturaufwand einer Lehrkraft kann auch beispielweise bei summarischem Feedback zu einer Art Abstumpfen führen. So schreiben Lipnevich et al. (2016, S. 178): »A teacher grading multiple essays over a weekend may well miss or misread an excellent point that a student has made, and make a comment on the paper that is not relevant, or perhaps is quite simply wrong« (Für eine Differenzierung zwischen Textbeurteilung und Bewertung vgl. Becker-Mrotzek & Böttcher, 2015, S. 131).
 - 6 Fremdquelle meint hierbei, dass das Feedback von außen zu dem*der Schüler*in kommt. Sich letztere*r also nicht selbst ein Feedback gibt, sondern hier eine fremde Instanz involviert ist.

Rahmen der Schreibförderung bzw. der Förderung von schriftsprachlichen Kompetenzen wirken kann.

Wird also von Feedback durch Fremdquellen im unmittelbaren Kontext des Schreibunterrichts gesprochen, muss als Alternative zum Feedback durch die Lehrkraft neben Peerfeedback auch KI-basiertes Feedback herangezogen werden. Diese beiden Feedbackarten sollen im Folgenden im Detail thematisiert werden.

2.3.1 Feedback durch die Peers

Peerfeedback ist häufig gängige Praxis in Unterrichtskontexten. Hier können

Schüler[*innen] ihren Mitschüler[*inne]n ein Feedback zu deren vorherigen Äußerungen geben und die Rückmeldungen des Lehrers so durch Schülerrückmeldungen ersetzt – zumindest aber ergänzt – werden. (Richter, 2006, S. 49)

Auf diese Weise lassen sich auch formative Feedbackpraktiken über die Schüler*innen im Unterricht etablieren (vgl. Philipp, 2023, S. 10).

Erfolgt das Feedback im Unterricht nicht durch die Lehrkräfte, sondern durch die Peers, kann dies vielfältige Auswirkungen auf den Feedbackprozess und auf die Rezipient*innen haben. Diese können beispielsweise in ihrer Autonomie gestärkt werden oder aber weniger Vertrauen in das ihnen vermittelte Feedback haben (vgl. Streber, 2019, S. 82). Die hier auftretenden Potentiale und Limitationen von Peerfeedback und dessen Rahmenbedingungen sollen nachfolgend im Kontext des Schreibunterrichts behandelt werden.

2.3.1.1 Potentiale von Peerfeedback im Schreibunterricht

Anders als es bei einem Feedback durch die Lehrkraft der Fall ist, begegnen sich bei einem Peerfeedback Schüler*innen als Sendende und Rezipierende auf Augenhöhe. Sie sind beide Lernende, gleichrangige Teilnehmer*innen im Dialog (vgl. Lehnen, 2023, S. 23). Während zwischen Lehrkraft und

Schüler*in die Vermittlung von Feedback häufig an eine Hierarchie oder ein Machtgefälle gebunden ist (vgl. Strijbos & Müller, 2014, S. 105), entfällt dies weitestgehend beim Peerfeedback.

Besonders ist jedoch auch die reziproke Natur, die ein Peerfeedback im Kontext des Schreibunterrichts annehmen kann. So definiert Peter (2023, S. 91) diese Form des Feedbacks als »Prozess, bei dem sich Lernende, die sich weitgehend auf der gleichen Entwicklungs- und Lernstufe befinden, gegenseitig Rückmeldung zur Erreichung des intendierten Ziels geben«. Es kommt hier also vor allem auf die Wechselseitigkeit an: Ein*e Schüler*in schreibt den eigenen und liest den anderen Text – dies nicht gleichzeitig, aber im unmittelbaren Bezug zueinander. Er*sie hat die Aufgabe bereits bearbeitet, gibt auf eine andere Lösung Feedback und erhält dann auch Feedback zu der eigenen Arbeit.

Peter (2023, S. 91) spricht hierbei auch von »Als-ob-Handlungen«, die die Schüler*innen ausführen

Die Schreiber:innen verfassen einen Text, ›als ob‹ sie einen Text mit echter kommunikativer Absicht schreiben würden, sie werden von der Lehrperson dazu aufgefordert, diesen Text einer Mitschülerin oder einem Mitschüler zu überlassen, ›als ob‹ sie Feedback zu einem echten Text wollten, und die Mitschüler:innen sollen Feedback auf den Text geben, ›als ob‹ sie echte Adressat:innen des Textes wären. (Peter, 2023, S. 91)

Die Etablierung von einer regelmäßigen Peerfeedbackpraxis, bei der sich an die Rückmeldung auch eine Überarbeitung schließt, sensibilisiert für den prozesshaften Charakter der Textproduktion. So kann das Bewusstsein bei Schüler*innen gefördert werden, dass Schreiben ein Prozess ist, der Überarbeitungen bedarf, und dies auch durch Unterstützung von anderen gelingen kann (vgl. Lehnen, 2023, S. 18f.).

Geben Schüler*innen einander Feedback, können sie sich gegenseitig dabei beobachten, wie ihre Texte gelesen, verstanden und im Hinblick auf das Lernziel eingeordnet werden (vgl. Sturm, 2022, S. 228). Sie nehmen dabei eine Vielzahl von Rollen ein: Erst sind sie Schreibende/Autor*innen, dann Beurteilende/Lesende und kehren schließlich – mit neuem Input – in ihre Ausgangsrolle zurück. Durch die Übernahme der Rollen, erhalten

Schüler*innen nicht nur die Möglichkeit eines Perspektivwechsels, sondern erfahren den Schreibprozess auf mehreren Ebenen. Lesen sie einen Text, können »die Lernenden ein Gespür für die Verständlichkeit und Angemessenheit von Texten und textsortenspezifischen Anforderungen« (Lehnen, 2023, S. 18) entwickeln. Erhalten sie das Feedback ihrer Mitschüler*innen, »erfahren die Lernenden Wertschätzung für ihre Formulierungsanstren-gungen und schärfen ihr Bewusstsein für die Wohlgeformtheit bzw. Schwächen des Textes« (Lehnen, 2023, S. 18f.). Im Schreibunterricht situiertes Peerfeedback kann somit nicht nur förderlich für die Rezipierenden, sondern auch für die Feedbackgebenden sein.

2.3.1.2 Limitationen von Peerfeedback im Schreibunterricht

Peerfeedback kann im Schreibunterricht somit mit vielen Vorteilen einhergehen: Schüler*innen können sich tiefer mit einer Aufgabe auseinandersetzen, wenn sie selber auch Fremdlösungen evaluieren, oder gemeinsam ihre Stärken im Schreibprozess bündeln (vgl. Banihashem et al., 2024, S. 2; Lehnen, 2023, S. 23f.). Gleichzeitig bedarf der Einsatz von Peerfeedback jedoch auch einer Vielzahl an Überlegung und Planung. Denn: Obwohl Schüler*innen von formativem Peerfeedback durchaus profitieren können, besteht das Risiko, dass Peerfeedback, sollte es unüberlegt und vor allem ungeübt eingesetzt werden, gegenteilige Auswirkungen erzielt.

Wie die Ausführungen zum formativen Feedback gezeigt haben, ist dieser Prozess sehr komplex und die Anforderungen an die Feedbackquelle hoch. Lehrkräfte müssen, soll das Feedback gelingen, in der Lage sein, adäquat auf Basis kontextueller und personaler Faktoren Anpassungen vorzunehmen. Dieses Wissen kann jedoch für den Peerfeedbackprozess nicht vorausgesetzt werden: Schüler*innen, besonders in der Grundschule und Sekundarstufe I, bleiben Lernende und können daher weder Lernstände ihrer Peers einschätzen noch gezielt deren Selbstregulationsprozesse fördern, wie es für effektives formatives Feedback gefordert ist (vgl. Lehnen, 2023, S. 22f.).

Das Geben von (formativem) Feedback muss erlernt werden. Die Überarbeitungskompetenz im Hinblick auf den eigenen oder den fremden Text kann nicht einfach vorausgesetzt werden, sondern muss sich entwickeln können. So gehören Beurteilungsfähigkeit, Diagnosefähigkeit und Alternativenkompetenz (vgl. Jantzen, 2003) im Überarbeitungsprozess zwar zusammen, sind dabei jedoch unterschiedlich anspruchsvoll und komplex. Wenn ein*e Schüler*in eine Problemstelle (oder Normabweichung) im Text erkennt »ist [dies] nicht gleichzusetzen mit der Fähigkeit, eine Erklärung für die Normabweichung geben und einen besseren Alternativvorschlag dafür machen zu können« (Peter, 2023, S. 95). Auch das präzise und klare Formulieren des Feedbacks kann Schüler*innen vor Hindernisse stellen (vgl. Peter, 2023, S. 95f.).

Diese »Schieflage [...] in der Interaktion« (vgl. Lehnen, 2023, S. 24) greift beim formativen Feedback jedoch nicht nur im Hinblick auf das Handeln im Prozess, sondern auch auf den Inhalt des Feedbacks selbst. Im Gegensatz zu der Lehrkraft handelt es sich bei den (Mit-)Schüler*innen nicht um Expert*innen. D. h. bei ihnen kann kein umfangreiches Wissen im Themengebiet vorausgesetzt werden, da sie selbst noch Lernende sind. Gerade ungeübte Schreiber*innen können dabei das Feedback durch die Lehrkraft dem der Peers vorziehen (vgl. Peter, 2023, S. 92).

Die Verortung der Schüler*innen auf der gleichen Ebene bzw. dem gleichen Wissensstand, die zuvor noch ein klarer Vorteil dieser Form des Feedbacks war, kann hier zu ihrer Limitation werden. Peers haben unter Umständen nicht ausreichendes Wissen über z. B. Textsorten oder Schreibstrategien. Aus der mangelnden Diagnosekompetenz können daher Feedbacks hervorgehen, die auf der Textoberfläche angesiedelt sind – z. B. im Bereich der Rechtschreibfehler –, anstatt sich auf den Text als solchen zu beziehen. Kurz gesagt: Damit Schüler*innen ihren Peers ein adäquates Feedback auf ihren Text zu einem bestimmten Kriterium geben können, brauchen sie selbst umfangreiches Wissen über dieses (vgl. Peter, 2023, S. 93–95).⁷

7 Hier gilt jedoch nicht, dass ein elaboriertes, gut ausgearbeitetes Peerfeedback zwingend zu einer Förderung des Lernprozesses führen muss. Strijbos et al. (2010) zeigen z. B. im Rahmen einer Untersuchung des Peer-feedbacks, dass Lernende, die Feedback eines weniger kompetenten Peers erhalten, bessere

Im Idealfall kann sich dieses Wissen jedoch mit der Zeit bei den Peers entwickeln, sodass sie selbst zu Expert*innen werden (vgl. Peter, 2023, S. 94). So stellen auch Pokorny und Pickford (2010) in ihrer Untersuchung (hier im Universitätskontext) fest, dass sich ein vielschichtiges Verständnis von Feedback mit der Zeit bei den Lernenden entwickeln kann (vgl. Pokorny & Pickford, 2010, S. 27).

Es zeigt sich somit, dass die Limitationen des Peerfeedbacks bei der Einführung und Umsetzung dieser Feedbackform im (Schreib-)Unterricht zwar Beachtung finden müssen, sie dabei aber nicht als Ausschlusskriterium behandelt werden dürfen. Vielmehr ist der Umgang mit ihnen entscheidend. So ist zum Beispiel zu prüfen, wie viel (handlungsspezifisches) Wissen Schüler*innen bereits aufweisen und wie im Unterricht Peerfeedbackverfahren eingeführt und -geübt werden können. Auch besteht die Möglichkeit, Schüler*innen durch entsprechende Hilfestellungen beim Geben von Peerfeedback zu unterstützen.

2.3.1.3 Rahmenbedingungen für Peerfeedback

Damit ein Peerfeedback gelingen kann, brauchen Schüler*innen Wissen aus zwei Domänen: (handlungspraktisches) Wissen über die Art und Weise der Rückmeldung und (inhaltliches) Wissen über die Standards, die mit der Schreibaufgabe verbunden sind (vgl. Streber, 2019, S. 83). Für ersteres benötigen die Schüler*innen ein Verständnis von ihrer Rolle im Prozess. Lehnen (2023) führt hier vier Vorstellungen an, die Lernende haben sollten: Sie sollten sich dem Text als Lesende nähern, anstatt ihn als Lektor*in zu beurteilen, die Rückmeldungen aus der eigenen Sicht geben, anstatt »allgemeine Grundsätze zu formulieren« (Lehnen, 2023, S. 24), neben Kritikpunkten auch positive Eigenschaften des Textes hervorheben und das Feedback klar halten und eine Überarbeitung ermöglichen (vgl. Lehnen, 2023, S. 24).

Leistungen erzielen. Eine mögliche Ursache dafür sehen sie darin, dass Lernende, die Feedback von einem leistungsstarken Peer bekommen haben, sich eher auf das Feedback verlassen haben und weniger selbst aktiv geworden sind (vgl. Strijbos et al., 2010, S. 300).

Wie hier deutlich wird, brauchen Schüler*innen, die Feedbacks für ihre Peers formulieren wollen, vor allem Sensibilität und ein Verständnis davon, dass es sich bei dem Fremdtext auch um einen Text von Lernenden handelt. Lehnen (2023, S. 21) begreift Elemente wie Anteilnahme, Produktivität oder Teamfähigkeit dabei »als Erwerbsprozesse [...], die unterrichtlich angeleitet und begleitet werden müssen, damit sie im Schreiben wirksam werden können«. Während Peerfeedback somit zur Stärkung der Autonomie von Schüler*innen durchaus genutzt werden kann, muss der Grundstein dafür im Klassenzimmer durch klare Vorgaben gelegt werden. So müssen Schüler*innen von der Lehrkraft angeleitet Feedbackregeln nicht nur erlernen, sondern durch Einsatz im Feedbackprozess immer wieder üben (vgl. Lehnen, 2023, S. 23).

Weitere Unterstützung können Schüler*innen auch im Rahmen der kommunikativ-sozialen Situation im Klassenzimmer benötigen. Mangelndes Vertrauen oder schlechte bzw. enge Beziehungen zwischen den Schüler*innen können nicht nur die Rezeption des Feedbacks beeinflussen, sondern auch dessen Formulierung. So können Freundschaften oder fehlendes Vertrauen oder Respekt dazu führen, dass Mitschüler*innen (auch in der Wahrnehmung ihrer Peers) beim Geben von Feedback nicht ehrlich sind (vgl. Gamlem & Smith, 2013, S. 160).

Hier empfiehlt es sich, die Ausgangstexte zu anonymisieren. Dies bewirkt zwar, dass die Schüler*innen einander nicht mehr beim Geben von Feedback beobachten, aber auch, dass sie möglichst objektiv Rückmeldungen geben können. In eine ähnliche Richtung geht auch Jantzen (2003). Laut ihm kann die Verfremdung von Texten die Überarbeitungsbereitschaft stützen. Beispielsweise kann ein handschriftlich geschriebener Text, in den Computer abgetippt, auf die Schüler*innen anders wirken und so ihre Bereitschaft zur Auseinandersetzung mit ihm fördern (vgl. Jantzen, 2003, S. 123).

Neben diesen Punkten, die primär die äußere Rahmung des Feedbacks betreffen, sind auch die Inhalte für ein gelingendes Peerfeedback relevant. Hier ist es notwendig, Schüler*innen zum Feedback objektive Kriterien im Hinblick auf das Lernziel zu geben, an denen sie sich orientieren können. So schreiben Gamlem und Smith (2013):

If there are no objective criteria for students to use when assessing their work, they end up with personal ›likes and dislikes‹, creating a dilemma for the students. The existence of objective criteria can lead to a perception of critical feedback as constructive and legitimate. The lack of such criteria can make perceptions of critical feedback negative and be understood more personally than substantively. (Gamlem & Smith, 2013, S. 161)

Feedbackkriterien nehmen also eine zentrale Rolle ein, wenn es darum geht, die Vertrauenswürdigkeit des Peerfeedbacks zu verbessern und das Feedback selbst objektiv und spezifisch zu halten. Feedbackgebende Schüler*innen können auf diese Weise dabei unterstützt werden, »bei ihren Bewertungen vermehrt auch inhaltliche oder textstrukturelle Aspekte in den Blick zu nehmen« (Peter, 2023, S. 92). Solche Kriterien können den Schüler*innen in einem Kriterienraster sowohl vor der Schreibaufgabe zur Planung als auch danach zum Ausfüllen für das Textfeedback zur Verfügung gestellt werden. Kriterienraster können dabei auch gemeinsam von Lehrkraft und Schüler*innen erarbeitet werden, damit sie besonders effektiv sind (vgl. Becker-Mrotzek & Böttcher, 2015, S. 133; Peter, 2023, S. 92).

Das Erstellen des Kriterienrasters ist jedoch nicht nur an die Interaktion zwischen Schüler*innen und Lehrkraft, sondern auch an die unmittelbare Schreibaufgabe gebunden (vgl. Peter, 2023, S. 96). Hier muss eine Passung zwischen Kriterien und Schreib- bzw. Lernziel gegeben sein. Finden Schüler*innen beispielsweise primär sprachformale Kriterien vor, »obwohl die kommunikative Funktion im Zentrum steht, dann richtet sich die Aufmerksamkeit der Lernenden auch auf sprachformale Aspekte« (Sturm, 2022, S. 237). Es gilt somit auch hier, als Lehrkraft eine steuernde und anleitende Position auszuüben.

2.3.2 Feedback durch die Künstliche Intelligenz

KI gewinnt in schulischen Lehr-Lernkontexten zunehmend an Relevanz. Der Einsatz von KI im konkreten Unterrichtssetting verspricht, »zentrale didaktische Funktionen in Einstiegs-, Aneignungs- und Nachbereitungsphasen von Lernprozessen [zu] unterstützen und didaktische Innovationen [zu] bringen« (Wrede et al., 2023, S. 133).

Gerade für den Feedbackprozess scheint der Einsatz von KI dabei besonders vielversprechend. *Human Feedback* – welches Katanneck und Suñer (2024, S. 16f.) in Abgrenzung zu KI-basiertem Feedback sehen – ist sowohl lokal als auch personal begrenzt. Anders ausgedrückt: Feedback kann hier nur dann gegeben werden, wenn auch eine Person vor Ort ist, die es geben kann.⁸ Anders als im *Human Feedback* kann auf eine KI von den Lernenden immer zugegriffen werden. So schreiben Steiss et al. (2024, S. 12) im Hinblick auf ChatGPT: »ChatGPT does not need to sleep, nor does it get tired of the same query«.

Wird KI als Feedback-Tool eingesetzt, bekommen Lernende – zumindest in der Theorie – somit allzeit die Möglichkeit, individuelles und unmittelbares Feedback abzurufen.⁹ Der Prozess ist dabei an eine Reihe von Potentiale und Limitationen gebunden. Diese sollen im Folgenden im Anschluss an eine Annäherung an den Begriff generativer KI und einer Verortung dieser im Schulkontext genauer thematisiert werden.

2.3.2.1 Eine Annäherung an und Verortung von generativer KI im schulischen (Schreib-)Kontext

Ray (2023, S. 121) definiert generative KI (engl. *generative AI*) wie folgt:

Generative AI refers to a class of artificial intelligence models that can create new data based on patterns and structures learned from existing data. These models can generate content across various domains, such as text, images, music, and more [...]. Generative AI models rely on deep learning techniques and neural networks to analyze, understand, and generate content that closely resembles human-generated outputs.

Generative KI bezeichnet somit die Art von Künstlicher Intelligenz, die auf Basis bereits existierender Datensätze neue Texte, Bilder u. Ä. hervorbringen kann. Die generierten Ausgaben weisen eine hohe Ähnlichkeit zu

8 Natürlich gibt es auch die Möglichkeit digitalen Feedbacks durch Lehrkräfte und Peers. Findet dies jedoch außerhalb des Unterrichts statt, geht das erneut mit zeitlichem Aufwand einher. Die personale Begrenzung bleibt dabei bestehen.

9 Auch die Notwendigkeit einer Internetverbindung kann dann wegfallen, wenn man beispielsweise auf lokale Sprachmodelle zurückgreift.

menschlichen Produkten auf (vgl. auch Neumann et al., 2024, S. 228f.). Große Sprachmodelle (engl. *Large Language Models*, LLM) finden dabei in Feldern wie Sprachübersetzungen oder Chatbots Anwendung und können darüber hinaus in einer Vielzahl an Bereichen eingesetzt werden – so auch in der Produktion von Texten (vgl. Ray, 2023, S. 133f.).

Wird ein LLM in der Textproduktion eingesetzt, so kommt es zu einer Verschmelzung von menschlichem und maschinellem Agieren. Chatbots wirken dabei nicht nur als Katalysatoren der Automatisierung des Schreibens. So schreibt Steinhoff bereits 2023: »Sie übernehmen Schreibaufgaben, die zuvor als genuin menschlich galten, und übertreffen viele gewöhnliche menschliche Schreibleistungen schon jetzt zeitlich und qualitativ« (Steinhoff, 2023a, S. 3).

Steinhoff (2023a; 2023b) führt hier den Begriff der »Postdigitale[n] Kommunikationssouveränität« (Steinhoff, 2023b, S. 10) ein. Dabei distanziert er sich von dem Verständnis des Computers¹⁰ als Werkzeug und bewegt sich hin zu einer Betrachtung »als komplexes und dynamisches Hardware-Software-Ensemble« (Steinhoff, 2023b, S. 9). Unter dieser Prämissen meint postdigitale Kommunikationssouveränität bei Steinhoff weder, dass der Computer nur als Tool betrachtet wird, welches man im Schreibprozess weglassen kann, noch meint es, dass der Computer alle Aufgaben übernimmt und der Mensch in seinem Arbeiten verloren geht. Postdigitale Kommunikationssouveränität zielt stattdessen auf ein Ko-Agieren zwischen beiden Instanzen ab (vgl. Steinhoff, 2023b, S. 10).

Zu den bekanntesten Chatbots im Diskurs um textgenerierende KI zählt ChatGPT von OpenAI, das auf einer Pre-trained Transformer (GPT)-Architektur basiert. Der Chatbot ermöglicht eine dialogähnliche Konversation – ein Ko-Agieren – mit den Nutzenden und kann dabei eine Vielzahl an

10 Steinhoff benutzt den Begriff *Computer* anstelle von *digitales Medium*, da es sich laut ihm hierbei um einen neutraleren Ausdruck handele und ausdrücke, »dass ein Desktop-PC, ein Laptop, ein Smartphone usw. ein *relativ autonomes Artefakt* ist, das mit dem Menschen in laufenden Schreibpraktiken im oben angedeuteten Sinne *zusammenspielt und -wirkt*« (Steinhoff, 2023a, S. 4).

kohärenten, kontextuell passenden und grammatisch korrekten Texten¹¹ generieren (vgl. SWK, 2024, S. 8f.; Ray, 2023, S. 121, 125).

Eine Möglichkeit, als Nutzer*in im Umgang mit textgenerierenden KI-Tools eigenaktiv vorzugehen und Einflüsse auf die finale Ausgabe zu nehmen, liegt im sog. *Prompting* bzw. *Prompt Engineering*. Das Verfassen geeigneter Prompts bezeichnet Wampfler (2024, S. 15) als »Schlüsselkompetenz im Umgang mit den verschiedenen Text-KI-Tools«.¹² Mit dieser Kompetenz gehen interaktives bzw. rekursives Nutzen der KI sowie eine kritisch-analytische Vorgehensweise einher (vgl. SWK, 2024, S. 14). So kann es beispielsweise hilfreich sein, genaue Angaben im Hinblick auf Format o. Ä. zu machen oder möglichst spezifisch in den Vorgaben zu sein: »Wer ideal prompten will, muss Kontext liefern« (Gredel et al., 2024, S. 401).

Ein effektives Nutzen generativer KI geht somit über eine simple Suchanfrage hinaus. Wird eine spezifische Ausgabe erwartet, müssen Nutzer*innen mindestens so spezifisch in ihrer Eingabe sein. *Prompting* wird so zu einem komplexen, rekursiven Prozess, der Übung und Anleitung braucht (vgl. Gredel et al., 2024, S. 386).

Künstliche Intelligenz und deren Anwendung ist dabei mittlerweile im Alltag vieler Jugendlicher angekommen. So zeigte bereits die JIM-Studie 2023, dass unter den Befragten ($n = 1200$) 38% ChatGPT nicht nur kannten, sondern es sogar schon ausprobiert hatten. Weitere 36% wussten, was ChatGPT ist und 11% kannten es, wussten aber nicht, was es bedeutet. Einzig 15% gaben an, noch nicht mit dem LLM in Kontakt gekommen zu sein (vgl. Medienpädagogischer Forschungsverbund Südwest, 2023, S. 30f.).

Mittlerweile (JIM-Studie 2024) ist allein die Zahl der Jugendlichen ($n = 1200$), die ChatGPT ausprobiert haben, auf 57% gestiegen. Insgesamt nutzen 62% der Befragten KI-Tools. Hauptmotivator ist die Verwendung für die Schule

11 Laut Bildungsministerium handelt es sich dabei um einen eher weiten Textbegriff, da der Chatbot auch in Bereichen wie Informatik oder Mathematik eingesetzt werden kann (vgl. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023, S. 4).

12 Offen bleibt hier natürlich, ob sich diese Notwendigkeit von Promptingkompetenz auch in Zukunft hält oder von einem gezielteren Vorgehen der KI – z. B. durch automatische Rückfragen an die Nutzenden – abgelöst wird.

oder Hausaufgaben (65%). Aber auch zum Spaß (52%) oder zur Beschaffung von Informationen (43%) wird KI benutzt (vgl. Medienpädagogischer Forschungsverbund Südwest, 2024, S. 59–61).

In schulischen Kontexten erhält KI so zunehmend größere Beachtung. Berger et al. (2024, S. 16) bezeichnen die Auswirkungen der Veröffentlichung von ChatGPT als »disruptive Entwicklung [...], die eine leichte Zugänglichkeit zu KI-Anwendungen sowohl für Lehrkräfte als auch für deren Schülerinnen und Schüler ermöglicht«. Denn: KI (und in diesem Fall ChatGPT) ist für alle Nutzenden prinzipiell immer verfügbar – außerhalb, aber auch und vor allem innerhalb von Bildungskontexten. Dies legt eine Beschäftigung mit Künstlicher Intelligenz bzw. LLMs auch in der Schule nahe.

Folgt man Berger et al. (2024), geht die KI-Nutzung außerhalb der Schule mit der Chance zur Nutzung innerhalb der Schule einher:

Gerade weil KI aus Sicht der Schülerinnen und Schüler die Möglichkeit bietet, sie bei Hausaufgaben, Präsentationen und Recherchen zu unterstützen und damit ihre Selbststeuerungskompetenzen zu fördern, ergibt sich nun die Chance, dass Schülerinnen und Schüler nun auch im Unterricht gezielt lernen könnten, KI-Anwendungen zugunsten einer an ihre Lernbedarfe angepassten Lernerfahrung zu nutzen. (Berger et al., 2024, S. 131)

Innerhalb ihrer Studie zur digitalen Bildung an bayrischen Schulen¹³ zeigen sie: Die meisten Schüler*innen sprechen über KI im Unterricht (74%), weniger (13%) verwenden es nach und noch weniger (9%) im Unterricht (vgl. Berger et al., 2024, S. 109). Geht es um die Anwendungsmöglichkeiten von KI, so liegen diese für die meisten Schüler*innen (33%) in der Bearbeitung der Hausaufgaben. 15% der Befragten nehmen KI als Möglichkeit zur generellen Arbeitserleichterung wahr (vgl. Berger et al., 2024, S. 108).

Das heißt: Während Schüler*innen KI¹⁴ bereits für die Schule nutzen, oder die Nutzung zumindest in Betracht ziehen, wird über KI im Unterricht selbst

13 Befragt wurden hierbei Schüler*innen an Mittelschulen, Realschulen und Gymnasien.

14 KI meint selbstverständlich nicht nur ChatGPT. So können auch Tools wie *DeepL* Anwendung im Unterricht finden (vgl. Gredel et al., 2024, S. 387). Dennoch ist

meist nur gesprochen. Statt kritisches Denken und den zielgerichteten Umgang mit dem LLM zu üben, wird er nur mündlich thematisiert. Schüler*innen sind so einer starke Diskrepanz zwischen privater Nutzung von KI und begleiteter Nutzung im Unterricht ausgesetzt.

Eine wichtige Komponente für die Realisierbarkeit von KI-Anwendungen im Unterricht stellt die Datensicherheit von Schüler*innen dar (vgl. Kasneci et al., 2023, S. 6). Hinzutritt die allgemeine Einstellung an Schulen. So stehen auf der einen Seite die Ressourcen, die durch den Einsatz von KI freigelegt werden können. Hierzu zählen Zeitersparnisse bei beispielsweise der Unterrichtsvorbereitung oder der Erstellung von Materialien (vgl. Mühlhoff & Henningsen, 2024, S. 6), Binnendifferenzierungen, bei der sowohl lernschwächere als auch lernstärkere Schüler*innen individuell und adaptiv begleitet werden können (vgl. SWK, 2024, S. 17) oder die Förderung von Problemlösestrategien der Schüler*innen (vgl. Kasneci et al., 2023, S. 2).

Auf der anderen Seite herrschen jedoch auch seitens der Lehrkräfte Bedenken bzgl. der Nutzung von KI durch die Schüler*innen. Wie Berger et al. aufzeigen, nehmen einige Lehrkräfte an weiterführenden Schulen die Herausforderungen von KI spezifischer als deren Chancen wahr. Hierzu zählen vor allem Zweifel im Hinblick auf die Kreativität oder das selbstständige Lernen (vgl. Berger et al., 2024, S. 90). Darüber hinaus kann es für Schüler*innen schwierig sein, geprüftes echtes Wissen von nicht verifizierten Inhalten in der Ausgabe der KI zu unterscheiden (vgl. Kasneci et al., 2023, S. 7). Oder sie schätzen ihre Fähigkeiten im Umgang mit der KI als hoch ein, beziehen sich dabei jedoch nur auf »die grundlegende Bedienung und Anwendung« (Berger et al., 2024, S. 131), zeigen also kein gezieltes Einsetzen der KI und hinterfragen ihre Funktionsweise nicht.

In einem Beitrag in *DIE ZEIT* ruft Peter Salden (2025) schon im Titel auf: »Schützt Jugendliche vor der Manipulation durch KI!«. Salden warnt davor, dass u. a. junge Nutzende die Ausgaben der KI unreflektiert rezipieren – auch dann, wenn es um die eigene Meinung geht:

ChatGPT das LLM, welches meist ein Großteil der Schüler*innen kennt (vgl. Berger et al., 2024, S. 106).

Denn den wenigsten Menschen ist bewusst, dass die Modelle keine ›Faktenmaschinen‹ sind. Ganz besonders fehlt dieses Bewusstsein denjenigen, die sich mitten im Bildungsprozess befinden.

Zu den Gefahren der Meinungsmanipulation treten finanzielle Hürden mancher LLMs, die zu einem Ungleichgewicht zwischen den Schüler*innen führen können (vgl. SWK, 2024, S. 17).

Dennoch bleibt KI eine technische Neuerung, die durch ihre alltägliche Relevanz zwangsläufig auch im schulischen Kontext ankommen muss.¹⁵ Verwiesen sei hier beispielsweise auf Wampfler (2023, S. 4), der die Möglichkeiten von KI mit denen des Autofokus in der Fotografie verknüpft. Während jedoch in dieser Annäherung von einem Werkzeug gesprochen wird, das nur in seiner Domäne greift, handelt es sich bei KI um eine Neuerung, die domänenübergreifend wirken kann.

Auch die Bundesländer haben reagiert und in Handreichungen Implikationen für den Umgang mit KI in der Schule aufgesetzt (vgl. hierzu auch Gredel et al., 2024, S. 382–385). Hierbei machen sie Schule zu einem wichtigen Akteur in der Auseinandersetzung mit KI. So formuliert beispielsweise das Bildungsministerium des Landes NRW, dass es

auch Aufgabe von Schule [ist], die Schülerinnen und Schüler im Rahmen des Unterrichts mit KI vertraut zu machen und gemeinsam im geschützten Raum zu erfahren, wie KI-basierte Textgeneratoren funktionieren, welche Potentiale, aber auch welche Risiken damit verbunden sein können. (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023, S. 5)

Folgt man diesen Handreichungen, sollen Lernende im Kontext Schule nicht nur über KI informiert und für sie sensibilisiert werden, sondern auch Gelegenheiten bekommen, sie aktiv zu erfahren, d. h. zu nutzen.¹⁶ Dies steht im

15 Auch oder gerade, weil erst durch das aktive Thematisieren und Experimentieren von und an Problemfeldern wie der von Salden (2025) angesprochenen Meinungsmanipulation ein reflektierter Umgang mit KI gewährleistet werden kann.

16 Gredel et al. (2024) kritisieren hierbei den Mangel an Anleitung der Lehrkräfte durch das Ministerium: »[D]as Angebot [an digitalen Ressourcen] ist groß, es bedarf einer Orientierung und einer Einschätzung der Qualität, zudem sind nicht alle Ressourcen für alle Lernenden bzw. alle Phasen des Schreibens geeignet und

Gegensatz zu den oben aufgeführten Studienergebnissen, bei denen die meisten Schüler*innen nur angaben, über KI zu sprechen.

Folgt man den Empfehlungen des Bildungsministeriums weiter, so kann »[e]in lernförderlich gestalteter Einsatz von KI [...] dazu beitragen, Sprach-, Schreib- und Beurteilungskompetenzen zu erweitern« (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023, S. 7). Wird KI im Unterricht eingesetzt, so ließen »sich z. B. die von der KI erstellten Texte auf Richtigkeit, Konsistenz, Machart etc. untersuchen« (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023, S. 7). Was hier dargelegt wird, ist ein Arbeiten an KI. Bereits generierte Texte werden einer (scheinbar) abschließenden Untersuchung unterzogen. Anstatt KI im Prozess zu nutzen, arbeiten Schüler*innen mit dem bereits erstellten Produkt.

Dies ist zur Förderung der oben genannten Kompetenzen zwar durchaus geeignet, aber nicht das einzige mögliche Verfahren. Auch die aktive Nutzung von KI durch die Schüler*innen kann fokussiert werden (vgl. Gredel et al., 2024, S. 400; Schindler, 2024, S. 35–38). Um den Einsatz von KI im eigenen Schreibprozess begreifen zu können, ist es notwendig, eigene Erfahrungen in dieser Hinsicht zu machen.

Ändert man die Mittel, die zur Textproduktion genutzt werden, so beeinflusst dies zwangsläufig auch den Schreibprozess (vgl. Wendt, 2023, S. 38f.). Eine Einführung von KI im Schreiben kann so zu einer Vielzahl an Veränderungen führen, denn der Computer ist, wie Steinhoff (2023a) darlegt, nicht mehr nur das Mittel, mit dem ein Text fixiert wird, sondern kann einen Text maßgeblich verändern. Steinhoff (2025) unterscheidet hier beispielsweise bei dem Umgang mit ChatGPT zwischen den Einsätzen als *Ghostwriter*, *Writing Tutor* oder *Writing Partner*. Je nachdem, wie LLMs von Schüler*innen eingesetzt werden, können sie so nach einem oder mehreren Prompts eher große Teile des Schreibens übernehmen (*Ghostwriter*), Schüler*innen in einer lehrkraftähnlichen Rolle beim Schreiben unterstützen (*Writing Tutor*)

außerdem verlangen viele der Anwendungen eine Anmeldung (also z. B. über eine E-Mail-Adresse oder Telefonnummer), die Daten werden zudem auf Servern außerhalb der EU gesichert. Das bedeutet für Lehrende, vielfältige – auch rechtliche – Überlegungen anzustellen« (Gredel et al., 2024, S. 384).

oder mit ihnen koagieren (Writing Partner). Letzteres verlangt von dem es nutzenden Menschen

relativ hohe Schreib- und Lesekompetenzen [...]. Er muss z. B. in der Lage sein, die Koaktivität metakognitiv zu steuern, also eine Vogelperspektive auf sie einzunehmen und zu entscheiden, wann und wie ein Einbezug eines LLM sinnvoll ist. (Steinhoff, 2025, S. 95)

Werden niedrigschwellige KI-Tools zum vollständigen Generieren eines Textes genutzt (Ghostwriter), so besteht die Gefahr, das Eigene im Produktionsprozess zu verlieren. Neben den Schwierigkeiten, denen sich Lehrkräfte bei der Bewertung von Texten, deren Ursprung sie nicht kennen, gegenüberfinden (vgl. Wampfler, 2024, S. 15), kann dies auch deutlich negative Implikationen für die Schüler*innen haben. Baron (2023) führt hierzu beispielsweise an, dass die eigene Schreibstimme durch die Nutzung von KI beeinflusst werden kann (vgl. Baron, 2023). Hinzu tritt das Risiko, dass Lernende sich mit dem eigentlichen Inhalt des Textes nicht mehr gänzlich auseinandersetzen. Die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) (2024, S. 10) spricht hier davon, »dass Lernende fertige Produkte erstellen lassen, deren inhaltliche Qualität sie nicht einschätzen können, und dabei die eigentlich lernförderlichen Aktivitäten auslagern«.

Dies kann Auswirkungen auf die eigene Schreibmotivation haben: Warum noch selbst schreiben, wenn man auch schreiben lassen kann? (vgl. Wampfler, 2024, S. 15) –, aber auch das Denken und den Wissenserwerb der Schüler*innen beeinflussen. So bedeutet eigenständig zu schreiben auch, im Schreibprozess Gedanken zu entwickeln und Zusammenhänge zu verstehen. Wird die KI ausschließlich als Ghostwriter verwendet, fallen diese Zugänge zumindest in Teilen weg und der Schreibprozess wird stark verkürzt (vgl. Wampfler, 2023, S. 4).

Die Einführung von KI ins schulische Schreiben sollte also wohl durchdacht und auf diese möglichen Risiken hin sensibilisiert geschehen. Schüler*innen sollten angeleitet werden, LLMs reflektiert als Schreib-Tutor*in oder -Partner*in zu nutzen.

Zentral für dieses Vorhaben, ist die Annahme, dass KI nicht ersetzen, sondern vorwiegend stützen soll – die Handlungs- und Entscheidungsvollmacht also weiterhin bei den Schreibenden liegt¹⁷. Dies kann auf der Textoberfläche geschehen – z. B. im Bereich von Orthografie- und Grammatikkorrektur –, aber auch bei tieferliegenden Prozessen (vgl. auch Wendt, 2023, S. 39). Generell gilt: KI-Tools können in jeder Phase des Schreibprozesses eingesetzt werden, sollten dabei aber nicht als Globallösung begriffen werden. D. h., das KI-Tool sollte nicht zur Generierung des Gesamttextes hinzugezogen werden, kann aber gezielte Schritte im Schreibprozess wie die Planung, das Formulieren oder das Überarbeiten unterstützen (vgl. Wampfler, 2023, S. 5).

2.3.2.2 Potenziale von KI-basiertem Feedback im Schreibunterricht

Wie bereits bei Steinhoff (2025) deutlich wurde, lässt sich KI, neben der Tätigkeit als Ghostwriter, auch gewinnbringend und lernförderlich in ihrer Rolle als Schreibpartner*in oder -tutor*in für den Schreibprozess bzw. seine Komponenten einsetzen. Dies gilt auch für die Überarbeitung des Schüler*innentextes. Hier kann KI als Feedbackquelle in Erscheinung treten.

KI kann Schüler*innen im Überarbeitungsprozess unterstützen, Fehlerstellen diagnostizieren und Alternativen aufstellen (vgl. Gredel et al., 2024, S. 387). Dieser Gedanke wird auch vom Bildungsministerium NRW aufgegriffen. Hier heißt es, dass die zuvor »festgelegten Bewertungskriterien für ein Lernprodukt [...] mithilfe der KI für ein jederzeit verfügbares, zeitnahe formatives Feedback genutzt werden [können]« (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023, S. 7).

Zu den großen Vorteilen von KI-basiertem Feedback zählt die Möglichkeit, auf die Lernenden zugeschnittene Rückmeldungen zu geben. So können digitale Tools den Schreibprozess unterstützen, indem sie individuelles Feedback anbieten, welches an den Lernstand der Schüler*innen angepasst ist

¹⁷ Wrede et al. (2023) differenzieren hierbei auch zwischen Autonomie und Automatisierung. Während bei ersterem das System allein entscheidet, bleibt bei der Automatisierung der Mensch, bzw. der*die Schreibende Steuerinstanz (vgl. Wrede et al., 2023, S. 140f.).

(vgl. Wendt, 2023, S. 40)¹⁸. Dadurch kann es prozessorientiert und adressatengerecht das Schreiben der Schüler*innen unterstützen – etwas, das beispielsweise bei intelligenten Tutorsystemen nicht möglich ist, da die individuellen Lösungen stark von den in das System eingelesenen Mustern abweichen können (vgl. Neumann et al., 2024, S. 230).

Zeit ist ein weiterer entscheidender Faktor, der hier beachtet werden sollte. Während das Sichten von einzelnen Schüler*innentexten durch die Lehrkraft bei zunehmender Lerner*innenzahl mit einem hohen zeitlichen Aufwand einhergeht, kann ein KI-Feedback innerhalb von Sekunden generiert werden. Dies verspricht eine Entlastung für die Lehrkräfte, indem sie das Geben von Feedback auslagern können und gleichzeitig sicherstellen, dass Schüler*innen ein auf sie zugeschnittenes Feedback erhalten (vgl. Fleckenstein et al., 2023, S. 4; vgl. Kasneci et al., 2023, S. 2f.).

2.3.2.3 Limitationen und Rahmenbedingungen von KI-basiertem Feedback im Schreibunterricht

Wird Feedback im Unterricht auf KI-Systeme ausgelagert, bedeutet dies nicht, dass Lehrkräfte im Feedbackprozess gänzlich überflüssig werden. Vielmehr verschiebt sich durch die KI als Feedback gebende Instanz ihr Aufgabenbereich. So müssen Lehrkräfte auf der einen Seite ein KI-Tool auswählen, das für den konkreten Feedbackprozess geeignet und dabei datenschutzkonform ist. Auf der anderen Seite müssen sie jedoch auch während die KI ihr Feedback gibt, diesen Prozess beobachten und gegebenenfalls einschreiten. Denn: Die Limitationen, die für KI im Bereich der Texterstellung greifen – beispielsweise Halluzinationen und Bias – können in dieser Form natürlich auch im Feedback auftauchen (vgl. SWK, 2024, S. 9f., 12f.).

Hinzukommt, dass viele KI-Tools, so auch ChatGPT, in erster Linie nicht für Feedbacks im Schulunterricht entwickelt wurden (vgl. Haverkamp et al.,

¹⁸ Wendt (2023) bezieht sich in ihren Ausführungen vorwiegend auf *Automated Essay Scoring*, *Automated Writing Evaluation* und *Intelligent Tutoring Systems*.

2024, S. 64). Um in diesem Kontext wirken zu können, wird ein klares *Prompt Engineering* benötigt. So schreiben Haverkamp et al. (2024, S. 69):

Damit ein LLM gutes Feedback zu einer Abgabe geben kann, ist es wichtig, dass dem LLM der Kontext der Aufgabe ›bekannt‹ ist. Kontext meint in diesem Fall den inhaltlichen Kontext, im Falle einer Gedichtinterpretation z. B. das konkrete Gedicht, und den Kontext, in dem die Aufgabe gestellt wird, also welche Schulform, Klasse etc., also den Anforderungskontext.

Doch auch bei für den schulischen Kontext ausgewiesenen KI-Feedback-Tools ist eine reflektierte, kritische Sichtweise vonnöten. So untersuchen Mühlhoff und Henningsen (2024) das KI-Korrekturtool des Unternehmens *Fobizz*, welches auf ChatGPT aufbaut. Im Rahmen ihrer Untersuchungen werden zehn simulierte Schüler*innenarbeiten fünfmal durch das Tool geprüft. Im Anschluss werden zwei aufgrund des Feedbacks iterativ verbessert und dann erneut geprüft (vgl. Mühlhoff & Henningsen, 2024, S. 13–16).

Mühlhoff und Henningsen zeigen, dass das Korrektur-Tool eine Reihe von Problemstellen aufweist. Es suggeriert nicht nur, dass jedes denkbare Bewertungskriterium von der Lehrkraft festgelegt werden kann¹⁹, sondern ist auch in seinem Feedback inkonsistent. So zeigt es für einen Teil der Texte bei wiederholtem Eingeben unterschiedliche Noten, die auch zwischen Bestehen und Nichtbestehen schwanken können (vgl. Mühlhoff & Henningsen, 2024, S. 18). Hinzu tritt, dass nur ein Teil der Texte eine Gesamtnote von der KI bekommt, die Benennung der Fehler variiert und eine Befolgunig des Feedbacks nicht automatisch dazu führt, dass das Folgefeedback besser ausfällt (vgl. Mühlhoff & Henningsen, 2024, S. 20–22). Die hier identifizierten Problemstellen unterstreichen die Notwendigkeit der Lehrkraft, auch bei KI-basiertem Feedback weiterhin als kontrollierende Instanz zu agieren und sich mit der Funktionsweise des KI-Tools auseinanderzusetzen.

Damit eine KI das in 2.3.2.2 beschriebene Feedback geben kann, sind also eine Vielzahl an Vorarbeiten zu erledigen, die sowohl die Auswahl des Tools, den Umgang mit ihm und ggf. das Prompten des LLM betreffen (vgl. hier

19 Mühlhoff und Henningsen zeigen, dass es möglich ist, ein Kriterium zu erstellen, nach dem geprüft wird, ob der eingegebene Text mittels KI erstellt wurde – obwohl *Fobizz* eigentlich keine KI-Texte erkennen kann (vgl. Mühlhoff & Henningsen, 2024, S. 20).

auch Latif/Zhai, 2024). Dieser Prozess ist komplex und aufwändig (vgl. Ray, 2023, S. 126) und kann daher gerade für jüngere bzw. ungeübte Schüler*innen schwierig sein. Auch hier ist somit erneut Sensibilität von der Lehrkraft gefragt, die – sollte das KI-Tool nicht von vornherein für den ihr vorliegenden Feedback- und Unterrichtskontext ausgestattet sein – selbst aktiv werden muss.

2.4 Vom Feedback zu den Lernenden: Feedbackrezeption

Neben der Feedbackbotschaft und der Quelle bzw. Form des Feedbacks ist auch der*die Feedbackempfänger*in fest im Feedbackprozess verankert. Bei ihm*ihr laufen die einzelnen Komponenten des Feedbacks zusammen, er*sie ist für die Verarbeitung und Umsetzung zuständig. Wie aus den theoretischen Überlegungen in Kapitel 2.1 bereits hervorgegangen ist, kann nur das Feedback wirken, das auch bei dem*der Rezipient*in ankommt bzw. seine*ihrre Aufmerksamkeit beansprucht (vgl. Kluger & DeNisi, 1996). Und so wird auch für Textfeedback im Schreibunterricht der*die Rezipierende zu der zentralen Größe, die im Feedbackprozess fokussiert werden muss (vgl. auch Brooks et al., 2021).

Schreibende müssen das Feedback, das sie erhalten, nicht zwingend umsetzen. Sie können es auch ablehnen oder modifizieren (vgl. Graham, 2018, S. 146). Das übermittelte Feedback zu verstehen, ist dabei für Lernende ein komplexer Vorgang (vgl. Pokorny & Pickford, 2010, S. 22). Umso entscheidender ist daher deren Wahrnehmung²⁰ bzw. Rezeption – die von der der Sendenden durchaus abweichen kann.

20 Das Verständnis von Feedbackwahrnehmung orientiert sich dabei an dem englischen Pendant *Feedback Perception*, das von Strijbos et al. (2021, S. 2) als »outcomes of how recipients spontaneously experience the feedback content as provided by an external source—or the feedback process as a whole—in terms of cognitive, metacognitive, motivational, and/or affective reactions« definieren. Eine ähnliche Begriffbestimmung findet sich auch bei van der Kleij (2019, S. 177): »The term feedback perception captures how students receive, interpret, and value a feedback message«. *Feedback Perception* wird so zu einem ganzheitlichen Ausdruck, der sowohl das Erfahren und Aufnehmen der Feedbackbotschaft

Dass die von den Empfangenden verstandene Botschaft nicht zwingend mit der der Sendenden übereinstimmen muss, zeigt ein Rahmenmodell von Strijbos und Müller (2014), welches den Feedbackprozess als interaktiven Vorgang begreift. Hier stehen die Akteur*innen des Prozesses im Fokus und ver- bzw. entschlüsseln die Feedbackbotschaft. Strijbos und Müller orientieren sich dabei an der Vier-Seiten-Metapher Schulz von Thuns (2010). Diese besagt, dass jede kommunikative Äußerung simultan vier Dimensionen umfasst – den Sachinhalt, die Selbstoffenbarung, den Beziehungshinweis und den Appell – wodurch sich Kommunikation als ein vielschichtiger Prozess darstellt, in dem Sender*in und Empfänger*in Botschaften auf unterschiedlichen Ebenen codieren und decodieren.

Die hier sichtbar werdenden Kommunikationsstörungen zwischen Sender*in und Empfänger*in, erkennen Strijbos und Müller auch im Feedbackprozess:

Ein Feedback, das vom Sender mit bestimmten Intentionen gegeben wird (z.B. ein Hinweis auf eine inhaltliche Verbesserungsmöglichkeit), muss vom Empfänger nicht unter derselben Zielperspektive interpretiert werden (z.B. kann der Hinweis auch auf der Beziehungsebenen als Abwertung der eigenen Person verstanden werden). (Strijbos & Müller, 2014, S. 87)

Ausschlaggebend für den Feedbackprozess sind situative und personale Faktoren wie persönliche Variablen, Repräsentationen des*r jeweils anderen und Kontextbedingungen (vgl. Strijbos & Müller, 2014, S. 88). Die mentalen Repräsentationen sind dabei auch von Vorerfahrungen geprägt (vgl. Strijbos & Müller, 2014, S. 91) und treten beispielsweise bei Feedbackgebenden durch Anpassung des Feedbacks an die Rezipierenden in Erscheinung.²¹

Doch was bedeutet dies konkret für die Feedbackwahrnehmung bzw. -rezeption? Einen ersten Einblick kann das *Student-Feedback Interaction Model* von Lipnevich et al. (2016) geben. Die Grundannahme dieses Modells ist dabei, dass Feedback immer in einem gewissen Kontext rezipiert wird, in

beinhaltet als auch das An- und Übernehmen dieser – in anderen Worten: die Rezeption von Feedback.

21 Hier kann es u. a. dazu kommen, dass der*die Sender*in in Erwartung einer bestimmten Resonanz »prospektiv eine Anpassung der Nachricht vornimmt, um negative Konsequenzen für sich selbst zu vermeiden« (Strijbos & Müller, 2014, S. 89).

welchem die Lernenden sich wohl bis unbehaglich fühlen können. Das hier gegebene Feedback bringt Konsequenzen mit sich, die ebenfalls als positiv oder negativ wahrgenommen werden können. Ähnliches gilt für das Feedback selbst, es kann z. B. oberflächlich oder detailreich, verallgemeinert oder individualisiert sein. Zusätzlich bringen die Lernenden persönliche Eigenschaften und Erfahrungen mit. Sie sind unterschiedlich stark in dem Fach, in dem Feedback gegeben wird, oder haben bereits unterschiedliche Erfahrungen mit Feedback gemacht (vgl. Lipnevich et al., 2016, S. 176–180, 182).

Bei der Rezeption von Feedback spielen nun all diese Faktoren mit und lösen dabei affektive und kognitive Reaktionen aus. Lernende können sich u. a. stolz oder entmutigt fühlen, sie verstehen das Feedback vollständig, zu Teilen oder haben hierbei große Probleme. Dies führt zu einer Vielzahl an Aktionen seitens der Lernenden, die sich bezogen auf das Feedback adaptiv (die Lernenden setzen das verstandene Feedback um) oder maladaptiv (die Lernenden setzen das (nicht) verstandene Feedback nicht um) auswirken können (vgl. Lipnevich et al., 2016, S. 180–182).

In Anlehnung an dieses Modell und die Ausführungen von Jonsson (2012), der im Kontext höherer Bildung Herausforderungen für den produktiven Umgang mit Feedback identifiziert, sowie unter Berücksichtigung weiterer empirischer Befunde lassen sich fünf konkrete Ansprüche festmachen, die für die Rezeption von Feedback durch Lernende zentral sind:

- (1) Das Feedback muss für die Lernenden nützlich und nutzbar sein

Egal, ob Lernende Feedback auf Ebene der Aufgabe, des Prozesses oder der Selbstregulation erhalten, müssen sie dieses nicht nur umsetzen können (vgl. Winstone et al., 2017, S. 22), sondern die Umsetzung auch als sinnvoll erachten, d. h. akzeptieren. Die einzelnen Komponenten des Feedbacks sollten für Schüler*innen sowohl sichtbar als auch nachvollziehbar sein und dabei als hilfreich wahrgenommen werden. Je nach Aufgaben- bzw. Lernumgebung kann daher eine andere Art des Feedbacks von den Lernenden präferiert werden (Jonsson, 2012, S. 66f.).

Dies geht einher mit einer Klarheit und Unmittelbarkeit im Feedback. So zeigt van der Kleij (2019, S. 182), dass Schüler*innen individualisiertes

Feedback am nützlichsten wahrnehmen, wenn es ihnen zeigt, wie sie sich verbessern können, dabei möglichst konkret ist und zeitnah gegeben wird. Bei Pokorny und Pickford (2010) wird im universitären Kontext von Lernenden das Feedback als hilfreich wahrgenommen, welches Hinweise zum erwarteten Lernziel enthält – auch in Form von Randnotizen an Entwürfen. Dabei diskutieren Studierende Feedback, das sie dazu befähigt, Diskrepanzen zwischen Soll und Ist zu erkunden und so auch ein gemeinsames Besprechen in der Peer-Group oder mit den Tutor*innen ermöglicht (vgl. Pokorny & Pickford, 2010, S. 25).

(2) Das Feedback muss auf sprachlicher Ebene verständlich sein

Neben Strategien, die Lernende beherrschen sollten, um produktiv mit Feedback umzugehen und so die Anwendung von ad hoc Strategien oder das passive Nutzen von Feedback zu vermeiden (Jonsson, 2012, S. 69), muss das Feedback auch im Hinblick auf die verwendete Sprache verständlich sein (Jonsson, 2012, S. 69f.). Erst wenn sowohl die sprachliche Verständlichkeit des Feedbacks als auch die Verfügbarkeit geeigneter Verarbeitungsstrategien gegeben sind, können Lernende Rückmeldungen als hilfreich wahrnehmen und gewinnbringend nutzen.

(3) Das Feedback muss für die Lernenden konkret sein und von ihnen akzeptiert werden, gleichzeitig aber auch die eigenaktive Auseinandersetzung mit dem (Text-)Produkt fördern

Folgt man Jonsson (2012) ist die inhaltliche Gestaltung des Feedbacks entscheidend. Hier ist es wichtig, bei den Lernenden die aktive Auseinandersetzung mit dem (Text-)Produkt zu fördern, statt sie vorgegebene Antworten aus dem Feedback übernehmen zu lassen. Es kann daher das Feedback hilfreich sein, das weniger präzise bzw. gezielt-vorgegebend ist, dafür aber detailliert und persönlich ist (vgl. Jonsson, 2012, S. 67f.). Auch van der Kleij (2019) hält fest, dass Schüler*innen sich vermehrt individualisiertes Feedback wünschen, dies jedoch häufig einen Bruch mit der Realität darstellt: »Students' comments indicated the desire for more individualised feedback, whilst teachers' comments reflected experiencing difficulties in implementing quality feedback practices due to time constraints« (van der Kleij, 2019, S. 186).

Bei Gamlem und Smith (2013) geht der Wunsch nach einem konkreten und greifbaren Feedback gleichzeitig auch mit der Akzeptanz von letzterem einher. Die Autor*innen untersuchen die Rezeption von Feedback bei 13- bis 15-Jährigen. Hier stellen sie fest, dass für Schüler*innen positives Feedback nicht nur die Anerkennung von Leistung oder Anstrengung ist, sondern auch das genaue Bestimmen von Punkten, die zur Verbesserung überarbeitet werden können. Auf der anderen Seite empfinden sie unspezifische und verallgemeinernde Ausdrücke wie, dass sie härter arbeiten müssten, negativ vor dem Hintergrund, dass sie das Gefühl haben, ihr Wissen bereits vollständig zu Papier gebracht zu haben (vgl. Gamlem & Smith, 2013, S. 159).

Ebenfalls unterstreichen Gamlem und Smith die Notwendigkeit, dass ein Feedback nicht am Ende des Lernprozesses gegeben wird, sondern in diesem verankert ist:

we found that students were not given time and opportunities to work with feedback from teachers, which causes feedback initially to increase learning (e.g. specifying improvement feedback) to be perceived as negative and abandoned. (Gamlem & Smith, 2013, S. 160)

Hier verstehen Schüler*innen Feedback als negativ, wenn sie keine Chance bekommen, mit diesem zu arbeiten oder es anzuwenden. Anders sieht es bei dem Feedback aus, das in den Lernprozess inkludiert wird und ein weiteres Arbeiten ermöglicht. Dieses verstehen Schüler*innen als positives Feedback, welches gleichzeitig das Verständnis stützt, dass sie durch die Rezeption und Umsetzung des Feedbacks ihr eigenes Lernen verbessern können (vgl. Gamlem & Smith, 2013, S. 160).

- (4) Lernende müssen sich im Feedbackprozess wohl und gerecht behandelt fühlen

Dies meint nicht, dass im Feedback nicht kritisiert werden darf. Vielmehr soll hierbei auf positive Affektwirkungen seitens der Lernenden und ein Empfinden von Fairness abgezielt werden. So begreift beispielsweise Jonsson autoritäres Feedback als nicht produktiv für die Lerner*innen. Stattdessen sollte sich der*die Feedbackgebende im

Rahmen der Möglichkeiten in die Rolle eines*r Dialogpartner*in begeben (vgl. Jonsson, 2012, S. 68f.).

Pokorny und Pickford (2010) stellen in ihrer Studie fest, dass Studierende, obwohl sie die Chance auf Feedback haben, dieses nicht immer in Anspruch nehmen: »Some students in the study reported being offered one-to-one feedback opportunities with tutors but also not feeling confident enough to take them, particularly if they had fallen behind« (Pokorny & Pickford, 2010, S. 26). Die Beziehung zum*r Tutor*in ist hier für die tatsächliche Inanspruchnahme von Feedback essentiell. Dies ist eng geknüpft an Gefühle wie Scham. Ein*e Lernende*r kann Feedback nur aktiv rezipieren, wenn er*sie sich während dieses Prozesses nicht schämt oder unterdrückt fühlt. So nennen Pokorny und Pickford als weitere einschränkende Faktoren seitens der Studierenden ein wahrgenommenes Machtgefälle, das Gefühl, selbst Interaktionen initiieren zu müssen oder das Bewusstsein, dass der eigene Bedarf mit dem zeitlichen Mehraufwand einhergeht (vgl. Pokorny & Pickford, 2010, S. 26).

(5) Lernende müssen der Feedbackquelle vertrauen

Die feedbackgebende Instanz nimmt eine große Rolle bei der Rezeption und der Verarbeitung des Feedbacks ein. So ist es – auch im Hinblick auf die Akzeptanz und Fairness – entscheidend, ob Lernende der Feedbackquelle vertrauen oder nicht (vgl. Winstone et al., 2017, S. 23).

Schüler*innen zu aktiven Rezipierenden zu machen, bedeutet also nicht nur, ihnen Feedback zu vermitteln, sondern sie auch als aktiv Teilnehmende in das Zentrum der Bemühungen zu setzen. Die hier aufgezeigten Ansprüche sind dabei nicht klar voneinander abgrenzbar, sie sind aneinander gebunden und gehen ineinander über. Übergreifende Rezeptionsdimensionen sind dabei Spezifität bzw. die konkrete Identifikation von Fehlern oder Verbesserungsvorschlägen (Anspruch 1, 3), Akzeptanz (Anspruch 1, 3, 5), Nützlichkeit (Anspruch 1, 2), aber auch Affektwirkungen (Anspruch 4), Vertrauen (Anspruch 5) und wahrgenommene Fairness (Anspruch 4, 5). Wird Feedback im konkreten Lehr-Lern-Setting untersucht, können diese Dimensionen Aufschlüsse im Hinblick auf dessen Rezeption durch Lernende geben.

2.5 Zusammenfassung und Einblick in den aktuellen Forschungsstand

Zusammenfassend lässt sich Feedback somit als complexes, aber auch wirksames Tool in Lehr-Lernkontexten begreifen. Auch für die Entwicklung von Schreibkompetenz kann es gewinnbringend eingesetzt werden. Durch formatives Feedback können Schüler*innen so auf ihrem Lernweg begleitet und unterstützt werden. Dabei müssen die Feedback gebenden Lehrkräfte u. a. darauf achten, das Feedback spezifisch und klar zu formulieren und es im Lernprozess passend zu positionieren (vgl. Shute, 2008). Gerade im Deutschunterricht (aber auch in jedem anderen unterrichtlichen Setting, in dem vermehrt Langtexte produziert werden) stellt dies Lehrkräfte vor kognitive, aber vor allem auch vor zeitliche Herausforderungen. Es empfiehlt sich daher, Feedback zu Teilen auszulagern.

Im theoretischen Teil dieser Masterarbeit wurden dafür zwei mögliche Formen von externem Feedback vorgestellt: das Feedback durch die Peers und das durch die KI. Diese Quellen sind dabei an eine Reihe unterschiedlicher Potenziale, Limitationen und Rahmenbedingungen geknüpft, die einen möglichst produktiven Umgang gewährleisten. So ermöglicht Peerfeedback beispielsweise durch seine Natur ein Feedback auf Augenhöhe, was jedoch gleichzeitig auch ein Feedback unter Noviz*innen bedeutet. Eine Anonymisierung der Texte oder das Geben von objektiven Feedbackkriterien kann hier helfen. KI-basiertes Feedback hat den Vorteil, immer und überall gegeben werden zu können, ist aber ggf. an Bias gebunden und unter Umständen nicht direkt für den schulischen Kontext entwickelt. Hier bedarf es der aufmerksamen Sichtung des Tools, bevor es im Unterricht eingesetzt werden kann.

Während diese Überlegungen hauptsächlich an die Lehrkraft gebunden sind, die die Feedbackform im Unterricht einsetzt, darf die Perspektive der Schüler*innen nicht vernachlässigt werden. Schließlich sind sie diejenigen, die das Feedback schlussendlich rezipieren. Damit Schüler*innen das Feedback verwenden können – und wollen –, müssen eine Vielzahl an Faktoren erfüllt sein. U. a. sollte das Feedback möglichst konkret sein und die Lernenden sollten sich im Feedbackprozess wohlfühlen. In Kapitel 2.4 konnten

folgende übergreifende Rezeptionsdimensionen identifiziert werden: Spezifität bzw. die konkrete Identifikation von Fehlern oder Verbesserungsvorschlägen²², Akzeptanz, Nützlichkeit, Affektwirkungen, Vertrauen und wahrgenommene Fairness. Im Hinblick auf die vorgestellten Feedbackformen und ihren Einsatz im Unterricht ist es daher zentral, ebendiese Dimensionen in der Schüler*innenrezeption genauer zu fokussieren.

In der aktuellen Forschung werden Peer- und KI-Feedback vergleichend, aber auch einander ergänzend, vermehrt im hochschulischen bzw. universitären Kontext untersucht. Banihashem et al. (2024) behandeln beispielsweise in diesem Setting die Qualität von Peer- und KI-Feedback (hier: ChatGPT) beim Schreiben von Essays (vgl. Banihashem et al., 2024, S. 4f.). Die Ergebnisse der Autor*innen zeigen einen signifikanten Unterschied in der Feedback-Qualität zwischen Peer-Feedback und dem von ChatGPT generierten Feedback. Peers liefern qualitativ hochwertigeres Feedback als ChatGPT, was – folgt man Banihashem et al. – hauptsächlich auf die besseren Beschreibungen und die genauere Identifikation von Problembereichen im Feedback zurückzuführen sei (vgl. Banihashem et al., 2024, S. 7f.).

Weiter zeigen die Ergebnisse keine signifikanten Zusammenhänge zwischen der Qualität der Essays und dem Feedback der Peers und dem von ChatGPT. Allerdings wird eine positive Korrelation zwischen der Essayqualität und dem affektiven Aspekt des von ChatGPT erzeugten Feedbacks festgestellt, während eine negative Beziehung zwischen der Essayqualität und dem affektiven Aspekt des von Peers gegebenen Feedbacks beobachtet wird (vgl. Banihashem et al., 2024, S. 8). Statt eine dichotome Sichtweise anzunehmen, die das Peerfeedback vor dem KI-basierten Feedback favorisiert, verweisen Banihashem auf eine komplementäre Nutzung durch das KI-Tool: ChatGPT soll als Feedback-Ressource bei Essay-Aufgaben im Hochschulbereich neben Peer-Feedback integriert werden, um dieses insgesamt zu bereichern (vgl. Banihashem et al., 2024, S. 12).

Einen ähnlichen Schwerpunkt nehmen auch Guo et al. (2024) ein. Sie fokussieren, ebenfalls im universitären Kontext, die Unterstützung von Peerfeedback durch ein KI-Tool (hier: der KI Chatbot *Eva* auf *EvaluMate*): Ein*e

22 Diese Rezeptionsdimension wird im Folgenden mit dem Begriff *Identifikation* (bzw. *Identifizierung* in Anhang 2) abgekürzt.

Student*in gibt Peerfeedback zu einem Aufsatz. Zu diesem Feedback erhält er*sie Rückmeldungen von dem KI-Tool und kann daraufhin das eigene Feedback überarbeiten (vgl. Guo et al., 2024, S. 3). Die Nutzung des KI-Tools geschieht dabei iterativ, die Studierenden können sich also mehrfach Rückmeldungen zu ihren Feedbacks einholen. Guo et al. stellen fest, dass die Nutzung des KI-Tools sowohl die Qualität des Feedbacks in Bereichen wie Beschreibung oder Rechtfertigung stärkt als auch die Schreibleistung (vgl. Guo et al., 2024, S. 7f.). Die Rückmeldung des KI-Tools bezieht sich bei Guo et al. jedoch einzig auf das Feedback der Peers und nicht auf den Ausgangstext (vgl. Guo et al., 2024, S. 9). Es fehlt somit die Passung zum Ausgangstext und damit dem eigentlichen Inhalt des Feedbacks. Eine Förderung der Studierenden ist daher zumindest im Framework von Guo et al. nur bedingt möglich.

Während Banihashem et al. (2024) und Guo et al. (2024) Peer- und KI-Feedback bei älteren, schreib-erprobteren Lernenden untersuchen, wenden sich Steiss et al. (2024) US-amerikanischen Schüler*innen der Klassen sechs bis zwölf zu. Auch hier steht die Qualität des formativen Feedbacks im Fokus. Statt das KI-Feedback (hier: ChatGPT) mit dem eines Peers zu vergleichen, wird ihm jedoch die Expertise einer Lehrkraft gegenübergestellt. Das menschliche Feedback zeigt bessere Ergebnisse im Hinblick auf Aspekte wie Klarheit, Richtigkeit oder unterstützenden Tonfall. Das KI-Feedback ist jedoch stärker kriterienbasiert (vgl. Steiss et al., 2024, S. 7). Zudem ist die Qualität des menschlichen Feedbacks zwar höher als die des KI-Feedbacks, letztere ist aber dennoch gut (vgl. Steiss et al., 2024, S. 7). Je nach Qualität des Essays verändert sich bei der KI (hier nur manchmal beim Menschen) das Feedback. So nimmt z. B. die Genauigkeit des Feedbacks mit steigender Qualität des Essays ab (vgl. Steiss et al., 2024, S. 7). Steiss et al. (2024) schlagen vor, KI-basiertes Feedback vor allem für frühere Textentwürfe zu nutzen und auf menschliches Feedback erst in den letzten Zügen des Schreibens zurückzukommen. So behalte die Lehrkraft die Kontrolle, es komme zum Ergänzen, aber nicht zum Ersetzen der Korrekturtätigkeit (vgl. Steiss et al., 2024, S. 10).

Wie an der hier aufgeführten Literatur deutlich wird, werden bei einem Vergleich der Feedbacks von KI und Peer (bzw. Lehrkraft bei Steiss et al., 2024) häufig Unterschiede in der Qualität fokussiert und untersucht. Die

eigentliche Rezeption durch die Lernenden wird dabei eher weniger beachtet. Dennoch gibt es Studien, die ebendies in den Vordergrund stellen und die Rezeption des Feedbacks beleuchten. Dabei werden sowohl Aspekte wie Vertrauen in die jeweilige Feedbackquelle untersucht als auch affektive Variablen oder die Wahrnehmung von Nützlichkeit oder Fairness.

Ruhe und Mayweg-Paus (2023) zeigen im Rahmen ihrer Untersuchung von non-kognitiven Effekten von KI-basiertem Feedback, dass das KI-System von teilnehmenden Lehramtsstudierenden hinsichtlich seiner Expertise im Vergleich zu Lehrenden oder Peers als vertrauenswürdiger eingeschätzt wird (vgl. Ruwe & Mayweg-Paus, 2023, S. 7). Demgegenüber können sie keine Unterschiede in der Wahrnehmung von Fairness, Nützlichkeit und Akzeptanz zwischen den verschiedenen Feedbackquellen feststellen (vgl. Ruwe & Mayweg-Paus, 2023, S. 8).

Auch Jansen et al. (2024) führen ihre Studie an Lehramtsstudierenden durch. Hierbei folgen sie der Fragestellung, wie diese die Nützlichkeit eines KI-generierten Feedbacks gegenüber einem Expert*innenfeedback einschätzen – ohne die Quelle zu kennen (vgl. Jansen et al., 2024, S. 83). Die textuelle Basis dafür sind Schüler*innentexte, auf die sowohl eine Fachlehrkraft als auch ChatGPT ein Feedback gibt. Nach Sichtung der Feedbacks ermitteln Studierende deren Nützlichkeit, vergleichen sie und entscheiden dann, welches sie dem*r jeweiligen Schüler*in geben würden. Im Anschluss erfolgt ein Debriefing und die Studierenden müssen entscheiden, welches Feedback mit der KI erstellt wurde (vgl. Jansen et al., 2024, S. 83).

Eine einfache und verständliche Sprache wird beim Feedback präferiert (vgl. Jansen et al., 2024, S. 88). In 77,3% der Fälle bevorzugten bei Jansen et al. (2024) die Studierenden das Feedback der Expert*innen (vgl. Jansen et al., 2024, S. 86). 85% der Teilnehmenden ordneten das KI-Feedback richtig zu (vgl. Jansen et al., 2024, S. 86). Auch Jansen et al. (2024) sehen Potential in dem KI-generierten Feedback und bezeichnen es als vermutlich lernförderlich, vor allem für die Schüler*innen, die sonst kein Feedback bekommen würden (vgl. Jansen et al., 2024, S. 87).

Dabei erlebt die Studie jedoch eine starke Limitation im Hinblick auf die Teilnehmenden: Die Textgrundlage kommt von Schüler*innen, das Feedback ist an Schüler*innen gerichtet, doch die final Rezipierenden sind

Studierende. So schreiben Jansen et al. (2024, S. 89) selbst: »While we focused on student teachers assessing the quality of feedback, an essential perspective – that of the students receiving the feedback – was not included«. Die Schüler*innen als Rezipierende in die Studie zu inkludieren, ist – folgt man dem Wortlaut der Autor*innen – eine essentielle Perspektive, die hier jedoch vernachlässigt wurde. (Lehramt-)Studierende weisen eine andere Schreibbiografie auf als Schüler*innen der Sekundarstufe I und II. Sie sind geübter im Schreibprozess, haben bereits eine differenziertere Wahrnehmung von Feedback (vgl. Ortner, 2006; Pokorny & Pickford, 2010). Nur weil Schreibfortgeschrittene ein Feedback als beispielsweise nützlich für Schüler*innen einstufen, bedeutet dies nicht, dass letztere ebenso empfinden, das Feedback also auf die gleiche Art und Weise rezipieren.

Um ein vollständiges Bild im Rezeptionsvergleich von KI-basiertem Feedback und Peer-Feedback erzielen zu können, müssen somit auch Schüler*innen in ihren Einschätzungen zu Wort kommen. Die im Folgenden dargelegte Studie verfolgt den Anspruch, sich ebendieser Forschungslücke zu nähern, indem in Anlehnung an die identifizierten Rezeptionsdimensionen und die Bestrebungen von Jansen et al. (2024) und Ruwe und Mayweg-Paus (2023)²³ die Feedbackrezeption von Schüler*innen der Sekundarstufe I untersucht wird.

23 Sowohl Jansen et al. (2024) als auch Ruwe und Mayweg-Paus (2023) nutzen in ihren Studien zur Erfassung der Daten Auszüge aus dem *Feedback Perception Questionnaire* (FPQ; Strijbos et al., 2010). Dieses legen sie jedoch an einen Rezipient*innenkreis an, der nicht dem der Schüler*innen der Sekundarstufe I entspricht. Daher ist es sinnvoll, in der folgenden Studie u. a. ebendieses Messinstrument im Kontext der Sekundarstufe I anzuwenden (s. hierzu Kapitel 3.3.3.4).

3 Rezeption von KI- und Peerfeedback im Deutschunterricht der Klasse 6 und 7

Um an die Forschungsdesiderate aus 2.5 anzuknüpfen, wird in der vorliegenden Studie die Rezeption von Peerfeedback und KI-basiertem Feedback bei Schüler*innen der Sekundarstufe I, genauer der Klassen 6 und 7, vergleichend untersucht. Die Schüler*innen wurden dabei im Vorhinein nicht über die Quelle der Feedbacks informiert. Die Feedbackarten werden hinsichtlich der wahrgenommenen Nützlichkeit, der affektiven Wirkungen, der Identifikation von Unterstützung, der wahrgenommenen Fairness und Akzeptanz der Schüler*innen analysiert. Ergänzend soll dabei ermittelt werden, welche Komponenten der Feedbacks bei den Schüler*innen verstärkt Beachtung finden. Von zentraler Relevanz ist zudem, welches Feedback von den Schüler*innen präferiert wird und ob eine Veränderung ihrer Meinung eintritt, wenn sie über die Feedbackquelle in Kenntnis gesetzt werden.

Im Folgenden soll dafür in einem ersten Schritt das KI-Tool vorgestellt werden. Im Anschluss werden die Forschungsfragen und Hypothesen dargelegt, bevor dann das methodische Vorgehen erläutert und die Ergebnisse aufgezeigt werden. An eine Diskussion dieser schließt sich eine kritische Be- trachtung der Studie im Hinblick auf ihre Limitationen an.

3.1 FelloFish als Tool für KI-basiertes Feedback

Das KI-basierte Feedback-Tool *FelloFish* (früher *fiete.ai*²⁴) wurde 2023 von Malte Hecht und Hendrick Haverkamp veröffentlicht und arbeitet auf Basis

24 Die Umbenennung des Tools erfolgte Anfang 2025 und lag damit im Zeitraum der hier behandelten empirischen Erhebung. Daher verwenden sowohl einige der ge- listeten Quellen als auch die erstellten Materialien in Anhang 1 noch den alten Namen *fiete.ai*.

von ChatGPT. Als Tool wurde es dabei speziell für Bildungskontexte entwickelt und ist noch recht neu in der (deutschen) Schullandschaft.²⁵

FelloFish knüpft an Schreibaufgaben bzw. unterrichtliche Kontexte an. In einem Interview mit dem Bildungsmagazin News4teachers definiert der Ko-Entwickler Haverkamp das Tool als »KI-Schreibbegleiter [...], der Schülerinnen und Schülern lernförderliche Rückmeldungen gibt – und zwar immer dann, wenn sie es brauchen und ganz personalisiert« (News4teachers, 2025). Das DSGVO-konforme Tool sei dabei nicht an Fächer gebunden. So könne es prinzipiell immer dann eingesetzt werden, wenn Texte produziert werden. Auch multimodale Aufgaben seien möglich, sodass beispielsweise Diagramme oder Karikaturen eingestellt werden können (vgl. News4teachers, 2025).

Mittlerweile gibt es online bereits eine Vielzahl an Praxisbeispielen und -berichten von Lehrkräften, die das Tool im Unterricht ausprobiert haben und nun auf privaten Websites (vgl. z. B. Falck, 2024) oder dem Blog von *FelloFish* (<https://www.fellofish.com/blog>) berichten. Hinzu gesellen sich Tutorials von Lehrkräften auf YouTube, in denen Aufgabenerstellung und -bearbeitung erklärt werden (z. B. [unterrichten.digital-tutor125](#), 2024).

FelloFish setzt dabei im Passungsproblem zwischen dem zeitlichen Aufwand und der gleichzeitigen Notwendigkeit von formativem Feedback an: Es verspricht Entlastung für Lehrkräfte und unmittelbares formatives Feedback für Schüler*innen, könnte also in Zukunft zu einem integralen Bestandteil des Schreibprozesses werden. Um das Tool für die folgende Erhebung besser einordnen zu können, sollen im Folgenden sowohl seine Funktionsweise und der Aufbau seines Feedbacks als auch seine Einsatzmöglichkeiten im (Schreib-)Unterricht dargelegt werden.

25 Im Rahmen eines Pilotprojektes des Ministeriums für Bildung und Kindertagesförderung des Landes Mecklenburg-Vorpommern soll beispielsweise die Funktionsweise des Tool an 16 Schulen zukünftig erprobt werden (Bewerbungsschluss war am 14. April 25) (vgl. Ministerium für Bildung und Kindertagesförderung Mecklenburg-Vorpommern, o. D.).

3.1.1 Funktionsweise und Aufbau

Das Ziel von *FelloFish* ist es, sowohl den Schüler*innen passgenaue Rückmeldungen zu geben, als auch den Lehrkräften ebendiese Rückmeldungen anzuzeigen. Haverkamp et al. (2024, S. 66) sprechen hierbei von einem Feedback »in zwei Richtungen«:

Die Schüler:innen erhalten eine Rückmeldung zu ihren Lösungen im Hinblick auf die jeweiligen Aufgaben und die Lehrkraft zur Wirksamkeit des eigenen Unterrichts bzw. zum Lernstand der Klasse. (Haverkamp et al., 2024, S. 66)

Damit dies funktionieren kann, muss die Lehrkraft in einem ersten Schritt eine Aufgabe in *FelloFish* anlegen. Beim Anlegen können Lehrkräfte zusätzlich zur Aufgabenstellung und der Überschrift die Jahrgangsstufe auswählen, für die die Aufgabe erstellt wurde, und bis zu sieben Feedbackkriterien festlegen. Die Kriterien sind essentiell für den anschließenden Feedbackprozess und geben der KI Anhaltspunkte für ihre Rückmeldungen. Neben dem eigenen Erstellen von Aufgaben können Lehrkräfte auch vorgefertigte Aufgabentypen (z. B. Vorgangsbeschreibungen, Gedichtinterpretation, materialgestütztes Schreiben) und vorgefertigte Aufgaben nutzen oder Aufgaben miteinander teilen.

Nach Fertigstellung der Aufgabe können Lehrkräfte auswählen, ob Schüler*innen zur besseren Transparenz die Feedbackkriterien angezeigt bekommen sollen oder nicht. Darüber hinaus können sie ihnen Tools wie eine Rechtschreib-/Grammatikprüfung, eine Übersetzung des anschließenden Feedbacks in eine andere Sprache oder in Einfache und Leichte Sprache oder das Vorlesen des Feedbacks zur Verfügung stellen (Stand: Februar 2025).

Möchte die Lehrkraft die Aufgabe im Anschluss verteilen, geschieht dies u. a. über einen QR-Code, den Schüler*innen scannen können. Die Aufgabe können sie dann mit einem selbstgewählten Nutzer*innennamen entweder direkt in *FelloFish* bearbeiten, den getippten Text später einfügen oder Handgeschriebenes abfotografieren. Sobald sie den Button *Gib mir Feedback* betätigen, wird von der KI eine individuelle Rückmeldung erstellt.

Das anschließende Feedback erfolgt dann anhand der vorher definierten Feedbackkriterien²⁶ – dies sowohl über Statusbalken, die die Leistung des*r Schüler*in knapp verorten, als auch über ein ausformuliertes Feedback, welches konkret auf die Schüler*innentexte Bezug nimmt. An dieser Stelle ist für die Lernenden eine Feedbackschleife möglich: Nach dem ersten erhaltenen Feedback, können die Schüler*innen auf dieses digital eine Rückmeldung geben und so auch die Lehrkraft in den Feedbackprozess einbetten (vgl. Haverkamp et al., 2024, S. 66). Darüber hinaus können die Lernenden nach der Überarbeitung ihres ersten Textentwurfs die neue Version ihres Textes hochladen und danach ein erneutes Feedback mittels KI erhalten.

Das Textfeedback von *FelloFish* bietet auf diese Weise das Potential, formativ und prozessorientiert zu wirken, indem es nicht summativ am Ende des Schreibprozesses ansetzt, sondern in diesen eingebunden wird. Die erste Abgabe der Schüler*innen wird hier (anders als bei einer Klassenarbeit oder Klausur) zu einem Entwurf, der – angeleitet durch die KI – überarbeitet werden kann. Die Lehrkraft kann die einzelnen Schritte dabei überwachen, sieht also sowohl die Feedbacks als auch die abgegebenen Text(entwürf)e, und kann entsprechend im Unterricht reagieren (vgl. Haverkamp et al., 2024, S. 66).

3.1.2 Einsatzmöglichkeiten im Schreibunterricht

Um im Schreibprozess der Schüler*innen zu unterstützen, ist es wichtig, prozessorientiert zu arbeiten. Das bedeutet auch, ihre Texte als Entwürfe wahrzunehmen und die aktive Auseinandersetzung mit und die Überarbeitung von diesen zu unterstützen (vgl. Becker-Mrotzek & Böttcher, 2015, S. 135). Hier verspricht *FelloFish* anzusetzen, indem Textentwürfe der Schüler*innen hochgeladen werden, sie durch das Feedback im Arbeitsprozess begleitet und dadurch formativ gefördert werden können.

Haverkamp et al. (2024) identifizieren dabei verschiedene Lern-/Arbeitsfelder, in denen *FelloFish* eingesetzt werden kann: bei der Nutzung von Aufgaben aus Lehrwerken, bei der Vorbereitung auf Klausuren und

26 *FelloFish* arbeitet dabei auf Basis der erstellten Aufgaben und Kriterien, wurde »also nicht auf spezifische Texte vortrainiert« (Haverkamp et al., 2024, S. 67).

Klassenarbeiten durch z. B. Probeklausuren, Berichtigungen, Hausaufgaben und in gemeinsamen Lernsituationen (vgl. Haverkamp et al., 2024, S. 66f.; vgl. auch News4teachers, 2025). Wie sich hier zeigt, kann *FelloFish* auch abseits des unmittelbaren schulischen Unterrichtskontextes eingesetzt werden und so eine Begleitung im häuslichen Umfeld ermöglichen. Neben der Entlastung der Lehrkräfte kann dies auch Chancengleichheit unter den Lernenden bedeuten. So können nicht alle Schüler*innen im häuslichen Kontext auf die gleiche Unterstützung bei z. B. dem Bearbeiten der Hausaufgaben zurückgreifen. Durch die Nutzung eines KI-Tools wie *FelloFish* könnte dieses Ungleichgewicht verringert werden (vgl. News4teachers, 2025).

Beispiele für die konkrete Nutzung von *FelloFish* finden sich in den bereits angesprochenen Blogs und Websites, in denen Lehrkräfte von ihren Erfahrungen berichten. So schreiben bei Falck (2024) Schüler*innen der Jahrgangsstufe 8 einer Mittelschule im Rahmen einer Deutsch-Schulaufgabe eine Inhaltsangabe zu einer Kurzgeschichte. Das von *FelloFish* erhaltene Feedback kann (und soll) dabei von den Schüler*innen an zwei Stellen (jeweils nach den Feedbacks) eingearbeitet werden, bevor der Text final abgegeben wird. Der Lehrer merkt bei der Sichtung der finalen Produkte, »dass die Texte der Schüler*innen umfangreicher, inhaltlich gehaltvoller und sprachlich höherwertiger waren« (Falck, 2024). Er geht zudem davon aus, dass eine stärkere Auseinandersetzung mit den eigenen Texten stattgefunden habe, die mit einer Reflexion des Feedbacks und einer Überarbeitung der Texte einhergegangen sei (vgl. Falck, 2024).

Die Lehrkraft erkennt aber auch Limitationen für *FelloFish* im Schreibunterricht: Einerseits erkennt Falck mögliche Schwierigkeiten bei der Wortwahl von *FelloFish* – beispielsweise scheinen die verwendeten Fremdwörter des KI-Tools für die Schüler*innen zu schwierig gewesen zu sein –, andererseits müssten die Fortschrittsbalken, die *FelloFish* bei den Schüler*innenlösungen zeigt, kritisch betrachtet werden. So schreibt Falck (2024):

Als Lehrkraft sollte man sich nicht auf die Anzeige der Fortschrittsbalken verlassen. Teilweise wurden Verbesserungen angezeigt, obwohl sich der Schüler*innen-Text nicht verändert hat.

Auch spricht Falck Probleme in der Texterkennung durch das KI-Tool an. So habe *FelloFish* an manchen Stellen Fehler vermerkt, die der*die Schüler*in nicht begangen habe. Er führt hier das angebliche Fehlen der Autorinnen-nennung oder eine mangelnde Chronologie in den produzierten Texten an.

Bei Pölert (2023) werden die Schüler*innen angehalten, über das KI-Feedback zu sprechen. Hier prüfen sie seine Korrektheit und betrachten dabei ihren Ausgangstext. Sie setzen sich mit Kriterien einer guten Zusammenfassung auseinander und reflektieren den Umgang mit KI. Pölert bezeichnet *FelloFish* »[a]ls sowohl im Unterricht als auch zuhause einsetzbaren Schreibbegleiter« (Pölert, 2023). *FelloFish* ermögliche individualisiertes, auf Kriterien basiertes Feedback, könne als Diagnosetool eingesetzt werden und als Gesprächsanlass fungieren. Doch auch Pölert zeigt Limitationen des Tools auf. So erkennt auch er Halluzinationen, die im Feedback von *FelloFish* vorliegen können. Dem gegenüber setzt er jedoch, dass *FelloFish* gerade für sprachliche Korrekturen eingesetzt werden und Lehrkräfte so in ihrem Feedback entlasten könne.

Die hier angeführten Stimmen aus der Praxis weisen im Generellen positiv auf das KI-Tool. Das hier gegebene Feedback ordnen sie als lernförderlich ein. *FelloFish* gebe kriterienbasierte Rückmeldungen, die Produkte der Schüler*innen werden von zumindest Falck (2024) als gehaltsvoller und sprachlich besser eingestuft. Mit den Limitationen des Tools gehen beide unterschiedlich um. Während Halluzinationen bei Pölert (2023) nur in geringem Maße angesprochen werden, zeigt Falck (2024) eine Reihe von Problemstellen auf.

Entscheidend ist dabei jedoch, dass es sich sowohl bei den wahrgenommenen Chancen als auch den Unterschieden um die Wahrnehmung der Lehrkräfte handelt. Die Rezeption durch die Schüler*innen wird nur indirekt – über Erzählungen der Lehrkraft – weitergegeben. Gerade um jedoch den vollen Umfang der Einsatzmöglichkeiten von *FelloFish* im (Schreib-)Unterricht begreifen zu können, ist es notwendig, auch die Schüler*innenperspektive als vollwertig zu verstehen.

3.2 Forschungsfragen und Hypothesen

In Anknüpfung an die Überlegungen aus 2.5 lassen sich folgende (globale) Forschungsfragen formulieren: Inwiefern unterscheidet sich die Feedbackrezeption der Schüler*innen im Hinblick auf die Feedbackquelle (KI vs. Peer), wenn sie diese nicht kennen? Ändert sich ihre Ausgangswahrnehmung, wenn sie im Anschluss über die Quelle informiert werden? Zur weiteren Präzisierung wird die erste der beiden leitenden Fragestellungen angelehnt an die identifizierten Rezeptionsdimensionen in detailliertere Subfragen überführt. Folgende Fragestellungen ergeben sich so für die Studie:

*Wie unterscheiden sich die Feedbackarten hinsichtlich ihrer Affektwirkung, der Akzeptanz durch die Schüler*innen, der wahrgenommenen Fairness und Nützlichkeit und der Identifikation von Fehlerstellen und Verbesserungsvorschlägen?* (F1a)

*Welches Feedback wird von den Schüler*innen bevorzugt?* (F1b)

Im Hinblick auf die Eigenschaften des Peerfeedbacks ist anzunehmen, dass Schüler*innen eine positivere Affektwirkung bei dieser Form des Feedbacks verspüren. Sind die Schüler*innentexte anonymisiert, entfallen die von Gamlem und Smith (2013) angeführten Verzerrungseffekte. Auf der einen Seite kann dies dazu führen, dass freundschaftliche Beziehungen das jeweilige Feedback nicht positiv, auf der anderen Seite aber Feindseligkeiten das Feedback auch nicht negativ beeinflussen. Dennoch bewegen Peers sich – anders als die KI – im gleichen Lernkontext wie die Schüler*innen, die das Feedback erhalten. Mitschüler*innen haben daher andere Einblicke als die KI, zeigen ggf. mehr Empathie, was zu positiveren Emotionen führen könnte.²⁷

Im Gegensatz zu Peers kann die KI im Feedback auf eine große Datenmenge zugreifen. Im Vergleich mit der Expertise einer Lehrkraft stellen Steiss et al. (2024) zwar einen qualitativen Unterschied fest, ordnen das KI(ChatGPT)-

27 Gleichzeitig besteht durch den Novizenstatus der Peers natürlich auch die Möglichkeit, dass Fehler übersehen oder Verbesserungsaufforderungen nicht gegeben werden, sodass den Autor*innen der Texte statt konstruktiver Kritik, häufiger Lob begegnet. Auch hier wäre als Folge mit einer positiveren Affektwirkung zu rechnen.

Feedback dennoch als gut ein. Da die hier verwendete KI (*FelloFish*) den Anspruch verfolgt, gezieltes Feedback für den schulischen Kontext zu geben, kann auch hier in erster Linie von einem (zumindest teilweise) treffenden Feedback ausgegangen werden.

Ruhe und Mayweg-Paus (2023) konnten in der Wahrnehmung von Fairness, Nützlichkeit und Akzeptanz zwischen den Feedbackquellen Lehrkraft, Peers und KI keine Unterschiede feststellen. Im Gegensatz zu ihren Teilnehmenden, die im universitären Kontext zu verorten waren, stammen die Partizipierenden in dieser Studie jedoch aus dem schulischen Kontext – genauer gesagt der Sekundarstufe I. Es ist also anzunehmen, dass sie weniger schreiberprobt als die Studierenden bei Ruwe und Mayweg-Paus sind, was zu qualitativen Unterschieden im Feedback führen kann. Für die Forschungsfrage wird daher angenommen, dass das Feedback durch die KI im Hinblick auf Akzeptanz, Identifikation, Wahrnehmung von Fairness und Nützlichkeit in der Rezeption der Schüler*innen positiver wahrgenommen wird. Aufgrund der fachlichen »Expertise« der KI ist zudem davon auszugehen, dass ein Großteil der Schüler*innen dieses Feedback präferiert.

*Welche Komponenten der einzelnen Feedbacks finden verstärkt Beachtung bei den Schüler*innen? (F2)*

Wie die Untersuchungen von Pokorny und Pickford (2010) und van der Kleij (2019) zeigen, empfinden Lernende Spezifität und konkrete Hinweise im Feedback als besonders hilfreich. Es ist daher anzunehmen, dass ebendiese Komponenten – bzw. das Fehlen von ihnen – auch bei den Schüler*innen der Sekundarstufe I verstärkt Beachtung finden. Hinzu tritt – folgt man Johnsson (2013) – die Notwendigkeit, dass Schüler*innen das gegebene Feedback auch nutzen können. Möglich ist also auch, dass Komponenten wie (mangelnde) Passung zum eigenen Text, (un-)verständliche Verbesserungsvorschläge o. Ä. berücksichtigt werden. Auch Fehlerstellen im Feedback (wie z. B. Halluzinationen durch die KI, vgl. Falck 2024) könnten hier fokussiert werden.

Folgt man den Überlegungen und Befunden hinsichtlich der Qualität des KI-Feedbacks, so ist zu vermuten, dass hier gerade ein starker Kriterienbezug, aber auch eine Klarheit bzw. Spezifität im Feedback positiv hervorgehoben

wird, während im Peerfeedback häufiger emotionale Komponenten, aber auch Fehlerstellen genannt werden.

Ändert sich ihre Meinung, wenn sie über die Quelle des Feedbacks informiert werden? (F3)

Wie Winstone et al. (2017) zeigen, spielt Vertrauenswürdigkeit bei der Rezeption von Feedback eine große Rolle. Dieser Aspekt greift in der folgenden Studie in dem Moment, in dem die Quelle des Feedbacks bekannt gegeben wird. Die Ausführungen von Ruwe und Mayweg-Paus (2023) zeigen im hochschulischen Kontext, dass der KI eine höhere Vertrauenswürdigkeit zugeschrieben wird. Auch für den schulischen Kontext zeigt beispielsweise die JIM-Studie (2024), dass eine große Zahl an Schüler*innen KI zur Beschaffung von Informationen oder für die eigenen Hausaufgaben benutzt. Gepaart mit der Bemerkung Strebers (2019), dass Schüler*innen Peerfeedback ggf. weniger vertrauen, wird angenommen, dass sich die Wahrnehmung der Schüler*innen, sobald sie die Quelle des Feedbacks kennen, zugunsten der KI verändert.

3.3 Methodisches Vorgehen

Im Folgenden soll das methodische Vorgehen dieser Studie dargelegt werden. Zu Beginn wird dafür die Stichprobe beschrieben und der zeitliche Ablauf thematisiert. Im Anschluss werden Material und Messinstrumente vorgestellt. Hierbei sollen sowohl die Textauswahl, die Aufgabenstellungen und die jeweiligen Kriterien, als auch die Eingaben bei *FelloFish*, der genutzte Fragebogen und Anpassungen zwischen den Durchgängen beleuchtet werden.

3.3.1 Beschreibung der Stichprobe

Die vorliegende Studie wurde mit Sechst- und Siebtklässler*innen eines Gymnasiums in NRW durchgeführt. In Klasse 6 nahmen 21 Schüler*innen ($n_6 = 21$) und in Klasse 7 26 Schüler*innen ($n_7 = 26$) an allen Treffen teil.²⁸ Die

²⁸ Die Studie wurde pro Klasse an je drei Terminen durchgeführt. Einige Schüler*innen, die beim ersten Termin fehlten, waren ab dem zweiten Termin anwesend. In diesem Fall bearbeiteten diese Schüler*innen im Fortlauf der Studie Duplikate

Studienintervention fand bei beiden Klassen im Deutschunterricht statt. In Klasse 6 war die reguläre Deutschlehrkraft vor Ort, in Klasse 7 wurde die Verfasserin der Masterarbeit von Vertretungslehrkräften begleitet. Nach Auskunft der jeweiligen Deutschlehrkräfte, hatten beide Klassen bereits grundlegende Erfahrungen mit Peerfeedback gesammelt.

Die Lehrkräfte wurden im Vorhinein über das genaue Forschungsziel der Studie aufgeklärt. Den Schüler*innen gegenüber wurde kommuniziert, dass sich die Studie mit dem Thema Feedback auseinandersetze, die Schüler*innen dabei eine Zusammenfassung selbstständig produzieren und im Anschluss zu einer anderen Feedback geben würden. Auch wurde ihnen mitgeteilt, dass sie am Ende Feedbacks zu ihrem eigenen Text bekommen und diese dann anhand von Fragebögen evaluieren würden. Um Verzerungseffekte durch eventuelles Vorwissen und Vorstellungen über KI-generierte Texte zu vermeiden, wurde zu Beginn der Studie nicht mitgeteilt, dass eins der Feedbacks von *FelloFish* hervorgebracht werden würde.

3.3.2 Zeitlicher Ablauf

Die Erhebungen fanden Ende 2024 (Klasse 7) und Anfang 2025 (Klasse 6) statt. Pro Klasse wurden dabei je drei Treffen durchgeführt (vgl. Abb. 1): Zu Beginn des ersten Treffens erfolgte die Vorstellung der Verfasserin der Masterarbeit und eine eingeschränkte Einführung ins Studienvorhaben. Die Verfasserin stellte klar, dass die produzierten Texte nicht zur Bewertung und Benotung an die regulären Deutschkräfte weitergegeben werden würden.

ihrer Mitschüler*innen. Um jedoch möglichst authentische Ergebnisse zu erzielen, wurden in der Studienauswertung nur die Teilnehmenden berücksichtigt, die an allen drei Terminen zugegen waren.

Vor der Studiendurchführung wurden für den Verlauf relevante Begriffe wie *Anonymität* thematisiert und gemeinsam mit der Klasse definiert.

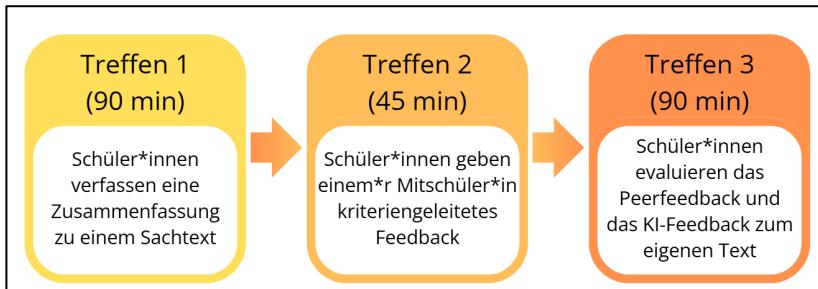


Abb. 1: Zeitlicher Ablauf

Im Anschluss wurden die Sachtexte ausgeteilt. Nach Besprechung der Aufgabenstellung wurde den Schüler*innen ein Zeitfenster von 45 Minuten gegeben, um eine Sachtextzusammenfassung anzufertigen – eine frühere Abgabe war ebenfalls möglich. Die Schüler*innen nutzten für das Schreiben der Zusammenfassung Schul-iPads²⁹ sowie persönliche Codes, die ein anonymes Vorgehen gewährleisteten. Während der Arbeitsphase stand die Verfasserin der Masterarbeit den Schüler*innen bei technischen Problemen oder Begriffsdefinitionen zu Verfügung, intervenierte darüber hinaus jedoch nicht.

Im Anschluss an das erste Treffen wurden die Texte hinsichtlich des Layouts normiert und ggf. eingefügte Bilder entfernt. Folgt man Steiss et al. (2024), so behindern weitere Korrekturen am und im Text der Schüler*innen die Authentizität der Feedbacksituationen im prozessbasierten Schreiben. In ihrer eigenen Studie zu ChatGPT schreiben sie:

Student writing was not corrected for grammar or spelling prior to evaluation by humans or ChatGPT so the study would most closely emulate authentic feedback situations in process-based writing. Because many teachers or evaluators are responding to student writing as written, not

29 Eine Ausnahme ist bei einem*r Schüler*in der Klasse 6 zu vermerken. Hier äußerte der*die Schüler*in Unsicherheiten beim Tippen mit dem iPad und durfte die Zusammenfassung daraufhin händisch notieren. Später wurde der Text durch die Verfasserin der Masterarbeit ins Digitale überführt.

correcting texts for errors presents a more ecologically valid test situation for the quality of ChatGPT and human evaluators. (Steiss et al., 2024, S. 3)

Daher wurden auch hier die Schüler*innentexte bis auf die Änderungen im Layout in ihrem Originalzustand gelassen.

Die normierten Texte wurden für das zweite Treffen mit den Schüler*innen ausgedruckt. In einer 45-minütigen Unterrichtsstunde wurden Letztere gebeten, den Text eines*r Mitschülers*in zu sichten – die Zuteilung erfolgte dabei zufällig und anonymisiert – und auf Basis eines Rückmeldebogens kriteriengeleitetes Feedback zu geben (vgl. Anhang 1.3). Um hier ein möglichst authentisches Peerfeedback zu erzeugen, wurden Feedbackregeln zu Beginn wiederholt (vgl. auch Lehnen, 2023, S. 24f.), darüber hinaus intervenierte die Verfasserin der Masterarbeit jedoch nicht. Des Weiteren wurden die Schüler*innen gebeten, ganze Sätze zu formulieren, um eine spätere Vergleichbarkeit mit dem Feedback durch *FelloFish* zu gewährleisten. Während des zweiten Treffens nutzte Klasse 6 vorbereitete Feedbackbögen auf iPads. Bei Klasse 7 lagen die Feedbackbögen analog vor.

Im dritten und letzten Treffen erhielten die Schüler*innen ihre eigenen Texte zurück. Beigefügt waren sowohl zwei Feedbackbögen (Peer und *FelloFish*, vgl. hierzu auch die Bögen von ELNA_1 exemplarisch in Anhang 1.3.2 und 1.3.3) als auch der erste Teil der Evaluations- bzw. Fragebögen (vgl. Anhang 1.4). Nach Besprechung des Ablaufes sichteten die Schüler*innen zuerst ihre eigenen Texte, verglichen diese dann mit den Feedbacks und füllten die Evaluationsbögen aus. Die Schüler*innen wurden dabei nicht darüber in Kenntnis gesetzt, welche Quelle welchem Feedback zugrunde lag: Für die eine Hälfte der Klasse war das Feedback durch *FelloFish* als Feedback 1 und das Peerfeedback als Feedback 2 aufgeführt. Die andere Hälfte der Klasse erhielt das Peerfeedback als Feedback 1 und das Feedback durch *FelloFish* als Feedback 2. Zudem wurden die jeweiligen Feedbacks und die Evaluationsbögen farblich codiert, um seitens der Schüler*innen Verwechslungen beim Ausfüllen zu vermeiden.

Nach Abgabe der ersten Evaluationsbögen wurde aufgeschlüsselt, welches Feedback von den Peers und welches von einer KI verfasst wurde. *FelloFish* (zu diesem Zeitpunkt noch *fiete.ai*) wurde den Schüler*innen knapp vorgestellt, bevor dann der letzte Evaluationsbogen (vgl. Anhang 1.5) verteilt

wurde. Nach Bearbeitung durch die Schüler*innen wurde dieser ebenfalls eingesammelt.

3.3.3 Material und Messinstrumente

Im Folgenden sollen sowohl das Material als auch die in der Studie angewandten Messinstrumente vorgestellt werden. Dabei soll in einem ersten Schritt die Textauswahl und die Aufgabenstellung und Kriterienauswahl behandelt werden. Im Anschluss werden die Eingaben bei *FelloFish*, der verwendete Fragebogen und die Anpassung zwischen den Durchgängen thematisiert.

3.3.3.1 Textauswahl

Im Deutschunterricht begegnen Schüler*innen einer Reihe von Textsorten, von Kurzgeschichten und Gedichten bis hin zu Zeitungsartikeln oder Blogbeiträgen (vgl. Rezat, 2022, S. 190f.). Sachtexte sind ein Teil eben dieses Spektrums an Textsorten, durch die und mit denen Schüler*innen über Jahrgangsstufen hinweg arbeiten und lernen. Unterschieden werden können Sachtexte in kontinuierliche und nicht-kontinuierliche bzw. diskontinuierliche Texte. Letztere stellen dabei zumeist eine Kombination aus Text im engeren Sinne und ergänzenden oder weiterführenden Grafiken oder Abbildungen dar (vgl. Gierlich, 2005, S. 29–31). Sachtexte unterscheiden sich darüber hinaus in ihrer Intention. Sie können als Sachbuchtexte informieren und darstellen, als Rede oder Essay argumentieren und überzeugen oder als Gesetze regulativ wirken (vgl. Gierlich, 2005, S. 27).

Schüler*innen begegnen Sachtexten im unterrichtlichen Kontext oftmals im Rahmen einer Sachtextzusammenfassung. Im Kernlehrplan des Landes NRW tauchen Sachtexte beispielsweise im Inhaltsfeld 2 *Texte* auf (vgl. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019, S. 18). Hier wird das Zusammenfassen schriftlicher (und mündlicher) Texte als Kompetenzerwartung für das Ende der Erprobungsstufe und das Ende der Sekundarstufe I formuliert (vgl. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019, S. 16, 23). Die Bedeutung dieser Kompetenz kann jedoch auch darüber hinausgehen. So schreibt Berkemeier (2010, S. 213):

Beim schriftlichen Zusammenfassen von Sachtexten handelt es sich ohne Frage um eine Qualifikation, die zumindest bis zum Ende der Schul- und Ausbildungszeit, im Hinblick auf viele Berufe sogar lebenslang von Bedeutung ist. Es handelt sich um eine Text-, Lern- und Lehrform und dient somit epistemischen wie auch kommunikativen Zwecken.

Das Verfassen einer Sachtextwiedergabe oder -zusammenfassung kann gerade für jüngere Schüler*innen eine beachtliche Herausforderung darstellen. So müssen sie den Sachtext nicht nur lesen und verstehen, sondern auch den gedanklichen Aufbau, die Sprache und die Gestaltung beachten, relevante Wissenselemente erkennen und entnehmen und schließlich auf Basis dessen einen eigenen Text produzieren und überarbeiten (vgl. Berkemeier, 2010, S. 213f.; Gierlich, 2005, S. 38f.; Jünger & Escher, 2023, S. 62). Hinzu treten Herausforderungen in den Bereichen Lexik und Syntax, die vor allem den Verstehensprozess beeinflussen. Dazu gehören Fremd- oder Fachwörter, Komposita oder Attributreihungen (vgl. Gierlich, 2005, S. 31f.). Diese können von den Schüler*innen zwar unter Anwendung diverser Strategien erarbeitet werden (vgl. Gierlich, 2005, S. 36), stellen aber im ersten Moment eine Hürde dar.

Auch kann es bei Sachtextzusammenfassungen in der eigenen Textproduktion zu Komplikationen kommen. So schreibt Berkemeier (2010, S. 215) im Kontext der Erstellung von Stichwörtern für Exzerpte, dass

[b]ei SchülerInnen [...] erfahrungsgemäß häufig festzustellen [ist], dass dabei genau die mentale Rekonstruktion der argumentativen Strukturen nicht gelingt: Es werden mitunter ganze Sätze kopiert, falsche oder zu wenige Wissenselemente getilgt, ausgewählte Wissenselemente werden falsch oder gar nicht verknüpft, sprachliche Mittel werden ohne grammatisch oder lexikalisch ökonomische Umformungen übernommen

Diese potenziellen Problemstellen machen deutlich, dass auch bei der Zusammenfassung von Sachtexten die Begleitung und Stützung unerlässlich ist. Diese erfolgt auf der einen Seite präventiv. Die Lehrkraft muss sicherstellen, dass der von ihr ausgewählte Sachtext auf sprachlicher Ebene für die Lernenden verständlich ist und ihnen dabei auf inhaltlicher Ebene ein möglichst vollständiges Bild des Sachverhaltes bietet. Auf der anderen Seite kann die Begleitung der Lehrkräfte aber auch während des Schreibens und

beim Überarbeiten des Geschriebenen erfolgen (vgl. Jünger & Escher, 2023, die in ihrem Beitrag erläutern, wie sie Schüler*innen mittels des Instruments *Code-Knacker* bei der Sachtextwiedergabe fördern).

Es handelt sich bei der Sachtextzusammenfassung somit um eine Herausforderung für Schüler*innen, der sie im Unterrichtskontext über die Jahrgangsstufen hinweg begegnen, und die auch im Überarbeitungsprozess Unterstützung durch die Lehrkraft bzw. durch adäquates Feedback braucht. Daher empfiehlt es sich, das Schreiben einer Sachtextzusammenfassung ins Zentrum des hier beschriebenen Forschungsinteresses zu setzen und den Schüler*innen für ihre Zusammenfassungen einen auf die jeweilige Lerner*innengruppe abgestimmten Sachtext vorzulegen.

Die Auswahl der Texte erfolgte in Absprache mit der jeweiligen Deutschlehrkraft: Für die Klasse 7 wurde ein Sachtext aus dem im Unterricht genutzten Deutschbuch *Deutsch kompetent 7* (Blatt et al., 2020, S. 22) verwendet. Für Klasse 6 wurde das dem Deutschbuch *Deutsch kompetent 6* zugewiesene Zusatzmaterial der Online-Ressource gewählt (vgl. Ernst Klett Verlag GmbH, 2020). Bei beiden Sachtexten – »Der GeckoFuß« für Klasse 7 und »Rollschuhe« für Klasse 6 (vgl. Anhang 1.1 und 1.2) – handelte es sich um Texte, die zuvor noch nicht im Unterricht besprochen wurden und so für die Schüler*innen neu waren.

Wie bereits angesprochen wurde, zählt zu den Haupteinsatzmöglichkeiten von *FelloFish* die Verknüpfung mit Lehrwerk-Aufgaben, um so den Feedbackprozess für Schüler*innen individuell stützen zu können (vgl. Haverkamp et al., 2024, S. 66). Die Auswahl der Sachtexte bietet eben dieses Umfeld. Die Texte stammen unmittelbar aus Ressourcen, die auch im analogen Unterricht genutzt werden könnten. Durch die Verwendung dieser Ausgangstexte kann ein authentisches Arbeiten der Schüler*innen im Rahmen der Studie gewährleistet und gleichzeitig Feedbacks durch die Peers und *FelloFish* in einer Umgebung angewendet werden, die so auch im realen Klassenraum möglich wäre.

3.3.3.2 Kriterienauswahl und Aufgabenstellung

Um Schüler*innen bei einer komplexen Aufgabe wie der Sachtextzusammenfassung zu stützen, ist es notwendig, ihnen mittels eines Kriterienkatalogs eine Orientierung an die Hand zu geben. Bei der vorliegenden Studie greifen die hier aufgestellten Kriterien dabei auf drei verschiedenen Ebenen: Einmal dienen sie den Schüler*innen als Leitfaden zur Erstellung der Zusammenfassung. Zusätzlich formen sie ein objektives Kriterienraster, dem sich Schüler*innen bei der Erstellung des Peerfeedbacks bedienen können (vgl. Gamlem & Smith, 2013, S. 161). Und darüber hinaus bilden sie die Grundlage für das Feedback durch *FelloFish*.

Diese Verzahnung von (Bewertungs-)Kriterien im Arbeitsprozess der Schüler*innen mit der späteren Eingabe in ein KI-System findet auch beim Bildungsministerium NRW Berücksichtigung:

Die festgelegten Bewertungskriterien für ein Lernprodukt können mithilfe der KI für ein jederzeit verfügbares, zeitnahe formatives Feedback genutzt werden. [...] Notwendig ist [...], dass die Kriterien für die Schülerinnen und Schüler nicht nur transparent, sondern auch verständlich sind, um mit dem Ergebnis sinnvoll weiterarbeiten zu können. (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023, S. 7)

Um diesem Anspruch gerecht zu werden, wurden die (Feedback-)Kriterien im Vorhinein mit den Schüler*innen besprochen und wiederholt. Die Kriterien selbst sind dabei ebenfalls im regulären Lehr-Lernmaterial der Schüler*innen verankert. In ihrem Deutschbuch *Deutsch kompetent 7* begegnen Schüler*innen beispielsweise in einer sogenannten »Kompetenzbox« (Blatt et al., 2020, S. 21) einer Reihe von Anforderungen, die sie bei der schriftlichen Wiedergabe eines Sachtextes beachten sollen. Hierzu zählen sowohl Inhalt und Aufbau der Zusammenfassung als auch die sprachliche Gestaltung.

Diese Anforderungen wurden im Hinblick auf die Lerner*innengruppe und den Ausgangstext in einzelne Kriterien umformuliert und gemeinsam mit einem zugehörigen Arbeitsauftrag an die Schüler*innen weitergegeben. Exemplarisch ist im Folgenden der Arbeitsauftrag für die Klasse 7 aufgeführt:

(1) Aufgabe Klasse 7

Fasse den Text in eigenen Worten zusammen, indem du dich auf die wichtigsten Aussagen konzentrierst. Beachte dabei die Kriterien zum Schreiben einer Zusammenfassung:

- In der Einleitung der Zusammenfassung sollten Titel, Autorin, Textart und Thema (TATT³⁰) genannt werden.
- Im Hauptteil liegt der Fokus auf den wichtigsten Aussagen und jeder Abschnitt des Textes wird kurz zusammengefasst.
- Die Zusammenfassung sollte knapp, informativ und gut verständlich sein.
- Als Zeitform wird das Präsens (bei Vorzeitigkeit das Perfekt) verwendet.
- Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz – der Text bleibt sachlich und neutral.
- Fachbegriffe und präzise Formulierungen werden verwendet.
- Es werden geeignete Satzverknüpfungen und Überleitungen genutzt, um Zusammenhänge zu verdeutlichen (z. B. Ursachen, zeitliche Abläufe ...).

3.3.3.3 Eingaben bei FelloFish und Übertragung des Outputs

Damit *FelloFish* den Schüler*innenprodukten ein Feedback geben kann, ist es notwendig, eine Aufgabe in dem KI-Tool anzulegen. Hierfür wurde eine neue Aufgabe bei *FelloFish* ohne Vorlage erstellt und der Aufgabentext, der den Schüler*innen gegeben wurde, in das entsprechende Feld kopiert. Zudem wurden die Kriterien der Sachtextzusammenfassung in ihrer

30 Das Akronym TATT setzt sich aus den Anfangsbuchstaben der Komponenten des Einleitungssatzes der Zusammenfassung zusammen: T-Titel, A-Autor, T-Textart und T-Thema (vgl. Blatt et al., 2020, S. 21).

Originalfassung übernommen und in das Tool mit einer normalen Gewichtung übertragen (vgl. Abb. 2). Am Schluss wurde der Ausgangstext hochgeladen.

FEEDBACKKRITERIEN (BIS ZU 7, JEWELLS 10 BIS 500 ZEICHEN)	
BESCHREIBUNG	GEWICHTUNG
1. In der Einleitung der Zusammenfassung sollten Titel, Autorin, Textart und Thema (TATT) genannt werden.	<input checked="" type="radio"/> Normal ▾ 
2. Im Hauptteil liegt der Fokus auf den wichtigsten Aussagen und jeder Abschnitt des Textes wird kurz zusammengefasst.	<input checked="" type="radio"/> Normal ▾ 
3. Die Zusammenfassung sollte knapp, informativ und gut verständlich sein.	<input checked="" type="radio"/> Normal ▾ 
4. Als Zeitform wird das Präsens (bei Vorzeitigkeit das Perfekt) verwendet.	<input checked="" type="radio"/> Normal ▾ 
5. Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz – der Text bleibt sachlich und neutral.	<input checked="" type="radio"/> Normal ▾ 
6. Fachbegriffe und präzise Formulierungen werden verwendet.	<input checked="" type="radio"/> Normal ▾ 
7. Es werden geeignete Satzverknüpfungen und Überleitungen genutzt, um Zusammenhänge zu verdeutlichen (z. B. Ursachen, zeitliche Abläufe...).	<input checked="" type="radio"/> Normal ▾ 

Abb. 2: Eingabe der Kriterien für Klasse 7 (Zeitpunkt: 20.11.2024)

Nachdem die Schüler*innen ihre Zusammenfassung verfasst und einge-reicht hatten, legte die Verfasserin der Masterarbeit anhand der Schüler*innencodes Nutzer*innenprofile an, über welche dann die jeweiligen Texte hochgeladen wurden. Es wurden dabei keine Veränderungen an den Schü-le-r*innentexten vorgenommen. Auch Emojis, die ein paar der Schüler*innen genutzt hatten, wurden übertragen. Im Anschluss wurde ein Feedback durch *FelloFish* angefordert.

Da es in der vorliegenden Studie um die Rezeption des ersten Feedbacks geht, d. h., um die Rezeption des Feedbacks, welches Schüler*innen auf die erste Fassung ihres Textes erhalten, wurde auch nur das erste Feedback von *FelloFish* generiert. Um eine Vergleichbarkeit mit dem Peerfeedback zu gewährleisten und gleichzeitig auch eine visuelle Ähnlichkeit der Feedbacks herzustellen, wurde das KI-basierte Feedback in das Feedbackraster über-führt, welches auch die Schüler*innen im Peerfeedback nutzten (vgl. Abb. 3). Im Anschluss wurden die KI-Feedbackbögen ausgedruckt und gemein-sam mit den Peerfeedbackbögen an die Schüler*innen verteilt.

<p>Deine Zusammenfassung ist gut gelungen. Du hast die wichtigsten Punkte genannt und die richtige Zeiform verwendet. Achte darauf, die Textart zu erwähnen und mehr Satzverknüpfungen zu nutzen. Weiter so!</p> <p></p> <p><input type="button" value="Up"/> <input type="button" value="Down"/></p> <p>1. In der Einleitung der Zusammenfassung sollten Titel, Autorin, Textart und Thema (TATT) genannt werden. • Normal gewichtet</p> <p>Du hast den Titel, die Autorin und das Thema gut genannt. Aber du hast die Textart vergessen. Schreibe zum Beispiel: Der Sachtext 'Der Gecko Fuß' von Sigrid Belzec aus dem Jahr 2010 beschreibt die besonderen Haftfähigkeiten der Geckofüße.</p> <p><input type="button" value="Up"/> <input type="button" value="Down"/></p> <p>2. Im Hauptteil liegt der Fokus auf den wichtigsten Aussagen und jeder Abschnitt des Textes wird kurz zusammengefasst. • Normal gewichtet</p>	<p style="text-align: center;">- Feedbackbogen -</p> <p>Feedback für AFAJ1 (Code des Textes)</p> <p>Hinweis: Euer Feedback sollte einen Mitschüler*innen dabei helfen, ihren Text zu verbessern. Vermeide negativen Text, der die eigene Meinung oder Einstellung zum Text ausdrückt. Bezeichne die Stärken des Textes und nenne konkrete Vorschläge, wie der Text gestrickt hat und nun eure Unterstützung möchte. Lebt daher auch, was euch gefällt, damit die Person weiß, was schon gut läuft, bevor ihr Kritik äußert. Gebt schließlich ehrliches, höfliches Feedback, indem ihr offen und freundlich beschreibt, was euch stört oder unklar ist, und erklärt genau, warum.</p> <p>Kriterium In der Einleitung der Zusammenfassung einen Titel, Autorin, Textart und Thema (TATT) genannt werden.</p> <p>Kommentar Du hast den Titel, die Autorin und das Thema gut genannt. Aber du hast die Textart vergessen. Schreibe zum Beispiel: Der Sachtext 'Der Gecko Fuß' von Sigrid Belzec aus dem Jahr 2010 beschreibt die besonderen Haftfähigkeiten der Geckofüße.</p>
--	---

Abb. 3: Übertragung des Outputs von FelloFish in das Kriterienraster des Feedbackbogens am Beispiel von 7_AFAJ1

3.3.3.4 Erstellung und Auswertung des Fragebogens

Zur Untersuchung der Feedbackrezeption von Klasse 6 und 7 wurde ein Mixed-Methods-Design in Form eines quantitativen und qualitativen Fragebogens angewendet. Dieser baut auf dem *Feedback Perception Questionnaire* (FPQ) von Strijbos et al. (2010) auf und wurde u. a. durch offene Fragen thematisch und inhaltlich ergänzt.

Bei dem FPQ handelt es sich um einen bereits in Peer-Kontexten validierten Fragebogen, der u. a. für die multidimensionale Messung und Untersuchung von (Peer-)Feedback-Wahrnehmungen eingesetzt werden kann (vgl. Strijbos et al., 2010; Strijbos et al., 2021). Strijbos et al. (2021) führen dabei auch an, dass Lehrende den Fragebogen nutzen können, um die Reaktionen von Schüler*innen auf (Peer-)Feedback zu erfassen und die unterrichtliche Unterstützung entsprechend anzupassen (vgl. Strijbos et al., 2021: 10).

Das FPQ untersucht die Rezeption des Feedbacks dabei ganz konkret. Dies bedeutet, dass eine direkte Anbindung des zur Erhebung genutzten Feedbacks zu einer Aufgabe in einer bestimmten Situation vorliegt, statt die allgemeine Wahrnehmung von Feedback über eine Reihe von Aufgaben und Situationen zu erfassen (vgl. Strijbos et al., 2021, S. 4). Bei der Anwendung des vollständigen FPQ werden insgesamt fünf Subskalen erhoben: *Fairness*, *Nützlichkeit*, *Akzeptanz*, *Bereitschaft zur Verbesserung* und *Affektwirkung*. Gemessen wird dabei auf einer 10cm langen analogen Skala, die eine Auswahl

von 0 (stimme nicht zu) bis 10 (stimme vollständig zu) ermöglicht (vgl. Strijbos et al., 2021, S. 5, Strijbos et al., 2010, S. 295).

In Abstimmung mit den Forschungsfragen aus 3.1 wurden die Subskalen *Fairness*, *Nützlichkeit*, *Akzeptanz* und *Affektwirkung* übernommen. Die zugehörigen Items (vgl. Strijbos et al., 2010, S. 295, 297) wurden vollständig übertragen und ins Deutsche übersetzt, sodass der Fragebogen folgende Items enthielt:

- *Mit dem Feedback bin ich zufrieden, Ich betrachte dieses Feedback als fair, Ich betrachte dieses Feedback als gerechtfertigt* (*Fairness*);
- *Ich betrachte dieses Feedback als nützlich, Ich finde dieses Feedback hilfreich, Dieses Feedback unterstützt mich* (*Nützlichkeit*);
- *Dieses Feedback akzeptiere ich, Diesem Feedback widerspreche ich, Dieses Feedback lehne ich ab* (*Akzeptanz*);
- *Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich: Zufrieden, Zuversichtlich, Erfolgreich, Beleidigt, Verärgert, Frustriert* (*Affektwirkung*).

In Abgleich mit den Forschungsfragen, den Rezeptionsdimensionen und in Anlehnung an die Befunde von Gamlem und Smith (2013) wurden zwei weitere Items ergänzt: *Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann* (*Identifikation*) und *Welches Feedback würdest du bevorzugen?*³¹.

In ihrer ersten Studie wenden Strijbos et al. (2010) bei der Erfassung von Peerfeedback das FPQ im Kontext akademischen Schreibens an.³² Die in dieser Arbeit an- und durchgeführte Studie hat jedoch eine deutlich jüngere Teilnehmer*innengruppe (11–14 Jahre). Vor dem Einsatz des FPQs in der Sekundarstufe I muss beachtet werden, dass es zu Passungsproblemen kommen kann, da sich die jüngeren Schüler*innen beispielsweise in einem anderen Stadium ihrer kognitiven Entwicklung befinden oder noch

31 Dieses Item tauchte erst später im Fragebogen auf und ermöglichte – im Gegensatz zu den anderen Items – nur die Auswahl *Feedback 1* und *Feedback 2*. Da es enge Bezüge zu der finalen Auflösung hat, wird es in Anhang 2.1.2 direkt neben den Schüler*innenaussagen nach Auflösung in den qualitativen Daten gelistet.

32 Und auch bei Strijbos et al. (2021) sind die Teilnehmenden im Durchschnitt ältere Schüler*innen.

Schwierigkeiten haben, logische Entscheidungen zu treffen (vgl. Omrani et al., 2019, S. 329f.). Daher muss das FPQ über die sprachliche Ebene hinaus auch weiter an die Lerner*innen- bzw. Teilnehmer*innengruppe angepasst werden.

Folgt man Omrani et al. (2019), so sollte bei der Erstellung von Fragebögen o. ä. *Satisficing* (d. h. das oberflächliche Beantworten von Fragen, beeinflusst durch Motivation, Schwierigkeit der Aufgabe oder kognitive Fähigkeiten der Teilnehmenden) möglichst vermieden werden (vgl. Omrani et al., 2019, S. 330). Ein Problem können auch zu viele Antwortmöglichkeiten im Fragebogen darstellen. Jüngere Teilnehmende können Unterschiede in ihren Sprach- oder Lesefähigkeiten aufweisen oder sich eine Vielzahl an Auswahloptionen schwieriger merken (vgl. Omrani et al., 2019, S. 333).

Borgers et al. (2004) zeigen, dass ein höheres Maß an Auswahlmöglichkeiten für Schüler*innen mit einer größeren Variabilität oder Inkonsistenz in ihren Antworten bei zwei getrennten Befragungszeitpunkten korreliert. Eine Erhöhung der Antwortoptionen auf bis zu sechs Optionen verbessert die interne Stabilität der Antworten, also die Konsistenz der Antworten innerhalb einer Befragung. Dies deutet darauf hin, dass eine moderate Anzahl von Wahlmöglichkeiten den Befragten hilft, ihre Präferenzen oder Einstellungen klarer und differenzierter auszudrücken. Werden zu viele Optionen angeboten (sieben oder mehr), nimmt die Zuverlässigkeit der Antworten ab. Gegenteiliges verzeichnen die Autor*innen bei Erhebungen über einen gewissen Zeitraum: Hier könnte eine Erhöhung der Antwortoptionen auf bis zu sieben zu einer Steigerung der absoluten Differenz führen (vgl. Borgers et al., 2004, S. 30): »This is logical, since offering more response options provides room for larger absolute differences.« (Borgers et al., 2004, S. 30). Borgers et al. (2004, S. 30) empfehlen für Studien mit Kindern schließlich um die vier Antwortmöglichkeiten.

Um die Ausprägungen der Skala des FPQ in ihren Grundzügen zu behalten, wurden die zuvor elf verschiedenen Antwortmöglichkeiten auf fünf gekürzt. Zusätzlich wurde jede Antwortoption beschriftet (»ALL-Form«, Omrani et al., 2019, S. 333f.), um das Verständnis der Antworten zu verbessern. Hierdurch können Antwortverzerrungen verringert und die Zuverlässigkeit der Antworten erhöht werden (vgl. Omrani et al., 2019, S. 333f.). Folgende

Antwortoptionen wurden den Schüler*innen auf der 5-stufigen Likert-Skala gegeben: *Stimme nicht zu* (1), *Stimme eher nicht zu* (2), *Neutral* (3), *Stimme eher zu* (4) und *Stimme vollständig zu* (5).

Um ein komplexes Bild der Feedbackrezeption der Schüler*innen zu erhalten, wurde das FPQ thematisch und inhaltlich durch offene Fragen ergänzt. Hier wurden die Teilnehmenden detaillierter bzgl. ihrer Nützlichkeitswahrnehmung befragt, aber auch einzelne Komponenten des Feedbacks konnten hier von den Schüler*innen angesprochen werden (»Welche Teile des Feedbacks empfindest du als besonders wichtig für die Überarbeitung deines Textes, und warum?«). Weitere Schwerpunkte lagen zudem auf wahrgenommenen Unterschieden und Gemeinsamkeiten der beiden Feedbacks (»Welche Unterschiede oder Gemeinsamkeiten hast du in Feedback 1 und Feedback 2 festgestellt?«) und der finalen Auflösung (»Schau dir deine Antworten auf den anderen Seiten an. Ändert sich etwas an deiner Wahrnehmung des Feedbacks? Warum (nicht)?«).³³

Zur Auswertung der qualitativen Daten wird die inhaltlich strukturierende qualitative Inhaltsanalyse nach Kuckartz (hier: Kuckartz & Rädiker, 2022; vgl. auch Kirschbaum & Rothärmel, 2017) unter Einsatz der Software *MAXQDA* angewendet. Die Hauptkategorien wurden deduktiv auf Basis der Fragestellungen entwickelt und umfassen je Feedback dessen wahrgenommene Nützlichkeit, Zustimmung (Fairness & Akzeptanz) und Affektwirkung, Feedbackkomponenten, die bei den Schüler*innen besondere Beachtung finden, und darüber hinaus die Wahrnehmung der Feedbacks nach Kenntnis der Quelle.

Nach Codierung wurden in einem nächsten Schritt induktiv Subkategorien gebildet. Für die Hauptkategorien *Nützlichkeit*, *Zustimmung* und *Affektwirkung* wurde dafür zwischen negativer und positiver Einschätzung durch die Schüler*innen differenziert. Bei den Feedbackkomponenten wurden je Feedbackart Subkategorien gebildet, die u. a. die Passung des Feedbacks zum Ausgangstext, die Nennung von Beispielen, Verbesserungsvorschlägen oder Tipps und Verweise auf die Knappheit bzw. die Detailliertheit des Feedbacks umfassten. Für die Hauptkategorie *Wahrnehmung der Feedbacks*

33 Für einen detaillierteren Einblick in die Fragebögen vgl. Anhang 1.4 und 1.5.

nach Kenntnis der Quelle wurden Subkategorien gebildet, die eine bleibende oder sich verändernde Wahrnehmung kennzeichnen. Und in einem nächsten Schritt wurden – erneut induktiv – Rechtfertigungsgründe in den Schüler*innentexten ausgemacht.³⁴

3.4 Ergebnisse

Im Folgenden werden die Befunde der Studie für sowohl die sechste als auch die siebte Klasse dargelegt. Die Darstellung erfolgt in Anlehnung an die Zeitpunkte der Durchführung der Studie. Daher werden zuerst die Resultate der Klasse 7 abgebildet und im Anschluss die der Klasse 6.

3.4.1 Klasse 7

Nachfolgend werden zunächst die quantitativen und qualitativen Resultate der Peerfeedback-Intervention in Klasse 7 beschrieben. Im Anschluss werden die Resultate des KI-basierten Feedbacks durch *FelloFish* thematisiert, bevor schließlich sowohl die Ergebnisse miteinander verglichen als auch die Schüler*innantworten nach der Auflösung behandelt werden.

3.4.1.1 Peerfeedback

In Klasse 7 geben siebzehn der Schüler*innen an, das erhaltene Peerfeedback zu präferieren. Das macht 65,38% der Schüler*innantworten in dieser Klasse aus. Im Bereich Fairness liegen die quantitativen Schüler*innantworten bei einem Mittelwert (M) von 3,84–4,08; die Schüler*innen stimmen dem Feedback hinsichtlich der Fairness also eher zu: Wie aus Abb. 4 hervorgeht, stimmen dem Item *Mit dem Feedback bin ich zufrieden* dabei achtzehn Schüler*innen eher oder vollständig zu, fünf sind neutral eingestellt und zwei geben an, ihm nicht zuzustimmen. Ein ähnliches Bild ergibt sich bei den Items *Ich betrachte das Feedback als fair* und *Ich betrachte das Feedback als gerechtfertigt*. Bei ersterem geben neunzehn Schüler*innen an, dem Feedback eher oder vollständig zuzustimmen, drei Antworten liegen im neutralen Bereich und zwei stimmen (eher) nicht zu. Letzterem Item stimmen siebzehn Schüler*innen eher oder vollständig zu, fünf sind neutral

34 Eine vollständige Darlegung des Codier-Leitfadens ist in Anhang 3 zu finden.

und drei geben an, eher oder nicht zuzustimmen. Insgesamt zeigt sich somit, dass die Schüler*innen der Klasse 7 mit dem Peerfeedback eher zufrieden sind und es als fair und gerechtfertigt empfinden.

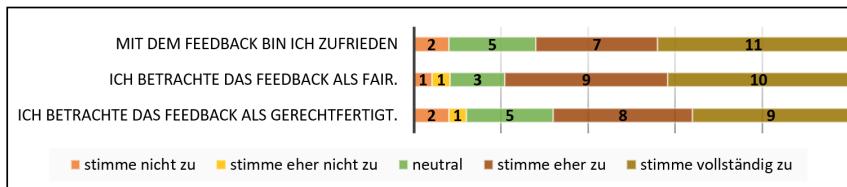


Abb. 4: Wahrnehmung der Fairness im Peerfeedback der Klasse 7

Auch im Rahmen der Nützlichkeit stimmen die Schüler*innen dem Feedback eher zu ($M = 3,54-3,7$): Die Schüler*innen nehmen es als nützlich und hilfreich wahr und fühlen sich durch das Feedback unterstützt. So geben im Hinblick auf das Item *Ich betrachte das Feedback als nützlich* fünfzehn der Teilnehmenden an, eher oder vollständig zuzustimmen. Neben vier neutralen Antworten, stimmen sechs Schüler*innen nicht oder eher nicht zu. Fünfzehn der Schüler*innen nehmen das Peerfeedback als *hilfreich* wahr. Vier stimmen dem nicht oder eher nicht zu und weitere sechs ordnen sich als neutral ein. Dreizehn Schüler*innen geben an, dass sie durch das Feedback eher oder vollständig *unterstützt* werden. Fünf Schüler*innen sind neutral und sechs stimmen nicht oder eher nicht zu.

Ähnlich positive Ergebnisse zeigen sich auch im Bereich Akzeptanz. Hier stimmen neunzehn Schüler*innen dem Item *Dieses Feedback akzeptiere ich* eher oder vollständig zu. Drei stimmen ihm (eher) nicht zu und zwei sind neutral. Zwölf der Schüler*innen *widersprechen* dem Feedback (eher) nicht, stimmen dem zugehörigen Item also nicht oder eher nicht zu, und weitere neunzehn Schüler*innen *lehnen* das Feedback (eher) nicht ab. Auf der anderen Seite widersprechen acht der Teilnehmenden dem Feedback (vier ordnen sich als neutral ein) und fünf lehnen es ab (hier ordnet sich eine*r Schüler*in als neutral ein). Insgesamt zeigen die Schüler*innen somit eine tendenziell höhere Akzeptanz gegenüber dem Feedback und neigen weniger dazu, es abzulehnen oder ihm zu widersprechen.

Auch im Bereich Identifikation der Fehlerstellen und Verbesserungsvorschläge erkennen die Schüler*innen Potenziale im Peerfeedback. Zwölf Schüler*innen geben hier an, durch das Feedback erkennen zu können, wo

und wie Verbesserungen möglich sind. Acht der Befragten stimmen dieser Aussage nicht zu. Vier zeigen sich unentschieden und bleiben neutral.

Im Rahmen der Affektwirkung evoziert das Peerfeedback vermehrt positive Emotionen (vgl. Abb. 5). So geben zwanzig der Schüler*innen an, dem Item *Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich zufrieden* eher oder vollständig zuzustimmen. Drei Schüler*innen stimmen ihm nicht zu und zwei sind neutral.

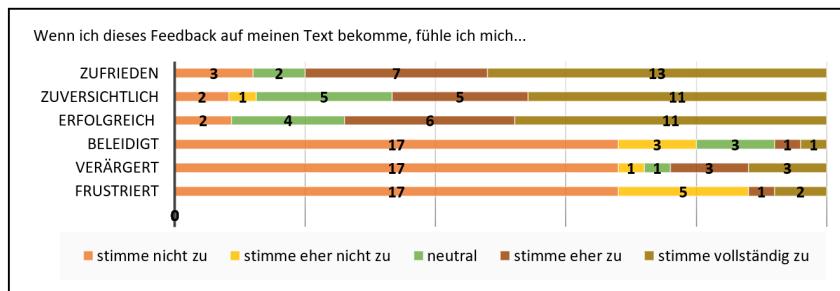


Abb. 5: Affektwirkung im Peerfeedback der Klasse 7

Jeweils sechzehn und siebzehn Schüler*innen fühlen sich durch das Peerfeedback zuversichtlich und erfolgreich. Hier geben drei bzw. zwei der Teilnehmenden an, den Items nicht oder eher nicht zuzustimmen. Ein Großteil der Schüler*innen fühlt sich durch das Peerfeedback weder beleidigt (20) noch verärgert (18) oder frustriert (22). Ein kleiner Teil stimmt diesen Emotionen eher oder vollständig zu (beleidigt: zwei, verärgert: sechs, frustriert: drei).

Im quantitativen Teil der Erhebung zeigt sich zusammenfassend eine überwiegend positive Rezeption des Peerfeedbacks durch die Schüler*innen. Dies betrifft sowohl die Dimensionen Akzeptanz, Nützlichkeit, Affektwirkung und Fairness als auch die angebotenen Hilfestellungen im Rahmen des Feedbacks (Identifikation).

Bestärkung finden die positiven Befunde der Rezeptionsdimensionen auch im qualitativen Teil der Erhebung. Hier erwähnen insgesamt neunzehn Schüler*innen die Nützlichkeit des Feedbacks. Vierzehn dieser Schüler*innen (53,8 % der gesamten Teilnehmenden) empfinden das Feedback dabei als nützlich bzw. hilfreich. So schreibt eine*r der Schüler*innen als Reaktion

auf das Peerfeedback: »War hilfreich, weil dort stand was gut war und was man besser machen kann« (7_AFAJ1). Ein*e andere*r Schüler*in schreibt: »Feedback 2 [Peerfeedback] fande ich am hilfreichsten, weil es auch Punkte waren, die ich verstehe. Mir hat geholfen, dass ich Absätze machen soll« (7_ELNA1).

Es gibt jedoch auch Gegenstimmen. Fünf Schüler*innen (19,2 % der gesamten Teilnehmenden) bezeichnen die Nützlichkeit als negativ. 7_AFEM0 schreibt beispielsweise

Allerdings konnte mir dieses Feedback nicht wirklich weiterhelfen, weil bei mehreren Teilen des Feedbacks nichts geschrieben wurde. Deshalb weiß ich nicht was ich aus diesen Teilen rausinterpretieren kann.

und 7_ILRJ1 notiert:

Ich finde dieses Feedback nicht besonders hilfreich, da ich in der Einleitung zwar etwas verbessern kann aber ich kann mir nicht erschließen wie ich es verbessern kann. Außerdem fehlt mehr als die Hälfte und mir hilft es nicht weiter wenn ich die Zahlen nicht ausschreiben darf/soll. Ich weiß aber, dass ich mehr Fachbegriffe verwenden sollte.

Während die Aussagen von 7_AFAJ1 und 7_ELNA1 noch recht vage sind (7_ELNA1 verweist an einer Stelle auf den Vermerk, dass er*sie Absätze im Text machen solle), verweisen 7_AFEM0 und 7_ILRJ1 gleichzeitig auf konkrete Fehler- oder Problemstellen im Peerfeedback: Das Feedback sei zu kurz oder unvollständig.

Vierzehn Schüler*innen äußern sich in ihren Antworten zur Akzeptanz oder Fairness des Feedbacks. Elf der Teilnehmenden (42,3 % der gesamten Teilnehmenden) stimmen dem Peerfeedback zu. Hier wird das Feedback als »sehr fair« (7_AFSO1) oder »gerechtfertigt« (7_ELNA1) verstanden und als »gutes feedback« (7_IFSF1) bzw. als – im Vergleich zum KI-basierten Feedback – »besser« (7_LFAM1) gesehen. Drei Schüler*innen äußern einen Widerspruch. So spricht beispielsweise 7_AFGJ1 von Unstimmigkeiten bei der Rezeption des Feedbacks: »Ein paar Sachen stimmen nicht so ganz, wie z. B. das ich Willkommen schreiben soll oder das ich mehr Fröhlichkeit und Gedanken rein bringen soll«.

Vier der Schüler*innen zeigen im qualitativen Teil der Erhebung eine Affektwirkung als Reaktion auf das Peerfeedback. Drei dieser Reaktionen fallen positiv aus: »Bin super zufrieden!« (7_IFZD2), »Es freut mich das auch positive Sachen geschrieben wurden« (7_LFAM1) und »Danke, für die tolle und ehrliche Meinung. :)« (7_OFNA0). Nur eine Schüler*innenstimme ist negativ: »Ich bin nicht zufrieden« (7_IFGM0). Somit zeigt sich auch hier ein überwiegend positiver emotionaler Effekt.

Die Komponenten des Peerfeedbacks, die bei den Schüler*innen der Klasse 7 besondere Beachtung finden, sind Lob bzw. Nennung von Positivem, die Kombination von Lob und Kritik, die positive und/oder sachliche Formulierung des Feedbacks, die Passung zur jeweiligen Zusammenfassung, die Nennung von Grammatik- und Rechtschreibfehlern im Text, das Aufzeigen von Verbesserungsvorschlägen und Tipps, sowie der Verweis, dass das Peerfeedback als weniger streng aufgefasst wird und kurz und prägnant sei.

Auf der anderen Seite wird auch auf ein zu kurzes Feedback verwiesen, in dem Teile fehlen. Schüler*innen schreiben, dass eine Überarbeitung mit dem erhaltenen Feedback nicht möglich sei, das Feedback Fehler enthalte oder nur auf der Textoberfläche greife. Die Anzahl der Schüler*innen, die auf die jeweiligen Feedbackteile verweisen, ist dabei in Abbildung 6 dargestellt.

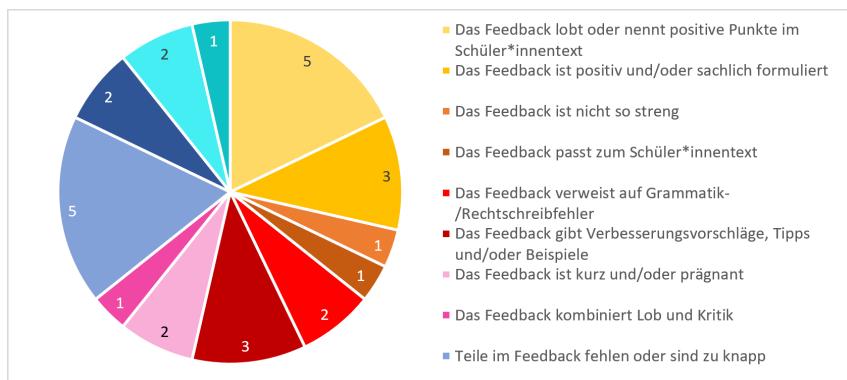


Abb. 6: Verteilung Komponenten des Peerfeedbacks in der Wahrnehmung der Klasse 7

Ein großer Teil der Schüler*innen kritisiert die Komplexität des Peerfeedbacks. Fünf Schüler*innen nehmen das Feedback als zu knapp oder lückenhaft wahr – darunter 7_IFGM0 und 7_AFEM0, die das Peerfeedback als »sehr kurz« (7_IFGM0) beschreiben und aufzeigen, dass hier »manche Teile nicht vorhanden und wenn eher knapp beschrieben [sind] im Vergleich zu Feedback 1 [KI-Feedback]« (7_AFEM0). Zwei weitere Schüler*innen schreiben, dass eine Überarbeitung mit dem Feedback nicht möglich sei, z. B. »Ich kann leider nichts überarbeiten.« (7_EFIS1). Ein Teil der Schülerinnen erfährt folglich durch das Peerfeedback eher eine eingeschränkte Unterstützung. Dies steht im Widerspruch zu den Ergebnissen des quantitativen Teils, in dem ein Großteil der Schüler*innen angibt, Potenziale im Bereich der Fehleridentifikation und Verbesserungsvorschläge zu erkennen.

Ein*e Schüler*in verweist darüber hinaus auf die oberflächliche Natur des Peerfeedbacks:

Bei Fb 1 [Peerfeedback] ist das ›obere‹, ›äussere‹ Feedback und bei Fb 2 [KI-Feedback] ist es detaillierter. Ich sehe, dass bei Fb 1 das wichtige steht und bei Fb 2 das etwas detaillierter steht. Also im Prinzip ist beides wichtig, aber bei Fb 2 sind mehr Infos und mehr Meinung erhalten. (OFNA0)

Während OFNA0 das Feedback der Peers zwar als »obere[s]« oder »äussere[s]« Feedback bezeichnet, ordnet er*sie ihm gleichzeitig eine gewisse Relevanz zu: Laut dem*der Schüler*in steht »bei Fb 1 das wichtige«. Laut OFNA0 greift das Feedback der Peers eher auf der Textoberfläche, was es für ihn*sie aber ebenso wichtig macht, wie das Feedback durch die KI, welches er*sie als detaillierter beschreibt (vgl. für das KI-Feedback Kapitel 3.4.1.2).

Bei zwei weiteren Schüler*innen werden in den erhaltenen Feedbacks mit Grammatik- und Rechtschreibfehlern Rückmeldungen auf der Textoberfläche positiv hervorgehoben. So schreibt beispielweise 7_IFHM4: »Bei Feedback Nummer 1 [Peerfeedback] waren z. B. dass ich Rechtschreibfehler gemacht habe oder grammatischen Fehler die ich gemacht habe.«

Daneben dominieren in der Schüler*innenrezeption die im Feedback gegebenen Tipps, eine positive und/oder sachliche Formulierung des Feedbacks – »Feedback 2 [Peerfeedback] war super nett« (7_ALNF2) –, sowie Lob

durch die Mitschüler*innen: »Ich finde es sehr gut weil ich oft gelobt werde« (7_AFSM0). Letzteres nimmt – genauso wie die Verweise auf ein lückenhaftes Feedback – ca. 18% der Schüler*innenäußerungen ein.

3.4.1.2 Feedback durch FelloFish

Sieben der Schüler*innen der Klasse 7 präferieren das Feedback durch *FelloFish*. Das macht 26,92 % der Schüler*innenantworten in dieser Klasse aus. Hinsichtlich der wahrgenommenen Fairness stehen die Schüler*innen dem KI-Feedback eher neutral bis positiv gegenüber ($M = 2,95\text{--}3,91$). In Abbildung 7 sind die Verteilungen der Items *Mit dem Feedback bin ich zufrieden*, *Ich betrachte das Feedback als fair* und *Ich betrachte das Feedback als gerechtfertigt* dargestellt.

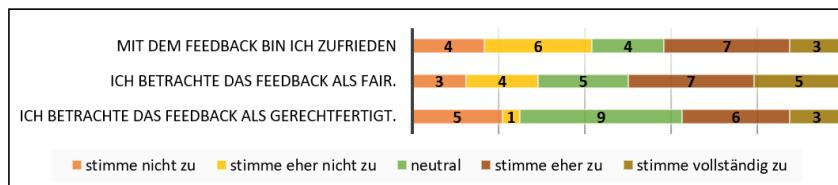


Abb. 7: Wahrnehmung der Fairness im KI-Feedback der Klasse 7

Zehn Schüler*innen geben an, dem Item *Mit dem Feedback bin ich zufrieden* nicht oder eher nicht zuzustimmen. Neben vier neutralen Einstellungen, stimmen zehn Schüler*innen der Aussage eher oder vollständig zu. Sieben Schüler*innen betrachten das Feedback als (eher) nicht fair, fünf sind neutral und zwölf stimmen dem Item *Ich betrachte das Feedback als fair* eher oder vollständig zu. Neun Schüler*innen geben an, dass sie das Feedback (eher) als *gerechtfertigt* empfinden, neun sind neutral und sechs nehmen eine Gegenposition ein. Insgesamt zeigen sich im Bereich der Zufriedenheit eher neutrale Rückmeldungen. Im Hinblick auf Fairness und das Empfinden des Feedbacks als gerechtfertigt überwiegen zwar die zustimmenden Stimmen gegenüber den ablehnenden. Die Unterschiede sind jedoch nur gering.

Negativer fallen die Befunde im Hinblick auf die Nützlichkeit aus. Hier sind die Schüler*innenstimmen eher neutral bis hin zur leichten Ablehnung zu verorten ($M = 2,63\text{--}3,16$). Sowohl dem Item *Dieses Feedback unterstützt mich* als auch dem Item *Ich betrachte das Feedback als nützlich* stimmen dabei

neun Schüler*innen nicht oder eher nicht zu. Auf der anderen Seite geben nur vier Schüler*innen an, dass das Feedback sie unterstützt (elf sind neutral) und zehn Schüler*innen nehmen das KI-basierte Feedback als (eher) nützlich war (sechs sind neutral) und. Dem Item *Ich finde dieses Feedback hilfreich* stimmen elf Schüler*innen eher oder vollständig zu. Acht stimmen dem Item (eher) nicht zu und sechs sind neutral.

Ähnlich fallen die Ergebnisse im Rahmen der Akzeptanz aus. Dreizehn Schüler*innen *akzeptieren* das Feedback (eher), sieben tun dies nicht und fünf ordnen sich als neutral ein. Acht der Teilnehmenden *widersprechen* dem Feedback. Zwölf stimmen dem zugehörigen Item (eher) nicht zu – widersprechen dem Feedback also nicht – und vier Schüler*innen sind neutral. Dem Item *Dieses Feedback lehne ich ab* stimmen sieben der Schüler*innen zu, während sich fünf neutral einordnen und dreizehn nicht oder eher nicht zustimmen. Das Feedback wird somit von einem großen Teil der Schüler*innen akzeptiert, es gibt jedoch auch vermehrt Stimmen, die das Feedback ablehnen und ihm widersprechen.

Auch im Bereich Identifikation der Fehlerstellen und Verbesserungsvorschläge zeigen sich zu Teilen negative Ausprägungen: Dem Item *Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann* stimmen elf der Schüler*innen eher oder vollständig zu. Während fünf neutral eingestellt sind, stimmen neun dem Item (eher) nicht zu – Geben also an, auf Basis des KI-Feedbacks nicht zu wissen, wo und wie sie in ihrem Textprodukt Verbesserungen vornehmen können.

Eine Übersicht der Affektwirkungen ist in Abbildung 8 dargestellt. Hierbei zeigen sich eher gemischte Emotionen als Reaktion auf das Feedback durch *FelloFish*.

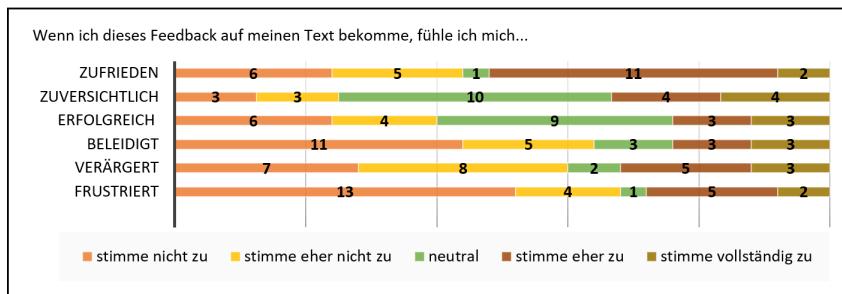


Abb. 8: Affektwirkung im KI-Feedback der Klasse 7

Dreizehn der Schüler*innen geben an, dass sie dem Item *Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich zufrieden* eher oder vollständig zustimmen. Auf der anderen Seite stimmen elf der Schüler*innen dem Item (eher) nicht zu. Während acht Schüler*innen dem Item *zuversichtlich* eher oder vollständig zustimmen, sind zehn Schüler*innen neutral und sechs stimmen ihm nicht oder eher nicht zu. Ein Großteil der Schüler*innen gibt an, den Items *beleidigt* (sechzehn), *verärgert* (fünfzehn) und *frustriert* (siebzehn) nicht oder eher nicht zuzustimmen. Daneben fühlen sich sechs Schüler*innen beim Rezipieren des KI-basierten Feedbacks beleidigt, acht verärgert und sieben frustriert – stimmen den jeweiligen Items also eher oder vollständig zu.

Für einen großen Teil der Schüler*innen geht die Rezeption des Feedbacks mit positiven Gefühlen einher. Dennoch treten auch deutliche negative Ausprägungen bei bestimmten Items auf: Mehr als ein Drittel der Befragten fühlen sich mit dem *FelloFish*-Feedback nicht zufrieden oder erfolgreich. Über 23 % geben außerdem negative Gefühle an wie Beleidigtsein, Verärgerung oder Frustration.

Die Befunde der quantitativen Befragung finden auch hier zu großen Teilen Bestätigung im qualitativen Teil des Fragebogens. Hier tägten sechzehn Schüler*innen Äußerungen hinsichtlich der Nützlichkeit des Feedbacks. Darunter finden sich jeweilige Extreme, die die Nützlichkeit des Feedbacks als klar negativ oder klar positiv einstufen, z. B. »Bei Feedback 1 [KI-Feedback] hat mir kein bisschen geholfen!« (7_ALNF2) und »Das Feedback ist gut und hilfsreich« (7_IFSF1). Die rein positive Wahrnehmung des Feedbacks ist häufig an Feedbackkomponenten wie Verbesserungsvorschläge, Tipps oder

eine generelle Spezifität bzw. Detailliertheit geknüpft. So schreibt beispielsweise 7_EFTS2: »Ich fande das Feedback 1 [KI-Feedback] als hilfreicher denn dort wurde mehr Verbesserungsmaterial gefunden.« Klar negative Äußerungen hinsichtlich der Nützlichkeit werden auf der anderen Seite an Fehlerstellen im Feedback gebunden: »Ich finde das Feedback nicht sehr hilfreich, weil dort Sachen genannt werden die ich im Text habe und im Feedback steht das ich es nicht habe« (7_ILNJ1).

Neben diesen klar negativen bzw. klar positiven Fällen, gibt es zusätzlich zu uneindeutigen Aussagen – z. B. »Ich finde es ist hilfreich und nicht hilfreich« (7_DFIS2) – auch Schüler*innenäußerungen, die eine differenziertere Sichtweise erkennen lassen. So identifiziert beispielsweise 7_AFGD1 die Fehlerstellen im KI-basierten Feedback, klammert diese jedoch im Hinblick auf die Gesamtbewertung aus, bzw. lässt die Fehlerpunkte nicht seine*ihrer Einstellung hinsichtlich der Nützlichkeit beeinflussen:

Ich finde, dass der der das Feedback geschrieben hat manche Punkte angesprochen die ich richtig gemacht habe. (z. B. dass ich die Haftlamellen beschreiben soll, obwohl ich es gemacht hatte.) Das Feedback ist aber bis auf da alles hilfreich. (7_AFGD1)

Auch 7_ILRJ1 findet Fehlerstellen in dem erhaltenen Feedback:

Dieses Feedback hilft mir nicht weiter, weil vieles von dem Kriterium nicht stimmt und ich daraus nichts lernen oder verbessern kann. Dennoch kann ich aus dem Feedback schließen das mein Text nicht schlecht ist, da das was im Feedback steht nicht richtig ist sondern in meinem Text vorhanden ist. (7_ILRJ1)

Bei 7_ILRJ1 löst der Fund der Fehlerstellen eine eigenaktive Auseinandersetzung mit dem Feedbacktext aus: D. h., der*die Schüler*in erfasst das gegebene Feedback und erkennt, dass es fehlerhaft ist, da die kritisierten, vermeintlich fehlenden Punkte in seinem*ihrer Text enthalten sind. Daraufhin erkennt er*sie, dass die eigene Lösung richtig ist: »ich [kann] aus dem Feedback schließen das mein Text nicht schlecht ist« (7_ILRJ1).

Fünfzehn Schüler*innen äußern sich in ihren Antworten auf die offenen Fragen zu der Fairness und Akzeptanz bzw. Zustimmung des Feedbacks. Während vier Schüler*innen dem KI-generierten Feedback zustimmen und es u.

a. als »gut« (z. B. 7_AFEM0 oder 7_IFSF1) beschreiben, akzeptieren acht das Feedback nicht bzw. stimmen ihm nicht zu. Auch hier ist eine starke Verknüpfung zwischen Ablehnung und Fehlerstellen im Feedback zu verzeichnen. 7_ELNA1 schreibt: »Ich bin nicht so richtig zufrieden, da ich viele Punkte nicht richtig verstehe. Ich finde das auch nicht richtig fair, weil viel dort anders steht« (7_ELNA1) und auch 7_AFGJ1 notiert:

Ich finde den Text eher nicht so verständlich. Bei dem Tatt habe ich die Textart in meinem Text beschrieben. Und das Thema wird ja eigentlich auch durch den Einleitungssatz klar. Den zweiten Teil versteh ich. Bei dem dritten habe ich das eigentlich schon so ausgedrückt. Das sechste versteh ich wieder gar nicht da ich das eigentlich so geschrieben habe. (7_AFGJ1)

Drei Schüler*innenäußerungen zeigen eine Affektwirkung beim Rezipieren des Feedbacks. Eine davon ist positiv: »Ich bin auch sehr zufrieden« (7_EFIS1). Die anderen beiden sind leicht bis stärker negativ. So äußert 7_ALNF2, dass er*sie mit dem Feedback »nicht sehr frustriert« sei, während 7_IFGM0 die positiven Formulierungen im Feedback kritisiert:

Feedback 2 [KI-Feedback]. Man sollte nicht so unnötige Sachen schreiben, wie ›das war super. Dann fühlt man sich noch blöder als eigentlich schon, weil dann das Schlecht kommt. Wenn man direkt schlecht macht, fühlt man sich vielleicht sogar besser.

7_IFGM0 nimmt hier Passungsprobleme zwischen Feedbackformulierung und -inhalt wahr: Laut ihm*ihr würde ein Feedback, dass zwar positiv beginnt (»das war super«), dann aber starke Kritik am Schüler*innentext äußert, dazu führen, dass sich die Schüler*innen schlecht fühlen. Stattdessen, so 7_IFGM0, wäre es sinnvoller, das Feedback direkt mit dem Negativen zu beginnen, damit die Schüler*innen unmittelbar auf ihre Fehlerstellen aufmerksam gemacht werden und entsprechend reagieren können. Dieses konkretere und weniger beschönigende Vorgehen könne dann dazu führen, dass »man sich vielleicht sogar besser« fühle.

Neben der positiven Formulierung des Feedbacks und dem Lob bzw. der Nennung von Positivem gehören zu den Komponenten des KI-Feedbacks, die bei den Schüler*innen der Klasse 7 besondere Beachtung finden, auch das Geben von Kritik bzw. das Aufzeigen von Fehlern im Schüler*innenprodukt, die Nennung von Verbesserungsvorschlägen oder Tipps und die ausführlichen bzw. detaillierten Ausführungen im Feedback. Hinzu treten der konkrete Bezug des Feedbacks auf einzelne Komponenten des Feedbackbogens, sowie die Verbindung des KI-Feedbacks mit Expert*innenfeedback (hier: dem der Lehrkraft). Andererseits finden jedoch auch Fehler im Feedback bei den Schüler*innen Beachtung.

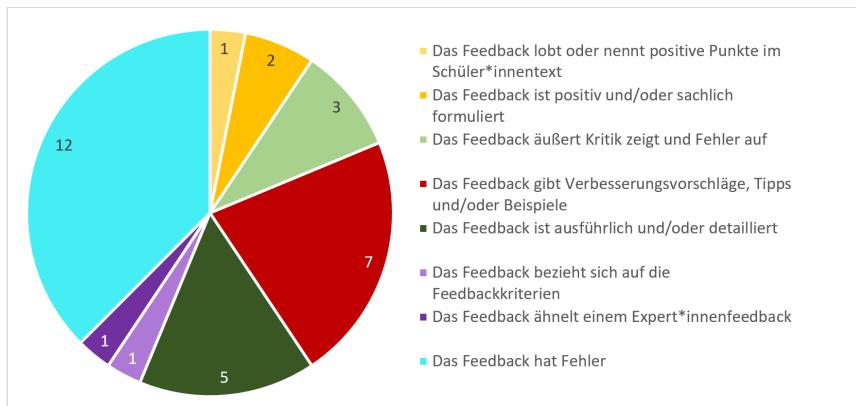


Abb. 9: Verteilung der Komponenten des KI-Feedbacks in der Wahrnehmung der Klasse 7

Wie aus Abbildung 9 ersichtlich wird, nehmen Äußerungen zu Fehlern im Feedback einen großen Teil in den Schüler*innenantworten ein. Insgesamt zwölf Schüler*innen verweisen auf Fehler, dies häufig sogar mehrmals. So schreibt 7_ILNJ1, wie oben bereits angesprochen: »Ich finde das Feedback nicht sehr hilfreich, weil dort Sachen genannt werden die ich im Text habe und im Feedback steht das ich es nicht habe.« Und wiederholt später: »Die KI sagt falsche Sachen. Dadurch finde ich Feedback zwei [KI-Feedback] nicht hilfreich.«

Häufig identifizieren Schüler*innen Fehler im Feedback bei ihrem Einleitungssatz. 7_AFGJ1 schreibt: »Bei dem Tatt habe ich die Textart in meinem Text beschrieben. Und das Thema wird ja eigentlich auch durch den Einleitungssatz klar.« Und 7_UJHJ1 notiert: »In meinem Feedback steht ich könnte

›In dem Sachtext mit dem Titel schreiben aber ich habe es geschrieben.‹ Für 7_IFRA2 bedeuten die Fehler im Feedback, dass sich nicht richtig mit dem Schüler*innenprodukt auseinandergesetzt wurde: »Ich glaube, dass der Text gar nicht richtig gelesen wurde«. Dieser Gedanke lässt sich unterschwellig auch in anderen Schüler*innenäußerungen wiederfinden, beispielsweise bei 7_IFZD2: »Feedback 2 [KI-Feedback] hat überhaupt nicht zu meinem Text gepasst«.

Über die Fehlerstellen hinaus heben sieben Schüler*innen die Verbesserungsvorschläge und Tipps des KI-Feedbacks hervor – »In dem 2. Feedback [KI-Feedback] waren viele Verbesserungstipps« (7_IFSF) – und fünf Schüler*innen erwähnen die Präzision und Ausführlichkeit des Feedbacks – »Bei Feedback 1 [KI-Feedback] wurde sehr ausführlich beschrieben und alles war vorhanden« (7_AFEM0). Zusätzlich vermerken drei Schüler*innen, dass in dem Feedback Kritik geübt wird, bzw. die Fehlerstellen im Schüler*innenprodukt aufgezeigt werden. So schreibt 7_AFSO1: »Feedback 1 [KI-Feedback] wurde viel mehr kritisiert als in Feedback 2 [Peerfeedback].«.

Neben Lob im Feedback (ein*e) und Verweisen auf die positive und/oder sachliche Formulierung des Feedbacks (zwei), findet sich auch ein*e Schüler*in, der*die die Lehrkraft als Sender*in des Feedbacks sieht. So schreibt 7_IFGM0 in einem Zusatz: »(Aber da sie eine (fast) Lehrerin sind, ist das ja klar)« und notiert später »man hat gesehen, dass es eine Erwachsene war.«. Das Feedback durch die KI wird hier als Expert*innenfeedback verstanden.

3.4.1.3 Vergleich der Ergebnisse

Zur Überprüfung der Mittelwertsdifferenzen der quantitativen Erhebung zwischen den Feedbackquellen wurden zunächst parametrische Analysen mittels gepaarter t-Tests durchgeführt.³⁵ Dabei wurde ein Ergebnis als signifikant gewertet, wenn der p-Wert unter 0,05 lag (entsprechend einem Signifikanzniveau von 5%).

Im Rahmen der Fairness kann dabei für das Item *Ich betrachte das Feedback als fair* zwischen den Mittelwerten von KI-basiertem Feedback ($M_{KI} = 2,958$; $SD = 1,334$) und Peerfeedback ($M_{PEER} = 4$; $SD = 1,19$) eine signifikante

35 Für eine genaue Auflistung vgl. Anhang 2.1.1 und 2.2.1.

Differenz ($p = 0,027$) festgestellt werden. Dies bedeutet, dass die Siebtklässler*innen die Fairness im KI-basierten Feedback signifikant schlechter einschätzen. Auch hinsichtlich der Affektwirkung können signifikante Unterschiede bei den Items *erfolgreich* und *verärgert* nachgewiesen werden. Für das Item *erfolgreich* ergeben sich dabei Mittelwerte von $M_{KI} = 2,72$ ($SD = 1,308$) und $M_{PEER} = 4,043$ ($SD = 1,224$) mit $p < 0,01$. Das Item *verärgert* weist mit den Mittelwerten $M_{KI} = 2,56$ ($SD = 1,417$) und $M_{PEER} = 1,96$ ($SD = 1,541$) einen p -Wert von 0,036 auf. Bei der Rezeption des Peerfeedbacks fühlen sich die Schüler*innen im Vergleich zum KI-Feedback somit erfolgreicher, während das KI-Feedback bei den Siebtklässler*innen vor allem Verärgerung auslöst.

Da die Annahme der Normalverteilung bei einigen Items nicht erfüllt war, wurden ergänzend nichtparametrische Wilcoxon-Tests angewendet. Im Bereich Fairness kann so für das Item *Mit dem Feedback bin ich zufrieden* zwischen den Medianen (Md) von KI-basiertem Feedback ($Md_{KI} = 3$) und Peerfeedback ($Md_{PEER} = 4$) eine signifikante Differenz ($p = 0,01$) festgestellt werden. Bei dem Item *Dieses Feedback unterstützt mich* (Nützlichkeit) zeigt sich ein signifikanter Unterschied ($p = 0,033$) zwischen KI-basiertem Feedback ($Md_{KI} = 3$) und Peerfeedback ($Md_{PEER} = 4$). Folglich sind die Schüler*innen der Klasse 7 mit dem Peerfeedback zufriedener und fühlen sich durch dieses stärker unterstützt.

Im Rahmen der Affektwirkung können signifikante Unterschiede bei den Items *zufrieden*, *zuversichtlich* und *beleidigt* nachgewiesen werden. Für das Item *zufrieden* ergeben sich dabei Mediane von $Md_{KI} = 4$ und $Md_{PEER} = 5$ mit $p = 0,002$. Das Item *zuversichtlich* weist mit den Medianen $Md_{KI} = 3$ und $Md_{PEER} = 4$ einen p -Wert von 0,013 und das Item *beleidigt* mit $Md_{KI} = 2$ und $Md_{PEER} = 1$ einen p -Wert von 0,034 auf. Dies weist darauf hin, dass die Siebtklässler*innen sich bei der Rezeption des Peerfeedbacks sowohl zufriedener als auch zuversichtlicher fühlen. Das KI-Feedback löst demgegenüber ein Gefühl von Beleidigung aus.

In den freien Antworten äußern sich die Schüler*innen sowohl im KI-Feedback als auch im Peerfeedback eher positiv zur Nützlichkeit der Rückmeldungen. Auch die Affektwirkung der beiden Feedbacks ähnelt sich: Im Hinblick auf die Affektwirkung äußern zwei Schüler*innen gegenüber des KI-Feedbacks negative und ein*e Schüler*in positive Emotionen; beim

Peerfeedback sind es drei positive und eine negative Stimme. Ein anderes Bild ergibt sich im Hinblick auf die Akzeptanz des jeweiligen Feedbacks: Während elf Schüler*innen auch in der qualitativen Befragung dem Peerfeedback zustimmen, sind es beim KI-basierten Feedback vier. Hier dominieren vor allem negative Stimmen (insgesamt acht Schüler*innen).

Beim Rezipieren der Feedbacks identifizieren Schüler*innen sowohl bei dem KI-Feedback als auch bei dem Peerfeedback Verbesserungsvorschläge oder Tipps, Lob, Besonderheiten in der Formulierung und Fehler im Feedback. Das KI-Feedback überwiegt dabei sowohl in der Nennung von Verbesserungsvorschlägen oder Tipps (KI: sieben Nennungen, Peer: drei Nennungen) als auch in der Nennung von Fehlern im Feedback (KI: zwölf Nennungen, Peer: zwei Nennungen).

Vergleicht man unter den Schüler*innen die Antworten, die u. a. Fehlerstellen im Feedback fokussieren, mit den Feedbackpräferenzen, so wird deutlich, dass eine Fehlerstelle im KI-Feedback häufig mit einer Präferenz für das Peerfeedback einher geht:

Tab. 1: Fehler im Feedback und Präferenz bei Klasse 7

Schüler*in	Das Feedback hat Fehler	Präferenz
7_AFGD1	KI	KI
7_AFGJ1	KI, Peer	Peer
7_ALNF2	KI	Peer
7_ELNA1	KI	Peer
7_IFHM4	KI	Peer
7_IFRA2	KI	Peer
7_IFZD2	KI	Peer
7_ILNJ1	KI	Peer
7_ILRJ1	KI	Peer
7_LFAM1	KI	Peer
7_MFRJ1	KI	Peer
7_UJHJ1	KI, Peer	Peer

Elf der zwölf Schüler*innen, die Fehler im KI-basierten Feedback fokussieren, bevorzugen das Peerfeedback (wenngleich zwei dieser Schüler*innen auch hier Fehlerstellen festmachen).

Weniger prominent sind die Fehlerstellen im Peerfeedback. Hier identifizieren die Siebtklässler*innen u. a. ein als oberflächlich oder mangelhaft eingeschätztes Feedback. Demgegenüber werden aber auch die Passung zum Sachtext oder die Kürze des Feedbacks positiv hervorgehoben. Positiv werden beim KI-Feedback u. a. der Bezug zu den Kriterien, die Präzision und Ausführlichkeit des Feedbacks sowie die Verknüpfung des Feedbacks mit Expert*innenfeedback erwähnt.

3.4.1.4 Veränderungen in der Feedbackrezeption nach der Auflösung

Nach der Auflösung lassen sich keine Änderungen in den Wahrnehmungen der Schüler*innen feststellen³⁶: Die Schüler*innen, die zuvor das Peerfeedback bevorzugten, tun dies auch, nachdem sie über die Feedbackquelle informiert werden. Gleiches gilt für die Schüler*innen, die zuvor das KI-Feedback durch *FelloFish* präferierten. Die Schüler*innenantworten unterscheiden sich dabei jedoch in ihren Rechtfertigungen, warum sie ihre Antworten beibehalten.³⁷

Bleibt die Präferenz bei dem Peerfeedback, so führen Schüler*innen u. a. an, dass sie dem Menschen mehr vertrauen würden (drei Nennungen) oder dass sie schon vermutet haben, dass das Feedback von einer KI komme (eine Nennung). Gerade der Aspekt des Vertrauens ist dabei an bereits identifizierte Feedbackkomponenten gebunden. So schreibt 7_IFHM4 auf die Frage, ob sich seine*ihr Wahrnehmung ändern würde:

Bei mir ändert sich nichts. Ich finde das erste Feedback besser weil es bessere Tipps gibt. Außerdem lobt es mich (das erste Feedback was ich immer noch sehr gut finde. Also finde ich es viel besser und werde meine Meinung nicht ändern! Auf die Person kann ich mich wirklich verlassen!

36 Hier gibt es jedoch auch uneindeutige Schüler*innenaussagen. Beispielsweise erwähnen Schüler*innen, dass sie ein bestimmtes Feedback gut fanden. Zuvor hatten diese Schüler*innen jedoch keine Präferenz angegeben.

37 Die Rechtfertigungsgründe sind dabei nicht exklusiv zu betrachten. So kann ein*e Schüler*in beispielsweise sowohl Vertrauen in die Feedbackquelle als auch die Verwendbarkeit des Feedbacks anführen.

Neben einer Stimme, die aufzeigt, dass das KI-Feedback für sie nicht schlecht sei, jedoch nicht an das durch die Peers herankomme – »Ich fand beide gut aber Feedback 2 [Peerfeedback] Besser!« (7_ALUD1) –, begründet ein großer Teil der Schüler*innen (insgesamt neun Nennungen) seine Entscheidung mit den Fehlern im KI-Feedback. 7_ILNJ1 schreibt hierzu: »Bei mir ändert sich nichts. Das erste Feedback [Peerfeedback] gibt mir verständliche Tipps. Die KI sagt falsche Sachen. Dadurch finde ich Feedback zwei [KI-Feedback] nicht hilfreich.« Und auch 7_MFRJ1 hält fest: »Meine Meinung ändert sich nicht. Ich bevorzuge weiterhin die Meinung der Schülerin. Die KI hat sehr viele Fehler gemacht die, die Schülerin nicht gemacht hat.«

Bleibt die Präferenz der Schüler*innen bei der KI, wird die Entscheidung auch hier bei zwei Schüler*innen damit begründet, dass sie schon vermutet hätten, dass das Feedback von einer KI komme:

Weil man das schon vorher gesehen hat, da Feedback 2 [KI-Feedback] vieeeeeeeeeel ausführlicher, genauer und detaillierter, und sonst joa. :) Ich finde, meine Wahrnehmung ändert sich nichts. (7_IFGM0)

Andere Schüler*innen führen an, dass die Quelle für sie keine Rolle spielt (zwei Nennungen) – »Meine Wahrnehmung ändert sich nicht, weil es für mich keinen Unterschied macht, wer es geschrieben hat« (7_AFGD1) – die KI ihnen konstruktive Kritik gegeben hätte (eine Nennung) oder sie der KI vertrauen würden (zwei Nennungen):

Nein denn eine KI ist ein Roboter und Roboter machen oft keine Fehler. Aber ein Mensch macht Fehler was auch normal ist. Bei der KI stande halt alles ausführlich und der von meinem Mitschüler ist halt nicht ausführlich. Ich vertraue mehr der KI bei solchen Texten. (7_EFTS2)

Auch bei 7_EFTS2 tritt Vertrauen als Rechtfertigungsgrund nicht in Isolation auf, sondern ist mit einer Feedbackkomponente (*das Feedback ist ausführlich und/oder detailliert*) verknüpft.

Ein großer Teil der Schüler*innen rechtfertigt seine Entscheidung mit der Vielfalt an Verbesserungsvorschlägen bzw. der Verwendbarkeit des Feedbacks für die eigene Überarbeitung: »Die KI ist ausführlicher und gute Verbesserungsvorschläge wurden genannt.« (7_AFAJ1) oder »Nein, weil ich

vertraue KI und da ich bei Feedback 1 [KI-Feedback] mehr sachen verbessern konnte» (7_EFIS1).

Es zeigt sich somit, dass die Feedbackform eher wenige bis keine Einflüsse auf die Schüler*innenwahl hat. Stattdessen werden neben dem Vertrauen in die jeweilige Feedbackquelle, die Feedbackkomponenten herangezogen, die auch zuvor schon von den Schüler*innen identifiziert worden waren.

3.4.2 Klasse 6

Im Folgenden werden die quantitativen und qualitativen Befunde der Schüler*innenbefragung der Peerfeedback-Intervention in Klasse 6 beschrieben. Daran folgt die Darstellung der Resultate des KI-basierten Feedbacks durch *FelloFish*. Schließlich werden sowohl die Ergebnisse miteinander verglichen als auch die Schüler*innantworten nach der Auflösung behandelt.

3.4.2.1 Peerfeedback

In Klasse 6 geben neun der Schüler*innen an, das erhaltene Peerfeedback zu bevorzugen. Dies macht 42,86 % der Schüler*innenantworten in dieser Klasse aus. Im quantitativen Teil der Erhebung liegen die Schüler*innenantworten im Bereich Fairness bei $M = 4,4,1$. Die Schüler*innen empfinden das Peerfeedback also als eher fair. In Abbildung 10 sind die Verteilungen der Items *Mit dem Feedback bin ich zufrieden*, *Ich betrachte das Feedback als fair* und *Ich betrachte das Feedback als gerechtfertigt* dargestellt: Dem ersten Item stimmen fünfzehn Schüler*innen eher oder vollständig zu, sechs verorten sich im neutralen Bereich. Bei dem Item *Ich betrachte das Feedback als fair* geben siebzehn Schüler*innen an, eher oder vollständig zuzustimmen. Zwei sind neutral und zwei stimmen nicht oder eher nicht zu. Vierzehn Schüler*innen ordnen das Feedback als gerechtfertigt ein, stimmen ihm also vollständig oder eher zu. Sechs sind hier neutral.

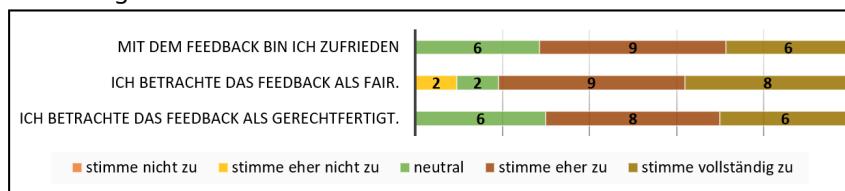


Abb. 10: Wahrnehmung der Fairness im Peerfeedback der Klasse 6

Es lassen sich somit vermehrt positive Reaktionen auf das Peerfeedback erkennen. Einzig im Hinblick auf die Fairness gibt es zwei negative Stimmen.

Ähnliche Befunde zeigen sich auch im Bereich der Nützlichkeit. Auch hier nehmen die Schüler*innen das Feedback eher positiv wahr ($M = 3,61\text{-}3,81$). Dreizehn Schüler*innen geben dabei an, dem Item *Ich betrachte dieses Feedback als nützlich* eher oder vollständig zuzustimmen. Jeweils vier Schüler*innen sind neutral oder stimmen ihm nicht oder eher nicht zu. Als hilfreich und unterstützend wird das Feedback von dreizehn Schüler*innen wahrgenommen – sie stimmen den zugehörigen Items eher oder vollständig zu. Zwei Schüler*innen stimmen dem Item *Ich finde dieses Feedback hilfreich* nicht oder eher nicht zu, sechs sind neutral und vier stimmen dem Item *Dieses Feedback unterstützt mich* nicht oder eher nicht zu. Hier sind vier Schüler*innen neutral. Die Sechstklässler*innen nehmen das Feedback somit zu großen Teilen als nützlich und hilfreich wahr. Auch fühlen sich viele Schüler*innen durch das Feedback unterstützt.

Im Rahmen der Akzeptanz geben achtzehn Schüler*innen an, dem Item *Dieses Feedback akzeptiere ich* eher oder vollständig zuzustimmen. Zwei sind neutral und eine*r stimmt dem Item eher nicht zu. Siebzehn Schüler*innen widersprechen dem Feedback nicht und lehnen es auch nicht ab. Ein*e Schüler*in stimmt dem Item *Diesem Feedback widerspreche ich* nicht zu, drei sind neutral. Dem Item *Dieses Feedback lehne ich ab* stimmt kein*e Schüler*in nicht oder eher nicht zu, hier sind vier neutral. Das Feedback durch die Peers wird folglich von der Mehrheit der Schüler*innen akzeptiert. Nur ein*e Schüler*in widerspricht dem Feedback, kein*e Schüler*in lehnt es ab.

Im Bereich der Identifikation von Fehlerstellen und Verbesserungsvorschlägen lassen sich gemischte Befunde feststellen: Neun Schüler*innen geben hier an, dass sie durch das Feedback wissen, wo und wie sie etwas verbessern können. Vier stimmen dem zugehörigen Item nicht oder eher nicht zu und acht sind neutral.

Ein Überblick über die Affektwirkung in der quantitativen Erhebung ist in Abbildung 11 gegeben:

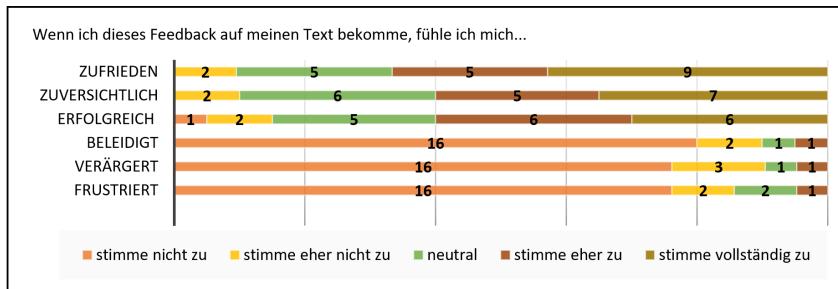


Abb. 11: Affektwirkung im Peerfeedback der Klasse 6

Hier zeigt sich, dass das Peerfeedback vermehrt positive Emotionen bei den Schüler*innen hervorruft. So stimmen vierzehn dem Item *zufrieden* eher oder vollständig zu. Fünf Schüler*innen sind neutral und zwei geben an, dem Item eher nicht zuzustimmen. Je zwölf Schüler*innen fühlen sich beim Rezipieren des Feedbacks *zuversichtlich* oder *erfolgreich* – dem Item *zuversichtlich* stimmen dabei zwei Schüler*innen eher nicht zu, sechs sind neutral und dem Item *erfolgreich* stimmen drei (eher) nicht zu, hier sind fünf neutral. Ein*e Schüler*in fühlt sich *beleidigt* (ein*e Schüler*in ist neutral, achtzehn stimmen (eher) nicht zu) und ein*e Schüler*in *frustriert* (zwei Schüler*innen sind neutral, achtzehn stimmen (eher) nicht zu). Ein*e Schüler*in ist *verärgert*, neunzehn Schüler*innen stimmen dem Item eher (nicht) zu, ein*e andere*r ist neutral.

Zusammenfassend lässt sich also im quantitativen Teil eine überwiegend positive Rezeption des Feedbacks beobachten. Die Sechstklässler*innen empfinden es zu großen Teilen als fair und nützlich und akzeptieren es. Auch die Affektwirkung ist positiv: Vermehrt geben Schüler*innen an, sich *zufrieden*, *zuversichtlich* und *erfolgreich* zu fühlen. Negative Emotionen (z. B. Verärgerung) sind nur in kleinen Teilen vorhanden.

Im qualitativen Teil der Befragung findet ein Teil der quantitativen Befunde Bestätigung. Wieder andere Feedbackeigenschaften oder -komponenten werden hier jedoch in ihrer Wahrnehmung eher kritisch eingestuft bzw. behandelt. Insgesamt zwölf Schüler*innen erwähnen die Nützlichkeit des Feedbacks. Acht der Schüler*innen (38,1 % der gesamten Teilnehmenden

aus Klasse 6) verweisen hier auf eine positive Nützlichkeit des Feedbacks. So schreibt 6_SMRA1: »Ich finde das Feedback hilfreich und es ist verständlich formuliert«. Die wahrgenommene Nützlichkeit ist dabei häufig an ein Feedback auf hierarchieniedrige Prozesse gebunden. 6_ALAF1 schreibt: »Das Feedback ist hilfreich und ich weiß das ich meine rechtschreibung verbessern sollte.« Und auch 6_AFNA1 notiert »Ich fand sehr hilfreich das mir gesagt wurde das ich vom Präsents abgeweicht bin«. Daneben werden auch die Fülle an Informationen, der Kriterienbezug oder eine positive und/oder sachliche Formulierung als Gründe für die Nützlichkeit angeführt, beispielsweise 6_TGZD2: »Ich fand Feedback 2 [KI-Feedback] und 1 [Peerfeedback] beide hilfreich, aber Feedback 2 war dann doch ein wenig hilfreicher, weil es mehr auf die Kriterien eingegangen ist«.

Daneben gibt es auch vier Schüler*innenstimmen, die das Feedback als nicht nützlich einstufen. Hier werden Gründe wie mangelnde Informationen oder fehlende Verbesserungsvorschläge aufgeführt: »Ich habe keine Verbesserungsvorschläge bekommen deswegen werde ich anhand des ersten Textfeedbacks meinen Text nicht verbessern können« (6_RFEM3).

Sieben Schüler*innen verweisen in ihren Aussagen auf eine Akzeptanz bzw. wahrgenommene Fairness im Peerfeedback. Insgesamt sechs dieser Schüler*innen ordnen das Feedback als fair ein bzw. akzeptieren es: »Ich finde das Feedback ist sehr freundlich und fair« (6_AGNO0). 6_TGZD2 konkretisiert seine*ihrre Aussage dabei und bezieht sich auf eine spezifische Aussage der Peers: »Ich finde, dass es richtig war, dass die Person gesagt hat, dass ich hätte schreiben sollen, dass der Spiegel beim ersten Auftritt kaputt gegangen ist« (6_TGZD2).

Das Rezipieren des Feedbacks löst in der qualitativen Befragung bei den Schüler*innen vermehrt positive Emotionen aus. So drückt 6_ELHA1 seine*ihrre Freude über die Präzision des Feedbacks aus: »Du hast sehr viele Hinweise gegeben und trotzdem nicht so viel geschrieben. Das finde ich toll« und 6_AFNA1 lobt die Ehrlichkeit im Feedback: »Sie hat auch gesagt das nicht alles was sie schreibt richtig ist und das fand ich echt cool und vor allem mutig das zu sagen«. Auch wird hier eine wertschätzende und empathische Haltung der feedbackgebenden Person deutlich: 6_AFNA1 nimmt

den*die Sender*in im Feedback wahr und äußert Anerkennung dafür, dass dieser auch auf Fehler im Ausgangstext hingewiesen hat.

Zu den Komponenten des Peerfeedbacks, die bei den Schüler*innen der Klasse 6 besondere Beachtung finden, gehören Lob bzw. die Nennung von Positivem im Schüler*innenprodukt, eine positive und/oder sachliche Formulierung des Feedbacks, das Bereitstellen vieler Informationen im Feedback, Rückmeldungen zu Grammatik-/Rechtschreibfehlern und das Geben von Verbesserungsvorschlägen und Tipps. Hinzu treten Vermerke, dass das Feedback kurz und/oder prägnant sei, Lob und Kritik kombiniere oder nur Kritik gebe. Auch vermerken Schüler*innen, dass Teile im Feedback fehlen würden, eine Überarbeitung nicht möglich oder das Feedback fehlerhaft sei.

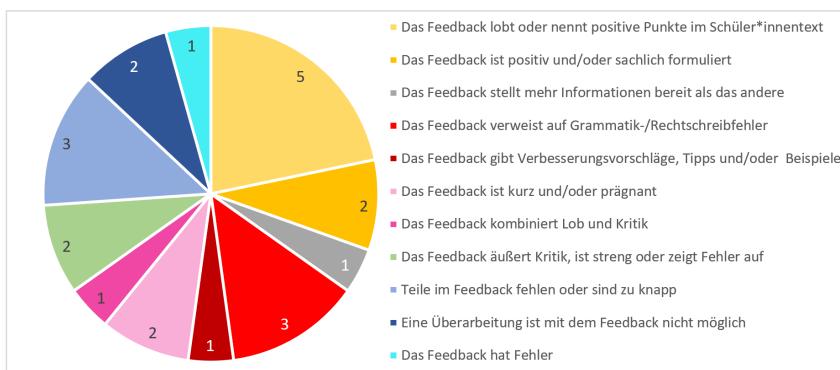


Abb. 12: Verteilung der Komponenten des Peerfeedbacks in der Rezeption der Klasse 6

Wie aus Abbildung 12 hervorgeht, nimmt Lob einen großen Teil der Schüler*innenäußerungen ein. Die Schüler*innen rezipieren das Lob im Feedback jedoch auch kritisch. So schreibt 6_RNX1: »Feedback 1 [Peerfeedback] wurde nicht so viel geschrieben wie Feedback 2 [KI-Feedback]. In Feedback 2 wurde immer was kommentiert in Feedback 1 wurde einfach gelobt.« Darüber hinaus bemängeln die Schüler*innen, dass reines Lob ihnen nicht bei der Überarbeitung des Textes helfen: »In Feedback 2 [Peerfeedback] wurde geschrieben das fast alles gut war außer eine Sache und es wurden keine Fehler genannt sodass man sich nicht verbessern kann« (6_IFIN2) und »Ich wurde in allen Teilen gelobt und ich finde die Person die das Feedback

geschrieben hat, hat an keiner stelle etwas schlechtes geschrieben. Deswegen ist kein Teil wichtig« (6_SMRA1).

Einen weiteren größeren Teil nehmen mit drei Nennungen Grammatik- oder Rechtschreibfehler ein – »ich weiß das ich meine rechtschreibung verbessern sollte« (6_ALAF1) – und auch auf die Kürze bzw. Prägnanz des Feedbacks verweisen die Schüler*innen: »Du hast sehr viele Hinweise gegeben und trotzdem nicht so viel geschrieben.« (6_ELHA1).

Darüber hinaus wird aber auch hervorgehoben, dass das Feedback zu kurz sei. Drei Schüler*innen kritisieren, dass das Feedback ihnen nicht genug Informationen gebe: »Der oder diejenige hat immer nur ja oder nein als Antwort geschrieben und nicht gesagt warum« (6_AMRA1) und zwei Schüler*innen notieren, dass eine Überarbeitung mit dem Feedback nicht möglich sei: »Bei Feedback 2 [Peerfeedback] konnte man nicht viel finden um sich zu verbessern.« (6_IFIN2).

3.4.2.2 Feedback durch FelloFish

In Klasse 6 geben zwölf Schüler*innen an, das Feedback von *FelloFish* zu präferieren. Dies entspricht einem Anteil von 57,14%. Hinsichtlich der Fairness des Feedbacks liegen die Schüler*innenantworten dabei bei neutral bis eher zustimmend ($M = 3,3 - 3,5$). Wie aus Abbildung 13 hervorgeht, stimmen dem Item *Mit dem Feedback bin ich zufrieden* zehn Schüler*innen eher oder vollständig zu, sieben sind neutral und vier stimmen nicht oder eher nicht zu.

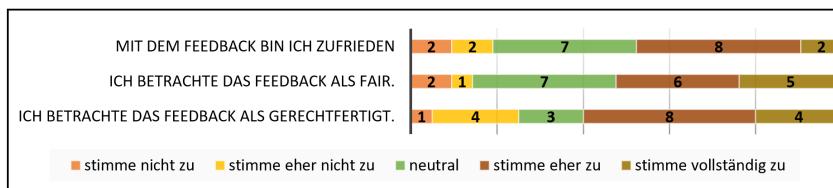


Abb. 13: Wahrnehmung der Fairness im KI-Feedback der Klasse 6

Elf Schüler*innen stimmen dem Item *Ich betrachte das Feedback als fair* eher oder vollständig zu, sieben sind neutral und drei stimmen ihm nicht oder eher nicht zu. Dem Item *Ich betrachte das Feedback als gerechtfertigt* stimmen zwölf Schüler*innen eher oder vollständig zu. Hier sind drei neutral und fünf stimmen nicht oder eher nicht zu. Folglich zeigen sich in allen drei Items

eher positive Ausprägungen. Ein großer Teil der Schüler*innen ist mit dem Feedback zufrieden und empfindet es als fair und gerechtfertigt. Die Gegenstimmen fallen eher klein aus.

Auch im Bereich der Nützlichkeit äußern die Schüler*innen ähnliche Einschätzungen. Sie nehmen das Feedback vielfach als hilfreich und nützlich wahr, allerdings treten auch neutrale Bewertungen auf. So stimmen elf Schüler*innen dem Item *Ich betrachte dieses Feedback als nützlich* zu. Sechs ordnen sich als neutral ein und vier Schüler*innen stimmen dem Item (eher) nicht zu. Zehn Schüler*innen verstehen das Feedback als *hilfreich*, acht Schüler*innen sind neutral und drei stimmen dem zugehörigen Item (eher) nicht zu. Hinsichtlich der empfundenen Unterstützung durch das Feedback zeigen sich eher gemischte Rückmeldungen: Acht Schüler*innen fühlen sich durch das Feedback *unterstützt*, sechs fühlen sich nicht unterstützt und sieben sind neutral.

Ein Großteil der Schüler*innen akzeptiert das Feedback. Dabei geben zwölf Schüler*innen an, dem Item *Dieses Feedback akzeptiere ich* eher oder vollständig zuzustimmen. Vier Schüler*innen stimmen hier (eher) nicht zu und fünf sind neutral. Über die Hälfte der Schüler*innen lehnen das Feedback weder ab noch widersprechen sie ihm. Dem Item *Diesem Feedback widerspreche ich* stimmen fünf Schüler*innen eher oder vollständig zu, zwei sind neutral und vierzehn stimmen dem (eher) nicht zu. Dem Item *Dieses Feedback lehne ich ab* stimmen zwei der Sechstklässler*innen vollständig oder eher zu, fünf sind neutral und dreizehn Schüler*innen stimmen (eher) nicht zu.

Im Bereich der Identifikation von Fehlerstellen und Verbesserungsvorschlägen lassen sich vermehrt positive Befunde feststellen: Hier geben elf Schüler*innen an, dass sie durch das Feedback wissen, wo und wie sie etwas verbessern können. Je fünf Sechstklässler*innen stimmen dem zugehörigen Item (eher) nicht zu oder sind neutral.

Ein Überblick über die Affektwirkungen des Feedbacks durch *FelloFish* ist in Abbildung 14 dargestellt. Hier zeigt sich, dass das KI-basierte Feedback vermehrt positive Emotionen auslöst.

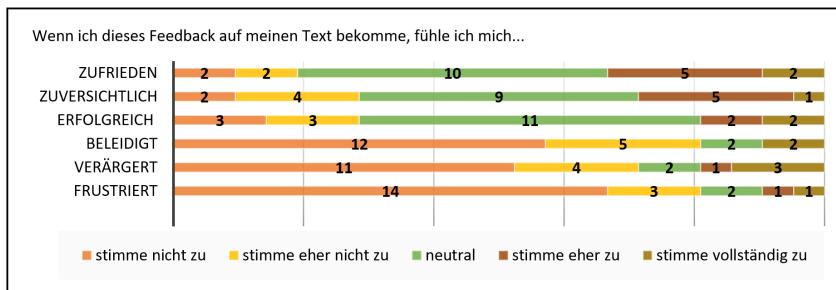


Abb. 14: Affektwirkung im KI-Feedback der Klasse 6

Bei dem Rezipieren des Feedbacks stimmen sieben Schüler*innen dem Item *zufrieden* eher oder vollständig zu. Zehn sind neutral und vier stimmen nicht oder eher nicht zu. Sechs Schüler*innen fühlen sich *zuversichtlich*, neun Schüler*innen sind neutral und sechs Schüler*innen geben an, dem zugehörigen Item nicht oder eher nicht zuzustimmen. Vier Schüler*innen fühlen sich *erfolgreich*. Hier sind elf neutral und sechs stimmen nicht oder eher nicht zu.

Den Items *beleidigt*, *verärgert* und *frustriert* stimmen die Schüler*innen vermehrt nicht oder eher nicht zu. So fühlen sich nur zwei Schüler*innen durch das Feedback beleidigt. Siebzehn Schüler*innen empfinden dies nicht und zwei sind neutral. Dem Item *verärgert* stimmen vier der Sechstklässler*innen zu. Hier sind zwei neutral und fünfzehn fühlen sich nicht verärgert. Zwei Schüler*innen fühlen sich durch das Feedback frustriert. Siebzehn empfinden dies nicht und zwei sind neutral.

Es lassen sich somit für das KI-basierte Feedback durch *FelloFish* für Klasse 6 vermehrt positive und weniger negative Emotionen feststellen. Die Schüler*innenstimmen für die einzelnen Rezeptionsdimensionen liegen dabei meist im neutralen bis positiven Bereich.

Im Rahmen der qualitativen Befragung können dreizehn Stellen identifiziert werden, an denen sich die Schüler*innen zur Nützlichkeit des Feedbacks äußern. Elf Schüler*innen empfinden das Feedback dabei hinsichtlich seiner

Nützlichkeit als positiv. So schreibt 6_AMRA1: »Das Feedback ist sehr hilfreich weil es sehr ausführlich ist« und 6_IFIN2 notiert: »Dieses Feedback war hilfreich da es an manchen Sachen die man nicht richtig gemacht hat genannt wurde was man machen kann um sich zu verbessern«.

Wie diese beiden Schüler*innenäußerungen erkennen lassen, kann die wahrgenommene Nützlichkeit des KI-Feedbacks mit dessen Ausführlichkeit und den gegebenen Verbesserungsvorschlägen einhergehen. Es finden sich aber auch Verweise auf den Kriterienbezug des Feedbacks. So schreibt 6_TGZD2 als Antwort auf die Fragen *Welches Feedback empfandest du als hilfreicher für die Verbesserung deines Textes, und warum? Welche Teile dieser Rückmeldung haben dir besonders geholfen?*: »Ich fand Feedback 2 [KI-Feedback] und 1 [Peerfeedback] beide hilfreich, aber Feedback 2 war dann doch ein wenig hilfreicher, weil es mehr auf die Kriterien eingegangen ist.«.

Zwei Schüler*innenäußerungen thematisieren, dass das Feedback eher nicht hilfreich ist. Dies ist dabei an Fehlerstellen im Feedback geknüpft:

Ich empfinde das Feedback nicht als sehr hilfreich. Es war zwar verständlich formuliert, jedoch gibt es einen Fehler. Es wurde gesagt, dass ich den Titel im Einleitungssatz nicht in Anführungszeichen gesetzt habe, obwohl ich das gemacht hab. Außerdem wurden Wiederholungen erwähnt, obwohl im Text keine Wiederholungen vorkommen. (6_OSBNM1)

Zur Akzeptanz bzw. Fairness äußern sich sieben Schüler*innen – sechs davon negativ. 6_ASML1 bezeichnet das Feedback als »sehr schlecht«, 6_RNXJ1 als »total falsch«. Aufschlüsse über die Ursachen bieten dabei Äußerungen von 6_AFNA1: »Du musst den Text noch besser lesen, die Autorin wurde genannt. Außerdem hast du selber gesagt, man soll nur das wichtigste benutzen.«. Auch hier wird die Akzeptanz des Feedbacks mit dessen Fehlerstellen verknüpft.

Während im Rahmen der Affektwirkung ein*e Schüler*in von positiven Emotionen spricht – »ich fühle mich dadurch glücklich« (6_AGNO0) – finden sich auch drei Schüler*innen, die sich negativ äußern: »Ich finde dieses Feedback für mich sehr frustrierend. Außerdem bin ich ziemlich wütend darauf« (6_ASML1). Auch hier werden die negativen Emotionen oftmals mit Fehlern im Feedback verbunden:

Ich finde es absurd man siet es klar und deutlich. Ich sage nur dazu das der spint der das feedback schreibt. Was mir helfen könnte wäre bei manchen richtige Fehler zu beachten. Bei Kommentar 3 werde ich gelobt was ich schön finde aber da steht auch das ich das und das nicht geschrieben habe. Obwohl ich das geschrieben habe. Das andere finde ich oke. (6_RNXJ1)

Fehler im Feedback gehören zu den Komponenten des KI-Feedbacks, die bei den Schüler*innen besondere Beachtung finden. Insgesamt fünf Schüler*innen erwähnen Fehler im KI-Feedback. So schreibt 6_RNXJ1: »Bei Feedback 2 [KI-Feedback] ist total falsch. Ich habe die Autorin hingeschrieben.« Auch im Vergleich mit dem Peerfeedback, werden diese Fehlerstellen hervorgehoben. 6_LMSJ1 schreibt hierzu, dass »in Text zwei [KI-Feedback] viel falsch ist und in Text eins [Peerfeedback] so gut wie alles gut ist«.

Wie aus Abbildung 15 hervorgeht, nehmen weitere große Teile der Komponenten des KI-Feedbacks seitens der Schüler*innen die Formulierung des Feedbacks (fünf Nennungen) und die Verbesserungsvorschläge, Tipps oder Beispiele (sechs Nennungen) ein. Im Rahmen der Verbesserungsvorschläge, heben vier Schüler*innen die Vielzahl an Beispielen hervor. So schreibt 6_EFNA3: »Ich finde Feedback 1 [KI-Feedback] besser weil dort Beispiele genannt werden und so kann ich gezielter meine Fehler bearbeiten.« und 6_HFRM1 notiert »Ich fande Feedback 1 [KI-Feedback] viel besser weil beispiele darin erwähnt waren.. Ein*e weitere*r Schüler*in (6_SMRA1) erzählt, dass das KI-Feedback sie*ihn im Hinblick auf präzisere Formulierungen unterstützt habe.

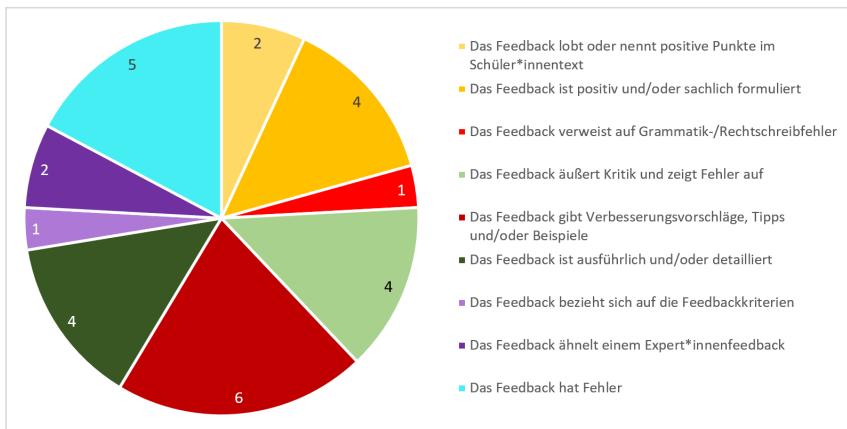


Abb. 15: Verteilung der Komponenten des KI-Feedbacks in der Wahrnehmung der Klasse 6

Die Formulierung des Feedbacks bezeichnen einige Schüler*innen als »netter und schöner formuliert« (6_AMRA1) oder als »ehrlicher und besser« (6_EKTS1). Es gibt jedoch auch Schüler*innenstimmen, die darüber hinausgehen. So erkennt 6_ELHA1 in der generischen Formulierung einen Hinweis auf KI: »Ist genau das gleiche wie bei Schüler X [Änderung LP] und anderen deshalb denke ich dass KI im Spiel ist.«.

Je vier Schüler*innen verweisen darüber hinaus darauf, dass im Feedback Kritik enthalten und das Feedback ausführlich und detailliert ist. Beispielsweise schreibt 6_SMRA1: »Feedback 2 [KI-Feedback] wurden mehr Sachen geschrieben die nicht so gut sind.« Und 6_RFEM3 notiert: »Feedback 1 [Peer-feedback] war viel kürzer und unausführlich im Gegensatz zu Feedback 2 [KI-Feedback].«.

Daneben verknüpfen zwei Schüler*innen das erhaltene *FelloFish*-Feedback mit Expert*innenfeedback. So notiert 6_AMRA1: »Das Feedback ist sehr hilfreich weil es sehr ausführlich ist und Frau Pitzen [Verfasserin der Masterarbeit] immer Beispiele gemacht hat.« Und auch 6_ILHJ1 schreibt: »Feedback 1 [KI-Feedback] war hilfreicher, aber auch weil es von einem Erwachsenen geschrieben wurde und nicht von einem Kind.« Auch heben die Schüler*innen Lob durch das Feedback hervor (zwei Nennungen) und den Bezug zu den einzelnen Feedbackkriterien (eine Nennung).

3.4.2.3 Vergleich der Ergebnisse

Zur Überprüfung der Mittelwertsdifferenzen der quantitativen Erhebung zwischen den Feedbackquellen werden auch hier parametrische Analysen mittels gepaarter t-Tests durchgeführt. Im Rahmen der Fairness kann für das Item *Mit dem Feedback bin ich zufrieden* zwischen den Mittelwerten von KI-basiertem Feedback ($M_{KI} = 3,286$; $SD = 1,102$) und Peerfeedback ($M_{PEER} = 4$; $SD = 0,775$) eine signifikante Differenz ($p = 0,044$) festgestellt werden. Im Bereich der Nützlichkeit des jeweiligen Feedbacks können keine signifikanten Unterschiede festgestellt werden. Bei dem Item *Dieses Feedback akzeptiere ich* (Akzeptanz) zeigt sich ein signifikanter Unterschied ($p = 0,034$) zwischen KI-basiertem Feedback ($M_{KI} = 3,524$; $SD = 1,250$) und Peerfeedback ($M_{PEER} = 4,333$; $SD = 0,856$). Diese Befunde deuten darauf hin, dass die Schüler*innen der Klasse 6 im Vergleich zufriedener mit dem Feedback der Peers sind und es bevorzugt akzeptieren.

Im Rahmen der Affektwirkung des Feedbacks können signifikante Differenzen bei den Items *zufrieden*, *zuversichtlich* und *erfolgreich* nachgewiesen werden. Für das Item *zufrieden* ergeben sich dabei Mittelwerte von $M_{KI} = 3,143$ ($SD = 1,062$) und $M_{PEER} = 4$ ($SD = 1,049$) mit $p = 0,027$. Das Item *zuversichtlich* weist mit den Mittelwerten $M_{KI} = 2,952$ ($SD = 1,024$) und $M_{PEER} = 3,85$ ($SD = 1,04$) einen p-Wert von 0,025 auf und das Item *erfolgreich* mit den Mittelwerten $M_{KI} = 3$ ($SD = 1,108$) und $M_{PEER} = 4$ ($SD = 1,174$) einen p-Wert von 0,028. Die Schüler*innen zeigen folglich beim Peerfeedback signifikant häufiger positive Emotionen. Sie fühlen sich zufriedener, zuversichtlicher und erfolgreicher.

Im Rahmen der qualitativen Erhebung finden sich beim Rezipieren des Peerfeedbacks vermehrt positive Stimmen im Bereich Nützlichkeit, Affektwirkung und Akzeptanz bzw. Fairness. Bei dem Feedback durch *FelloFish* zeigen sich bei der wahrgenommenen Nützlichkeit deutlich positive Reaktionen, während in den Bereichen Affektwirkung und Akzeptanz bzw. Fairness negative Stimmen dominieren.

Sowohl im Peer- als auch im KI-Feedback erkennen die Schüler*innen lobende bzw. positive Rückmeldungen oder eine unterstützende Formulierung. Während Lob im Peerfeedback von einigen Schüler*innen kritisch bewertet wird, ist eine solche Einschätzung beim KI-Feedback nicht zu

finden. Zudem wird beim KI-Feedback die Vielzahl an Verbesserungsvorschlägen und Beispielen positiv hervorgehoben (sechs Nennungen), wohingegen dies beim Peerfeedback lediglich von einer Person erwähnt wird. Den Aussagen der Schüler*innen zufolge scheint *FelloFish* somit sowohl ehrliches Lob als auch konkrete Verbesserungsvorschläge zu enthalten.

Demgegenüber werden beim KI-Feedback häufig Fehlerstellen hervorgehoben. Diese gehen jedoch nicht zwingend mit der anschließenden Präferenz des Peerfeedbacks einher. So präferieren nur zwei der fünf Schüler*innen, die Fehler identifizieren, am Ende des ersten Fragebogens das Feedback der Peers.

Tab. 2: Fehler im Feedback und Präferenz bei Klasse 6

Schüler*in	Das Feedback hat Fehler	Finale Präferenz
6_IFIN2	KI	KI
6_AFNA1	KI	Peer
6_LMSJ1	KI	Peer
6_OSBNM1	KI, Peer	KI
6_RNXJ1	KI	KI

Beim Peerfeedback finden sich auf der anderen Seite vermehrt Äußerungen hinsichtlich eines zu knappen oder unvollständigen Feedbacks (einige Schüler*innen geben auch an, dass eine Überarbeitung nicht möglich sei).

3.4.2.4 Veränderungen in der Feedbackrezeption nach der Auflösung

Nach der Auflösung lassen sich keine Änderungen in den Wahrnehmungen der Schüler*innen feststellen.³⁸ So bleiben Schüler*innen, die zuvor angegeben hatten, dass sie das Peerfeedback bevorzugen würden, bei dieser Einstellung. Fünf Schüler*innen führen dabei an, dass sie sich schon vorher gedacht hätten, dass das andere Feedback von einer KI verfasst wurde – implizieren somit also ein Vertrauen in das *Human Feedback*.

38 Auch hier gibt es jedoch wieder uneindeutige Aussagen.

Letzterer Gedanke wird dabei auch von einem*r weiteren Schüler*in aufgegriffen, als er*sie gefragt wird, ob sich die Wahrnehmung des Feedbacks ändere: »Nein weil eine künstliche Intelligenz ist nicht so gut wie ein mensch und man kann auch sehen das ich Feedback 2 [Peerfeedback] gewählt habe!!!« (6_EKTS1). Drei Schüler*innen rechtfertigen ihre Wahl mit den Fehlern im KI-Feedback. So schreiben beispielsweise 6_ASML1 und 6_AFNA1:

Bei Feedback 2 [Peerfeedback] war alles genau so wie ich mir eigentlich forgestellt hatte z.B. So wie ich mir erhofft hatte. Bei Feedback 1 was von der KI (fiete.ai) geschrieben wurde sind ein paar Fehler ihm unterlaufen, zudem wendet die KI nur kritische Fehler an. Was ziemlich blöd ist, weil man dann richtig sauer und frustriert ist. (6_ASML1)

Nein, den Ich habe es schon vermutet das Feedback 1 von einem Mitschüler war. Außerdem kann es mir merkwürdig vor das gesagt wurde: >Du hast den Titel und das Thema gut genannt. Aber du hast die Autorin vergessen< obwohl da also im Text die Autorin stand. Außerdem würde das ein Schüler niemals machen. (6_AFNA1)

Im weiteren Verlauf ihrer*seiner Antwort lässt 6_AFNA1 zudem eine Interaktion mit den Peers erkennen: »Und wenn die Person die den zweiten Text geschrieben hat das hier sieht: >Ich fand dein Feedback mega nett, du hast mich auch sehr, sehr, sehr nett auf meine Fehler hingewiesen.<<.

Bleiben die Schüler*innen bei ihrer Präferenz des Feedbacks durch *FelloFish*, werden auch hier verschiedene Rechtfertigungen angeführt. Am häufigsten werden dabei die bereitgestellten Verbesserungsvorschläge oder die Ausführlichkeit des Feedbacks genannt:

Meine Meinung ändert sich nicht, weil ich jetzt weiß dass Feedback 2 von einer KI geschrieben wurde. Ich finde nachwie vor das mir die KI ein besseres Feedback gegeben hat und gute Verbesserungsvorschläge gegeben hat. (6_RFEM3)

Hinzutritt ein differenzierteres Verständnis der Feedbackquellen, d. h., dass beispielsweise die Fehlerstellen der KI zwar angesprochen werden, die Schüler*innen aber dennoch angeben, das KI-Feedback zu präferieren. So schreibt 6_IFIN2:

Sich ändert nichts an meiner Wahrnehmung da ich finde das das feedback von ai richtig ist. Es hat ein gutes, vollständiges und richtiges Feedback gegeben was auch stimmt. Es hat ein paar Fehler gemacht (nicht alle) und hat genannt was man nächstes mal besser machen kann.

6_TGZD2 gibt an, der KI zwar nicht komplett zu vertrauen, unterstreicht aber gleichzeitig auch den Novizenstatus der Peers im Feedbackprozess:

Nein, eigentlich nicht, weil ich das eigentlich mir schon vorher vorgestellt habe. Aber trotzdem nehme ich das Feedback vom Schüler oder von der Schülerin anders war. Beim einen weiß ich, dass es vielleicht ein wenig ungenauer oder nicht so viele Tipps gibt, weil diese Person das ja auch selber neu gemacht hat und wahrscheinlich noch nicht so oft gemacht hat. Ich vertraue der KI aber trotzdem nicht 100%, weil die KI kann auch Fehler machen.

Auch **6_ILZD0** bezieht die Peers in seine*ihre Überlegungen ein: »Bei mir endert sich nichts wobei der Schüler wahrscheinlich länger überlegen muste außerdem ist die Ki schlauer als der Schüler.« In diesen Aussagen scheint der Noviz*innenstatus der Peers gepaart mit den Unterstützungsmöglichkeiten der KI bedeutsamer gewesen zu sein, als die Fehler im Feedback von FelloFish.

Ein Teil der Schüler*innen gibt an, dass für sie die Feedbackquelle wichtig ist oder sie schon Vorahnungen hatten, kombinieren dies für ihre Recht fertigung jedoch häufig mit anderen Komponenten der jeweiligen Feedbacks. So vermerkt **6_RNXJ1**, dass, ihm*ihr die Kenntnis über die Feedbackquelle wichtig ist, obwohl er*sie die Ausgangspräferenz beibehält, weil u. a. das Lob im Peerfeedback wenig förderlich für ihn*sie war:

Nein es ändert nichts weil ich schon forher wusste das feedback 1 [Peer-feedback] von einem Klassenkamerad*in geschrieben. Trotzdem finde ich das beruhigender das man weiß wer das geschrieben hat. Feedback 2 [KI-Feedback] war lange und hat feedback gegeben und bei Feedback 1 wurde nur gelobt. Das findet man ja schön doch das ist kein Feedback.

6_EFNA3 führt an, dass sie sich schon gedacht habe, dass es sich um ein KI-Feedback gehandelt hat. Greift aber gleichzeitig die vielen Beispiele im Feedback auf – und bleibt so bei seiner*ihrer Präferenz:

Ich habe mir schon gedacht das Feedback 1 von der KI geschrieben wurde weil man es beim lesen einfach gemerkt hatwie schon bereits erwähnt zum Beispiel die ganzen Beispiele. An meiner wahrnehmung ändert sich daher auch nichts.

3.5 Diskussion

In der Untersuchung der quantitativen Daten der Klassen 6 und 7 zeigen sich Ähnlichkeiten im Hinblick auf die Rezeptionsdimensionen (F1a). In den Bereichen Akzeptanz, Fairness und Affektwirkung wird das Peerfeedback im Vergleich zum Feedback durch *FelloFish* in beiden Klassen positiver eingestuft. So geben die Schüler*innen beispielsweise an, sich nach dem Rezipieren des Feedbacks erfolgreich und unterstützt zu fühlen, auch seien sie in Bezug auf die Fairness zufriedener mit dem Feedback.

Bestätigung finden diese Befunde zu Teilen in den qualitativen Ergebnissen. Auch hier wird das Feedback der KI eher mit negativen Emotionen rezipiert, während das Peerfeedback eher mit einer positiven Affektwirkung einhergeht. Noch größer wird die Schere zwischen den Feedbackformen im Bereich Akzeptanz bzw. Fairness: Im qualitativen Teil ordnen hier elf (Klasse 7) und sechs (Klasse 6) der Schüler*innen das Peerfeedback positiv ein. Beim KI-Feedback sind es nur vier (Klasse 7) und ein*e (Klasse 6) Schüler*innen.

Ein anderes Bild ergibt sich für die wahrgenommene Nützlichkeit: Hier sind in der quantitativen Untersuchung eher weniger signifikante Unterschiede festzustellen³⁹. Auch im qualitativen Teil zeigen sich Ähnlichkeiten zwischen den beiden Feedbackformen: Sowohl die Schüler*innen der Klasse 6 als auch der Klasse 7 äußern sich vermehrt positiv zu sowohl KI- als auch Peerfeedback.⁴⁰

39 Einzig bei dem Item *Dieses Feedback unterstützt mich* lassen sich bei Klasse 7 signifikante Unterschiede zugunsten des Peerfeedbacks feststellen.

40 Die Schüler*innen der Klasse 7 äußern sich in zwölf Stimmen positiv (negativ: vier) zur Nützlichkeit des KI-Feedbacks und in 14 Stimmen positiv zur Nützlichkeit des Peerfeedbacks (negativ: fünf) Stimmen. Die Schüler*innen der Klasse 6 äußern sich in elf Stimmen positiv und in zwei Stimmen negativ zur Nützlichkeit des KI-Feedbacks. Beim Peerfeedback sind es hier acht positive und vier negative Stimmen.

Den Hypothesen zu F1a kann somit nur in Teilen zugestimmt werden: Wie erwartet, ordnen die Schüler*innen das Peerfeedback im Bereich Affektwirkung positiver ein. Statt jedoch das Feedback durch die KI im Hinblick auf Zustimmung, Wahrnehmung und Nützlichkeit zu favorisieren, bleibt diese Form des Feedbacks hinter dem Peerfeedback zurück (Akzeptanz und Fairness) oder wird von den Schüler*innen ähnlich eingeordnet (Nützlichkeit).

Da das Peerfeedback in den Rezeptionsdimensionen häufig besser bewertet wird als das der KI, könnte man annehmen, dass es von den Schüler*innen bevorzugt wird. Diese Annahme trifft jedoch nicht auf beide Klassen gleichmaßen zu. Ein Vergleich der Schüler*innengruppen hinsichtlich ihrer Feedbackpräferenzen zeigt deutliche Unterschiede auf: Während in Klasse 7 siebzehn Schüler*innen das Peerfeedback präferieren und sieben das von *FelloFish*, bevorzugen in Klasse 6 zwölf Schüler*innen das Feedback der KI und sechs das der Peers (F1b).

Ein erster möglicher Erklärungsansatz für diesen Unterschied liegt in der Identifikation von Fehlerstellen und Verbesserungsvorschlägen innerhalb der jeweiligen Feedbackkarten (Item: *Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann*, F1a). Im quantitativen Teil des Fragebogens zeigt sich, dass in Klasse 6 das KI-Feedback dominiert, während in Klasse 7 das Peerfeedback bevorzugt wird.⁴¹ Einen tiefergehenden Erklärungsansatz liefert jedoch die Analyse der von den Schüler*innen benannten Feedbackkomponenten (F2).

Die Feedbackkomponenten in den Klassen 6 und 7 weisen auf den ersten Blick Parallelen auf. In beiden Fällen werden in den Rückmeldungen Elemente wie Lob, positive Formulierungen sowie fehlerhafte Inhalte identifiziert. Was sich jedoch unterscheidet, ist die Intensität und die Rahmung, in denen die Schüler*innen von den Feedbackkomponenten berichten. Zum besseren Verständnis sollen die Feedbackkomponenten im Folgenden in Kleingruppen systematisch besprochen werden.

41 Die Unterschiede sind zwar nicht signifikant, können jedoch als erste Indikatoren dienen.

Tab. 3: Feedbackkomponenten zu wertschätzender Rückmeldung und Identifikation von Fehlern in ihrer Verteilung und Häufigkeit (Nennungen: **** (>40%), *** (40–20 %), ** (19–10 %), * (<10%)

Feedbackkomponente	Klasse 7		Klasse 6	
	Peerfeedback	FelloFish-Feedback	Peerfeedback	FelloFish-Feedback
Das Feedback lobt oder nennt positive Punkte im Schüler*innentext	***	*	***	**
Das Feedback ist positiv und/oder sachlich formuliert	**	*	**	***
Das Feedback ist nicht so streng	*			
Das Feedback kombiniert Lob und Kritik	*		*	
Das Feedback äußert Kritik zeigt und Fehler auf		**	**	***

Sowohl im Peer- als auch im *FelloFish*-Feedback verweisen beide Klassen auf das Lob oder die positive und/oder sachliche Formulierung im Feedback. In Klasse 7 finden sich neutrale bis positive Reaktionen auf das Lob von sowohl *FelloFish* als auch den Peers – wobei Lob im Peerfeedback verstärkt wahrgenommen wird und dabei mit einer positiven Affektwirkung oder dem Gefühl von Zustimmung und Akzeptanz einhergeht: »Ich finde es sehr gut weil ich oft gelobt werde« (7_AFSM0).

In der Rezeption der Klasse 6 zeichnet sich ein gegensätzliches Bild im Hinblick auf die Wahrnehmung von Lob ab. Dieses wird ebenfalls gerade im Peerfeedback verstärkt hervorgehoben. Anders als in Klasse 7 ordnen einige Schüler*innen der Klasse 6 dieses Lob jedoch eher kritisch oder negativ ein. So schreibt beispielsweise 6_SMRA1: »Ich wurde in allen Teilen gelobt und ich finde die Person die das Feedback geschrieben hat, hat an keiner stelle etwas schlechtes geschrieben. Deswegen ist kein Teil wichtig«. Während in Klasse 7 Lob etwas durchweg Positives ist, verweisen einige

Schüler*innen in Klasse 6 darauf, dass reines Lob ihnen nicht bei ihrer Überarbeitung helfe – kritisieren das Peerfeedback also stark. Dieser Befund zeigt Schnittstellen zu den Ergebnissen von Gamlem und Smith (2013) auf: Auch hier gehört für die Lerner*innen zu einem guten Feedback nicht nur Lob, sondern auch die Konkretisierung von Fehlerstellen.

Neben der Identifikation von Lob wird das Peerfeedback in Klasse 7 als weniger streng aufgefasst. In beiden Klassen wird zudem darauf verwiesen, dass das Peerfeedback Lob und Kritik kombiniert. Demgegenüber stellt Klasse 7 nur im *FelloFish*-Feedback Kritik in Isolation fest – Klasse 6 findet diese in beiden Feedbacks.

Wie bereits in den Vorüberlegungen in Kapitel 2.4 aufgeführt wurde, ist es für die Feedbackrezipient*innen auch wichtig, sich im Feedbackprozess wohlzufühlen. Beispielsweise autoritäres Feedback sei hier zu vermeiden (vgl. Jonsson, 2012). Auch wenn das Aufzeigen von Kritik in Isolation nicht mit autoritarem Feedback gleichzusetzen ist, kann doch – gerade, weil das Peerfeedback den Schüler*innen viel Lob bietet – angenommen werden, dass sich die Schüler*innen der Klasse 7 weniger wohl mit dem KI-basierten Feedback fühlen. Denkbar ist auch, dass sie bisher weniger oder andere (Vor-)Erfahrung mit Feedback sammeln konnten als die Schüler*innen der Klasse 6 und so reine Kritik im Feedback für sie weniger gut umzusetzen ist oder eher mit Ablehnung einhergeht (vgl. (1) in Kap. 2.4).

Tab.4: Feedbackkomponenten zu Qualität und Passung des Feedbacks in ihrer Verteilung und Häufigkeit (Nennungen: **** (>40%), *** (40–20 %), ** (19–10 %), * (<10%))

	Klasse 7		Klasse 6	
Feedbackkomponente	Peer-feedback	<i>FelloFish</i>-Feedback	Peer-feedback	<i>FelloFish</i>-Feedback
<i>Das Feedback stellt mehr Informationen bereit als das andere</i>			•	
<i>Das Feedback passt zum Schüler *innentext</i>	•			

<i>Das Feedback verweist auf Grammatik-/Rechtschreibfehler</i>	•		••	•
<i>Das Feedback gibt Verbesserungsvorschläge, Tipps und/oder Beispiele</i>	••	•••	•	•••
<i>Das Feedback ist ausführlich und/oder detailliert</i>		•••		•••
<i>Das Feedback ist kurz und/oder prägnant</i>	•		••	
<i>Das Feedback bezieht sich auf die Feedbackkriterien</i>		•		••
<i>Das Feedback ähnelt einem Expert*innenfeedback</i>		•		•

Hinsichtlich der Qualität und Angemessenheit des Peerfeedbacks erkennen Schüler*innen beider Klassen Hinweise auf Grammatik- und Rechtschreibfehler sowie eine kompakte bzw. prägnante Ausdrucksweise im Peerfeedback. Eine durchgängig in allen Rückmeldungen auftretende Komponente sind Verbesserungsvorschläge, Tipps oder Beispiele.

Wie aus Tabelle 4 ersichtlich wird, zeigen diese die stärksten Ausprägungen im *FelloFish*-Feedback. Dies bedeutet, dass die meisten Schüler*innen (ungeachtet der Klasse) beim Rezipieren des KI-basierten Feedbacks Verbesserungsvorschläge, Tipps und/oder Beispiele identifizieren. Auch erkennen die Schüler*innen klassenübergreifend im Feedback durch *FelloFish* einen Bezug zu den (Feedback-)Kriterien, empfinden es als detailliert und/oder ausführlich und verweisen darauf, dass es dem Feedback eines*r Expert*in ähnelt.

Aus den Überlegungen in Kapitel 2.4 ist deutlich geworden, dass es gerade die Detailliertheit oder die übermittelten Verbesserungsvorschläge sind, die Schüler*innen im Feedback brauchen bzw. bevorzugen, um Verbesserungen an ihrem eigenen Text vornehmen zu können. Auch ein individuelles Feedback, welches konkrete Fehlerstellen am Text herausarbeitet, wird hier

präferiert (vgl. Gamlem & Smith, 2013; Jonsson, 2012; Pokorny & Pickford, 2010). Folgt man diesen Annahmen, so bietet – zumindest in der Wahrnehmung der Schüler*innen – das Feedback von *FelloFish* genau diese zentralen Komponenten. Dies steht im Widerspruch zur abschließenden Präferenz der Klasse 7 für das Peerfeedback. Um die Entscheidungsgründe der Schüler*innen vollständig nachvollziehen zu können, ist es daher notwendig, neben dem oben thematisierten Lob und der Qualität und Passung der Feedbacks auch die in ihnen identifizierten Problemstellen zu berücksichtigen.

*Tab. 5 Feedbackkomponenten zu Problemstellen im Feedback in ihrer Verteilung und Häufigkeit (Nennungen: **** (>40%), *** (40–20 %), ** (19–10 %), * (<10%))*

Feedbackkomponente	Klasse 7		Klasse 6	
	Peer-feedback	<i>FelloFish</i>-Feedback	Peer-feedback	<i>FelloFish</i>-Feedback
<i>Teile im Feedback fehlen oder sind zu knapp</i>	***		**	
<i>Eine Überarbeitung ist mit dem Feedback nicht möglich</i>	*		**	
<i>Das Feedback hat Fehler</i>	*	****	*	***
<i>Das Feedback greift nur auf der Oberfläche</i>	*			

Sowohl in Klasse 6 als auch in Klasse 7 weisen die Schüler*innen darauf hin, dass im Peerfeedback teilweise Inhalte fehlen oder zu knapp formuliert sind. Zudem geben einige Schüler*innen beider Klassen an, dass eine Überarbeitung auf Grundlage des Peerfeedbacks nicht möglich sei.

Den größten Anteil in den identifizierten Problemstellen nehmen jedoch die Schüler*innenaussagen ein, die auf Fehler im Feedback hinweisen. Zwar werden dabei auch im Peerfeedback kleinere Fehler identifiziert, jedoch ist die Anzahl der Fehlerfunde im KI-Feedback deutlich höher. Insgesamt stellen sie in der Feedbackrezeption der Klasse 7 mit über 40% die am häufigsten genannte Feedbackkomponente dar.

Gerade in der Wahrnehmung dieser Klasse scheinen die Fehlerfunde alle weiteren Feedbackeigenschaften zu überstrahlen. Erkennt ein*e Siebtklässler*in Fehler im Feedback der KI, lassen sie deutliche Überschneidungen mit der finalen Feedbackpräferenz erkennen: Hier geben von den zwölf Schüler*innen, die im Feedback von *FelloFish* Fehler gefunden haben, elf an, das Feedback der Peers zu bevorzugen. Ein großer Teil der Schüler*innen identifiziert Fehlerstellen dabei bereits im ersten Feedbackkriterium – hier halluziniert die KI. So schreibt 7_AFGJ1: »Bei dem Tatt habe ich die Textart in meinem Text beschrieben. Und das Thema wird ja eigentlich auch durch den Einleitungssatz klar.«

Je früher die Schüler*innen Fehler im Feedback finden, desto länger begleiten diese sie in ihrer Feedbackrezeption. Tauchen Fehlerstellen bereits zu Beginn des Feedbacks auf, so kann dies folglich den restlichen Feedbackindruck färben. Der Fund von Fehlern geht dabei in Klasse 7 mit einer mangelnden Akzeptanz des Feedbacks und einem Vertrauensverlust einher: Fehler müssen hier zwar nicht zwingend zu einem Ausschluss des Feedbacks führen, können jedoch zumindest das Vertrauen in dieses hemmen, was wiederum Konsequenzen für dessen Rezeption hat (vgl. Winstone et al., 2017).

In der Rezeption der Klasse 6 zeichnet sich ein gegensätzliches Bild im Hinblick auf die Fehlerfunde ab. Diese werden auf der einen Seite von Klasse 6 nicht so häufig hervorgehoben wie von Klasse 7 (vgl. Tab. 5)⁴². Gleichzeitig gehen die Erwähnungen aber auch mit einer weniger negativen Haltung bei zumindest einigen der Schüler*innen einher: In Klasse 6 geben nur zwei der fünf Schüler*innen, die einen Fehler im *FelloFish*-Feedback gefunden haben, am Ende an, das Peerfeedback zu bevorzugen.

Die Schüler*innen der Klasse 6 zeigen hier ein differenzierteres Verständnis. Einen Einblick hierzu gibt beispielsweise 6_IFIN2 nach der Auflösung. Die Fehler im Feedback werden hier als solche erkannt, ändern jedoch nichts am Nützlichkeitsempfinden:

42 Eine mögliche Ursache hierfür kann selbstverständlich auch die unterschiedliche Aufgabenumgebung sein.

Sich ändert nichts an meiner Wahrnehmung da ich finde das das feedback von ai richtig ist. Es hat ein gutes, vollständiges und richtiges Feedback gegeben was auch stimmt. Es hat ein paar Fehler gemacht (nicht alle) und hat genannt was man nächstes mal besser machen kann. (6_IFIN2)

Es ist anzunehmen, dass die unterschiedliche Präferenz seitens der Feedbackformen auch an deren emotionale Rezeption gekoppelt ist. Das positive Empfinden des Lobs im Peerfeedback und das negative Empfinden von Fehlerstellen im *FelloFish*-Feedback können mit einer Präferenz des Peerfeedbacks einhergehen, während negatives oder kritisches Empfinden von Lob im Peerfeedback oder ein differenzierterer Blick auf Fehlerstellen im KI-Feedback zu einer Präferenz des Feedbacks durch *FelloFish* führen können.

Bestätigung findet die Annahme der Bedeutsamkeit negativer emotionaler Rezeption auch in den bereits angesprochenen quantitativen Ergebnissen der Affektwirkung. Sowohl in Klasse 6 als auch in Klasse 7 gaben die Schüler*innen positivere Emotionen als Reaktion auf das Peerfeedback an. Signifikante Unterschiede hinsichtlich der negativen Affektwirkung (*verärgert* und *beleidigt*) ließen sich jedoch nur bei Klasse 7 im KI-Feedback feststellen. Auch die quantitativen Ausprägungen der Akzeptanz und Fairness der jeweiligen Feedbackformen fielen im Bereich KI-Feedback der Klasse 7 negativer aus – auch, wenn kein Item der Akzeptanz signifikante Unterschiede aufzeigte. Diese Beobachtungen verweisen auf die Ansprüche (3) und (4) des Kapitels 2.4: Damit Lernende ein Feedback umsetzen möchten, muss es ihnen nicht nur als nützlich erscheinen, sie müssen sich auch im Feedbackkontext und in der Feedbackbotschaft selbst wohl und gerecht behandelt fühlen (vgl. Jonsson, 2012, Lipnevich et al., 2016).

Auch die Reaktionen nach der Auflösung verstärken die hier formulierten Annahmen (F3). Anders als angenommen, wechseln die Schüler*innen nicht in ihrer Präferenz zum *FelloFish*-Feedback, nachdem ihnen mitgeteilt wurde, dass es sich um das Feedback durch eine KI handelt. Während hier nur ein Teil der Schüler*innen Vertrauen in jeweils Mensch oder KI als Rechtfertigungsgrund anführt, verweist ein großer Teil der Schüler*innen zurück auf die Feedbackkomponenten. Besonders in Klasse 7 wird erneut häufig auf die Fehler im KI-Feedback hingewiesen. Gleichzeitig erwähnen mehrere

Schüler*innen, dass die Verbesserungsvorschläge im KI-Feedback für sie ein entscheidendes Kriterium seien, dieses weiterhin zu präferieren.

Für die Schüler*innen, die die Feedbackkomponenten ins Zentrum ihrer Begründung stellen, scheint die Quelle des Feedbacks keine Auswirkungen auf ihre Wahrnehmung zu haben. Es gibt jedoch auch Schüler*innen, die angeben, bereits eine Vorahnung im Hinblick auf die Feedbackquelle gehabt zu haben, und führen dies als Rechtfertigungsgrund für ihre finale Wahl an⁴³. Der Hypothese aus F3 ist somit nicht vollständig zu widersprechen, da anzunehmen ist, dass die Vorahnung der Schüler*innen zumindest zu Teilen ihre Feedbackpräferenz beeinflusst hat.

3.6 Limitationen

Die aufgeführten Befunde stellen erste indikative Ergebnisse für die Rezeption von Peer- und *FelloFish*-Feedback dar. Eine Generalisierung ist dabei u. a. aufgrund der kleinen Stichprobengröße nicht möglich. Zudem wird in der Studie eher eine Momentaufnahme abgebildet. Dies liegt primär an dem allen KI-bezogenen Studien inhärenten Problem: Ergebnisse veralten schnell, KI entwickelt sich weiter (vgl. auch Schindler, 2024, S. 35). Und so stagniert auch das hier eingesetzte Feedbacktool *FelloFish* nicht – wie alleine an seiner Namensänderung ersichtlich wurde – und ermöglicht bereits jetzt neue Funktionen, die zum Zeitpunkt der ersten Erhebung noch nicht möglich waren.

Daneben wirkt die angewandte Methodik in der Studiendurchführung zu Teilen limitierend. In der Studie wurden Schüler*innen rein schriftlich befragt. Sie lasen die Fragen und notierten daraufhin ihre Antworten. Dies kann dazu geführt haben, dass Schüler*innen in ihren Antworten weniger ausführlich waren, um den Fragebogen schneller beenden zu können (vgl. *Satisficing*). Auf der anderen Seite wurden individuelle Faktoren der Schüler*innen nicht berücksichtigt. So wurden eventuelle positive oder negative Vorerfahrungen mit Feedback nicht erhoben (vgl. Lipnevich et al., 2016), aber auch schreibprozessbezogene Kompetenzen der einzelnen

43 Auch dies geschieht häufig in Kombination mit bereits identifizierten Feedbackkomponenten.

Schüler*innen nicht erfasst. Durch Kenntnis über diese individuellen Faktoren hätten die Ergebnisse weiter zugespitzt und spezifiziert werden können.

Die Schüler*innen erhielten in der Studie auf ihren ersten Textentwurf jeweils ein Feedback pro Feedbackquelle. Um zwischen den Feedbacks eine Vergleichbarkeit zu ermöglichen, wurden beide in Kriterienrastern angelegt und in ausgedruckter Form weitergegeben. Auch, wenn der Inhalt der jeweiligen Feedbacks bestehen blieb, wurde der in den beiden Feedbackformen eigentlich verankerte Kontext verändert und die Feedbackformen konnten nicht in ihrer vollen Leistungskraft wirken: Zu Peerfeedback gehört neben dem Tausch der Rückmeldungen auch der Austausch über diese. Schüler*innen stellen Rückfragen, kommen interaktiv über ihr eigenes Textprodukt ins Gespräch (vgl. Sturm, 2022). Dem gegenüber ermöglicht *Fello-Fish* als Feedbacktool Feedbackschleifen. Das überarbeitete Textprodukt kann hier wieder in das KI-Tool eingefüttert werden, um erneutes Feedback zu generieren.

Sowohl die Interaktionen als auch die Feedbackschleifen wurden in der Studie stark eingeschränkt. Dies ermöglichte zwar einen Fokus auf die konkreten Elemente in der Feedbackbotschaft. Es darf jedoch nicht vernachlässigt werden, dass durch den Verlust des eigentlichen Kontextes auch die Interpretation der Feedbackbotschaft unter Umständen beeinflusst wurde.

Über diese Limitationen hinweg bietet die hier durchgeföhrte Studie jedoch wichtige und richtungsweisende Erkenntnisse, die sowohl an die Grundlagen in der Feedbackforschung anknüpfen als auch Einblicke in das Potenzial von KI-basiertem Feedback geben. In zukünftiger Forschung kann so unter Beachtung der angeführten Limitationen an die hier aufgeführten Ergebnisse angeknüpft werden.

4 Fazit

Textbezogenes Feedback ist im Schreibunterricht – aber auch darüber hinaus – zentral. Nur durch geeigneten Input können Schüler*innen ihre Schreibfähigkeiten am eigenen Text erweitern und in der Entwicklung ihrer Überarbeitungskompetenz gestützt werden. Während in der Forschung vor allem das Feedback als solches untersucht und verschiedene Formen von ihm besprochen werden, kommen die eigentlichen Rezipient*innen des Feedbacks häufig zu kurz. Dabei sind gerade die Rezipierenden als Schaltstelle zwischen Feedbackbotschaft und -umsetzung zentral. Dies gilt sowohl für traditionellere, analoge Formen des Feedbacks wie das Peerfeedback als auch für technologische Neuerungen wie das Feedback durch die KI. Ziel dieser Masterarbeit war es daher, die Feedbackrezeption von KI-basiertem Feedback (*FelloFish*) und Peerfeedback vergleichend bei Schüler*innen der Sekundarstufe I (Klasse 6 und 7) zu untersuchen.

Wie die dargestellten Ergebnisse und die anschließende Diskussion gezeigt haben, werden Peer- und KI-Feedback von den Schüler*innen unterschiedlich wahrgenommen. Ersteres akzeptieren die Schüler*innen dabei eher, sie empfinden es als fairer und zeigen positivere Affektwirkungen. Im Hinblick auf die identifizierten Feedbackkomponenten wird hier vermehrt Lob im Feedback hervorgehoben. Beim KI-Feedback sind es Verbesserungsvorschläge, Tipps oder Beispiele und eine ausführliche Darstellung – aber auch Fehler im Feedback werden hier vermehrt gefunden.

Ein Vergleich unter den Klassen hat gezeigt, dass ein Fehlerfund andere mögliche Feedbackkomponenten überstrahlen kann, dies aber zwingend nicht der Fall sein muss. So kann ein reflektierter Umgang mit dem Feedback auch dazu führen, trotz Fehlerstellen beispielsweise die Nützlichkeit der Rückmeldung anzuerkennen. Der Einsatz von Feedbacktools wie *FelloFish*, denen mögliches Halluzinieren inhärent ist, sollte also im Vorhinein mit den Schüler*innen besprochen und gerahmt werden, denn: Über die Fehlerstellen hinaus bietet das KI-Tool (auch im Vergleich mit dem Feedback durch die Peers) deutliche Potenziale, die auch von den Schüler*innen erkannt und als solche verstanden werden.

Die Studie öffnet ein kleines Blickfenster in den komplexen Feedbackprozess: Sie setzt nach der Feedbackbotschaft und vor der -umsetzung an. In künftiger Forschung könnte dieses Fenster erweitert werden, indem neben der Sicht der Schüler*innen auch die Ausgangstexte und die Feedbackbotschaften fokussiert werden. Auch eine Untersuchung der tatsächlichen Überarbeitung durch die Schüler*innen ist sinnvoll. So schreibt 6_OFSF3 in Bezug auf das erhaltene Peerfeedback: »Ich finde wichtig bei der Überarbeitung das da viel schon gesagt wurde als Verbesserung das ich nicht mehr so viel machen musste.«

Es erscheint daher plausibel, dass auch in diesem Fall zwischen unterschiedlichen Formen der Übernahme des gegebenen Outputs differenziert werden kann. Ähnlich wie bei Steinhoff (2025) könnte eine Feedbackbotschaft so als *Ghostinstructor* verstanden werden und das eigene Denken abnehmen. Gleichzeitig ist aber auch ein interaktiver oder kritisch-reflektierter Umgang mit den im Feedback angegebenen *Feed-Forwards* denkbar. Beide Feedbacks haben im Rahmen der Studie wenig bis keine Interaktionen und Adaptivität ermöglicht. Auch diese Punkte sollten in zukünftigen Untersuchungsvorhaben inkludiert werden.

5 Literaturverzeichnis

- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, Hendrik (2024). Feedback Sources in Essay Writing: Peer-generated or AI-generated Feedback? *International Journal of Educational Technology in Higher Education*, 21, 1–15. <https://doi.org/10.1186/s41239-024-00455-4>
- Baron, N. S. (2023, 19. Januar). *How ChatGPT robs students of motivation to write and think for themselves*. The Conversation <https://theconversation.com/how-chatgpt-robs-students-of-motivation-to-write-and-think-for-themselves-197875>
- Becker-Mrotzek, M., & Böttcher, I. (2015). *Schreibkompetenz entwickeln und beurteilen*. Cornelsen.
- Berger, S., Vejvoda, J., Schultz-Pernice, F., Li, W., Sailer, M., & Fischer, F. (2024). *Digitale Bildung an bayerischen Schulen zwischen Pandemie und KI*. vbw. https://www.vbw-bayern.de/Redaktion/Frei-zugaengliche-Medien/Abteilungen-GS/Bildung/2024/Downloads/vbw_Studie_Digitale-Bildung-III_2024_final.pdf
- Berkemeier, A. (2010). Das Schreiben von Sachtextzusammenfassungen lernen, lehren und testen. In T. Pohl & T. Steinhoff (Hrsg.), *Textformen als Lernformen* (S. 211–232). Gilles & Francke.
- Blatt, M., Henniger, H., & Zdrallek, A. (Hrsg.). (2020). *Deutsch kompetent 7. Ausgabe Nordrhein-Westfalen Gymnasium (G9) ab 2019*. Ernst Klett Verlag.
- Borgers, N., Hox, J., & Sikkel, D. (2004). Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. *Quality & Quantity*, 38, 17–33. <https://doi.org/10.1023/B:QUQU.0000013236.29205.a6>
- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, John, & Neill, S. (2021). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education*, 105, S. 1–10. <https://doi.org/10.1016/j.tate.2021.103387>
- Decker, L., Guschner, B., Hensel, S., & Schindler, K. (2021). *Wissenschaftliches Schreiben lernen in der Sekundarstufe II. Fachdebatte und Praxisprojekte*. wbv Publikation.

- Ditton, H., & Müller, A. (2014). Feedback. Begriff, Formen und Funktionen. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 11–28). Waxmann.
- Ernst Klett Verlag GmbH (2020). *Deutsch kompetent 6 - Lehrerausgabe Zusatzmaterial*. Klett. <https://www.klett.de/inhalt/lehrwerk-online/126655>
- Falck, J. (2024, 29. Februar). *Fiete.ai in der Deutsch-Schulaufgabe*. Blog für Schulentwicklung. Digitalisierung, Fortbildung, Unterricht. <https://joschafalck.de/fiete-in-der-deutsch-schulaufgabe/>
- FelloFish GmbH (2025). *FelloFish. Feedback für alle*. <https://www.felofish.com/>
- Fleckenstein, J., Reble, R., Meyer, J., Jansen, T., Liebenow, L. W., Möller, J., & Köller, O. (2023). Digitale Schreibförderung im Bildungskontext. Ein systematisches Review. In K. Scheiter & I. Gogolin (Hrsg.), *Bildung für eine digitale Zukunft* (S. 3–25). Springer VS.
- Gamble, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150–169. <http://dx.doi.org/10.1080/0969594X.2012.749212>
- Gierlich, H. (2005). Sachtexte als Gegenstand des Deutschunterrichts. Einige grundsätzliche Überlegungen. In M. Fix & R. Jost (Hrsg.), *Sachtexte im Deutschunterricht. Für Karlheinz Fingerhut zum 65. Geburtstag* (S. 25–46). Schneider-Verl. Hohengehren.
- Graham, S. (2018). Instructional Feedback in Writing. In A. A. Lipnevich & J. K. Smith (Hrsg.), *The Cambridge Handbook of Instructional Feedback* (S. 145–168). Cambridge University Press.
- Graham, S., Harris, K., & Hebert, M. (2011). *Informing Writing. The Benefits of Formative Assessment. A Report from Carnegie Corporation of New York*. Alliance for Excellent Education.
- Gredel, E., Pospiech, U., & Schindler, K. (2024). Künstliche Intelligenz und Schreiben in (hoch-)schulischen Kontexten. *Zeitschrift für germanistische Linguistik*, 52(2), 378–404. <https://doi.org/10.1515/zgl-2024-2018>
- Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education*, 63, 1–15. <https://doi.org/10.1016/j.iheduc.2024.100962>
- Haag, L., & Götz, T. (2019). Was wissen wir über Feedback? Synopse des aktuellen Forschungsstandes. *Friedrich Jahresheft »Feedback«*, S. 14–17.
- Hattie, J. (2015). *Lernen sichtbar machen*. Schneider Verl. Hohengehren.

- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
<https://doi.org/10.3102/003465430298487>
- Haverkamp, H., Hecht, M., & Schindler, K. (2024). Lernförderliches Feedback KI-basiert vermittelt. Erfahrungen mit der Lernumgebung Fiete. *Der Deutschunterricht*, 5, 60–71.
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, 29(3), 369–388. <https://doi.org/10.1177/0741088312451260>
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Hrsg.), *Cognitive Processes in Writing* (S. 3–30). Rouledge.
- Jäger, R. S. (2019). Diagnostik, Evaluation und Feedback. In der Schule zusammenfügen, was zusammengehört. *Friedrich Jahresheft »Feedback«*, 10–11.
- Jansen, T., Höft, L., Bahr, L., Fleckenstein, J., Möller, J., Köller, O., & Meyer, J. (2024). Comparing Generative AI and Expert Feedback to Students' Writing: Insights from Student Teachers. *Psychologie in Erziehung und Unterricht*, 71(2), 80–92. <https://doi.org/10.2378/peu2024.art08d>
- Jantzen, C. (2003). Eigene Texte in der Schule überarbeiten. beobachten - verstehen - lernen. In: E. Brinkmann, N. Kruse, & C. Osburg (Hrsg.), *Kinder schreiben und lesen. Beobachten - verstehen - lehren* (S. 111–126) Fillibach Verlag.
- Jonsson, A. (2012). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63–76.
<https://doi.org/10.1177/1469787412467125>
- Jünger, A.-M., & Escher, K. (2023). Sachtexte wiederzugeben ist gar nicht so leicht! Lernende gezielt mit dem Beratungsinstrument Code-Knacker bei der Sachtextwiedergabe unterstützen. *ide. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule*, 47(2), 61–69.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 1–9. <https://doi.org/10.1016/j.lindif.2023.102274>

- Katannek, S., & Suñer, F. (2024). Selbstreguliertes Lernen mit Feedback von ChatGPT. Die Rolle individueller Einstellungen und Wahrnehmungen beim selbstregulierten Lernen mit KI-basiertem Feedback. *KONTEXTE: Internationales Journal zur Professionalisierung in Deutsch als Fremdsprache* 2(2), 14–32. <https://doi.org/10.24403/jp.1394575>
- Kirschbaum, A., & Rothärmel, A. (2017). Die inhaltlich strukturierende qualitative Inhaltsanalyse nach Kuckartz. Dargestellt am Beispiel des Rollenverständnisses von Writing Fellows. In M. Brinkschulte & D. Kreitz (Hrsg.), *Qualitative Methoden in der Schreibforschung* (S. 227–248). W. Bertelsmann Verlag.
- 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.
- Kuckartz, U., & Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Beltz Juventa.
- Kultusministerkonferenz (KMK) (2022, 23. Juni). *Bildungsstandards für das Fach Deutsch. Erster Schulabschluss (ESA) und Mittlerer Schulabschluss (MSA). Beschluss der Kultusministerkonferenz vom 15.10.2004 und vom 04.12.2003, i.d.F. vom 23.06.2022*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
- https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-ESA-MSA-Deutsch.pdf
- Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education: Artificial Intelligence*, 6, 1–10.
- Lehnen, K (2023). Peerfeedback beim schulischen Schreiben. Grundlagen, Methoden, Modellierung. *ide. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule* 47(2), 18–30.
- Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a Model of Student Response to Feedback. In G. T. L. Brown & L. R. Harris (Hrsg.), *Handbook of Human and Social Conditions in Assessment* (S. 169–185). Routledge.
- Medienpädagogischer Forschungsverbund Südwest (2023). *Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Mpfs.
- https://www.mpfs.de/fileadmin/files/Studien/JIM/2022/JIM_2023_web_final_kor.pdf

Medienpädagogischer Forschungsverbund Südwest (2024). *Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Mfps.

https://mpfs.de/app/uploads/2024/11/JIM_2024_PDF_barrierearm.pdf

Ministerium für Bildung und Kindertagesförderung Mecklenburg-Vorpommern (o. D.). Ausschreibung »Pilotvorhaben Feedbacktool«, Ministerium für Bildung und Kindertagesförderung. Mecklenburg-Vorpommern. <https://www.bildung-mv.de/export/sites/bildungsserver/galleries/dokumente/fobi/Ausschreibung-fellofish.pdf>

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2019). *Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen. Deutsch*. https://www.schulentwicklung.nrw.de/lehrplene/lehrplan/196/g9_d_klp %203409_2019_06_23.pdf

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2023). *Umgang mit textgenerierenden KI-Systemen. Ein Handlungsleitfaden*. https://www.schulministerium.nrw/system/files/media/document/file/handlungsleitfaden_ki_msbt_nrw_230223.pdf

Mühlhoff, R., & Henningsen, M. (2024). Chatbots im Schulunterricht. Wir testen das Fobizz-Tool zur automatischen Bewertung von Hausaufgaben. *arXiv*, 1–33. <https://doi.org/10.48550/arXiv.2412.06651>

Neumann, K., Kuhn, J., & Drachsler, H. (2024). Generative Künstliche Intelligenz in Unterricht und Unterrichtsforschung – Chancen und Herausforderungen. *Unterrichtswissenschaft*, 52(2), 227–237. <https://doi.org/10.1007/s42010-024-00212-6>

News4teachers (2025, 20. Februar): »Feedback jederzeit und zu jedem Schüler«. Wie Künstliche Intelligenz Lehrkräften konkret beim Unterrichten helfen kann. News4teachers. Das Bildungsmagazin. <https://www.news4teachers.de/2025/02/feedback-jederzeit-und-zu-jedem-schueler-wie-kuenstliche-intelligenz-lehrkraeften-konkret-beim-unterrichten-helfen-kann/>

Omrani, A., Wakefield-Scurr, J., Smith, J., & Brown, N. (2019). Survey Development for Adolescents Aged 11–16 Years: A Developmental Science Based Guide. *Adolescent Research Review*, 4(4), 329–340. <https://doi.org/10.1007/s40894-018-0089-0>

Ortner, H. (2006). Spontanschreiben und elaboriertes Schreiben. Wenn die ursprüngliche Lösung zu einem Teil des Problems wird. In W. Kissling & G. Perko (Hrsg.), *Wissenschaftliches Schreiben in der Hochschullehre. Reflexionen, Desiderate, Konzepte* (S. 77–101). StudienVerlag.

- Peter, K. (2023). Blinde Flecken des Peerfeedbacks im Schreibunterricht. *ide. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule*, 47(2), 89–97.
- Philipp, M. (2015). *Schreibkompetenz. Komponenten, Sozialisation und Förderung*. A. Francke Verlag.
- Philipp, M. (2023). Formatives Feedback aus der Sicht des selbstregulierten Lernens. Grundlagen und Grundsätze förderlicher Rückmeldungen. *ide. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule*, 47(2), 8–17.
- Pokorny, H., & Pickford, P. (2010). Complexity, cues and relationships: Student perceptions of feedback. *Active Learning in Higher Education*, 11(1), 21–30. <https://doi.org/10.1177/1469787409355872>
- Pölert, Hauke (2023, 4. September). *Fiete.ai – Passgenaues, lernförderliches KI-Feedback mit GPT / Sprachmodell*. Unterrichten. Digital. Lernen mit digitalen Medien. <https://unterrichten.digital/2023/09/04/fiete-ai-feedback-chatgpt/>
- Ray, P. P. (2023). ChatGPT. A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Rezat, S. (2022). Texte im Deutschunterricht. In A. Franken, A. U. & Pertzel, Eva (Hrsg.), *12 Perspektiven auf den Deutschunterricht. Wissenswertes für Lehrkräfte* (S. 177–214). Jambus. QUA-LiS NRW.
- Richert, P. (2006). Lehrerfeedback. Kommunikation im Unterricht. *Lernchancen* 9(54), 46–49.
- Rupe, T., & Mayweg-Paus, E. (2023). «Your argumentation is good», says the AI vs humans – The role of feedback providers and personalised language for feedback effectiveness. *Computers and Education: Artificial Intelligence*, 5, 1–14. <https://doi.org/10.1016/j.caeari.2023.100189>
- Saldan, P. (2025, 31. März). *Schützt Jugendliche vor der Manipulation durch KI!* Die Zeit. <https://www.zeit.de/2025/13/kuenstliche-intelligenz-schule-neutral-jugendliche-fakten>
- Schindler, K. (2024). Schreiben mit, durch und über KI. Herausforderungen und Chancen für das Schreiben in der Schule. *ide. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule*, 48(2), 32–41.
- Schulz von Thun, F. (2010). *Miteinander reden. 1. Störungen und Klärungen. Allgemeine Psychologie der Kommunikation*. Rowohlt.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research* 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>

- Stanat, P., Sturm, A., & Becker-Mrotzek, M. (2019). Schreiben – gezielt fördern. *BiSS-Journal*, 10, 5–7. <https://www.biss-sprachbildung.de/wp-content/uploads/2019/11/biss-journal-10-juni-2019.pdf>
- Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) (Hrsg.) (2024). *Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz*. SWK. <https://doi.org/10.25656/01:28303>
- Steinhoff, T. (2023a). Der Computer schreibt (mit). Digitales Schreiben mit Word, Whatsapp, ChatGPT & Co. als Koaktivität von Mensch und Maschine. *MIDU - Medien im Deutschunterricht* 5(1), 1–16. <https://doi.org/10.18716/ojs/midu/2023.1.4>
- Steinhoff, T. (2023b). Literalität oder Digitalität? Sowohl als auch! Überlegungen zu einer postdigitalen Deutschdidaktik am Beispiel des Lese- und Schreibens unter besonderer Berücksichtigung Künstlicher Intelligenz. *leseforum.ch*, 3, 1–14. <https://doi.org/10.58098/lffi/2023/3/799>
- Steinhoff, Torsten (2025). Künstliche Intelligenz als Ghostwriter, Writing Tutor und Writing Partner. Zur Modellierung und Förderung von Schreibkompetenzen im Zeichen der Automatisierung und Hybridisierung der Kommunikation am Beispiel des Schreibens mit ChatGPT in der 8. Klasse. In C. Albrecht, J. Brüggemann, T. Kretschmann, & C. Meier (Hrsg.), *Personale und funktionale Bildung im Deutschunterricht. Theoretische, empirische und praxisbezogene Perspektiven* (S. 85–99). J.B. Metzler.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 1–15. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Streber, D. (2019). Peerfeedback zur Weiterentwicklung des Unterrichts. Schülerinnen und Schüler geben sich gegenseitig Rückmeldung. *Friedrich Jahresheft »Feedback«*, 82–83.
- Strijbos, J.-W., & Müller, A. (2014). Personale Faktoren im Feedbackprozess. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 83–134). Waxmann.

- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291–303. <https://doi.org/10.1016/j.learninstruc.2009.08.008>
- Strijbos, J.-W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, 68, 1–12. <https://doi.org/10.1016/j.stueduc.2021.100980>
- Sturm, A. (2022). Schreiben für Anfängerinnen und Anfänger. In A. Franken, A. U./Pertzel, Eva (Hrsg.), *12 Perspektiven auf den Deutschunterricht. Wissenswertes für Lehrkräfte* (S. 177–214). Jambus. QUA-LiS NRW.
- Underwood, J. S., & Tregidgo, A. P. (2006). Improving Student Writing Through Effective Feedback: Best Practices and Recommendations. *Journal of Teaching Writing*, 22(2), 73–98.
- unterrichten.digital-tutor125 (2024, 3. November): *FieteAI - KI-Feedback für Schülertexte - Funktionen und Praxisbeispiele* [Video]. Youtube. <https://www.youtube.com/watch?v=757RG5X7hII>
- van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education*, 85, 175–189. <https://doi.org/10.1016/j.tate.2019.06.010>
- Wampfler, P. (2023). Schreiben mit KI-Tools. Digital unterstützte Schreibprozesse gestalten und begleiten. *leseforum.ch*, 3, 1–8. <https://doi.org/10.58098/lffi/2023/3/806>
- Wampfler, P. (2024). Feedback und Schreibunterstützung durch Sprach-KI. *Pädagogik*, 3, 15–19.
- Wendt, C. (2023). Schreiben lernen mit intelligenter Hilfe. Wie computergestütztes Feedback Schreiblernprozesse verändern kann. *ide. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule* 47(2), 38–47.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes. *Educational Psychologist* 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>
- Wisniewski, B., Zierer, K., & Hattie, J. (2019). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, 10, 1–14. <https://doi.org/10.3389/fpsyg.2019.03087>

Wrede, S. E., Gloerfeld, C., & de Witt, C. (2023). KI und Didaktik – Zur Qualität von Feedback durch Recommendersysteme. C. de Witt, C. Gloerfeld, & S. E. Wrede (Hrsg.), *Künstliche Intelligenz in der Bildung* (S. 133–154). Springer Fachmedien Wiesbaden.
https://doi.org/10.1007/978-3-658-40079-8_7

Zierer, K. (2019). Erfolgreiches Feedback ist nicht einfach, aber es ist wirkmächtig. Klaus Zierer im Interview mit John Hattie. *Friedrich Jahrestheft »Feedback«*, 6–9.

6 Anhang

Anhang 1: Arbeitsmaterialien für die Schüler*innen	133
Anhang 1.1: Arbeitsblatt »Geckofuß« (Klasse 7)	133
Anhang 1.2: Arbeitsblatt »Rollschuhe« (Klasse 6)	135
Anhang 1.3: Feedbackbogen	136
Anhang 1.3.1: Feedbackbogen leer (exemplarisch für Kasse 7)	136
Anhang 1.3.2: Feedbackbogen ausgefüllt (Peer) (exemplarisch; ELNA_1)	139
Anhang 1.3.3: Feedbackbogen ausgefüllt (FelloFish) (exemplarisch; ELNA_1)	140
Anhang 1.4: Fragebogen – Teil 1	142
Anhang 1.5: Fragebogen – Teil 2	145
Anhang 2: Daten	146
Anhang 2.1: Klasse 7	146
Anhang 2.1.1: Quantitative Daten	146
Anhang 2.1.2: Qualitative Daten	149
Anhang 2.2: Klasse 6	156
Anhang 2.2.1: Quantitative Daten	156
Anhang 2.2.2: Qualitative Daten	160
Anhang 3: Codierleitfaden	163

Anhang 1: Arbeitsmaterialien für die Schüler*innen

Anhang 1.1: Arbeitsblatt »Geckofuß« (Klasse 7)

Sigrid Belzec **Mit besonderer Haftung: Der Geckofuß** (2010. Ausschnitt)

Der Gecko ist ein kleines Reptil, das in den warmen Gegenden unserer Erde lebt und ein hervorragender Insektenjäger ist. Es gibt viele verschiedene Geckoarten, sie sind zwischen vier und vierzig Zentimeter groß. Die größeren Arten werden bis zu 300 Gramm schwer, das ist mehr als ein Päckchen Butter! Das Spannende an diesen Tieren ist, dass auch die großen und schweren Arten auf allen möglichen Oberflächen laufen können, ohne abzurutschen. Um seine Beute zu verfolgen, kann der Gecko sogar über Kopf rennen! Und obwohl er dafür mit seinen Füßen fest an der Wand haften muss, kann er seine Füße auch blitzschnell wieder ablösen – sonst könnte er ja nicht weiterlaufen. So einen schnellen und präzisen Wechsel zwischen Haften und Lösen kann derzeit kein Klebstoff leisten, den wir Menschen entwickelt haben.

Der Gecko kann das, weil seine Fußsohlen eine extrem große Oberfläche haben. Jetzt denkst du vielleicht, dass so ein Gecko riesig große Füße hat, aber das stimmt nicht. Stattdessen haben die Geckos an ihren langen flachen Zehen viele kleine Läppchen, die man sogar mit bloßem Auge erkennen kann. Diese sogenannten Haftlamellen sehen flauschig aus und fühlen sich auch so an. Das kommt daher, dass die Haftlamellen ganz weich und anschmiegsam sind. Doch der eigentliche Trick des Geckos wird erst sichtbar, wenn man seine Fußsohlen unter einem starken Mikroskop untersucht. Die Haftlamellen sind mit vielen kleinen Härchen besetzt, die sich am Ende in noch kleinere Härchen aufspalten, die an den Enden platt sind. Weil die Unterseite des Geckofüßes sehr fein aufgespalten ist, ist seine Oberfläche

viel größer, als wenn die Fußsohle glatt wäre. Diese „riesige“ Oberfläche kann an jeder beliebigen anderen Oberfläche haften. Verantwortlich dafür sind Van-der-Waals-Kräfte. [...]

Die Van-der-Waals-Kräfte sind nach dem niederländischen Physiker Johannes Diderik van der Waals benannt. Seine Erkenntnis: Oberflächen, die sehr, sehr nah beieinander sind, ziehen sich gegenseitig an. Je größer die Flächen sind, desto größer ist auch die Anziehungskraft zwischen ihnen. [...]

Der Geckofuß klebt also sozusagen ohne Klebstoff auf der Oberfläche! Beim Beutefang ist es natürlich wichtig, dass das Tier seinen Fuß auch schnell wieder von der Oberfläche lösen kann, über die es rennt. Das klappt nur, weil der Gecko bei jedem Schritt den Fuß erst leicht anhebt und dann von der Zehenspitze aus stückchenweise nach oben aufrollt. Das funktioniert so ähnlich, wie du das kennst, wenn du ein Pflaster entfernst: Du fängst an einer Ecke an und ziehst das Pflaster von dort aus ab. Übrigens: Ein Gecko, der mit einem Fuß an der glatten Wand hängt, könnte theoretisch das Gewicht zweier erwachsener Menschen halten, bevor sich sein Fuß lösen würde. Von dieser haftenden Fußoberfläche versucht man deshalb einiges zu lernen!

Amerikanische Forscher haben nach dem Vorbild des Geckofüßes ein Klebeband entwickelt, das eine sehr große Haftkraft hat und immer wieder verwendbar ist.



Quelle: Blatt et al. (2020: 22)

CODE: _____

Aufgabe:

Fasse den Text in eigenen Worten zusammen, indem du dich auf die **wichtigsten Aussagen** konzentrierst. Beachte dabei die Kriterien zum Schreiben einer Zusammenfassung:

- In der Einleitung der Zusammenfassung sollten **Titel, Autorin, Textart und Thema (TATT)** genannt werden.
- Im Hauptteil liegt der **Fokus auf den wichtigsten Aussagen** und jeder Abschnitt des Textes wird kurz zusammengefasst.
- Die Zusammenfassung sollte **knapp, informativ und gut verständlich** sein.
- Als Zeitform wird das **Präsens** (bei Vorzeitigkeit das Perfekt) verwendet.
- Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz - der Text bleibt **sachlich und neutral**.
- **Fachbegriffe und präzise Formulierungen** werden verwendet.
- Es werden geeignete **Satzverknüpfungen** und **Überleitungen** genutzt, um Zusammenhänge zu verdeutlichen (z. B. Ursachen, zeitliche Abläufe...).

Anhang 1.2: Arbeitsblatt »Rollschuhe« (Klasse 6)

CODE: _____

Anette Bullach: **Rollschuhe** (2016)

Der erste Auftritt mit Rollschuhen ging ganz schön schief, seitdem wurden die „rollenden Schuhe“ optimiert und sind sicherer.

Einer der ersten öffentlichen Auftritte mit Rollschuhen vor 250 Jahren endete mit einer kleinen Katastrophe: Eigentlich wollte der belgische Instrumentenbauer John Joseph Merlin mit seiner Violine auftreten und dabei mit selbstgebauten „rollenden Schuhen“ in einem Londoner Ballsaal einen guten Eindruck hinterlassen. Doch Merlin fuhr ungebremst in einen teuren Spiegel, der in tausend Teile zersprang. Pech für Merlin, dass an seinen Rollschuhen noch keine Stopper waren! Überhaupt erinnerten seine „rollenden Schuhe“ mit den hintereinander angeordneten Metallrollen eher an Schlittschuhe. Daher hatte er seine Idee: Er wollte auch im Sommer Schlittschuh fahren können. Sich mit den „rollenden Schuhen“ fortzubewegen war sehr anstrengend. Die Rollen waren aus Holz oder Metall. Außerdem

waren die Straßen früher noch nicht so gut wie die heutigen.

Erst die Röllschuhe mit zwei Rollen vorne und zwei hinten und mit Gummigelenken machten dieses neue Fortbewegungsmittel stabil und lenkbar. Erfunden wurden sie 1863 von James L. Plimpton in New York. Wenige Jahre später tobten sich die Röllschuh-Begeisterten auf den ersten amerikanischen Röllschuhbahnen aus. Pro Abend kamen bis zu 8000 BesucherInnen! Mit der Zeit wurden Röllschuhe immer weiter technisch verbessert: Sie bekamen Kugellager, Plastikrollen und einen Stopper vorne.

Das machte sie attraktiver. Doch andere Trends wie zum Beispiel Inline-Skates verdrängten die Röllschuhe fast vom Markt. So laufen heute fast nur noch die RollkunstläuferInnen auf klassischen Röllschuhen. Allerdings treten diese im Gegensatz zu Merlin nicht ohne Stopper auf. Diese brauchen sie für ihre gedrehten Sprünge und natürlich auch zum Bremsen – falls mal ein Spiegel in der Nähe stehen sollte ;-)

Quelle: <https://www.klett.de/inhalt/lehrwerk-online/126655>

Aufgabe:

Fasse den Text in eigenen Worten zusammen, indem du dich auf die **wichtigsten Aussagen** konzentrierst. Beachte dabei die Kriterien zum Schreiben einer Zusammenfassung:

- In der Einleitung der Zusammenfassung sollten **Titel, Autorin und Thema** genannt werden.
- Im Hauptteil liegt der **Fokus auf den wichtigsten Aussagen** und jeder Abschnitt des Textes wird kurz zusammengefasst.
- Die Zusammenfassung sollte **knapp, informativ und gut verständlich** sein.
- Als Zeitform wird das **Präsens** (bei Vorzeitigkeit das Perfekt) verwendet.
- Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz - der Text bleibt **sachlich und neutral**.
- **Fachbegriffe** und **präzise Formulierungen** werden verwendet.
- Es werden geeignete **Satzverknüpfungen** und **Überleitungen** genutzt, um Zusammenhänge zu verdeutlichen (z. B. Ursachen, zeitliche Abläufe...)

Anhang 1.3: Feedbackbogen

Anhang 1.3.1: Feedbackbogen leer (exemplarisch für Kasse 7)

- Feedbackbogen -	
Feedback für	(Code des Textes)
<p><u>Hinweis:</u> Euer Feedback sollte euren Mitschüler*innen dabei helfen, ihren Text zu verbessern. Versucht daher, das Feedback möglichst konkret zu formulieren. Bedenkt, dass jemand Mühe in den Text gesteckt hat und nun eure Unterstützung möchte. Lobt daher auch, was euch gefällt, damit die Person weiß, was schon gut läuft, bevor ihr Kritik äußert. Gebt schließlich ehrliches, höfliches Feedback, indem ihr offen und freundlich beschreibt, was euch stört oder unklar ist, und erklärt genau, warum.</p> <p>Schreibt in ganzen Sätzen!</p>	
Kriterium	Kommentar
In der Einleitung der Zusammenfassung sollten Titel, Autorin, Textart und Thema (TATT) genannt werden.	
Im Hauptteil liegt der Fokus auf den wichtigsten Aussagen und jeder Abschnitt des Textes wird kurz zusammengefasst.	

<p>Die Zusammenfassung sollte knapp, informativ und gut verständlich sein.</p>	
<p>Als Zeitform wird das Präsens (bei Vorzeitigkeit das Perfekt) verwendet.</p>	
<p>Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz – der Text bleibt sachlich und neutral.</p>	

Fachbegriffe und
präzise Formulie-
rungen werden
verwendet.

Es werden geeigne-
te **Satzverknüp-**
fungen und **Über-**
leitungen genutzt,
um Zusammenhän-
ge zu verdeutlichen
(z. B. Ursachen,
zeitliche Abläufe...).

Anhang 1.3.2: Feedbackbogen ausgefüllt (Peer) (exemplarisch; ELNA_1)

<p>Die Zusammenfassung sollte kurz, informativ und gut verständlich sein.</p>	<p>Die Zusammenfassung wurde gut verständlich und knapp geschrieben. Der Text ist verständlich und man hat eine gute Vorstellung von dem Geckofuß.</p>
<p>Als Zeitform wird das Präsenz (bei Vorzeitigkeit das Perfekt) verwendet.</p>	<p>Die Zusammenfassung wurde in der richtigen Zeitform geschrieben</p>
<p>Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz - der Text bleibt sachlich und neutral.</p>	<p>In dem Text wurden keine eigenen Meinungen genannt und er wurde sachlich und neutral formuliert.</p>

2

- Feedbackbogen -	
<p>Feedback für ELNA1</p>	<p>(Code des Textes)</p>
<p>Hinweis: Euer Feedback sollte euren Mitschüler*innen dabei helfen, ihren Text zu verbessern. Versucht daher, Feedback möglichst konkret zu formulieren. Bedenkt, dass jemand Mühe in den Text gesteckt hat und nun eure Unterstützung möchte. Lobt daher auch, was euch gefällt, damit die Person weiß, was schon gut läuft, bevor ihr Kritik äußert. Gebt schließlich ehrlieches, höfliches Feedback - indem ihr offen und freundlich beschreibt, was euch stört oder unklar ist, und erklärt genau, wewomit.</p> <p>Schreibt in ganzen Sätzen!</p>	<p>In der Einleitung wurden alleine Punkte beachtet und in den Text geschrieben. Die Formulierung der Einleitung ist auch sehr schön. Ich würde vielleicht zwischen Einleitung und Hauptteil eine Zeile freilassen, damit man Einleitung und Hauptteil besser unterscheiden kann.</p>
<p>Kriterium</p>	<p>Kommentar</p>
<p>In der Zusammenfassung sollten Titel, Autorin, Textart und Thema (ATT) benannt werden.</p>	<p>Im Hauptteil liegt der Fokus auf den wichtigsten Aussagen und jeder Abschnitt des Textes wird kurz zusammengefasst.</p>

1

Fachbegriffe und präzise Formulierungen werden verwendet.	In der Zusammenfassung wurden Fachbegriffe benutzt und richtige und abwechslungsreiche Formulierungen verwendet.
---	--

Anhang 1.3.3: Feedbackbogen ausgefüllt (FelloFish) (exemplarisch; ELNA_1)

- Feedbackbogen -	
Feedback für [ELNA1]	[Code des Textes]
<p>Hinweis: Euer Feedback sollte euren Mitschüler*innen dabei helfen, ihren Text zu verbessern. Versucht daher das Feedback möglichst konkret zu formulieren. Bedenkt, dass jemand Mühe in den Text gesteckt hat und nun eine Unterstützung möchte. Lohnt daher auch, dass euch Gefällt; damit die Person weiß, was schön, gut läuft, bevor ihr Kritik äußert. Gebt schließlich ehrliches, höfliches Feedback, indem ihr offen und freundlich beschreibt, was euch steht oder unklar ist, und erklärt genau, warum.</p> <p>Schreibt in ganzen Sätzen!</p>	
Kriterium	Kommentar
In der Einleitung der Zusammenfassung sollten Titel, Autorin, Textart und Thema (TATT) genannt werden.	<p>Du hast den Titel, die Autorin und das Thema des Textes gut benannt. Aber du hast die Textart vergessen. Schreibe zum Beispiel: Der Sachtext. Mit besonderer Haftung: Der Gecko Fuß von Sigrid Beitz aus dem Jahr 2010 handelt vom Gecko Fuß.</p>
In Hauptteil liegt der Fokus auf den wichtigsten Aussagen und jeder Abschnitt des Textes wird kurz zusammengefasst.	<p>Du hast die wichtigsten Aussagen gut zusammengefasst. Aber achte darauf, dass du alle wichtigen Details erwähnst. Zum Beispiel könnetest du genauer erklären, wie die Van-der-Waals-Kräfte funktionieren und warum sie wichtig sind.</p>

<p>Fachbegriffe und präzise Formulierungen werden verwendet.</p> <p>Du hast einige Fachbegriffe gut verwendet. Achte darauf, immer präzise zu formulieren. Zum Beispiel kannst du 'Haftmellen' durch 'Haftmellen' ersetzen und 'Fußfläche' durch 'Fußflächen'.</p>	<p>Die Zusammenfassung ist gut verständlich und informativ. Versuche aber, noch knapper zu schreiben. Zum Beispiel könnetest du einige Wiederholungen vermeiden und Sätze kürzer fassen.</p>	<p>Deine Zusammenfassung sollte knapp, informativ und gut verständlich sein.</p>
<p>Es werden geeignete Satzverknüpfungen und Überleitungen genutzt, um Zusammenhänge zu verdeutlichen (z. B. Ursachen, zeitliche Abläufe...).</p>	<p>Du hast einige gute Satzverknüpfungen verwendet. Versuche, noch mehr Überleitungen zu nutzen, um die Zusammenhänge klarer zu machen. Zum Beispiel könnetest du Wörter wie 'deshalb', 'außerdem' oder 'daher' verwenden.</p>	<p>Als Zeitform wird das Präsens (bei Vorzeltigkeit das Perfekt) verwendet.</p>
		<p>Eigene Meinungen oder Bewertungen haben in der Zusammenfassung keinen Platz - der Text heißt sachlich und neutral.</p>

Anhang 1.4: Fragebogen – Teil 1

Fragebogen Feedback

Feedback 1

	Stimme nicht zu	Stimme eher nicht zu	Neutral	Stimme eher zu	Stimme vollständig zu
Mit dem Feedback bin ich zufrieden.	<input type="checkbox"/>				
Ich betrachte dieses Feedback als fair.	<input type="checkbox"/>				
Ich betrachte dieses Feedback als gerechtfertigt.	<input type="checkbox"/>				
Ich betrachte dieses Feedback als nützlich.	<input type="checkbox"/>				
Ich finde dieses Feedback hilfreich.	<input type="checkbox"/>				
Dieses Feedback unterstützt mich.	<input type="checkbox"/>				
Dieses Feedback akzeptiere ich.	<input type="checkbox"/>				
Diesem Feedback widerspreche ich.	<input type="checkbox"/>				
Dieses Feedback lehne ich ab.	<input type="checkbox"/>				
Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	<input type="checkbox"/>				
Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:					
• Zufrieden	<input type="checkbox"/>				
• Zuversichtlich	<input type="checkbox"/>				
• Erfolgreich	<input type="checkbox"/>				
• Beleidigt	<input type="checkbox"/>				
• Verärgert	<input type="checkbox"/>				
• Frustriert	<input type="checkbox"/>				

Empfindest du das Feedback als hilfreich und verständlich formuliert? Welche Teile des Feedbacks empfindest du als besonders wichtig für die Überarbeitung deines Textes, und warum?

Code: _____

Feedback 2

	Stimme nicht zu	Stimme eher nicht zu	Neutral	Stimme eher zu	Stimme vollstän- dig zu
Mit dem Feedback bin ich zufrieden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich betrachte dieses Feedback als fair.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich betrachte dieses Feedback als gerechtfertigt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich betrachte dieses Feedback als nützlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich finde dieses Feedback hilfreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dieses Feedback unterstützt mich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dieses Feedback akzeptiere ich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diesem Feedback widerspreche ich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dieses Feedback lehne ich ab.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:

• Zufrieden	<input type="checkbox"/>				
• Zuversichtlich	<input type="checkbox"/>				
• Erfolgreich	<input type="checkbox"/>				
• Beleidigt	<input type="checkbox"/>				
• Verärgert	<input type="checkbox"/>				
• Frustriert	<input type="checkbox"/>				

Empfindest du das Feedback als hilfreich und verständlich formuliert? Welche Teile des Feedbacks empfindest du als besonders wichtig für die Überarbeitung deines Textes, und warum?

Code: _____

Feedback 1 UND 2

Welche Unterschiede oder Gemeinsamkeiten hast du in Feedback 1 und Feedback 2 festgestellt?

**Welches Feedback empfandest du als hilfreicher für die Verbesserung deines Textes, und warum?
Welche Teile dieser Rückmeldung haben dir besonders geholfen?**

Gab es Teile in einem Feedback, die dir geholfen haben, über deinen Schreibprozess nachzudenken? Wie unterscheidet sich dies von der Rückmeldung der anderen Quelle?

Welches Feedback würdest du bevorzugen?

Feedback 1

Feedback 2

Anhang 1.5: Fragebogen – Teil 2

Code: _____

Feedback 1 wurde von einem*r Klassenkamerad*in geschrieben. Feedback 2 wurde von einer Künstlichen Intelligenz (fiete.ai) geschrieben.

Schau dir deine Antworten auf den anderen Seiten an. Ändert sich etwas an deiner Wahrnehmung des Feedbacks? Warum (nicht)?

Anhang 2: Daten

Anhang 2.1: Klasse 7

Anhang 2.1.1: Quantitative Daten

Code	Fairness				Nützlichkeit				Akzeptanz				Eigene Kategorie (Identifizierung)				Affektwirkung				
	Mit dem Feedback bin ich zufrieden	Ich betrachte das Feedback als fair.	Ich betrachte das Feedback als gerechtfertigt.	Ich betrachte dieses Feedback als nützlich.	Ich finde dieses Feedback hilfreich.	Dieses Feedback unterstützt mich.	Dieses Feedback akzeptiere ich.	Dieses Feedback widerspricht mir.	Dieses Feedback lehne ich ab.	Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	Zufrieden	Zuversichtlich	Erfolgreich	Beleidigt	Verärgert	Frustriert	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
AFAJ1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_AFAJ1	4	4	5	4	4	3	5	4	1	5	4	4	3	1	1	1					
> Peer_AFAJ1	4	5	4	5	4	3	4	5	1	2	4	4	3	1	1	1					
AEFM0																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_AEFM0	5	5	4	5	4	3	5	3	1	4	4	5	4	1	2	1					
> Peer_AEFM0	3	4	4	2	3	2	3	3	4	2	3	5	5	3	4	2					
AFGD1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_AFGD1	4	4	3	2	2	3	4	2	1	3	4	3	3	1	2	1					
> Peer_AFGD1	4	4	3	4	4	5	5	1	1	3	5	5	5	1	1	1					
AFG11																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_AFG11	2	2	1	3	3	2	2	4	3	2	2	3	3	3	4	4					
> Peer_AFG11	3	3	4	3	3	4	4	3	1	2	3	3	3	3	3	3					
AFM11																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_AFMS0	4	4	5	5	5	2	5	2	4	5	1	5	1	1	1	1					
> Peer_AFMS0	5	5	5	5	5	5	5	1	1	1	5	5	5	1	1	1					
AFS01																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_AFS01	4	5	3	3	4	4	4	2	2	2	4	4	4	3	2	2	3				
> Peer_AFS01	5	4	4	4	4	4	4	4	2	3	4	4	4	4	1	1	1				
ALNF2																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_ALNF2	2	3	1	1	1	1	4	5	2	3	1	1	1	1	4	4	4				
> Peer_ALNF2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	1	1	1				
ALUD1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_ALUD1	4	4	5	5	3	5	1	1	1	4	4	3	3	2	1	1	1				
> Peer_ALUD1																					
CLSF5																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_CLSF5	1	2	3	2	3	1	2	4	5	2	2	5	3	3	5	4	4				
> Peer_CLSF5	5	5	5	4	4	4	5	3	4	4	4	5	5	4	1	4	4				
DFIS2																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_DFIS2	5	5	4	3	4	4	4	4	3	2	4	4	4	3	3	2	2	1			
> Peer_DFIS2	5	5	5	5	5	5	5	1	1	1	5	5	5	5	1	1	1				
EFIS1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_EFIS1	4	3	4	5	5	3	4	1	1	5	4	3	4	1	1	1	1				
> Peer_EFIS1	5	4	3	2	2	2	5	4	1	3	5	5	5	5	1	1	2				
EFTS2																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_EFTS2	1	3	3	4	4	3	3	2	3	4	1	2	5	5	2	2	2				
> Peer_EFTS2	5	4	3	2	2	4	4	1	1	2	5	5	5	5	1	1	1				
ELNA1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_ELNA1	3	2	1	1	2	3	1	4	3	2	2	4	5	3	4	4	1				
> Peer_ELNA1	5	5	5	5	5	5	5	2	2	4	5	5	5	5	2	2	2				
IFGM0																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_IFGM0	2	5	3	2	2	2	2	3	4	3	1	2	1	3	5	4	4				
> Peer_IFGM0	1	1	1	1	1	1	1	5	5	1	1	1	1	1	4	5	5				
IHFM4																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_IHFM4	2	3	3	2	2	4	3	4	4	4	4	4	3	5	1	3	1				
> Peer_IHFM4	5	5	3	4	5	5	5	1	1	5	5	3	5	5	1	1	1				
IFRAZ																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_IFRAZ	1	1	1	1	1	1	1	5	5	1	1	1	1	1	1	5	5				
> Peer_IFRAZ	1	2	1	2	3	1	2	5	5	1	1	1	1	1	1	5	5				
IFS1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_IFS1	5	5	5	5	5	2	4	2	1	3	5	3	2	1	1	1	1				
> Peer_IFS1	4	5	3	4	3	3	5	1	1	4	5	4	4	1	1	1	1				
IFZD2																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_IFZD2	1	1	1	2	1	1	1	5	5	1	1	1	1	1	1	5	5				
> Peer_IFZD2	5	5	5	5	5	5	5	1	1	5	5	5	5	5	1	1	1				
ILNI1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_ILNI1	3	4	4	4	2	2	4	5	1	1	4	5	4	4	1	1	1				
> Peer_ILNI1	4	4	4	4	3	2	5	4	1	2	4	5	4	4	1	1	1				
ILR11																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_ILR11	3	3	3	2	2	2	5	4	1	1	4	4	3	3	1	2	2				
> Peer_ILR11	3	4	4	3	3	2	5	4	1	2	4	4	3	3	1	1	2				
LFAAM1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_LFAAM1	2	1	2	3	3	3	4	3	2	2	3	2	3	3	3	3	3				
> Peer_LFAAM1	5	5	5	5	4	5	5	1	1	4	4	4	4	3	4	1	1				
MFRU1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_MFRU1	4	4	4	4	4	3	4	4	2	4	4	4	4	4	5	2	2				
> Peer_MFRU1	4	3	2	4	3	3	4	5	1	4	4	5	4	4	5	3	4				
OFNAO																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_OFNAO	4	4	5	4	5	5	5	2	1	5	4	4	3	4	2	1	1				
> Peer_OFNAO																					
UFSM1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_UFSM1	3	2	3	3	3	3	3	3	3	2	2	3	3	2	2	3	1				
> Peer_UFSM1	3	3	3	3	3	2	3	3	3	4	1	1	5	1	2	2	5				
UHJ1																	Wenn ich dieses Feedback auf meinen Text bekomme, fühle ich mich:				
> Kl_UHJ1	2	4	3	4	4	4	4	4	1	1	5	4	4	4	2	2	2				
> Peer_UHJ1	3	4	4	4	4	3	4	4	1	1	5	4	4	4	1	1	1				

Descriptive Statistics_Fairness

Descriptive Statistics

	KI_Mit dem Feedback bin ich zufrieden	Peer_Mit dem Feedback bin ich zufrieden	KI_Ich betrachte das Feedback als fair.	Peer_Ich betrachte das Feedback als fair.	KI_Ich betrachte das Feedback als gerechtfertigt.	Peer_Ich betrachte das Feedback als gerechtfertigt.
Valid	24	25	24	24	24	25
Missing	4	3	4	4	4	3
Median	3.000	4.000	3.500	4.000	3.000	4.000
Mean	2.958	4.000	3.292	4.083	3.042	3.840
Std.	1.334	1.190	1.334	1.060	1.301	1.214
Deviation						
Minimum	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000

Descriptive Statistics_Nützlichkeit

Descriptive Statistics

	KI_Ich betrachte dieses Feedback als nützlich.	Peer_Ich betrachte dieses Feedback als nützlich.	KI_Ich finde dieses Feedback hilfreich.	Peer_Ich finde dieses Feedback hilfreich.	KI_Dieses Feedback unterstützt mich.	Peer_Dieses Feedback unterstützt mich.
Valid	25	25	25	25	24	24
Missing	3	3	3	3	4	4
Median	3.000	4.000	3.000	4.000	3.000	4.000
Mean	3.160	3.600	3.160	3.720	2.625	3.542
Std.	1.375	1.225	1.281	1.173	0.970	1.351
Deviation						
Minimum	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	4.000	5.000

Descriptive Statistics_Akzeptanz

Descriptive Statistics

	KI_Dieses Feedback akzeptiere ich.	Peer_Dieses Feedback akzeptiere ich.	KI_Diesem Feedback widerspreche ich.	Peer_Diesem Feedback widerspreche ich.	KI_Dieses Feedback lehne ich ab.	Peer_Dieses Feedback lehne ich ab.
Valid	25	24	25	24	25	25
Missing	3	4	3	4	3	3
Median	4.000	4.500	3.000	2.500	2.000	1.000
Mean	3.360	4.083	3.080	2.625	2.520	1.880
Std.	1.350	1.248	1.320	1.637	1.447	1.481
Deviation						
Minimum	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000

Descriptive Statistics_Identifizierung

Descriptive Statistics

	KI_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	Peer_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.
Valid	25	24
Missing	3	4
Median	3.000	3.600
Mean	3.120	3.292
Std. Deviation	1.301	1.398
Minimum	1.000	1.000
Maximum	5.000	5.000

Descriptive Statistics_Affekt

Descriptive Statistics

	KI_Zufrieden	Peer_Zufrieden	KI_Zuversichtlich	Peer_Zuversichtlich	KI_Erfolgreich	Peer_Erfolgreich	KI_Beidigt	Peer_Beidigt	KI_Verärgert	Peer_Verärgert	KI_Frustriert	Peer_Frustriert
Valid	25	25	24	24	25	23	25	25	25	25	25	25
Missing	3	3	4	4	3	5	3	3	3	3	3	3
Median	4.000	5.000	3.000	4.000	3.000	4.000	2.000	1.000	2.000	1.000	1.000	1.000
Mean	2.920	4.080	3.125	3.917	2.720	4.043	2.280	1.640	2.560	1.960	2.160	1.640
Std. Deviation	1.412	1.320	1.227	1.283	1.308	1.224	1.458	1.114	1.417	1.541	1.463	1.221
Minimum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000

Paired Samples T-Test

Paired Samples T-Test

Measure 1	Measure 2	Test	Statistic	z	df	p
KI_Mit dem Feedback bin ich zufrieden	- Peer_Mit dem Feedback bin ich zufrieden	Student	-3.037	23	0.006	
		Wilcoxon	18.500	-2.560		0.010
KI_Ich betrachte das Feedback als fair.	- Peer_Ich betrachte das Feedback als fair.	Student	-2.373	21	0.027	
		Wilcoxon	37.500	-2.314		0.018
KI_Ich betrachte das Feedback als gerechtfertigt.	- Peer_Ich betrachte das Feedback als gerechtfertigt.	Student	-2.105	22	0.047	
		Wilcoxon	23.000	-1.852		0.066
KI_Ich betrachte dieses Feedback als nützlich.	- Peer_Ich betrachte dieses Feedback als nützlich.	Student	-1.252	23	0.223	
		Wilcoxon	66.000	-1.167		0.246
KI_Ich finde dieses Feedback hilfreich.	- Peer_Ich finde dieses Feedback hilfreich.	Student	-1.574	23	0.129	
		Wilcoxon	51.500	-1.481		0.138
KI_Dieses Feedback unterstützt mich.	- Peer_Dieses Feedback unterstützt mich.	Student	-2.352	21	0.029	
		Wilcoxon	27.500	-2.094		0.033
KI_Dieses Feedback akzeptiere ich.	- Peer_Dieses Feedback akzeptiere ich.	Student	-2.011	22	0.057	
		Wilcoxon	37.500	-1.846		0.063
KI_Diesem Feedback widerspreche ich.	- Peer_Diesem Feedback widerspreche ich.	Student	1.546	22	0.136	
		Wilcoxon	95.500	1.422		0.157
KI_Dieses Feedback lehne ich ab.	- Peer_Dieses Feedback lehne ich ab.	Student	1.909	23	0.069	
		Wilcoxon	80.000	1.726		0.087
KI_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	- Peer_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	Student	-0.216	22	0.831	
		Wilcoxon	74.500	-0.095		0.942
KI_Zufrieden	- Peer_Zufrieden	Student	-3.866	23	0.001	
		Wilcoxon	10.000	-2.999		0.002
KI_Zuversichtlich	- Peer_Zuversichtlich	Student	-2.737	21	0.012	
		Wilcoxon	5.000	-2.490		0.013
KI_Erfolgreich	- Peer_Erfolgreich	Student	-3.855	21	<.001	
		Wilcoxon	8.500	-2.925		0.003
KI_Begeistert	- Peer_Begeistert	Student	2.378	23	0.026	
		Wilcoxon	86.000	2.103		0.034
KI_Verärgert	- Peer_Verärgert	Student	2.223	23	0.036	
		Wilcoxon	93.500	1.903		0.056
KI_Frustriert	- Peer_Frustriert	Student	1.877	23	0.073	
		Wilcoxon	52.000	1.689		0.093

Assumption Checks

Test of Normality (Shapiro-Wilk)

		W	p
KI_Mit dem Feedback bin ich zufrieden	- Peer_Mit dem Feedback bin ich zufrieden	0.907	0.030
KI_Ich betrachte das Feedback als fair.	- Peer_Ich betrachte das Feedback als fair.	0.918	0.068
KI_Ich betrachte das Feedback als gerechtfertigt.	- Peer_Ich betrachte das Feedback als gerechtfertigt.	0.893	0.018
KI_Ich betrachte dieses Feedback als nützlich.	- Peer_Ich betrachte dieses Feedback als nützlich.	0.961	0.464
KI_Ich finde dieses Feedback hilfreich.	- Peer_Ich finde dieses Feedback hilfreich.	0.957	0.381
KI_Dieses Feedback unterstützt mich.	- Peer_Dieses Feedback unterstützt mich.	0.893	0.021
KI_Dieses Feedback akzeptiere ich.	- Peer_Dieses Feedback akzeptiere ich.	0.957	0.412
KI_Diesem Feedback widerspreche ich.	- Peer_Diesem Feedback widerspreche ich.	0.961	0.484
KI_Dieses Feedback lehne ich ab.	- Peer_Dieses Feedback lehne ich ab.	0.915	0.045
KI_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	- Peer_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	0.953	0.341
KI_Zufrieden	- Peer_Zufrieden	0.854	0.003
KI_Zuversichtlich	- Peer_Zuversichtlich	0.826	0.001
KI_Erfolgreich	- Peer_Erfolgreich	0.916	0.062
KI_Begeistert	- Peer_Begeistert	0.889	0.013
KI_Verärgert	- Peer_Verärgert	0.925	0.077
KI_Frustriert	- Peer_Frustriert	0.784	<.001

Note. Significant results suggest a deviation from normality.

Anhang 2.1.2: Qualitative Daten

Code	Feedback einzeln	Weilche Unterschiede oder Gemeinsamkeiten hast du in Feedback 1 deines Textes, und warum? Und Feedback 2 für die Überarbeitung deines Textes, und warum?	Feedback Vgl.	Feedback X wurde von einem*r Klassenkamerad*in geschrieben. Feedback Y wurde von einer Künstlichen Intelligenz (Rete.ai) geschrieben. Schau dir deine Antworten auf den anderen Seiten an. Ändert sich etwas an deiner Wahrnehmung des Feedbacks? Warum (nicht)?
AFAF11	Ja, es ist sehr hilfreich, da dort viele Verbesserungsvorschläge genannt werden. War hilfreich, weil dort stand was gut war und was man besser machen kann.	In dem einen waren eher so die gelungenen Sachen und in dem anderen was man besser machen kann.	Das mit den Verbesserungsvorschlägen, da ich dann weiß wie ich es besser machen kann.	Die Kl ist ausführlicher und gäte Verbesserungsvorschläge wurden genannt.
AFEM0	Ich fände an dem Feedback gut, dass immer konstruktive Kritik angewendet wurde. Ich finde das Feedback gut und für mich gut verständlich. Außerdem wurde mir viele Tipps gegeben, wo vielleicht noch Fehler oder in der Art aufgetreten sind.	Bei Feedback 1 [Kl] wurde sehr ausführlich beschrieben und alles war vorhanden. Bei Feedback 2 [Peer] wiederum sind manche Teile nicht vorhanden und wenn eine knapp beschrieben im Vergleich zu Feedback 1.	Feedback 1 [Kl] war für mich hilfreicher, weil alles sehr gut beschrieben wurde und ich sehr viel Gutes daraus nehmen konnte. Mir hat besonders gut geholfen, wo ich noch kleine Fehler gemacht habe, dass sie mir konstruktiv beschrieben wurden.	Also an sich ändert sich nichts an meiner Meinung. Ich finde das KI generierte (Feedback 1) immer noch besser, weil ich alle vorhenden war und mir wirklich gute Tipps gegeben wurden. Außerdem fand ich die konstruktive Kritik der Kl ebenfalls sehr gut und für mich hilfreich.
AFGD1	Ich finde, dass der der das Feedback geschrieben hat manche Punkte angesprochen die richtig gemacht habe, z.B. dass ich die Haftrammen beschreiben soll, obwohl ich es gemacht hatte. Das Feedback ist aber bis auf da alles hilfreich.	In dem 1. Feedback habe ich eher Feedback 1 [Kl], weil in Feedback 2 [Peer] keine Verbesserungsvorschläge hatte.	Der 1. Text von Feedback 1 [Kl] hat mir dabei geholfen, meinen Schreibprozess zu verbessern	Meine Wahrnehmung ändert sich nicht, weil es für mich keinen Unterschied macht, wer es geschrieben hat.
> Peer_AFGD1	Es ist verständlich und hilfreich.			

AFG11	Ich finde den Text eher nicht so verständlich. Bei dem Tatt habe ich die Textart in meinem Text beschrieben. Und das Thema wird ja eigentlich auch durch den Einleitungssatz klar. Den zweiten Teil verstehe ich. Bei dem dritten habe ich das eigentlich schon so ausgedrückt. Das "sehrte" verstehe ich wieder gar nicht da ich das eigentlich so geschrieben habe.	Beide haben mir nicht so geholfen da ich viele Sachen genauso geschrieben habe. Aber der zweite Teil hat mir geholfen. Also das Feedback nach dem Tatt	Nein eigentlich nicht. Die zweite Frage verstehe ich nicht ganz.	Peer	Meine Meinung ändert sich nicht. Ich weiß zwar das in Feedback zweit [Peer] ein paar Sachen falsch sind aber ich behalte meine Meinung trotzdem, weil ich das einfach so empfinde. KI hat mehr Fehler gemacht.		
> KI_AFG11	Ein paar Sachen stimmen nicht so ganz, wie z. B., das ich Willkommen schreiben soll oder das ich mehr Fühllichkeit und Gedanken rein bringen soll. Aber sonst ist eigentlich alles verständlich.						
AFSM0	Ich finde schön das derjenige mir gezeigt hat wie es besser geht, jedoch hat er nur geschrieben wie es sich für ihn besser anhört	Feedback 1 [KI] weil ich jetzt weiß was ich besser machen kann. No geschrieben wurde wie ich es hätte besser formulieren können	Ja als der Leser geschrieben hat was ich gut gemacht hab. Es gibt keine Unterschiede.	Peer	Ich finde es sehr gut weil geschrieben wurde was ich verbessern kann und wie ich es schreiben könnte		
> KI_AFSM0	Peer_AFSM0	Keine!					
AFSO1	Ich finde es sehr gut weil ich oft gelobt werde						
> KI_AFSO1	Peer_AFSO1						
ALNF2	Ich finde dieses Feedback in ordnung. Es gibt viele Sachen die mir nicht gepasst haben. Doch bin ich nicht sehr frustriert. Z. B. "Geckos ernähren sich von Insekten" statt "Geckos ernähren sich von Insekten!" Ich finde dieses Feedback sehr hilfreich. Das Feedback finde ich auch sehr fair.	In Feedback 1 [KI] waren sachen die keinen Sinn ergaben. Feedback 2 [Peer] war super nett.	Der 1 Teil [KI] war ungerecht! Der 2 Teil [Peer] war gerecht und netter, Feedback 1 [KI] war etwas besser.	Bei Feedback 1 [KI] hat mir kein bisschen geholfen! Feedback 2 [Peer] war super hilfreich	Peer	In meiner Wahrnehmung ändert sich nichts weil ich meine Antwort richtig finde. KI ist dummi!	
> KI_ALNF2	> Peer_ALNF2						

AUD11	Ich finde alles besonders, weil ich weiß das nicht immer meine Deutschlehrerin meine Arbeit sieht sondern auch jemand anderes (Ich mache) Ich finde es gut das man auch gelobt wird und nicht nur schlechte dingé schreibt.	Es war sehr ähnlich nur in Feedback 1 [K] wurde alles ausführlich geschrieben, dennoch wurde Feedback 2 [Peer] kurz und versäumnisvoll geschrieben.	Eigentlich haben die Aussagen mir gehofft, dann wenn ich keinen Feedback bekommen hätte würde ich vielleicht immer noch Den Einleitungssatz falschschreiben oder ich würdestellen vergessen in Präzisen zu schreiben. Ich habe mir manche Sachen abgespeichert und kann jetzt nächstes mal einen besseren Text schreiben.	Nein, denn die Aussagen von beiden Seiten ähnlich sind und ich fand es Seitsam bei der KI was sie sage beim Präzens. Es ergab nämlich keinen Sinn. Ich fand beide gut aber Feedback 2 [Peer] besser!
CLSF5	Ich finde das Feedback ist gut formuliert. Ich würde aber mehr positivs ein schreiben. Das Feedback ist auch klar formuliert. Ich finde das Feedback erfolgreich und hilfreich. Das Feedback ist der Person sehr gut gelungen.	In einem war sehr viel Text und im anderen sehr wenig Text!	Ich finde das zweite [Peer] war hilfreich.	Ich bleib bei meiner Meinung weil der Text ist für mich genau gleich.
DFIS2	Ich finde es ist hilfreich und nicht hilfreich. Das Feedback ist der Person sehr gut gelungen.	Ich habe nicht so viele Gemeinsamkeiten gesehen weil im	Das 2te und mir haben fast alle gefallen	nein es ändert sich nichts
KI_DFS2	Ich finde das Feedback sehr hilfreig. EFIS2	Sehr viele Unterschiede	Feedback 1 [K], weil z. B. er überall was da verbessern kann.	Nein, weil ich vertraue KI und da ich bei Feedback 1 [K] mehr sachen verbessern konnte.
Peer_DFS2	Ich finde alles wichtig, weil darüber sind kleine Verbesserungen zu finden. Ich bin auch sehr zufrieden. Ich finds ok. Ich kann leider nichts überarbeiten.			Nein denn eine KI ist ein Roboter und Roboter machen oft keine Fehler. Aber ein Mensch macht Fehler was auch normal ist. Bei der KI stande halb etwas ausführlich und der von meinem Mitschüler ist halb nicht ausführlich. Ich vertraue mehr der KI bei solchen Texten.
KI_EFTS1	Ich finde das Feedback hilfreich um meine nächsten Sachterte zu verbessern. Ich finde dass dritte Feedback als sehr nützlich.	Also bei dem 1sten Feedback [K] wurde ausführlicher geschrieben als im 2.	Ich fand das Feedback 1 [K] als hilfreicher denn dort wurde mehr Verbesserungsmaterial gefunden.	Ja es gab Teile wo ich mir um meinen Text nachzudenken z. B. wieo habe ich diesen Vorn-der-Waals nicht erwähnt.
Peer_EFTS1	Ich finde diesen Feedbackbogen eher nicht hilfreich denn man hätte im Text mehr Fehler finden können. Ich finde kein Feedback als wichtig denn man könnte es ausführlicher erklären.	Im ersten Feedback [K] wurden mehr Fehler gefunden.		KI

ELNA1	ich bin nicht so richtig zufrieden, da ich viele Punkte nicht richtig verstehe, ich finde das auch nicht richtig fair, weil viel dort anders steht. Und in der Einleitung habe ich eigentlich auch alles geschrieben, was dort stehen sollte, aber jetzt steht es dort nicht mehr. Ich finde das Feedback sehr gut, da es es sachlich und gerechtfertigt verstanden. Verbeserungsvorschläge.	Im Feedback 1 [KI] wurden viele Sachen geschrieben, die ich eigentlich auch gemacht habe. Im Feedback 2 [Peer] habe ich die Habe ich die schreibe die Verbeserungsvorschläge.	Eigentlich nicht. Mir hat nur das mit den Abstraktion geholfen. Aber ich geholfen, dass ich Abstrakte machen verstehe die Frage auch nicht richtig, sage sehr gut verstanden.	Peer	Ich änderne meine Meinung wegen der KI nicht. Ich finde immer noch, dass Feedback 1 [KI] nicht gut ist. Die Vorschläge zum Verbesserem hatten ich schon dort stehen. Außerdem hatte ich nur einen Rechtschreibfehler, der direkt mit in die Bewertung genommen.
IFGM0	Das Feedback ist verständlich und gut formuliert. (Aber dass sie eine (fast) Lehrerin sind, ist das ja klar) Hilfreich finde ich es mittel. Man muss ja nicht auf das Feedback hören. Außerdem finde ich, dass Sie viel wiedeholt haben im Feedback. (Emojis und co. sind nur zur Verzierung da) Ich bin gaaaaaaaar nicht zufrieden.	Feedback 2 [KI] war sehr genau und man hat gesehen, dass es eine Erwachene war. Man hat bingere Sätze geschrieben bei Feedback 2. Feedback 1 ist sehr kurz und nicht hilfreich.	Feedback 2 [KI]. Man sollte nicht so unnötige Sachen schreiben, wie "das war super". Dann fühlt man sich noch blöder als eigentlich schon, weil dann das Schlecht kommt. Wenn man direkt schlecht macht, fühlt sich schlecht sogar besser.	KI	Nein. Weil man das schon vorher gesehen hat, da Feedback 2 [KI] wiederum ebenfalls, genauer und detaillierter, und sonst pos. :) Ich finde, meine Wahrnehmung ändert sich nichts.
KI_IFGM0	Ich empfinde das Feedback gaaaaaaaaaaaaaaar nicht hilfreich und blöd. Ich finde gar keinen Teil wichtig für die Überarbeitung meines Textes. Ich bin nicht zufrieden. Emojis und co sind nur zur Verschönerung da, !? (Ich habe irgendwie genommen)	Feedback 2. Feedback 1 ist sehr kurz und nicht hilfreich.	Bei mir ändert sich nichts. Ich finde das erste Feedback [KI] besser weil es bessere Tipps gibt. Außerdem lobt es mich das erste Feedback/ was ich immer noch sehr gut finde. Also finde ich es viel besser und würde meine Meinung nicht ändern! Auf die Person kann ich mich wirklich verlassen!	Peer	Bei mir ändert sich nichts. Ich finde das erste Feedback [KI] besser weil es bessere Tipps gibt. Außerdem lobt es mich das erste Feedback/ was ich immer noch sehr gut finde. Also finde ich es viel besser und würde meine Meinung nicht ändern! Auf die Person kann ich mich wirklich verlassen!
IFHM4	Ich finde das Feedback nicht so gut da es Fehler genannt hat, die ich nicht gemacht habe. Aber ich wurde auch gelobt, was ich gut finde. Die Bewertung ist auch sachlich.	Feedback 1 [Peer] war nicht so streng und hat nur Fehler genannt die ich gemacht habe!!! Nr. 2 [KI] hat Fehler genannt die ich nicht gemacht habe.	Bei Feedback Nummer 1 [Peer] waren mehr Tipps und auch direkt für die Fehler etc. die ich gemacht habe, nicht nur allgemein.	Peer	Bei Feedback Nummer 1 [Peer] waren z. B. das ich Rechtschreibfehler gemacht habe oder grammatische Fehler die ich gemacht habe.
Peer_IFHM4	Ich finde das Feedback gut weil nicht nur negative Kritik geschrieben wurde sondern auch positive. Außerdem wurden Tipps genannt um den Text zu verbessern was ich sehr wichtig finde da es ohne keinen Sinn macht				

I[FRA2]	Ich glaube, dass der Text gar nicht richtig gelesen wurde.	Bei Nummer 2 [Kl] wurde der Text garnicht richtig lesen.	Nichts hat mir richtig weitergeholfen. Aber wenn, dann das Feedback 1 [Peer].	Es gab nicht viele teile die mir geholfen haben und es gab kritik ohne mir zu sagen wie ich es besser machen könnte.	Peer	An meiner Meinung hat sich nichts verändert, da ich schon damit gerechnet habe. Ich finde, dass man schnell erkennt, wenn etwas Kl geschrieben ist.
Kl_I[FRA2]	Ich finde das Feedback nicht hilfreich.					
Peer_I[FRA2]	Ich finde, dass es schlechter bewertet wurde.					
I[FR11]						
Kl_I[FR11]	Den Titel besser machen.		Einleitungssatz	Alle außer ein.	Peer	Ich finde Feedback 1 [Peer] besser weil es keine Fehler hatte. Feedback 1 ist auch besser geschrieben.
Peer_I[FR11]	Den Einleitungssatz nehme ich mit. Und das ich mehr beschreibe.			Der Anfang hat mir geholfen. Erläuterungssatz.		
I[FSF1]						
Kl_I[FSF1]	Es wurden viele sachen verbessert und tips genannt. Das Feedback ist gut und hilfreich	In Feedback 1 [Peer] wurde sehr viel weniger genannt. Es gab auch viel weniger tips. In dem 2. gennant.	Der zweite Text [Kl] hat mir mehr geholfen.		Kl	Nein es ändert sich nichts. Weil mir ist egal ob ein echter Mensch das schreibt oder Kl.
Peer_I[FSF1]	Ich finde es ist ein gutes feedback. Es wurden positive und negative Punkte genannt.					
I[FZD2]						
Kl_I[FZD2]	Das Feedback ist nicht hilfreich!!! Bin super zufrieden!	Feedback 2 [Kl] hat überhaupt nicht zu Feedback 1 [Peer] da es einfach zu meinem Sachtext gepasst hat. Und Feedback [Kl] waren	Feedback 1 [Peer] da es einfach zu meinem Sachtext gepasst hat. Und Feedback 1 [Peer] wurde ich nochmal gelobt, sodass ich alles		Peer	Bei mir hat es sich nichts am meinen Meinung verändert. Da es ja trotzdem ein Feedback ist, habe damit schon gerechnet.
Peer_I[FZD2]						
I[LN11]						
Kl_I[LN11]	Ich finde das Feedback nicht sehr hilfreich, weil dort Sachien genannt werden die ich im Text habe und im Feedback steht das ich es nicht habe.	Feedback 1 [Peer] ist besser, weil dort nicht Sachien drin stehen	Das erste [Peer]. Das Feedback im Haupteil hilft mir wegen dem falsch geschriebenen Wort.		Peer	Bei Feedback zwei [Kl] musste ich kurz nachdenken, weil dort stand das ich Abschnitte machen soll [= abschnittweise zusammenfassen]. Ich finde das nicht nötig ist.
Peer_I[LN11]	Ich finde das Feedback gut, bis auf das im Hauptteil vom Feedback steht: "Ich wusste gar nicht, dass Geckos Flachsen an ihren Füßen haben". Wenn man Rechtschreibfehler schreibt reicht das.	doch ich angeblich nicht in den Text geschrieben habe, obwohl ich es gemacht hatte				

ILR1	Dieses Feedback hilft mir nicht weiter, weil vieles von dem Kriterium nicht stimmt und ich daraus nichts lernen oder verbessern kann. Dennoch kann ich aus dem Feedback schließen das mein Text nicht schlecht ist, da das was im Feedback steht nicht richtig ist, sondern in meinem Text vorhanden ist.	Feedback 1 [Peer] hat mir mehr geholfen da ich bei der Einleitung ungefähr weiß was ich verbessern kann und weiß dass ich mehr Fachbegriffe verwenden sollte. Feedback 2 [KI] hat mir aber auch geholfen, da ich weiß was richtig ist.	Durch Feedback 1 [Peer] wurden mir Dinge klar was ich verbessern kann und was schon gut ist.	Peer	Bei mir ändert sich nichts. Ich finde nämlich immer noch das Feedback 2 [KI] nicht so hilfreich wie Feedback 1 [Peer]. In dem Feedback der KI wird viel falsches erwähnt und es hilft mir nicht beim Überarbeiten.
LFAM1	Ich finde dieses Feedback nicht besonders hilfreich, da ich in der Einleitung zwar etwas verbessern kann aber ich kann mir nicht abschließen wie ich es verbessern kann. Außerdem fehlt mehr an die Hälfte und mir hilft es nicht weiter wenn ich die Zahlen nicht ausschreiben darf/soll. Ich weiß aber, dass ich mehr Fachbegriffe verwenden sollte.	Beide Feedbacks haben mir nur zur Hälfte weitergeholfen.		Peer	An meiner Meinung ändert sich nichts, denn in Fb 2 [KI] sind viele Fehler.
KI_LFAM1	Ich finde das mehrere Sachen genannt wurden bei denen ich nicht zustimmen kann. z.B. das ich anscheinend nicht geschrieben habe das es sich um einen Sachtext handelt.	FB 1 [Peer] ist besser, weil dort weit dorthin geschrieben wurden. FB 2 [KI] es wurde bei denen ich zustimmen kann FB 1 ist netter formuliert.	Das ich kürzer schreiben könnte	Peer	Meine Meinung ändert sich nicht. Ich bevorzuge weiterhin die Meinung der Schülerin. Die KI hat sehr viele Fehler gemacht die, die Schule in nicht gemacht hat.
Peer_LFAM1	Ich finde es hilfreich und verständlich formuliert. Ich stimme es gut, dass im Feedback geschrieben wurde, dass ich es kürzer formulieren könnte, dadurch kann ich meinen Text besser optimieren. Es freut mich das auch positive Sachen beschrieben wurden	Feedback 1 [Peer], weil dort nachvollziehbare Dinge schrieben wurden. Feedback 2 [KI] es wurde z.B. geschrieben das ich nicht erwähnt habe dass es ein Sachtext ist. Das stimmt nicht!!!		Peer	
MFR1	Ich finde Feedback 2 ebenfalls sehr hilfreich. In diesem Feedback wird ebenfalls gesagt, dass ich einen nicht "richtigen" Einleitungssatz geschrieben habe. Dem stimme ich nicht zu.	In beiden Feedbacks wird ein nicht ausreichender Einleitungssatz genannt. Dem stimme ich nicht zu.	Ich fand beide Texte/Feedbacks sehr hilfreich, jedoch stimme ich manchen Dingen zu, manchen aber auch nicht. Ich fand den Tipps, dass ich die "Van-de-Waals"-Kräfte zu nennen sehr hilfreich	Peer	
Peer_MFR1	Ich finde das Feedback sehr hilfreich, jedoch finde ich, dass ich einen informativen Einleitungssatz geschrieben habe, der alles nötige enthält. Ansonsten bin ich sehr zufrieden mit diesem verständlichen formulierten Sachtext.	Meine Meinung ändert sich nicht. Ich bevorzuge weiterhin die Meinung der Schülerin. Die KI hat sehr viele Fehler gemacht die, die Schule in nicht gemacht hat.			

OFNAO KI_OFNAO	Ich finde das Feedback Okay und ich bin auf jeden Fall, der in Zukunft weniger Emojis und weniger eigene Meinung reinschreiben. Sorry, aber bin Peer_OFNAO	Bei FB1 [Peer] ist das "obere", "ausseres" Feedback und bei FB2 [KI] ist es detaillierter. Ich sehe, dass bei FB1 das wichtigste steht und bei FB2 das etwas detailliertere steht. Also im Prinzip ist beides wichtig, aber bei FB2 sind mehr Infos und mehr Meinung erhalten.	Ja, ich finde, dass die Emojis zu viel sind, das werde ich verbessern. Und ich werde sachlicher schreiben mit den Emojis.	KI	Ah okey. Nein, nicht wirklich aber ich dachte, dass FB1 von einer Person war und FB2 von einer KI oder AI war.	
UFSM1 KI_UFSM1	Ja es ist hilfreich weil ganz viele Verbesserungsvorschläge enthalten sind. Ja, es ist hilfreich weil ganz viele Meinungen/Informationen/Verbesserungsvorschläge genannt wurden	Keine Unterschiede!	Alle Feedbacks waren hilfreich!	Peer	Nein, es ändert nichts, weil ich der KI nicht mehr vertraue als der echten Person.	
UJHJ1 Peer_UJHJ1	In meinem Feedback steht ich könnte "In dem Sachtext mit dem Titel" schreiben aber ich habe es geschrieben, in dem Titel wurde in Präsens geschrieben, aber da steht, das ich es an manchen Stellen anders geschrieben habe.	In beiden Feedbacks hab es ein oder zwei Sätze die mir geholfen haben aber auch Punkte die mir nicht ganz weitergeholfen haben.	Feedback 1 [Peer] hat mir ein bisschen mehr geholfen, jedoch hat mir die Rückmeldung zur Zusammenfassung von Feedback 2 [KI] mehr geholfen. Ich habe durch das Feedback ganz besonders auf unterschiedliche Punkte in meinem Text geachtet.	Peer	Bei mir ändert sich nicht viel, im Feedback 2 [KI] wurden nur manchmal Sachen genannt, die ich eigentlich schon im Text hatte.	
Peer_UJHJ1	Ich würde in einen Sachtext eher kein "Hallo" oder "Tschüss" schreiben, was auf meinem Gegen steht, da ich glaube, dass man es nicht dar einschreibt.					

Anhang 2.2: Klasse 6

Anhang 2.2.1: Quantitative Daten

Code	Frage	Nutzlichkeit	Akzeptanz	Affektivierung	
				Eigene Kategorie (Identifizierung)	Wann ich dieses Feedback auf meinem Test bekomme, fühle ich mich:
<u>Allgemein</u>					
PEER_ANNA1	Mit dem Feedback bin ich zufrieden	Ich betrachte das Feedback als gernartige Feedbacks.	Durch Feedback, das mir Feedback-Bilanzschlusswerte nicht entspricht.	Eigener Feedback.	Wenn ich das Feedback weiß, bin ich darüber informiert.
PEER_ANNA2	4	3	3	3	2
PEER_ANNA3	5	4	5	5	2
PEER_ANNO	5	4	4	5	4
PEER_ANNO2	4	4	5	4	4
PEER_ANNO3	4	4	4	5	4
PEER_ALAF1	5	5	5	5	5
PEER_ALAF2	5	5	5	5	5
PEER_ALAF3	5	5	5	5	5
PEER_ALAF4	5	5	5	5	5
PEER_ALAF5	5	5	5	5	5
PEER_ALAF6	5	5	5	5	5
PEER_ALAF7	5	5	5	5	5
PEER_ALAF8	5	5	5	5	5
PEER_ALAF9	5	5	5	5	5
PEER_ALAF10	5	5	5	5	5
PEER_ALAF11	5	5	5	5	5
PEER_ALAF12	5	5	5	5	5
PEER_ALAF13	5	5	5	5	5
PEER_ALAF14	5	5	5	5	5
PEER_ALAF15	5	5	5	5	5
PEER_ALAF16	5	5	5	5	5
PEER_ALAF17	5	5	5	5	5
PEER_ALAF18	5	5	5	5	5
PEER_ALAF19	5	5	5	5	5
PEER_ALAF20	5	5	5	5	5
PEER_ALAF21	5	5	5	5	5
PEER_ALAF22	5	5	5	5	5
PEER_ALAF23	5	5	5	5	5
PEER_ALAF24	5	5	5	5	5
PEER_ALAF25	5	5	5	5	5
PEER_ALAF26	5	5	5	5	5
PEER_ALAF27	5	5	5	5	5
PEER_ALAF28	5	5	5	5	5
PEER_ALAF29	5	5	5	5	5
PEER_ALAF30	5	5	5	5	5
PEER_ALAF31	5	5	5	5	5
PEER_ALAF32	5	5	5	5	5
PEER_ALAF33	5	5	5	5	5
PEER_ALAF34	5	5	5	5	5
PEER_ALAF35	5	5	5	5	5
PEER_ALAF36	5	5	5	5	5
PEER_ALAF37	5	5	5	5	5
PEER_ALAF38	5	5	5	5	5
PEER_ALAF39	5	5	5	5	5
PEER_ALAF40	5	5	5	5	5
PEER_ALAF41	5	5	5	5	5
PEER_ALAF42	5	5	5	5	5
PEER_ALAF43	5	5	5	5	5
PEER_ALAF44	5	5	5	5	5
PEER_ALAF45	5	5	5	5	5
PEER_ALAF46	5	5	5	5	5
PEER_ALAF47	5	5	5	5	5
PEER_ALAF48	5	5	5	5	5
PEER_ALAF49	5	5	5	5	5
PEER_ALAF50	5	5	5	5	5
PEER_ALAF51	5	5	5	5	5
PEER_ALAF52	5	5	5	5	5
PEER_ALAF53	5	5	5	5	5
PEER_ALAF54	5	5	5	5	5
PEER_ALAF55	5	5	5	5	5
PEER_ALAF56	5	5	5	5	5
PEER_ALAF57	5	5	5	5	5
PEER_ALAF58	5	5	5	5	5
PEER_ALAF59	5	5	5	5	5
PEER_ALAF60	5	5	5	5	5
PEER_ALAF61	5	5	5	5	5
PEER_ALAF62	5	5	5	5	5
PEER_ALAF63	5	5	5	5	5
PEER_ALAF64	5	5	5	5	5
PEER_ALAF65	5	5	5	5	5
PEER_ALAF66	5	5	5	5	5
PEER_ALAF67	5	5	5	5	5
PEER_ALAF68	5	5	5	5	5
PEER_ALAF69	5	5	5	5	5
PEER_ALAF70	5	5	5	5	5
PEER_ALAF71	5	5	5	5	5
PEER_ALAF72	5	5	5	5	5
PEER_ALAF73	5	5	5	5	5
PEER_ALAF74	5	5	5	5	5
PEER_ALAF75	5	5	5	5	5
PEER_ALAF76	5	5	5	5	5
PEER_ALAF77	5	5	5	5	5
PEER_ALAF78	5	5	5	5	5
PEER_ALAF79	5	5	5	5	5
PEER_ALAF80	5	5	5	5	5
PEER_ALAF81	5	5	5	5	5
PEER_ALAF82	5	5	5	5	5
PEER_ALAF83	5	5	5	5	5
PEER_ALAF84	5	5	5	5	5
PEER_ALAF85	5	5	5	5	5
PEER_ALAF86	5	5	5	5	5
PEER_ALAF87	5	5	5	5	5
PEER_ALAF88	5	5	5	5	5
PEER_ALAF89	5	5	5	5	5
PEER_ALAF90	5	5	5	5	5
PEER_ALAF91	5	5	5	5	5
PEER_ALAF92	5	5	5	5	5
PEER_ALAF93	5	5	5	5	5
PEER_ALAF94	5	5	5	5	5
PEER_ALAF95	5	5	5	5	5
PEER_ALAF96	5	5	5	5	5
PEER_ALAF97	5	5	5	5	5
PEER_ALAF98	5	5	5	5	5
PEER_ALAF99	5	5	5	5	5
PEER_ALAF100	5	5	5	5	5
PEER_ALAF101	5	5	5	5	5
PEER_ALAF102	5	5	5	5	5
PEER_ALAF103	5	5	5	5	5
PEER_ALAF104	5	5	5	5	5
PEER_ALAF105	5	5	5	5	5
PEER_ALAF106	5	5	5	5	5
PEER_ALAF107	5	5	5	5	5
PEER_ALAF108	5	5	5	5	5
PEER_ALAF109	5	5	5	5	5
PEER_ALAF110	5	5	5	5	5
PEER_ALAF111	5	5	5	5	5
PEER_ALAF112	5	5	5	5	5
PEER_ALAF113	5	5	5	5	5
PEER_ALAF114	5	5	5	5	5
PEER_ALAF115	5	5	5	5	5
PEER_ALAF116	5	5	5	5	5
PEER_ALAF117	5	5	5	5	5
PEER_ALAF118	5	5	5	5	5
PEER_ALAF119	5	5	5	5	5
PEER_ALAF120	5	5	5	5	5
PEER_ALAF121	5	5	5	5	5
PEER_ALAF122	5	5	5	5	5
PEER_ALAF123	5	5	5	5	5
PEER_ALAF124	5	5	5	5	5
PEER_ALAF125	5	5	5	5	5
PEER_ALAF126	5	5	5	5	5
PEER_ALAF127	5	5	5	5	5
PEER_ALAF128	5	5	5	5	5
PEER_ALAF129	5	5	5	5	5
PEER_ALAF130	5	5	5	5	5
PEER_ALAF131	5	5	5	5	5
PEER_ALAF132	5	5	5	5	5
PEER_ALAF133	5	5	5	5	5
PEER_ALAF134	5	5	5	5	5
PEER_ALAF135	5	5	5	5	5
PEER_ALAF136	5	5	5	5	5
PEER_ALAF137	5	5	5	5	5
PEER_ALAF138	5	5	5	5	5
PEER_ALAF139	5	5	5	5	5
PEER_ALAF140	5	5	5	5	5
PEER_ALAF141	5	5	5	5	5
PEER_ALAF142	5	5	5	5	5
PEER_ALAF143	5	5	5	5	5
PEER_ALAF144	5	5	5	5	5
PEER_ALAF145	5	5	5	5	5
PEER_ALAF146	5	5	5	5	5
PEER_ALAF147	5	5	5	5	5
PEER_ALAF148	5	5	5	5	5
PEER_ALAF149	5	5	5	5	5
PEER_ALAF150	5	5	5	5	5
PEER_ALAF151	5	5	5	5	5
PEER_ALAF152	5	5	5	5	5
PEER_ALAF153	5	5	5	5	5
PEER_ALAF154	5	5	5	5	5
PEER_ALAF155	5	5	5	5	5
PEER_ALAF156	5	5	5	5	5
PEER_ALAF157	5	5	5	5	5
PEER_ALAF158	5	5	5	5	5
PEER_ALAF159	5	5	5	5	5
PEER_ALAF160	5	5	5	5	5
PEER_ALAF161	5	5	5	5	5
PEER_ALAF162	5	5	5	5	5
PEER_ALAF163	5	5	5	5	5
PEER_ALAF164	5	5	5	5	5
PEER_ALAF165	5	5	5	5	5
PEER_ALAF166	5	5	5	5	5
PEER_ALAF167	5	5	5	5	5
PEER_ALAF168	5	5	5	5	5
PEER_ALAF169	5	5	5	5	5
PEER_ALAF170	5	5	5	5	5
PEER_ALAF171	5	5	5	5	5
PEER_ALAF172	5	5	5	5	5
PEER_ALAF173	5	5	5	5	5
PEER_ALAF174	5	5	5	5	5
PEER_ALAF175	5	5	5	5	5
PEER_ALAF176	5	5	5	5	5
PEER_ALAF177	5	5	5	5	5
PEER_ALAF178	5	5	5	5	5
PEER_ALAF179	5	5	5	5	5
PEER_ALAF180	5	5	5	5	5
PEER_ALAF181	5	5	5	5	5
PEER_ALAF182	5	5	5	5	5
PEER_ALAF183	5	5	5	5	5
PEER_ALAF184	5	5	5	5	5
PEER_ALAF185	5	5	5	5	5
PEER_ALAF186	5	5	5	5	5
PEER_ALAF187	5	5	5	5	5
PEER_ALAF188	5	5	5	5	5
PEER_ALAF189	5	5	5	5	5
PEER_ALAF190	5	5	5	5	5
PEER_ALAF191	5	5	5	5	5
PEER_ALAF192	5	5	5	5	5
PEER_ALAF193	5	5	5	5	5
PEER_ALAF194	5	5	5	5	5
PEER_ALAF195	5	5	5	5	5
PEER_ALAF196	5	5	5	5	5
PEER_ALAF197	5	5	5	5	5
PEER_ALAF198	5	5	5	5	5
PEER_ALAF199	5	5	5	5	5
PEER_ALAF200	5	5	5	5	5
PEER_ALAF201	5	5	5	5	5
PEER_ALAF202	5	5	5	5	5
PEER_ALAF203	5	5	5	5	5
PEER_ALAF204	5	5	5	5	5
PEER_ALAF205	5	5	5	5	5
PEER_ALAF206	5	5	5	5	5
PEER_ALAF207	5	5	5	5	5
PEER_ALAF208	5	5	5	5	5
PEER_ALAF209	5	5	5	5	5
PEER_ALAF210	5	5	5	5	5
PEER_ALAF211	5	5	5	5	5
PEER_ALAF212	5	5	5	5	5
PEER_ALAF213	5	5	5	5	5
PEER_ALAF214	5	5	5	5	5
PEER_ALAF215	5	5	5	5	5
PEER_ALAF216	5	5	5	5	5
PEER_ALAF217	5	5	5	5	5
PEER_ALAF218	5	5	5	5	5
PEER_ALAF219	5	5	5	5	5
PEER_ALAF220	5	5	5	5	5
PEER_ALAF221	5	5	5	5	5
PEER_ALAF222	5	5	5	5	5
PEER_ALAF223	5	5	5	5	5
PEER_ALAF224	5	5	5	5	5
PEER_ALAF225	5	5	5	5	5
PEER_ALAF226	5	5	5	5	5
PEER_ALAF227	5	5	5	5	5
PEER_ALAF228	5	5	5	5	5
PEER_ALAF229	5	5	5	5	5
PEER_ALAF230	5	5	5	5	5
PEER_ALAF231	5	5	5	5	5
PEER_ALAF232	5	5	5	5	5
PEER_ALAF233	5	5	5	5	5
PEER_ALAF234	5	5	5	5	5
PEER_ALAF235	5	5	5	5	5
PEER_ALAF236	5	5	5	5	5
PEER_ALAF237	5	5	5	5	5
PEER_ALAF238	5	5	5	5	5
PEER_ALAF239	5	5	5	5	5
PEER_ALAF240	5	5	5	5	5
PEER_ALAF241	5	5	5	5	5
PEER_ALAF242	5	5	5	5	5
PEER_ALAF243	5	5	5	5	5
PEER_ALAF244	5	5	5	5	5
PEER_ALAF245	5	5	5	5	5
PEER_ALAF246	5	5	5	5	5
PEER_ALAF247	5	5	5	5	5
PEER_ALAF248	5	5	5	5	5
PEER_ALAF249	5	5	5	5	5
PEER_ALAF250	5	5	5	5	5
PEER_ALAF251	5	5	5		

PEER_0001	4	4	5	5	4	1	1	4	3	3	1	1	1
PEER_0002	5	5	5	5	4	5	1	1	4	5	5	5	1
PEER_0003	4	5	5	4	4	5	1	1	3	5	4	5	1
PEER_0004	4	5	5	4	4	5	1	1	4	5	5	5	1
PEER_0005	2	3	2	3	2	3	4	3	2	2	3	3	1
PEER_0006	4	4	3	3	3	5	2	1	3	5	3	4	1
PEER_0007	4	2	2	3	2	2	4	3	2	2	2	1	3
PEER_0008	4	3	3	3	4	4	2	1	3	3	3	3	1
PEER_0009	4	3	2	3	4	3	2	2	5	4	2	1	2
PEER_000A	3	3	4	5	4	2	2	3	3	3	3	1	4
PEER_000B	4	4	3	3	4	3	3	2	3	4	4	3	1
PEER_000C	4	4	3	3	4	5	1	1	4	5	4	4	1
PEER_000D	5	5	4	5	4	5	1	1	4	5	4	4	1
PEER_000E	3	3	3	3	3	3	3	3	3	3	3	3	1
PEER_000F	4	4	3	3	3	3	3	3	3	3	3	3	1
PEER_000G	4	3	2	3	3	4	2	1	3	3	3	3	1
PEER_000H	4	4	3	5	5	4	1	3	5	3	3	3	3
PEER_000I	4	4	4	4	4	4	2	1	1	4	4	4	1
PEER_000J	1	1	1	1	1	1	1	1	3	1	1	1	1
PEER_000K	5	5	5	5	5	5	5	1	1	4	5	5	1
PEER_000L	5	5	5	5	5	5	5	5	1	1	4	5	1
PEER_000M	4	4	5	5	5	5	5	5	2	5	3	3	1
PEER_000N	4	4	4	4	4	4	4	2	2	2	3	3	2
PEER_000O	5	5	5	4	4	5	5	1	4	5	5	5	1
PEER_000P	5	5	5	5	5	5	5	5	1	4	5	5	1
PEER_000Q	3	3	4	3	3	3	3	2	2	3	3	3	2
PEER_000R	3	4	3	3	4	3	4	2	2	3	2	2	1

Descriptive Statistics_Fairness

Descriptive Statistics

	KI_Mit dem Feedback bin ich zufrieden	Peer_Mit dem Feedback bin ich zufrieden	KI_Ich betrachte das Feedback als fair	Peer_Ich betrachte das Feedback als fair	KI_Ich betrachte das Feedback als gerechtfertigt	Peer_Ich betrachte das Feedback als gerechtfertigt
Valid	21	21	21	21	20	20
Missing	0	0	0	0	1	1
Median	3.000	4.000	4.000	4.000	4.000	4.000
Mean	3.286	4.000	3.524	4.095	3.500	4.000
Std.	1.102	0.775	1.209	0.944	1.192	0.795
Deviation						
Minimum	1.000	3.000	1.000	2.000	1.000	3.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000

Descriptive Statistics_Nützlichkeit

Descriptive Statistics

	KI_Ich betrachte dieses Feedback als nützlich	Peer_Ich betrachte dieses Feedback als nützlich	KI_Ich finde dieses Feedback hilfreich	Peer_Ich finde dieses Feedback hilfreich	KI_Dieses Feedback unterstützt mich	Peer_Dieses Feedback unterstützt mich
Valid	21	21	21	21	21	21
Missing	0	0	0	0	0	0
Median	4.000	4.000	3.000	4.000	3.000	4.000
Mean	3.571	3.714	3.476	3.810	3.095	3.619
Std.	1.326	1.231	1.327	1.123	1.261	1.024
Deviation						
Minimum	1.000	1.000	1.000	1.000	1.000	2.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000

Descriptive Statistics_Akzeptanz

Descriptive Statistics

	KI_Dieses Feedback akzeptiere ich.	Peer_Dieses Feedback akzeptiere ich.	KI_Diesem Feedback widerspreche ich.	Peer_Diesem Feedback widerspreche ich.	KI_Dieses Feedback lehne ich ab.	Peer_Dieses Feedback lehne ich ab.
Valid	21	21	21	21	20	21
Missing	0	0	0	0	1	0
Median	4.000	5.000	2.000	2.000	2.000	1.000
Mean	3.524	4.333	2.190	1.810	2.100	1.524
Std.	1.250	0.856	1.436	1.030	1.294	0.814
Deviation						
Minimum	1.000	2.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	5.000	3.000

Descriptive Statistics_Identifizierung

Descriptive Statistics

	KI_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	Peer_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.
Valid	21	21
Missing	0	0
Median	4.000	3.000
Mean	3.619	3.143
Std. Deviation	1.203	1.108
Minimum	2.000	1.000
Maximum	5.000	5.000

Descriptive Statistics_Affekt

Descriptive Statistics

	KI_Zufrieden	Peer_Zufrieden	KI_Zuversichtlich	Peer_Zuversichtlich	KI_Erfolgreich	Peer_Erfolgreich	KI_Beidigt	Peer_Beidigt	KI_Verärgert	Peer_Verärgert	KI_Frustriert	Peer_Frustriert
Valid	21	21	21	20	21	20	21	20	21	21	21	21
Missing	0	0	1	0	1	0	1	0	0	0	0	0
Median	3.000	4.000	3.000	4.000	3.000	4.000	1.000	1.000	1.000	1.000	1.000	1.000
Mean	3.143	4.000	2.952	3.850	2.857	3.700	1.810	1.350	2.095	1.381	1.667	1.429
Std. Deviation	1.062	1.049	1.024	1.040	1.108	1.174	1.250	0.813	1.480	0.805	1.155	0.870
Minimum	1.000	2.000	1.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	4.000	5.000	4.000

Paired Samples T-Test

Paired Samples T-Test

	Measure 1	Measure 2	Test	Statistic	z	df	p
KI_Mit dem Feedback bin ich zufrieden	- Peer_Mit dem Feedback bin ich zufrieden		Student	-2.152	20	0.044	
			Wilcoxon	26.500	-1.903	0.054	
KI_Ich betrachte das Feedback als fair.	- Peer_Ich betrachte das Feedback als fair.		Student	-1.474	20	0.156	
			Wilcoxon	50.500	-1.524	0.120	
KI_Ich betrachte das Feedback als gerechtfertigt.	- Peer_Ich betrachte das Feedback als gerechtfertigt.		Student	-1.522	19	0.144	
			Wilcoxon	31.000	-1.350	0.174	
KI_Ich betrachte dieses Feedback als nützlich.	- Peer_Ich betrachte dieses Feedback als nützlich.		Student	-0.383	20	0.706	
			Wilcoxon	45.500	-0.439	0.677	
KI_Ich finde dieses Feedback hilfreich.	- Peer_Ich finde dieses Feedback hilfreich.		Student	-0.791	20	0.438	
			Wilcoxon	39.000	-0.847	0.408	
KI_Dieses Feedback unterstützt mich.	- Peer_Dieses Feedback unterstützt mich.		Student	-1.471	20	0.157	
			Wilcoxon	39.500	-1.474	0.139	
KI_Dieses Feedback akzeptiere ich.	- Peer_Dieses Feedback akzeptiere ich.		Student	-2.274	20	0.034	
			Wilcoxon	34.500	-1.988	0.044	
KI_Diesem Feedback widerspreche ich.	- Peer_Diesem Feedback widerspreche ich.		Student	1.000	20	0.329	
			Wilcoxon	69.000	1.036	0.307	
KI_Dieses Feedback lehne ich ab.	- Peer_Dieses Feedback lehne ich ab.		Student	1.748	19	0.097	
			Wilcoxon	68.000	1.572	0.118	
KI_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	- Peer_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.		Student	1.227	20	0.234	
			Wilcoxon	68.000	0.973	0.335	
KI_Zufrieden	- Peer_Zufrieden		Student	-2.378	20	0.027	
			Wilcoxon	21.500	-2.187	0.028	
KI_Zuversichtlich	- Peer_Zuversichtlich		Student	-2.438	19	0.025	
			Wilcoxon	21.000	-2.215	0.026	
KI_Erfolgreich	- Peer_Erfolgreich		Student	-2.378	19	0.028	
			Wilcoxon	14.500	-2.166	0.031	
KI_Beileidigt	- Peer_Beileidigt		Student	1.422	19	0.171	
			Wilcoxon	39.500	1.223	0.235	
KI_Verärgert	- Peer_Verärgert		Student	1.917	20	0.070	
			Wilcoxon	37.000	1.718	0.090	
KI_Frustriert	- Peer_Frustriert		Student	0.706	20	0.489	
			Wilcoxon	28.500	0.711	0.511	

Assumption Checks

Test of Normality (Shapiro-Wilk)

		W	p
KI_Mit dem Feedback bin ich zufrieden	- Peer_Mit dem Feedback bin ich zufrieden	0.926	0.113
KI_Ich betrachte das Feedback als fair.	- Peer_Ich betrachte das Feedback als fair.	0.908	0.051
KI_Ich betrachte das Feedback als gerechtfertigt.	- Peer_Ich betrachte das Feedback als gerechtfertigt.	0.939	0.230
KI_Ich betrachte dieses Feedback als nützlich.	- Peer_Ich betrachte dieses Feedback als nützlich.	0.954	0.404
KI_Ich finde dieses Feedback hilfreich.	- Peer_Ich finde dieses Feedback hilfreich.	0.951	0.352
KI_Dieses Feedback unterstützt mich.	- Peer_Dieses Feedback unterstützt mich.	0.958	0.484
KI_Dieses Feedback akzeptiere ich.	- Peer_Dieses Feedback akzeptiere ich.	0.943	0.254
KI_Diesem Feedback widerspreche ich.	- Peer_Diesem Feedback widerspreche ich.	0.901	0.037
KI_Dieses Feedback lehne ich ab.	- Peer_Dieses Feedback lehne ich ab.	0.928	0.139
KI_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	- Peer_Durch das Feedback weiß ich, wo und wie ich etwas verbessern kann.	0.847	0.004
KI_Zufrieden	- Peer_Zufrieden	0.942	0.241
KI_Zuversichtlich	- Peer_Zuversichtlich	0.942	0.259
KI_Erfolgreich	- Peer_Erfolgreich	0.916	0.084
KI_Beileidigt	- Peer_Beileidigt	0.860	0.008
KI_Verärgert	- Peer_Verärgert	0.815	0.001
KI_Frustriert	- Peer_Frustriert	0.866	0.008

Note. Significant results suggest a deviation from normality.

Anhang 2.2.2: Qualitative Daten

			Sie findet nicht sehr viel Wahrnehmung da ich finde das das Feedback von alrict st. Es hat einiges, vorstädiges und richtiges Feedback gegeben was auch stimmt. Es hat ein paar Fehler gemacht (nicht alle) und hat gezeigt was man nichts mit besser machen kann.
			Ich finde es etwas cool, dass Feedback 1 von einer KI generiert wurde, denn die "K" war vielleicht mal ein Programmierer. Ich finde es aber so eine gute Aufführung.
			Bei mir ändert sich nichts wobei der Schüler wahrscheinlich länger überlegen musste, außerdem sagt er, die KI schlägt als der Schüler meine Schwiegerin vor.
			ja auf Reit beim ersten Lesen sahen auf das von einer KI sein könnte, damit ändert sich meine Schwiegerin weiter. Außer dass ich der KI weniger vertraue.
			Es ändert sich nichts bei der Wahrnehmung wenn man nach einer Reittraining Sitzung kommt nur wenn er jetzt zum nächsten Reittraining kommt. Die KI kann sie nicht mehr einspielen weil sie zwei Mal auf den gleichen Autoren hin, die sie sagt dass kein Autoren da steht.
			Bei jedem Reit kann sie mir nicht mehr helfen weil sie nicht mehr weiß wo ich mich befinden soll.
			Es ändert sich bei mir nichts weil ich schon das Feedback 2 von der KI geschrieben wurde. Es hat sich keine Veränderung in dem Text. Es ändert sich etwas weil es auch nicht mit vielen Veränderungen kommt und wenn man auch nach mir was will kann sie nicht mehr helfen.
			Es ändert sich das sie für das erste von einer Klassenübung aufgeschrieben wurde. Und in Feedback 2 dargestellte Beispiele waren, und in Feedback 1 eher kurz gehalten wird.
U11W2	JFH2	In Feedback 1 [K] wurden manche Sachen die man nicht richtig gesehen hat genannt und eine Verbesserungsvorschlag während man machen kann um sich zu verbessern. Bei Feedback 2 [K] konnte man nicht finden um sich zu verbessern.	Man hat bei Feedback 1 [K] wahrscheinlich darüber gesprochen das man sich gut verbessern konnte. Bei Feedback 2 [K] konnte man nicht finden um sich zu verbessern.
JFH2	JFH2	Feedback 1 war hilfreicher da die Fehler genannt wurden und man weiß was man verbessern soll. Feedback 2 war nicht wirklich genannt und war eher eine Sache und es wurden keine Fehler gefunden.	Feedback 1 war hilfreicher, aber auch was es von einem Kind war. Es gab Teile in den Feedbacks die mir gehören haben zum Beispiel das ich nicht mehr eigene Meinung mitnehmen darf.
U11J1	JFH2	Die Feedbacks wurden von zweien unterschiedlichen Personen geschrieben und einem Kind. Das merkt man.	Feedback 1 [K] war hilfreicher, aber auch was es von einem Kind war. Es gab Teile in den Feedbacks die mir gehören haben zum Beispiel das ich nicht mehr eigene Meinung mitnehmen darf.
JFH2	JFH2	In Feedback 2 [Peter] wurde ein netterer Beurteilung in Feedback 1 [K] gegenübergestellt.	In Feedback 2 [Peter] wurde ein netterer Beurteilung in Feedback 1 [K] gegenübergestellt.
U11J1	JFH2	Eigentlich verstand ich es nicht so sehr, obwohl es aufschlussreich war. Ich finde es eine Menge Unterschiede zwischen den beiden Feedbacks.	Eine Gemeinsamkeit die mir aufgefallen ist ist, dass beide sachliches und neutrales Feedback in einem einzigen haben.
JFH2	JFH2	Es, dass sie aufschlussreich waren.	Ich fand auf jeden Fall einen [Peter] besser weil, in Test zwei [K] viel falsch war und in Test eins hat mir alles gut ist. Die Rückmeldung in Text eins hat mir gelohnt.
U11M1	JHM1	Ich empfinde das Feedback als nicht infrage viele Fehler sind und manches Stehen nicht so gut wie man das das schlechter geschrieben hat.	In Feedback eins [Peter] steht doch in der Einleitung also genau was habe dich in der Feedback so gut gemacht.
JHM1	JHM1	Ja es geht auch weich machen. Stehen nicht vorne wenn man das das schlechter geschrieben hat.	Ich fand Feedback eins [Peter] besser weil es mehr detailliert war. Aber auch nicht sohart ist wie bei Feedback 1 [Peter]. Weil Feedback 1 es kurz beschreibt und es kann man nicht mehr so gut wie der Text der Kommentare des mir ja gehören hätte.
U11S1	LMS1	Trotzdem finde ich hier es ein anderer Unterschied weil ich finde wichtig die Fehler daran waren.	Das ich den Einleitungsatz so kurz habe und der Haupttext nicht so flüssig ist.
LMS1	LMS1	Ich finde wichtig bei der Überprüfung ob da schon Fehler sind und manches Stehen nicht so gut wie man das das schlechter geschrieben hat.	Es hat geholfen, das ich so viele Erklärungen bekommen habe was ich bei sonst nichts und verloren habe.
OFE3	ER, OFE3	Das Feedback ist gut.	Feedback 2 [K] war mehr detailliert ist.
ER, OFE3	ER, OFE3	Ich finde das Feedback sehr gut und stimme binisch überall zu.	Mir hat alles geholfen
OHNA1	OHNA1	Das Feedback ist gut.	
ER, OHNA1	OHNA1	Ich finde das Feedback sehr gut und stimme binisch überall zu.	

OSENNA1	Ich empfände das Feedback nicht als sehr hilfreich. Es war zwar wertvoll, aber es gab einen Fehler, der mich aufmerksam machen sollte, was ich falsch gemacht habe. Außerdem wurde im Feedback 2 verständlicher, erhebt in Text Weisungen vornehmlich.	Beide waren nicht sehr hilfreich.	Nein,	KI
KI_OSENNA1	Es wurde lediglich dazu aufgerufen, den Test nach Schleichen nicht zu machen, aber ohne weitere Informationen darf stehen.			
PIER_OSENNA1	Das Feedback ist direkt und verständlich. Ich kann sehr viel Verständigungssprache rausbekommen.	Feedback 2 [Peer] war viel kürzer und unabsichtlich im Gegensatz zu Feedback 1.	Es waren beide unterschiedlich und umgeglichen.	KI
PIER_BFRM3	Ich habe keine Probleme mit dem Feedback, wenn es darum geht, was ich falsch gemacht habe.	Feedback 2 [Peer] war gegenüber Feedback 1 sehr viel kürzer und unabsichtlich im Gegensatz zu Feedback 1.		KI
ZINNA1	Bei Feedback 2 [KI] ist noch Fazit, ich habe die Autoren hingehoben, ich sage es Besser man liest es mir und deutet es mir auf, das ist der Fehler, der gemacht wurde.	Feedback 1 [Peer] wurde nichts so viel geschrieben wie Feedback 2 [KI], weil es direkt schreibt, was mir fehlt, wie man es machen kann.	Eigentlich Feedback 2 [KI] war dort konkretizierter gemacht wurden und bei dem anderen wurde nur gesagt, was falsch gemacht wurde.	KI
KI_RN012	Bei Kommentar 2 wurde ich darüber informiert, dass ich mich nicht gut verstanden habe. Obwohl ich das gemacht habe, das andere verstanden haben.	Feedback 1 wurde einfach gesagt.		
PIER_INNOL	Ich finde es richtig, dass man nur über Feedback [KI] ist.			
SIMAA1	Füllt das Feedback gerechtfertigt.			
KI_SHB41	Die Übersetzung des Textes finde ich den teil kann passen, aber es fehlt etwas, was ich nicht weiß.	Die beiden Feedbacks sind komplett unterschiedlich. In Feedback 2 [KI] ist es viel mehr kritik.	In Feedback 2 [KI] gibt es Teile die mir gefallen haben über den Schreibprozess nachzudenken.	Peer
PIER_SHB41	Ich finde, das Feedback ist sehr gut formuliert, ich würde es allen Freunden geben und ich finde die Person, die das Feedback geschrieben hat, hat an keiner Stelle etwas schlechtes geschrieben. Deswegen ist kein Fehler.			
TEG02	Ich fand nichts, dass es aber trotzdem gerechtfertigt.	Ich finde beide Feedbacks gut für die Verbesserung des Textes nur in Feedback 2 [KI] werden mehr Seiten geschrieben, die ich nicht gelesen habe.		
KI_TDZ02	Ich fand nichts, dass es aber trotzdem gerechtfertigt.	Es wurde geschrieben, dass es noch Künster halten soll, das möchte ich verbessern.	Beide haben gerechtfertigt, dass es noch Künster halten soll. Das werde ich mir zu Hause nehmen.	KI
PIER_TDZ02	Ich fand, dass es richtig war, dass die Person gesagt hat, dass ich hätte schreiben sollen, dass sie mich Künster halten soll, das möchte ich verbessern.	Es wurde geschrieben, dass es noch Künster halten soll. Das werde ich mir zu Hause nehmen.		

Anhang 3: Codierleitfaden

Hauptkategorie	Code-Memo/Definition	Subkategorie	Code-Memo/Definition	Beispiel	Teilnehmende berichten/fockussieren.	Das Feedback ist positiv und/oder sachlich formuliert	„Sie bezeichnet die Schwächen sehr freundlich und so das man mehr die Schüler*in angibt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)
Nützlichkeit Peer-feedback	Umfasst Äußerungen der Teilnehmer, die das erhaltenen Feedback durch Peers als nützlich oder hilfreich bezeichnen.	Nützlichkeit Peer-feedback > Positiv	Umfasst Peerfeedback, das als nützlich oder hilfreich bezeichnet.	„Ich empfinde dieses Feedback sehr hilfreich.“ (6_ASIM1)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)
Zustimmung	Umfasst Äußerungen der Teilnehmer, die dem Feedback durch die Peers klar zustimmen oder wider sprechen bzw. es als fair oder nicht fair einordnen.	Zustimmung Peer-feedback > Positiv	Umfasst Äußerungen, die das Feedback als fair bezeichnen.	„Ich finde es nicht hilfreich!!!! Der oder diejenige hat immer wieder rein als Antwort geschrieben, ob sie ja oder nein nicht gesagt waren.“ (6_AMRA1)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)
Zustimmung	Zustimmung Peer-feedback > Negativ	Zustimmung Peer-feedback > Negativ	Umfasst Äußerungen, die dem Feedback als fair bezeichnen.	„Ich finde es nicht hilfreich!!!! Der oder diejenige hat immer wieder rein als Antwort geschrieben, ob sie ja oder nein nicht gesagt waren.“ (6_AMRA1)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)	„Umfasst Äußerungen, die die Schüler*in nicht anzeigt, dass Formulierung im Feedback positiv und/oder sachlich ist.“ (6_AFN4)
Affektwirkung Peer-feedback	Umfasst Äußerungen der Teilnehmer, die in denen im Zusammenhang mit dem Feedback durch die Peers eine emotionale Reaktion oder ein Gefühl verbleibt.	Affektwirkung Peer-feedback > Positiv	Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine positive emotionale Reaktion oder ein positives Gefühl verbleibt	„Ich finde, dass es schlechter bewertet wurde.“ (7_FRA2)	„Umfasst Äußerungen, die dem Feedback nicht zustimmen.“	„Umfasst Äußerungen, die dem Feedback nicht zustimmen.“	„Umfasst Äußerungen, die dem Feedback nicht zustimmen.“
Affektwirkung Peer-feedback	Affektwirkung Peer-feedback > Negativ	Affektwirkung Peer-feedback > Negativ	Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine negative emotionale Reaktion oder ein negatives Gefühl verbleibt	„Du hast sehr viele Hinweise gegeben und trotzdem nicht so viel geschrieben. Das finde ich toll.“ (6_EHAA1)	„Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine positive emotionale Reaktion oder ein positives Gefühl verbleibt wird.“	„Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine negative emotionale Reaktion oder ein negatives Gefühl verbleibt wird.“	„Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine positive emotionale Reaktion oder ein positives Gefühl verbleibt wird.“
Feedback-eigenelemente Peer-feedback	Umfasst die Nennung von Elementen oder Merkmalen der Eigenchaften des Feedbacks	Das Feedback	Das Feedback im Punkt im Schüler*innenkontext	„Ich bin nicht zufrieden.“ (7_IFGM0)	„Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine positive emotionale Reaktion oder ein positives Gefühl verbleibt wird.“	„Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine negative emotionale Reaktion oder ein negatives Gefühl verbleibt wird.“	„Umfasst Äußerungen, die in denen im Zusammenhang mit dem Feedback durch die Peers eine positive emotionale Reaktion oder ein positives Gefühl verbleibt wird.“
Feedback-kombiniert	Das Feedback kombiniert Lob und Kritik	Das Feedback	Das Feedback, lobt und kritisiert Peers, die die	„Also Feedback 2 (Peer) hat alles sehr nett ausgedrückt und auch die Schwächen gesagt.“ (6_AFN4)	„Umfasst Äußerungen, die in einem Satz sowohl lob als auch Kritik nennen.“	„Umfasst Äußerungen, die in einem Satz sowohl lob als auch Kritik nennen.“	„Umfasst Äußerungen, die in einem Satz sowohl lob als auch Kritik nennen.“

	Das Feedback hat Fehler.	Umfasst Äußerungen der TeilnehmerInnen, die das Feedback nicht erkannt haben, obwohl alle wichtigen Informationen drin stehen. (6_CSBW1)	„Es wurde Geschriebenes das man nicht erfahren hat, obwohl alle wichtigen Informationen drin stehen.“ (6_CSBW1)
	Eine Überarbeitung ist mit dem Feedback nicht möglich	Umfasst Äußerungen, in denen das Feedback nur verwiesen auf etwas gemacht wird, was durch das Feedback keine Überarbeitung möglich ist.	„In Feedback 2 [Peer] wurde geschrieben das fast alles gut war aber eine Sache und es wurden keine Fehler genannt sodass man sie sich nicht verbessern kann.“ (6_FIN2)
Nützlichkeit KI-Feedback	Nützlichkeit KI-Feedback > Positiv	Nützlichkeit KI-Feedback > Negativ	Nützlichkeit KI-Feedback > Negativ
Zustimmung KI-Feedback & Akzeptanz	Zustimmung KI-Feedback > Positiv	Zustimmung KI-Feedback > Negativ	Zustimmung KI-Feedback > Negativ
Affektivierung KI-Feedback	Affektivierung KI-Feedback > Positiv	Affektivierung KI-Feedback > Negativ	Affektivierung KI-Feedback > Negativ

Kenntnis der Quelle	der Wahrnehmung nach Kommunikation mit der Quelle gefragt werden und eine Antwort notieren.	Peer-Feedback	wird, die Rechtfertigung jedoch nicht eindeutig ist.
	Teilnehmende vertrauen dem Menschen.	Umfasst Äußerungen, bei denen das Peer-Feedback weiterhin präferiert wird, weil die Teilnehmenden einem Human-Feedback mehr vertrauen.	„Nein, denn die Aussagen von beiden Seiten ähnlich sind und ich fände es Sitzsam bei den KI ligenz ist nicht so gut wie man kann auch sein das ich Feedback 2 [Peer] gewählt habe!!“ (6.-EKTSS)
Beide Feedbacks sind gut, das Feedback durch die Peers ist besser	Beide Feedbacks sind gut, das Feedback wird, obwohl das KI-Feedback auch als gut eingestuft wird.	Umfasst Äußerungen, bei denen das Peer-Feedback weiterhin präferiert wird, weil es von der KI (field) geschrieben wurde und ein paar Fehler kann unterlaufen, zudem werden die KI nur kritische Fehler an. Was ziemlich blöd ist,	„Nein, denn die Äußerungen, bei denen das Peer-Feedback weiterhin präferiert wird, weil das Feedback durch die KI fehlerhaft ist.“
	Die KI ist fehlerhaft	„Bei Feedback war alles genauso wie mir eigentlich erhofft hatte z.B. wie ich mir erhofft hatte. Bei Feedback 1 war von der KI (field) geschrieben wurde sind ein paar Fehler kann unterlaufen, zudem werden die KI nur kritische Fehler an. Was ziemlich blöd ist, weil man nicht richtig sauer und frustriert ist.“ (6.-ASLMI)	„Bei mir ändert sich nicht meine Wahrnehmung, da ich es schon wusste, dass das KI ist halb irgendeiner kein Mensch bzw. nicht auf der Mensch-Ebene deshalb finde ich somit manchmal sowieso nutzen wir mal müssen.“ (6.-ELHAL)
Teilnehmende haben schon vorher vermutet, dass andere Feedback durch die KI kommt	Differenzierter, vertieftes Verständnis und Empathie gegenüber der Peers	Umfasst Äußerungen, bei denen das Peer-Feedback weiterhin präferiert wird, weil die Teilnehmenden vorher schon vermutet haben, dass es sich bei dem anderen Feedback um ein KI-Feedback handelt.	„Nein eigentlich nicht, weil ich das eigentlich mir schon vorher vorgestellt habe. Aber trotzdem nehme ich das Feedback vom Schüler oder von der Schulleiter aber auch ein, gleichzeitig aber auch ein differenziertes Verständnis im Hinblick auf die Feedbackquellen und den Peers. Gegenüber Empathie gezeigt wird.“
	Präferenz bleibt bei KI-Feedback	Umfasst Äußerungen, bei denen das Peer-Feedback weiterhin präferiert wird, gleichzeitig aber auch ein differenziertes Verständnis im Hinblick auf die Feedbackquellen und den Peers. Gegenüber Empathie gezeigt wird.“	„Nein eigentlich nicht, weil ich das eigentlich mir schon vorher vorgestellt habe. Aber trotzdem nehme ich das Feedback vom Schüler oder von der Schulleiter anders war. Beim einen weiß ich, dass es vielleicht ein wenig ungenauer oder nicht so viele Tipps gibt, weil diese Person hat und wahrscheinlich noch nicht so oft gemacht hat. Ich vertraue der KI aber trotzdem nicht 100% weil die KI kann auch Fehler machen.“ (6.-TGZ2D)