

Muhammed Akbulut, Sandra Reitbrecht & Sabine Schmölzer-Eibinger

Erfassung wissenschaftlicher Textqualität in der Sekundarstufe II: Entwicklung und Validierung eines Ratinginstruments für die Schreibaufgabe *Kontroversenreferat*

Während für die Messung wissenschaftlicher Textqualität von Studierenden bereits entsprechende Ratinginstrumente entwickelt wurden, liegen für den wissenschaftspropädeutischen Kontext der Schule bisher nur wenige vergleichbare Produkte vor. In diesem Beitrag werden die Entwicklung, theoretische Fundierung und empirische Validierung eines Ratinginstruments für die Schreibaufgabe *Kontroversenreferat* zur Erfassung wissenschaftlicher Textqualität in der Sekundarstufe II vorgestellt. Die Ergebnisse zeigen, dass das Instrument gute bis hervorragende Reliabilitätswerte aufweist und wissenschaftliche Textqualität differenziert erfassen kann. Damit steht ein theoretisch und empirisch fundiertes Verfahren für Forschung und Unterricht zur Verfügung.

Schlagwörter: *Kontroversenreferat* – Textrating – Wissenschaftliche Textqualität – Instrumentenentwicklung

Assessing Academic Text Quality in Upper Secondary Education: Development and Validation of a Rating Scale for the Writing Task *Kontroversenreferat*

While suitable rating instruments have already been developed to assess the quality of university students' academic texts, only a few comparable tools are available for the pre-academic context of upper secondary education. This article presents the development, theoretical grounding, and empirical validation of a rating instrument for the writing task *Kontroversenreferat* to assess academic text quality at the upper secondary level. The results show that the instrument yields good to excellent reliability values and is able to capture academic text quality in a differentiated way. It thus provides an empirically grounded tool for both research and classroom practice.

Keywords: *Kontroversenreferat* – text rating – academic text quality – scale development

1 Einleitung

Die Auseinandersetzung mit dem Erwerb wissenschaftlicher Textkompetenz in der Schule ist ein noch relativ junges Forschungsgebiet (Bushati et al. 2018), gewinnt aber angesichts der Tatsache, dass die Aneignung wissenschaftlicher Textkompetenz als ein sich länger gestaltender Prozess zu verstehen ist (Pohl 2007; Steinhoff 2007), in der schreibdidaktischen Forschung im deutschsprachigen Raum zunehmend an Bedeutung (z. B. Decker et al. 2021; Akbulut & Schmölzer-Eibinger 2021; Pohl 2019). Wissenschaftliche Textkompetenz kann als die Fähigkeit verstanden werden, den domänenspezifischen Diskurspraktiken der

Wissenschaft entsprechend schriftsprachlich zu handeln (Bushati et al. 2018, 14). Dies umfasst auf Prozessebene auch die Fähigkeit, wissenschaftliche Texte zu lesen und daraus gewonnene Inhalte in einem Text zu verarbeiten. Dazu zählt auch, unterschiedliche Zugänge und Sichtweisen zu einem Thema zu identifizieren, zu vergleichen und beim Schreiben aufeinander zu beziehen (Segev-Miller 2007; Mateos & Solé 2009; Solé et al. 2013; Schüler 2017; Akbulut et al. 2022). Wissenschaftliche Diskurse durch Lektüre von Fachtexten kennenzulernen ist auch eine wichtige Voraussetzung für die im Studium zu erwerbenden Fachkompetenzen (Decker et al. 2021, 39) und sollte bereits in der Schule angebahnt werden.

Aus Perspektive der Kompetenzmodellierung gestaltet sich wissenschaftliche Textkompetenz als komplexes Phänomen (Steinhoff 2007; Bushati et al. 2018) und lässt sich in mehrere Kompetenzbereiche untergliedern (Knorr 2019). Studien, die die Entwicklung wissenschaftlicher Schreibkompetenz untersuchen, sind daher mit der Herausforderung konfrontiert, ausreichend differenzierte, empirisch sowie theoretisch fundierte und zugleich praktikable Verfahren für die Bestimmung von wissenschaftlicher Schreibkompetenz über die Bewertung von Textqualitäten zu entwickeln. Während für die Messung wissenschaftlicher Textkompetenz von Studierenden im internationalen akademischen Kontext bereits entsprechende Ratinginstrumente entwickelt wurden (Cumming et al. 2021; Ghanbari & Barati 2020; Knight et al. 2020; Margolin et al. 2013), liegen für den wissenschaftspropädeutischen Kontext der Sekundarstufe im deutschsprachigen Raum bisher nur wenige vergleichbare Produkte vor (Akbulut et al. 2021; Decker & Kaplan 2014). Im vorliegenden Beitrag werden die Entwicklung, theoretische Fundierung und empirische Validierung eines Ratinginstruments für die Schreibaufgabe *Kontroversenreferat* (Feilke & Lehnen 2011) mit insgesamt zehn verschiedenen Kriterien vorgestellt. Das Ratingverfahren kommt derzeit in einer longitudinalen Interventionsstudie in der Sekundarstufe II (FWF-Projekt Entwicklung wissenschaftlicher Textkompetenz in der Sekundarstufe II, kurz: ElWisS, Projektleitung: Sabine Schmölzer-Eibinger & Muhammed Akbulut, 2024-2028, <https://wissenschaftliche-textkompetenz.uni-graz.at/de/>) zum Einsatz.

Hierzu wird zunächst die Schreibaufgabe *Kontroversenreferat* skizziert und es wird herausgearbeitet, welche Anforderungen die Bewältigung dieser Schreibaufgabe an die Schüler:innen stellt. Danach wird der Prozess der Entwicklung und Validierung des Ratingverfahrens näher beschrieben. Anschließend werden die zehn Kriterien des Ratingverfahrens mit Fokus auf ihre theoretische Fundierung und konkrete Operationalisierung dargestellt und einer empirischen Überprüfung unterzogen. Abschließend wird ein Schüler:innentext anhand des Ratingverfahrens exemplarisch analysiert, um die Ausführungen der vorangegangenen Kapitel zu veranschaulichen.

2 Das Kontroversenreferat als wissenschaftspropädeutische Schreibaufgabe

In der Schreibdidaktik hat sich das *Kontroversenreferat* (Steinhoff 2008; Feilke & Lehnen 2011; Schüler 2017; Bushati et al. 2018, Steinseifer et al. 2019; Akbulut & Schmölzer-

Eibinger 2021) als wissenschaftspropädeutische Schreibaufgabe mittlerweile etabliert. Es handelt sich dabei wie auch beim sog. Diskursreferat (Decker et al. 2021) um eine Schreibaufgabe, die materialgestütztes Schreiben erfordert (Schüler & Lehnen 2014) und das Verfassen eines wissenschaftlichen Textes evozieren soll, um die Grundlagen wissenschaftlichen Schreibens zu vermitteln und Schüler:innen auf das Schreiben komplexerer akademischer Textsorten wie der Seminar- oder Bachelorarbeit vorzubereiten. Gerade vor dem Hintergrund des zunehmenden Einsatzes generativer KI im schulischen (Scheiter et al. 2025) und universitären Schreiben (Gredel et al. 2024) kommt wissenschaftspropädeutischen Schreibaufgaben wie dem Kontroversenreferat eine besondere Bedeutung zu. Denn ein kompetenter, reflektierter und verantwortungsvoller Umgang mit KI setzt voraus, dass grundlegende Praktiken wissenschaftlichen Schreibens zunächst eigenständig und schrittweise erworben werden.

Zur Bewältigung dieser Schreibaufgabe müssen Schüler:innen unterschiedliche Standpunkte bzw. Erkenntnisse aus vorgegebenen wissenschaftlichen Texten beim Schreiben aufeinander beziehen, um daraus eine schlüssige Position bzw. ein schlüssiges Fazit zu entwickeln. Dafür gilt es, die Texte zunächst zu lesen, darin enthaltene wissenschaftliche Standpunkte bzw. Erkenntnisse zu identifizieren und schließlich in die argumentative Struktur des eigenen Textes unter Berücksichtigung domänenspezifischer Schreibkonventionen der Wissenschaftssprache zu integrieren (Steinseifer et al. 2019; Akbulut et al. 2022). Die rezipierten Inhalte sollen dabei nicht bloß aggregativ aneinandergereiht, sondern synthetisch aufeinander bezogen werden (Schüler & Lehnen 2014; Schüler 2017; Akbulut et al. 2022). Auf sprachlicher Ebene sind Schüler:innen beim Kontroversenreferat zudem gefordert, bestimmte Schreibhandlungen (z. B. Referieren, Vergleichen) in alltäglicher Wissenschaftssprache zu formulieren. Das Verfassen eines Kontroversenreferats stellt somit für Schüler:innen eine komplexe Herausforderung dar, deren Kenntnis nicht einfach vorausgesetzt werden darf (Bushati et al. 2018, 10). Gleichzeitig zeigen Ergebnisse aus empirischen Studien (Schüler 2017; Akbulut & Schmölzer-Eibinger 2021), dass das Kontroversenreferat dazu beitragen kann, die wissenschaftliche Textkompetenz von Schüler:innen zu verbessern.

Im ELWiS-Projekt wird das Kontroversenreferat im Rahmen einer longitudinalen Interventionsstudie eingesetzt, um die Entwicklung wissenschaftlicher Textkompetenz in der Sekundarstufe II zu untersuchen. Im Fokus steht dabei die Wirksamkeit unterschiedlicher genrebasierter Schreibfördermaßnahmen (Akbulut et al. 2025). Genrebasierte Ansätze sind in der Didaktik wissenschaftlichen Schreibens weit verbreitet und zielen darauf ab, Lernenden den Zugang zu den konventionalisierten sprachlich-textuellen Strukturen und Diskurspraktiken der Wissenschaft zu ermöglichen (Swales 1990; Johns 2008). Gleichzeitig stehen insbesondere explizite Formen der Genrevermittlung (*explicit genre teaching*) in der Kritik, zu stark reproduktiv und präskriptiv zu sein (Clark & Hernandez 2011). Vor diesem Hintergrund wurden alternative Ansätze wie der *genre awareness approach* entwickelt, die darauf abzielen, ein metakognitives und kontextsensibles Verständnis von Genres zu fördern und die flexible Anpassung Genrewissens an diverse Schreibkontexte zu ermöglichen (Johns 2008). Vor dem Hintergrund bislang fehlender empirischer Befunde

zur relativen Wirksamkeit dieser Ansätze untersucht das Projekt zwei genrebasierte Treatmentgruppen (Explicit Genre Teaching und Genre Awareness Approach) sowie eine Vergleichsgruppe (Implicit Genre Acquisition), die ausschließlich Feedback erhält (Akbulut et al. 2025). Anhand einer Stichprobe von 240 Schüler:innen wird über einen Zeitraum von drei Jahren (9.–12. Schulstufe) mit insgesamt sechs Messzeitpunkten die Entwicklung wissenschaftlicher Textkompetenz längsschnittlich erfasst. Im Folgenden wird skizziert, wie die Kontroversenreferate im Rahmen des ElWisS-Projekts in der Schule erhoben und aufbereitet wurden.

Der konkrete Schreibauftrag, der für die Bearbeitung unterschiedlicher Kontroversen aus dem Bereich der Gesundheitsbildung eingesetzt wurde, lautet in Anlehnung an Schüler und Lehnen (2014) wie folgt:

Liebe Schüler:innen,
stellt euch folgende Situation vor: Ihr habt die Sekundarstufe II erfolgreich absolviert und möchtet euch an einer Universität einschreiben. Für die schriftliche Bewerbung sollt ihr als Schreibprobe einen wissenschaftlichen Artikel (mind. 350 Wörter) zu einem vorgegebenen Thema verfassen.
Die Aufgabenstellung der Universität lautet:
Stellen Sie bitte die wissenschaftliche Kontroverse dar, die in den drei vorliegenden Fachtexten zur Frage „Ist es beim Krafttraining sinnvoll, bis zum Muskelversagen zu trainieren?“ geführt wird. Informieren Sie über den Stand der Forschung, indem Sie die unterschiedlichen Positionen und Argumente der Autor:innen miteinander vergleichen und diskutieren. Fassen Sie abschließend Ihre wichtigsten Erkenntnisse in einem Fazit zusammen und ziehen Sie daraus nachvollziehbare Schlussfolgerungen für die Forschung.

Abb. 1: Schreibauftrag

Die Schüler:innen erhielten für die Bearbeitung der Schreibaufgabe drei wissenschaftliche Texte (jeweils ca. 500 Wörter lang) und hatten für das Schreiben der Kontroversenreferate insgesamt 135 Minuten zur Verfügung. Die Texte wurden handschriftlich verfasst, da nicht alle beteiligten Klassen mit Laptops ausgestattet waren und die Verwendung generativer KI ansonsten kaum hätte kontrolliert werden können. Anschließend wurden die Texte abgetippt und anhand eines Transkriptionsmanuals grammatisch und orthographisch normalisiert. Diese Normalisierung erfolgte, um Verzerrungen des inhalts-, struktur- und sprachbezogenen Ratings durch sprachformale Fehler zu vermeiden, also um einen sogenannten Halo-Effekt (Ausstrahlungseffekt) bei der Textbeurteilung zu reduzieren, der in der Textbeurteilungsforschung gut belegt ist (Jansen et al. 2021).

3 Entwicklung und Validierung des Ratingverfahrens

Zur Erfassung der wissenschaftlichen Textqualität der Kontroversenreferate im ElWisS-Projekt wurde ein neues Ratinginstrument entwickelt. Dieses baut auf dem Instrument von Akbulut et al. (2021) auf, mit dem die Qualität von Kontroversenreferaten entlang der vier

Dimensionen *Inhalt*, *Synthese*, *Referieren* und *Domänenspezifisch* beurteilt wird, und stellt eine Weiterentwicklung dessen dar. In Anlehnung an aktuelle Erkenntnisse der Textbeurteilungsforschung (s. zusammenfassend Lindauer & Sommer 2018; Akbulut et al. 2021) wurde das ursprünglich stärker holistisch angelegte Instrument aus folgenden Gründen zugunsten einer analytischeren Konzeption weiterentwickelt: Zum einen zielt das Instrument darauf ab, die Qualität von Kontroversenreferaten möglichst differenziert zu erfassen, da im eLWiS-Projekt nicht nur die Beurteilung der Effektivität genrebasierter Schreibinterventionen im Fokus steht, sondern auch die Analyse spracherwerbsbezogener Entwicklungsverläufe wissenschaftlicher Textkompetenz über mehrere Schuljahre hinweg. Zum anderen war es aus forschungsstrategischen Gründen notwendig, das Verfahren so zu konzipieren, dass es nicht nur als Ratinginstrument, sondern auch als differenziertes Feedbackinstrument für Schüler:innen eingesetzt werden kann. Zu diesem Zweck wurde zum einen die Anzahl der Beurteilungskriterien im Vergleich zum Vorgängerinstrument (Akbulut et al. 2021) von vier auf zehn erhöht (Kap. 4). Bestehende Skalen wie *Inhalt* und *Synthese* wurden hierzu weiter ausdifferenziert, andere Skalen wie *Einleitung* und *Fazit* wurden neu in das Verfahren aufgenommen. Die Ausdifferenzierung und Erweiterung erfolgten auf Grundlage eines deduktiv-induktiven Vorgehens, bei dem sowohl theoretische Erkenntnisse der Forschung zur Entwicklung wissenschaftlicher Textkompetenz als auch empirische Analysen eines Pilotierungskorpus, bestehend aus 71 Kontroversenreferaten zu den Fragen „Sind wir allein im Universum?“ und „Warum können Affen nicht sprechen?“ iterativ verschränkt wurden.

Zum anderen wurde ein schüler:innengerecht aufbereiteter Feedbackbogen entwickelt, der alle Ratingkriterien verständlich erläutert und den Schüler:innen visualisiert, auf welcher Stufe sie sich jeweils befinden. Der Feedbackbogen enthält auch konkrete Hinweise dazu, welche Schritte notwendig sind, um die nächste Stufe zu erreichen. Folgende Abbildung zeigt einen Auszug aus dem Feedbackbogen in Bezug auf das Kriterium *Einleitung*.


<p>Dein Text führt in der Einleitung in das Thema ein.</p> <div style="text-align: right;">  </div> <p>TIPP: Versuche beim nächsten Mal, zusätzlich einen klaren Bezug zum wissenschaftlichen Diskurs herzustellen und die Leser:innen über das Ziel deines Textes zu informieren. Dann sind alle wesentlichen Elemente einer Einleitung in einem wissenschaftlichen Text vorhanden.</p>

Abb. 2: Auszug aus dem Feedbackbogen für Schüler:innen (Kriterium Einleitung)

Ergänzend kommen bei der Beurteilung der sprachlichen Kriterien *Sachlichkeit*, *Präzision* und *Wissenschaftstypische Formulierungen* (s. Kap. 4) farbige Markierungen direkt im Text sowie bei der Erfassung des Kriteriums *Inhaltliche Vollständigkeit* eine tabellarische Übersicht (s. Kap. 4) zum Einsatz, die den Schüler:innen transparent macht, welche Inhalte aus den Bezugstexten in ihrem Kontroversenreferat aufgegriffen und weiterverarbeitet wurden. Durch diese Vorgehensweise wird das Ratingverfahren nicht nur für die

Schüler:innen nachvollziehbarer, sondern auch der Interpretationsspielraum für die Rater:innen wird reduziert.

Um die Objektivität des Verfahrens weiter zu erhöhen, steht den Rater:innen (im EIWisS-Projekt wissenschaftliche Projektmitarbeiter:innen) ein umfangreiches Ratingmanual zur Verfügung, das sämtliche Beurteilungskriterien detailliert erläutert und anhand von Benchmarktexten aus dem Pilotierungskorpus veranschaulicht, wie die einzelnen Kriterien zu beurteilen sind. Der Einsatz von Benchmarktexten gilt als zentrales Mittel zur Erhöhung der Beurteiler:innenübereinstimmung, da sie eine gemeinsame Referenzbasis für die Einschätzung von Textqualität schaffen (Lindauer & Sommer 2018; Akbulut et al. 2021). Darüber hinaus enthält das Manual eine präzise Schritt-für-Schritt-Anleitung, die den Ablauf des Ratings klar strukturiert und eine möglichst einheitliche Vorgehensweise sicherstellt. Ergänzend dazu wurden die Rater:innen im Rahmen einer Rater:innen-Schulung auf den Einsatz des Ratingverfahrens vorbereitet. In dieser Schulung wurden zunächst die Ratingkriterien besprochen und anhand eines Beispieltextes gemeinsam angewendet. Anschließend beurteilten die Rater:innen vier weitere Texte einzeln, bevor in einer zweiten Besprechung die Ergebnisse diskutiert und offene Fragen sowie Unklarheiten geklärt wurden. Die statistische Überprüfung der Interrater-Reliabilität (Kap. 5) als Maß für die Objektivität des Verfahrens erfolgte durch die Berechnung des Intraklassenkorrelationskoeffizienten (ICC).

4 Theoretische Fundierung und Operationalisierung der Ratingkriterien

Die zehn Kriterien des Ratingverfahrens (Tab. 1) können zwei übergeordneten Dimensionen wissenschaftlicher Textqualität zugeordnet werden: Die ersten sechs Kriterien erfassen primär die inhaltlich-argumentative Qualität der Texte (Einleitung, Inhaltliche Vollständigkeit, Inhaltliche Genauigkeit und Nachvollziehbarkeit, Eigenständige Struktur, Intertextuelle Verknüpfungen, Fazit), während die letzten vier Kriterien primär die sprachlich-stilistischen Anforderungen der Domäne Wissenschaft operationalisieren (Präzision, Sachlichkeit, Wissenschaftstypische Formulierungen, Quellenangaben). Dabei wurde bewusst darauf verzichtet, hoch-inferentielle Skalen (z.B. zu „argumentativer Entfaltung“ oder „domänentypischem Sprachgebrauch“) einzusetzen. Stattdessen wurden die beiden Dimensionen wissenschaftlicher Textqualität analytisch über mehrere niedrig-inferentielle Einzelkriterien gestreut, um die Objektivität und Reliabilität des Verfahrens zu erhöhen.

Tab. 1: Ratingkriterien mit vierstufiger Skalierung

Kriterium	0 Punkte	1 Punkt	2 Punkte	3 Punkte
Einleitung	keine Einleitung erkennbar, sprunghafter Einstieg	Einführung in das Thema	Einführung in das Thema + Bezug zum wiss. Diskurs ODER Information zu Ziel des Beitrags	Einführung in das Thema + Bezug zum wiss. Diskurs + Information zu Ziel des Beitrags
Inhaltliche Vollständigkeit	0-5 wesentliche Inhalte	6-8 wesentliche Inhalte	9-11 wesentliche Inhalte	12 oder mehr wesentliche Inhalte
Inhaltliche Genauigkeit und Nachvollziehbarkeit	Die Inhalte der Bezugstexte werden inhaltlich nicht genau und nachvollziehbar dargestellt.	Die Inhalte der Bezugstexte werden inhaltlich teilweise genau und nachvollziehbar dargestellt	Die Inhalte der Bezugstexte werden inhaltlich weitgehend genau und nachvollziehbar dargestellt	Die Inhalte der Bezugstexte werden inhaltlich durchgängig genau und nachvollziehbar dargestellt
Eigenständige Textstruktur	Serielle Textzusammenfassungen mit sprunghaften Übergängen	Serielle Textzusammenfassungen mit kohärenten Übergängen	Teilweise eigenständige, kohärente Textstruktur	Eigenständige, durchgängig kohärente Textstruktur
Intertextuelle Verknüpfungen	keine intertextuellen Verknüpfungen	1-2 intertextuelle Verknüpfungen	3-4 intertextuelle Verknüpfungen	5 oder mehr intertextuelle Verknüpfungen
Fazit	Kein Fazit oder lediglich eine subjektive Meinungsäußerung	Einzelne (1-2) zentrale Erkenntnisse werden zusammengefasst	Mehrere (mind. 3) zentrale Erkenntnisse werden zusammengefasst ODER einzelne (1-2) zentrale Erkenntnisse werden zusammengefasst + diskursbezogene Schlussfolgerung	Mehrere (mind.3) zentrale Erkenntnisse werden zusammengefasst + diskursbezogene Schlussfolgerung
Präzision	Der Text ist nicht präzise formuliert.	Der Text ist teilweise präzise formuliert.	Der Text ist weitgehend präzise formuliert.	Der Text ist durchgängig präzise formuliert.
Sachlichkeit	Der Text ist nicht sachlich formuliert.	Der Text ist teilweise sachlich formuliert.	Der Text ist weitgehend sachlich formuliert.	Der Text ist durchgängig sachlich formuliert.
Wissenschaftstypische Formulierungen	Der Text ist nicht wissenschaftstypisch formuliert.	Der Text ist teilweise wissenschaftstypisch formuliert.	Der Text ist weitgehend wissenschaftstypisch formuliert.	Der Text ist durchgängig wissenschaftstypisch formuliert.
Quellenangaben	Informationen aus Bezugstexten werden nicht anhand von Quellenangaben belegt	Informationen aus Bezugstexten werden teilweise anhand von Quellenangaben belegt	Informationen aus Bezugstexten werden weitgehend anhand von Quellenangaben belegt	Informationen aus Bezugstexten werden durchgängig anhand von Quellenangaben belegt

Im Folgenden werden die Kriterien im Einzelnen dargestellt.

Kriterium 1: Einleitung

Eine wissenschaftliche Einleitung zeichnet sich im Kern dadurch aus, dass sie eine Gegenstandsdimension (Was ist das Thema/der Gegenstand des Beitrags?), eine Diskursdimension (Wie verhält sich der Beitrag zum bestehenden Wissensstand?) und eine Argumentationsdimension (Wie wird das Thema/der Gegenstand im Beitrag behandelt?) aufweist (Pohl 2007, 246). Je nach Disziplin (z.B. Geistes- vs. Naturwissenschaften), methodischer Ausrichtung (z.B. empirisch vs. hermeneutisch), Textsorte (z.B. Forschungsartikel vs. Handbuchartikel) und Umfang der Arbeit (z.B. 300-seitige Monographie vs. 15-seitiger Forschungsartikel) können die einzelnen Dimensionen unterschiedlich stark ausgebaut sein. Für das wissenschaftspropädeutische Aufgabenformat Kontroversenreferat (ca. 350 Wörter) wurde im Rahmen dieses Textratings festgelegt, dass diese drei Dimensionen erfüllt sind, wenn

1. das Thema des Kontroversenreferats eingeführt wird (z. B. „In den letzten Jahren gewinnt Social Media im Leben von Jugendlichen zunehmend an Bedeutung.“) > *Gegenstandsdimension*,
2. ein Bezug zum wissenschaftlichen Diskurs hergestellt wird (z. B. „Im aktuellen Forschungsdiskurs ist die Frage, welchen Einfluss Social Media auf die psychische Gesundheit von Jugendlichen hat, umstritten.“) > *Diskursdimension*,
3. das eigene Vorhaben benannt bzw. das Ziel des Beitrags deklariert wird (z. B. „Im Folgenden werden aktuelle Studien miteinander verglichen und Lücken in der Forschung aufgezeigt.“) > *Argumentationsdimension*.

Dementsprechend wurden bei der Beurteilung der Einleitungen die Rater:innen dazu aufgefordert, in einem Abgleich mit Benchmarktexten zu beurteilen, wie viele dieser drei Dimensionen (0–3 Punkte) in den Einleitungen der Schüler:innen erfüllt sind.

Kriterium 2: Inhaltliche Vollständigkeit

Beim Kriterium Inhaltliche Vollständigkeit wird beurteilt, in welchem Ausmaß es den Schreibenden gelingt, die für die Darstellung des Forschungsüberblicks wesentlichen Inhalte aus den Bezugstexten in ihrem Kontroversenreferat zu verarbeiten. Theoretisch speist sich das Kriterium zum einen aus der Relevanz von Leseprozessen für das wissenschaftliche Schreiben (Philipp 2020; Schüler 2017; Schüler & Lehnen 2014). Zum anderen verlangt das Kriterium von den Lernenden auch eine Informationsverdichtung von den Bezugstexten zum eigenen Text, wie sie in der linguistischen Fachliteratur als typisch für die Wissenschaftssprache Deutsch besprochen und über ein gehäuftes (und musterhaftes) Vorkommen nominaler Strukturen belegt wird (Brommer 2018, 167–184; Gołębiowski 2015, 59–60; Kretzenbacher 1991).

Zur Beurteilung des Kriteriums wurden vom Forscher:innenteam 15 wesentliche Inhalte aus den insgesamt drei Bezugstexten intersubjektiv festgelegt und in einer Referenztafel zusammengefasst. Anhand dieser Tabelle beurteilten die Rater:innen im Ratingprozess, wie viele dieser wesentlichen Inhalte in den Kontroversenreferaten aufgegriffen wurden. Die Punktevergabe erfolgt in folgenden Stufen: 12 bis 15 wesentliche Inhalte (3 Punkte), 9 bis 11 (2 Punkte), 6 bis 8 (1 Punkt), 0 bis 5 (0 Punkte).

Für die Anwendung des Ratingverfahrens auf andere Themen bzw. Aufgabenstellungen ist eine entsprechende Referenztafel jeweils neu zu erstellen. Abhängig von der Anzahl der Bezugstexte und der darin enthaltenen wesentlichen Inhalte kann dabei auch eine Anpassung der Schwellenwerte sinnvoll sein. Ein intersubjektives Vorgehen wie im hier beschriebenen Projekt ist jedoch empfehlenswert, da es dazu beiträgt, subjektive Bewertungen der Relevanz einzelner Inhalte zu reduzieren.

Kriterium 3: Inhaltliche Genauigkeit und Nachvollziehbarkeit

Ergänzt wird die *Inhaltliche Vollständigkeit* im Ratingverfahren um ein zweites inhaltsorientiertes Kriterium, nämlich die *Inhaltliche Genauigkeit und Nachvollziehbarkeit*, um auch die Qualität der Informationsverarbeitung in den Schüler:innentexten angemessen berücksichtigen zu können. Es werden damit den Schreibprozessen zugrundeliegende Leseprozesse beim materialgestützten Schreiben im Kontext der Schreibaufgabe Kontroversenreferat über eine Beurteilung der Korrektheit/Genauigkeit der Informationsverarbeitung beleuchtet (Philipp 2020). Zudem spielt im Rahmen dieses Kriteriums auch die Frage eine Rolle, ob bzw. in welchem Ausmaß es den Lernenden gelingt, die Inhalte der Bezugstexte losgelöst von der Situation, also für Lesende auch ohne Kenntnis der Ausgangstexte nachvollziehbar darzustellen (vgl. dazu die Vorstellung einer zerdehnten Sprechsituation beim Schreiben nach Ehlich (1984) sowie auch den umfassenden Diskurs zu *Bildungssprache und konzeptioneller Schriftlichkeit*).

Im Ratingverfahren werden erneut vier Stufen unterschieden, die auch in den Deskriptoren zu den vier Stufen (0 bis 3 Punkte) mit graduellen Abstufungen verbalisiert werden (Die Inhalte der Bezugstexte werden inhaltlich nicht/teilweise/weitgehend/durchgängig genau und nachvollziehbar dargestellt.).

Kriterium 4: Eigenständige Struktur

Wie bereits in Kap. 2 eingehend beschrieben, ist das Kontroversenreferat eine Schreibaufgabe, die hohe Anforderungen an die Synthesefähigkeiten der Schüler:innen stellt. Synthese setzt dabei u. a. voraus, dass Schüler:innen eine eigenständige, von den Bezugstexten losgelöste argumentative Textstruktur entwickeln. Ziel ist die Bildung einer sogenannten Makroproposition (Segev-Miller 2007), bei der aus den Bezugstexten übernommene Inhalte in eine neue, übergeordnete Struktur überführt werden. Studien zur Ontogenese von Synthesefähigkeiten bei wissenschaftlichen Schreibnoviz:innen (Segev-Miller 2007; Schüler 2017; Steinhoff 2007; Pohl 2007) zeigen einen Entwicklungsverlauf von lokal ausgerichteten, an vorgegebenen Strukturen orientierten Textorganisationsformen hin zu zunehmend textübergreifend angelegten, eigenständigen Organisationsformen.

Zur Beurteilung dieses Kriteriums im Rahmen des Ratingverfahrens wurden die Rater:innen dazu aufgefordert, im Abgleich mit Benchmarktexten 0 bis 3 Punkte anhand folgender Deskriptoren zu vergeben:

- 0 Punkte: Die Bezugstexte werden seriell zusammengefasst. Die Übergänge von einer Zusammenfassung zur nächsten sind sprunghaft.
- 1 Punkt: Die Bezugstexte werden seriell zusammengefasst. Die Übergänge von einer Zusammenfassung zur nächsten sind kohärent.

- 2 Punkte: Der Text verfügt teilweise über eine eigenständige Strukturierung, ist jedoch noch nicht durchgängig kohärent nach textübergreifenden inhaltlichen Gesichtspunkten organisiert.
- 3 Punkte: Der Text verfügt über eine durchgängig kohärente, eigenständige Strukturierung entlang textübergreifender inhaltlicher Gesichtspunkte.

Es ist zu beachten, dass das Kriterium der eigenständigen Strukturierung nicht vollständig unabhängig vom nachfolgenden Kriterium der intertextuellen Verknüpfungen ist, da die Möglichkeit, solche Verknüpfungen herzustellen, bei Texten, die stark an den Bezugstexten orientiert sind, eingeschränkt ist. Gleichwohl erscheint eine getrennte Erfassung beider Kriterien sinnvoll, da eine eigenständige Textstruktur nicht notwendigerweise mit expliziter Intertextualität einhergeht: Ein Text kann über eine eigenständige, thematisch strukturierte Organisation verfügen, ohne dass dabei die Beziehungen zwischen den Bezugstexten explizit gemacht werden.

Kriterium 5: Intertextuelle Verknüpfungen

Synthese erfordert von Schüler:innen nicht nur, dass sie eine eigenständige Textstruktur entwickeln, sondern auch in der Lage sind, Inhalte aus unterschiedlichen Quellen innerhalb dieser Struktur aufeinander zu beziehen (Schüler & Lehnen 2014, 225; Schüler 2017, 258). Wissenschaftliche Texte weisen eine spezifische Form von Intertextualität auf, die als *explizite Intertextualität* bezeichnet wird (Steinhoff 2007). Explizite Intertextualität wird durch das Kontroversenreferat evoziert, da dieses von den Schüler:innen verlangt, divergierende Inhalte aus mehreren Bezugstexten aufzugreifen und unterschiedliche Sichtweisen im eigenen Text diskursiv zu verarbeiten (Steinhoff 2007, 346; Akbulut et al. 2022, 24).

Bei der Beurteilung der intertextuellen Verknüpfungen im Rahmen des Ratingverfahrens werden ausschließlich solche Verknüpfungen berücksichtigt, die von den Schüler:innen *eigenständig* hergestellt und nicht lediglich aus den Bezugstexten übernommen werden. Zudem müssen die Verknüpfungen *explizit*, z. B. durch Nennung der Bezugstexte oder eindeutige sprachliche Verweise, erfolgen; implizite inhaltliche Verknüpfungen ohne solche Verweise werden folglich nicht gewertet. Darüber hinaus wurde festgelegt, dass die Verknüpfungen *konkrete* inhaltliche Zusammenhänge herstellen müssen, bloß kursorische Feststellungen allgemeiner Gemeinsamkeiten – etwa die Aussage, dass alle drei Bezugstexte das Thema X behandeln – werden dementsprechend nicht als intertextuelle Verknüpfungen gewertet. Die Punktevergabe orientiert sich an der Anzahl der identifizierten Verknüpfungen: Wenn im Kontroversenreferat keine intertextuellen Verknüpfungen vorzufinden sind, werden 0 Punkte vergeben, für ein bis zwei Verknüpfungen 1 Punkt, für drei bis vier Verknüpfungen 2 Punkte und für fünf oder mehr Verknüpfungen 3 Punkte.

Kriterium 6: Fazit

Neben der Einleitung kommt dem Fazit als zweitem rahmendem Textelement eine zentrale Funktion in der Ausbildung der argumentativen Zielorientierung einer wissenschaftlichen Arbeit (und somit auch eines Kontroversenreferats) zu. Pohl (2007, Kap. 5) zeigt in diesem Zusammenhang, dass es erst in fortgeschrittenen Stadien wissenschaftlicher Schreibentwicklung gelingt, Texte in *konklusiv-epistemischen* Fazits münden zu lassen, in denen die

wichtigsten Erkenntnisse der argumentativen Auseinandersetzung mit dem Forschungsgegenstand zusammengefasst und daraus diskursbezogene Schlussfolgerungen abgeleitet werden. Auf dem Weg dorthin beinhalten die Texte wissenschaftlicher Schreibnoviz:innen oft keine Schlussteile oder enden mit *subjektiv-evaluierenden* Appendizes in Form von Empfehlungen oder persönlichen Meinungsäußerungen.

Anhand dieses theoretischen Bezugsrahmens und im Abgleich mit dem Pilotierungskorpus wurde daher festgelegt, dass 3 Punkte im Ratingverfahren nur dann vergeben werden, wenn mehrere (mind. 3) zentrale Erkenntnisse der argumentativen Auseinandersetzung mit dem Thema zusammengefasst und eine epistemische Konklusion (diskurs-bezogene Schlussfolgerung für die Forschung, z. B. durch Nennung eines Forschungsdesiderats oder Geben eines Forschungsausblicks) daraus gezogen werden. Eine Beurteilung des Fazits mit 2 Punkten bedeutet, dass sich die Lernenden einem konklusiv-epistemischen Fazit bereits annähern, aber entweder nur einzelne (1–2) zentrale Erkenntnisse zusammengefasst werden oder eine diskursbezogene Schlussfolgerung fehlt. Für knappe Fazits, die lediglich einzelne (1–2) zentrale Erkenntnisse zusammenfassen, wird 1 Punkt vergeben und rein subjektiv-evaluierende Fazits (Pohl 2007), in denen Lernende ausschließlich persönliche Meinungsäußerungen oder Empfehlungen kundtun, werden mit 0 Punkten beurteilt.

Kriterium 7: Präzision

Wissenschaftliche Texte erfordern ein hohes Maß an sprachlicher Präzision, um wissenschaftliche Sachverhalte klar, eindeutig und prägnant darzustellen (Kühtz 2011, 14). In Texten von wissenschaftlichen Schreibnoviz:innen ist die sprachliche Präzision häufig aus zwei Gründen beeinträchtigt. Zum einen neigen Noviz:innen dazu, wissenschaftlichen Sprachgebrauch nachzuahmen, etwa durch die Verwendung selbst kreierter Fachbegriffe, eines übersteigerten Nominalstils oder hochkomplexer Satzgefüge (Steinhoff 2007, 423). Zum anderen verfügen wissenschaftliche Schreibnoviz:innen, insbesondere solche im Sekundarstufenbereich, häufig über ein noch wenig entwickeltes lexikalisches Ausdrucksrepertoire im Bereich der Wissenschaftssprache, was zur Verwendung semantisch inadäquater Formen führen kann.

Auf Grundlage eines Pilotierungskorpus wurden unterschiedliche Formen unpräzisen Sprachgebrauchs in Kontroversenreferaten identifiziert und in einer Referenztabelle zusammengefasst. Im Folgenden ist ein gekürzter Auszug aus der Tabelle zur Veranschaulichung aufgeführt:

- *unpräzise Nomen*, z.B. die *Verbindung* zwischen Bildschirmzeit und psychischem Wohlbefinden ist negativ (präzise: der Zusammenhang ist negativ)
- *unpräzise Verben*, z.B. dass Online-Interaktionen weniger Stress *hervorholen* (präzise: Stress hervorrufen)
- *unpräzise Adjektive*, z.B. eine *große* Nutzung von sozialen Medien (präzise: häufige Nutzung)
- *unpräzise Referenz* (Pronomen, Artikel, Demonstrativa, etc.), z.B. Hierzu wurden Studien mit Menschen durchgeführt. Die zeigt, dass... (präzise: Diese zeigen, dass...)
- *unpräzise Fachbegriffe*, z.B. generalistische Angststörung (präzise: generalisierte Angststörung)

Zur Beurteilung der sprachlichen Präzision markieren die Rater:innen anhand der Referenztablelle zunächst alle unpräzisen Formulierungen in den Schüler:innentexten **lila**. Anschließend werden die Texte auf Basis des visuellen Gesamteindrucks sowie im Abgleich mit Benchmarktexten auf einer Skala von 0 bis 3 Punkten als nicht präzise (0 Punkte), teilweise präzise (1 Punkt), weitgehend präzise (2 Punkte) oder durchgängig präzise (3 Punkte) beurteilt.

Kriterium 8: Sachlichkeit

Ein wesentliches Kriterium wissenschaftlicher Texte ist ihre Sachlichkeit (Eco 1973/2020, 191), die sich zum einen aus ihrer Gegenstandsbindung (Steinhoff 2007, 10) ergibt und zum anderen aus der Notwendigkeit, wissenschaftliche Erkenntnisse objektiv, d.h. intersubjektiv nachvollziehbar und überprüfbar, darzustellen (Popper 1959/2002, 22). In Texten wissenschaftlicher Schreibnoviz:innen kann die Sachlichkeit beeinträchtigt sein, wenn aufgrund mangelnder Schreiberfahrung auf *Transpositionen* (Steinhoff 2007, 423-424), d. h. sprachlich-stilistische Mittel aus nichtwissenschaftlichen Domänen (z. B. Journalismus, Populärwissenschaft, Alltagssprache, schulischer Schreibunterricht), zurückgegriffen wird.

Auf Grundlage eines Pilotierungskorpus wurden unterschiedliche Formen von Transpositionen in Kontroversenreferaten identifiziert und in einer Referenztablelle zusammengefasst. Im Folgenden ist ein gekürzter Auszug aus der Tablelle zur Illustration aufgeführt:

- wertende Adjektive, z. B. es kann *schlimme* Folgen haben (sachlich: es kann negative Folgen haben)
- expressive Verben, z. B. soziale Medien *zerstören* die Psyche (sachlich: soziale Medien beeinflussen die Psyche negativ)
- alltagssprachliche Metaphern, z.B. übermäßiger Konsum kann einem viel *Zeit und Kraft stehlen* (sachlich: übermäßiger Konsum kann zeit- und kraftraubend sein)
- subjektive Meinungsäußerungen, z.B. *Ich finde, dass...* (sachlich: es kann davon ausgegangen werden, dass...)

Für das Rating der Sachlichkeit markieren die Rater:innen auf Grundlage der Referenztablelle alle unsachlichen Formulierungen in den Schüler:innentexten **gelb**. Anschließend wird auf Basis des visuellen Gesamteindrucks und im Abgleich mit Benchmarktexten auf einer Skala von 0 bis 3 Punkten beurteilt, ob der Text nicht sachlich (0 Punkte), teilweise sachlich (1 Punkt), weitgehend sachlich (2 Punkte) oder durchgängig sachlich (3 Punkte) formuliert ist.

Kriterium 9: Wissenschaftstypische Formulierungen

Wissenschaftliche Texte werden durch domänenspezifische sprachliche Handlungen konstituiert, wie z. B. das Referieren oder Kontrastieren (vgl. Bushati et al. 2018; Steinseifer 2018; Emmrich 2019). Eine besondere Rolle spielen dabei sog. Textprozeduren (Bachmann & Feilke 2014, Feilke & Rezat 2025) als konventionalisierte Handlungsmuster bzw. Moves als typische Abschnitte in wissenschaftlichen Genres (Swales 1990). Sie sind funktional mit dafür typischen Formulierungen verbunden und als textkonstituierende Elemente dafür entscheidend, ob ein Text als wissenschaftlich wahrgenommen wird (Steinhoff, 2007). Domänentypische Formulierungen indizieren somit einerseits wissenschaftliches Handeln in Texten, erlauben es andererseits aber auch, Texte als Resultat eines wissenschaftlichen

Sprachgebrauchs zu beschreiben und zu beurteilen. Das Vorhandensein oder Nichtvorhandensein wissenschaftstypischer Formulierungen bzw. deren lerner:innsprachliche Varianten können daher als Kompetenzindikatoren betrachtet und analysiert werden (Feilke & Rezat 2025; Steinhoff 2024).

Auf Grundlage handlungsorientierter theoretischer und empirischer Arbeiten zu wissenschaftstypischen Formulierungsmustern (z. B. Bushati et al. 2018; Kültz 2011; Steinhoff 2007; Swales 1990) und einer Analyse des Pilotierungskorpus wurden neun unterschiedliche Moves in Kontroversenreferaten identifiziert und sowohl domänentypische als auch lerner:innsprachliche Realisierungen dieser Moves in einer Referenztablelle zusammengefasst. Tab. 2 zeigt eine gekürzte Fassung der Referenztablelle.

Tab. 2: Domänentypische und lerner:innsprachliche Realisierung von Moves

Move	Domänentypische Realisierung	Lerner:innsprachliche Realisierung
Wissenschaftlichen Diskurs kommentieren	Die Frage ... wurde im Zuge vieler Studien untersucht.	Viele Forscher:innen haben sich die Frage gestellt, ob...
Eigenes Vorhaben beschreiben	Im Folgenden wird ... analysiert.	über ... werde ich Sie in diesem Text informieren.
Aussagen von Forscher*innen wiedergeben	Sharman et al. (2019) sind der Auffassung, dass...	Im 1. Absatz des 1. Fachtexes wird verdeutlicht, dass...
Aussagen und Erkenntnisse von Forscher*innen verknüpfen	Im Gegensatz dazu geht die Goldlöckchen-Hypothese davon aus, dass...	Im Gegensatz davon berichtet der 2. Text, dass...
Forschungsmethoden berichten	Auch untersucht wurden...	Um der Sache etwas näher zu kommen, wurden...
Forschungsergebnisse berichten	Es konnte gezeigt werden, dass...	Schlussendlich war es in der Studie so, dass...
Zentrale Erkenntnisse zusammenfassen	Zusammenfassend lässt sich sagen, dass...	Als Fazit kann man erkennen, dass...
Desiderate aufzeigen	... wurde bislang nicht ausreichend erforscht.	Es gibt bis jetzt fast keine Studien zu ...
Ausblick geben	Zukünftige Forschung sollte ... in den Blick nehmen.	In Zukunft sollte mehr darauf geschaut werden, ob...

Zur Beurteilung dieses Kriteriums markieren die Rater:innen anhand der Referenztablelle alle domänentypischen Move-Realisierungen grün und alle lerner:innsprachlichen Realisierungen türkis. Anschließend wird auf Basis des visuellen Gesamteindrucks und im Abgleich mit Benchmarktexten auf einer Skala von 0 bis 3 Punkten beurteilt, ob der Text nicht wissenschaftstypisch (0 Punkte), teilweise wissenschaftstypisch (1 Punkt), weitgehend wissenschaftstypisch (2 Punkte) oder durchgängig wissenschaftstypisch (3 Punkte) formuliert ist.

Kriterium 10: Quellenangaben

Mit dem Kriterium Quellenangaben wird beurteilt, ob die Lernenden in ihrem Kontroversenreferat explizit auf Quellen verweisen und in welchem Ausmaß (also wie konsequent und durchgängig) sie dies tun. Damit wird ermittelt, ob die Lernenden das Kriterium der expliziten Intertextualität als domänentypisches Merkmal wissenschaftlichen Schreibens umsetzen und wie sehr sie somit auch dem wissenschaftlichen Wert der Transparenz bzw. der intersubjektiven Überprüfbarkeit (Steinhoff 2007, 121-123) hinsichtlich der Bezüge zu fremden Texten in ihrem eigenen Text gerecht werden. Rein formale Aspekte (z. B. auch Uneinheitlichkeit), die einer eindeutigen Zuordnung von Informationen zum jeweiligen Bezugstext nicht zuwiderlaufen, führten nicht zu Punkteabzügen.

Beurteilt wird mit den vier Stufen des Kriteriums, ob die Informationen aus den Bezugstexten gar nicht (0 Punkte), teilweise (1 Punkt), weitgehend (2 Punkte) oder durchgängig (3 Punkte) mit Quellenangaben gekennzeichnet werden. Formale Aspekte stehen dabei im Hintergrund; maßgeblich ist, dass die zitierten Quellen im Sinne der oben genannten intersubjektiven Überprüfbarkeit eindeutig zugeordnet werden können.

5 Beurteiler:innenübereinstimmung

Die statistische Überprüfung der Interrater-Reliabilität (s. Kap. 4) als Maß für die Objektivität des Verfahrens erfolgte durch die Berechnung des Intraklassenkorrelationskoeffizienten (ICC). Zu diesem Zweck wurden insgesamt 121 Kontroversenreferate aus den Messzeitpunkten 1 bis 3 des ElWisS-Projekts zu drei unterschiedlichen Themen von zwei Rater:innen unabhängig voneinander beurteilt. Die Anzahl von 121 Texten ergibt sich daraus, dass zu jedem Messzeitpunkt angestrebt wurde, etwa 10–15 % des Gesamtkorpus parallel von zwei Rater:innen beurteilen zu lassen, was der gängigen Praxis entspricht, bei größeren Textkorpora die Interrater-Reliabilität anhand einer Teilstichprobe durch Double-Coding zu bestimmen (O'Connor et al. 2020). Da beide Rater:innen alle Texte beurteilt haben, die Auswahl der Rater:innen nicht zufällig erfolgte und um Milde-Strengere Tendenzen der Rater:innen berücksichtigen zu können, wurde der ICC (2) berechnet.

Die Ergebnisse sind in Tab. 3 dargestellt. ICC-Werte über 0,5 werden dem Interpretationsschema von Wirtz und Caspar (2002) folgend als ausreichend, Werte über 0,7 als gut und Werte über 0,9 als hervorragend betrachtet.

Tab. 3: Überprüfung der Interrater-Reliabilität (n=121 Texte)

Kriterium	Rater:in 1	Rater:in 2	ICC	Prozentuale Übereinstimmung
Einleitung	M=1,72; SD=0,77	M=1,72; SD=0,81	0,86	86,8%
Inhaltliche Vollständigkeit	M=1,68; SD=0,82	M=1,66; SD=0,79	0,89	85,1%
Inhaltliche Genauigkeit und Nachvollziehbarkeit	M=1,56; SD=0,62	M=1,45; SD=0,65	0,77	81,0%
Eigenständige Textstruktur	M=0,91; SD=0,74	M=0,82; SD=0,72	0,87	86,0%
Intertextuelle Verknüpfungen	M=0,42; SD=0,68	M=0,40; SD=0,68	0,93	93,4%
Fazit	M=1,12; SD=0,87	M=1,16; SD=0,93	0,84	76,0%
Präzision	M=2,08; SD=0,63	M=2,09; SD=0,56	0,80	86,0%
Sachlichkeit	M=1,99; SD=0,73	M=1,98; SD=0,64	0,84	85,2%
Wissenschaftstypische Formulierungen	M=1,42; SD=0,63	M=1,40; SD=0,65	0,81	84,3%
Quellenangaben	M=0,98; SD=0,83	M=0,98; SD=0,81	0,90	89,2%

Die Ergebnisse zeigen, dass das Ratinginstrument in allen Kriterien über eine gute bis hervorragende Inter-Rater-Reliabilität verfügt und somit das sprachdiagnostische Kriterium der Objektivität erfüllt.

Ergänzend zur Überprüfung der Interrater-Reliabilität wurde die Trennschärfe der einzelnen Kriterien analysiert. Die Trennschärfen für neun Kriterien liegen im Bereich von $r = .32$ bis $.57$ und sind damit als akzeptabel bis gut zu bewerten. Dies spricht dafür, dass die einzelnen Kriterien in hinreichendem Maße zur Erfassung des Gesamtkonstrukts beitragen.

Lediglich ein Kriterium (Inhaltliche Vollständigkeit) weist mit $r = .28$ eine geringere Trennschärfe auf, liegt jedoch noch im tolerierbaren Bereich und wurde vor dem Hintergrund seiner inhaltlichen Relevanz für die Erfassung wissenschaftlicher Textqualität im Instrument belassen.

Die in Tab. 3 ausgewiesenen Mittelwerte und Standardabweichungen zeigen zudem, dass die Texte der Messzeitpunkte 1–3 überwiegend im unteren bis mittleren Leistungsbereich liegen. Dies ist eine Folge der bewussten Konzeption des Instruments, das so angelegt wurde, dass zu Beginn der Intervention keine Deckeneffekte auftreten und Entwicklungsverläufe über die Zeit hinweg differenziert abgebildet werden können. Vorläufige Auswertungen der späteren Messzeitpunkte (4 und 5) weisen darauf hin, dass ein substantieller Anteil der Schüler:innen im Verlauf der Intervention höhere Punktbereiche (>20 Punkte) erreicht. Dies spricht dafür, dass das Instrument auch im oberen Leistungsbereich differenzierungsfähig ist. Eine detaillierte Analyse dieser Verteilungen wird im Rahmen zukünftiger Projektpublikationen erfolgen.

Im Folgenden wird nun ein Kontroversenreferat exemplarisch mithilfe des präsentierten Ratingverfahrens beurteilt, um die bisherigen theoretischen und empirischen Ausführungen zu veranschaulichen.

6 Rating eines Beispieltextes

Bei dem Beispieltext (abgebildet in der normalisierten Fassung) handelt es sich um ein Kontroversenreferat, das zum zweiten Messzeitpunkt zum Thema „*Ist es sinnvoll, beim Krafttraining bis zum Muskelversagen zu trainieren?*“ verfasst wurde. Der Text eignet sich gut zur Illustration, da er im Gesamtscore 14 von max. 30 Punkten erreicht und damit sehr nah am durchschnittlichen Leistungsniveau des zweiten Messzeitpunkts ($M=12,98$, $SD=3,86$, $n=247$) liegt.

Ist es sinnvoll, beim Krafttraining bis zum Muskelversagen zu trainieren? (02aen4_MZP2)

In den letzten Jahren ist die Zahl der Menschen, die sich **einen bestmöglichen Muskelaufbau wünschen**, gestiegen. **Es gibt nicht genügend relevante Studien, um festzustellen, was der beste Weg** zum Muskelaufbau ist. **Im derzeitigen wissenschaftlichen Diskurs ist man der Meinung, dass** es einen Unterschied zwischen dem Trainieren mit leichten oder schweren Gewichten **macht**. **Ebenso belegen Studien (Lasevicius, T., Schoenfeld, B. J., Silva Batista, C., De Souza Barros, T., Aihara, A. Y., Brendon, H., Longo, A. R., Tricoli, V., De Almeida Peres, B., & Teixeira, E. L. (2022)), dass** das Gesamttrainingsvolumen eine wichtige Rolle für Muskelaufbau und -kraft spielt.

Durch das Training bis zum Muskelversagen entsteht ein höherer metabolischer Stress, wobei allerdings nicht klar ist, ob es nicht einen **Schwellwert** für metabolischen Stress gibt und über diesem keine weiteren positiven Auswirkungen erzielt werden können. Auch bietet kontinuierliches Training bis zum Muskelversagen eine höhere Wahrscheinlichkeit, dass es zu einem Übertraining und auch einem Burnout kommen kann. Wenn man Training bis zum Muskelversagen in seinen Trainingsplan integrieren möchte, scheint es daher ratsam, dies mit Bedacht zu tun. Eine Empfehlung lautet daher, nur beim letzten Satz bis zum Muskelversagen zu **gehen** und in den **vorderen** Sätzen die geplanten Wiederholungen durchzuführen.

Beim Training mit leichten Gewichten sei es allerdings erforderlich, bis zum Muskelversagen zu trainieren, um so den Muskelaufbau zu stimulieren. **Dieselbe** Studie **kommt ebenso zu der Aussage, dass** bei Training mit schweren Gewichten das Training bis zum Muskelversagen keine positiven Vorteile mit sich bringt. Durch das Training bis zum Muskelversagen bei Anfängern kommt es zu einem subjektiv höheren Anstrengungsempfinden, was dazu führen kann, dass die Freude am Training **schrumpft**, dies wiederum steht der Aufrechterhaltung des Trainingsplans im Weg. Allerdings sollte das Training bis zum Muskelversagen individuell im Trainingsplan betrachtet werden. Bei höherer Frequenz im Training einer bestimmten Muskelgruppe sollte das Training bis zum Muskelversagen nur **bedacht** und sparsam eingesetzt werden. Diese Empfehlung ergibt sich aus der längeren Regenerationszeit zwischen 24 und 48 Stunden.

Durch die Forschung dieser Studien komme ich zu dem Schluss, dass das Training bis zum Muskelversagen für jede Person anders zu betrachten ist. Grundsätzlich würde **ich** Anfängern allerdings empfehlen, nicht bis zum Muskelversagen zu trainieren und zuerst **in die Routine hineinzukommen** und erst dann mit einem Coach abzuklären, ob das Training bis zum Muskelversagen persönlich sinnvoll wäre.

Abb. 3: Beispieltext

Kriterium 1: Einleitung

Der Text beginnt mit einer diskursbezogenen Einleitung, in der die Relevanz des Themas betont und ein Bezug zum aktuellen wissenschaftlichen Diskurs hergestellt wird. Das eigene Vorhaben bzw. das Ziel des Beitrags werden nicht explizit benannt. Da somit zwei der drei für Einleitungen festgelegten Dimensionen (Gegenstands- und Diskursdimension) erfüllt sind, erhält der Text für dieses Kriterium 2 Punkte.

Kriterium 2: Inhaltliche Vollständigkeit

Hinsichtlich der inhaltlichen Vollständigkeit zeigt sich, dass der Text nur einen begrenzten Teil der wesentlichen Inhalte aus den drei Bezugstexten aufgreift. Insgesamt werden sieben der vorab festgelegten 15 wesentlichen Inhalte verarbeitet, sodass das

Kontroversenreferat gemäß den Bewertungsstufen mit einem Punkt beurteilt wird. Konkret greift der Text folgende wesentlichen Inhalte in der dargestellten Reihenfolge auf:

- die begrenzte Anzahl aussagekräftiger Studien zum Training bis zum Muskelversagen (Bezugstext 1)
- den Unterschied zwischen Training mit leichten und schweren Gewichten (Bezugstext 3)
- die Bedeutung des Gesamttrainingsvolumens für Muskelaufbau und -kraft (Bezugstext 3)
- den durch Training bis zum Muskelversagen erzeugten metabolischen Stress sowie die Möglichkeit eines Schwellenwerts ohne zusätzliche Trainingseffekte (Bezugstext 2)
- das erhöhte Risiko von Übertraining und psychischem Burnout (Bezugstext 1)
- das erhöhte subjektive Anstrengungsempfinden, insbesondere bei Anfänger:innen und dessen mögliche Auswirkungen auf die langfristige Trainingsadhärenz (Bezugstext 3)
- die verlängerte Regenerationszeit nach Training bis zum Muskelversagen (Bezugstext 1)

Es fällt auf, dass die Auswahl der wesentlichen Inhalte einer gegenstandsorientierten Logik (Pohl 2007) folgt. Im Fokus steht die inhaltliche Beantwortung der Frage, ob Training bis zum Muskelversagen sinnvoll ist, während diskursive und argumentationsbezogene Aspekte weitgehend unberücksichtigt bleiben. Informationen zur Methodik der referierten Studien sowie zu Limitationen der bestehenden Forschung, die in den Bezugstexten enthalten sind und für einen umfassenden diskurs- und argumentationsorientierten Forschungsüberblick notwendig wären, werden weitgehend ausgeblendet.

Kriterium 3: Inhaltliche Genauigkeit und Nachvollziehbarkeit

Für das Kriterium der inhaltlichen Genauigkeit und Nachvollziehbarkeit erhält der Text 2 Punkte. Die aufgegriffenen Inhalte werden bis auf wenige Ausnahmen, die hier aus Platzgründen nicht näher ausgeführt werden können, in ausreichender Genauigkeit wiedergegeben. Auch die Nachvollziehbarkeit ist weitgehend gegeben, sie wird stellenweise dadurch beeinträchtigt, dass der Text abrupt zwischen unterschiedlichen Aspekten der Fragestellung hin- und herspringt und an einer Stelle (*Dieselbe Studie kommt ...*) nicht klar ist, auf welche Studie sich die referierte Aussage bezieht. Für Leser:innen ohne Kenntnis der Bezugstexte ist der Argumentationsgang daher insgesamt nicht durchgängig nachvollziehbar.

Kriterium 4: Eigenständige Struktur

Im Hinblick auf die eigenständige Strukturierung zeigt der Text Ansätze einer von den Bezugstexten losgelösten Organisation. Inhalte aus den unterschiedlichen Bezugstexten werden miteinander kombiniert und in eine neue Textstruktur überführt. Jedoch fehlt eine durchgängige Leser:innenführung entlang klarer, textübergreifender inhaltlicher Gesichtspunkte. Der Text erhält daher für dieses Kriterium 2 Punkte.

Kriterium 5: Intertextuelle Verknüpfungen

Für das Kriterium der intertextuellen Verknüpfungen werden 0 Punkte vergeben. Zwar weist der Text durch die Kombination von Inhalten aus unterschiedlichen Bezugstexten

eine ausgeprägte implizite Intertextualität auf, explizite, eigenständig hergestellte Verknüpfungen zwischen den Bezugstexten sind nicht erkennbar. An einer Stelle wird mit der Formulierung „Ebenso belegen Studien, dass ...“ der Versuch einer expliziten Verknüpfung unternommen, es bleibt jedoch implizit, worauf sich das Adverb „ebenso“ konkret bezieht. Zumal es sich hierbei um eine Verknüpfung handelt, die nicht textübergreifend ist, sondern aus einem der Bezugstexte übernommen wurde und somit nicht als eigenständige Leistung betrachtet werden kann.

Kriterium 6: Fazit

Das Fazit wird mit einem Punkt bewertet. Es folgt überwiegend einer subjektiv-evaluierenden Logik (Pohl 2007) und enthält dementsprechend persönlich formulierte Empfehlungen an Trainingsnoviz:innen. Nichtsdestotrotz greifen diese Empfehlungen zentrale Erkenntnisse der inhaltlichen Auseinandersetzung mit dem Thema auf, etwa die Einschätzung, dass Training bis zum Muskelversagen für Anfänger:innen nicht geeignet ist und grundsätzlich mit Bedacht eingesetzt werden sollte. Für eine höhere Bepunktung des Fazits wäre entweder eine etwas umfassendere Zusammenfassung zentraler Erkenntnisse oder die Ableitung einer expliziten diskursbezogenen Schlussfolgerung erforderlich gewesen.

Kriterium 7: Präzision

In Bezug auf die sprachliche Präzision ist der Text insgesamt als weitgehend präzise einzu-stufen (siehe **lila** Markierungen) und erhält hierfür 2 Punkte. Nur vereinzelt treten Unge-nauigkeiten im Ausdruck auf, etwa bei der Verwendung von „Schwellwert“ statt „Schwellenwert“ oder in Bezug auf die Referenz, da etwa nicht nachvollziehbar ist, worauf sich die Formulierung „dieselbe Studie“ bezieht.

Kriterium 8: Sachlichkeit

Der Text ist weitgehend sachlich formuliert (siehe **gelbe** Markierungen) und erhält für dieses Kriterium ebenfalls 2 Punkte. Lediglich an wenigen Stellen finden sich unsachliche sprachlich-stilistische Mittel, etwa expressive Verben wie *wünschen* oder *schrumpfen*, sowie alltagssprachliche Wendungen (*bis zum Muskelversagen gehen, in die Routine hineinkommen*).

Kriterium 9: Wissenschaftstypische Formulierungen

Für das Kriterium der wissenschaftstypischen Formulierungen (siehe **grüne** und **türkise** Markierungen) erhält der Text einen Punkt. Zwar wird auf einzelne wissenschaftstypische Moves zurückgegriffen, um den Diskurs zu kommentieren (*Im derzeitigen wissenschaftlichen Diskurs ist man der Meinung, dass ...*) oder um Aussagen von Forscher:innen wiederzugeben (*kommt zu der Aussage, dass ...*), die konkreten Realisie-rungen dieser Moves sind jedoch stellenweise lerner:innsprachlich. Die insgesamt ge-ringe Anzahl an Moves ist wie bereits angesprochen darauf zurückzuführen, dass der Text stark gegenstandsorientiert und nur in geringem Maße diskurs- und argumentations-orientiert konzipiert ist. Entsprechend fehlen insbesondere Moves wie das Berichten über Forschungsmethoden oder das Aufzeigen von Forschungsdesideraten, wie sie für einen wissenschaftlichen Forschungsüberblick üblich sind.

Kriterium 10: Quellenangaben

Im Hinblick auf das Kriterium der Quellenangaben erhält der Text einen Punkt, da zumindest an einer Stelle explizit auf eine Quelle verwiesen wird. Die formale Ausgestaltung der Quellenangabe, die im Rahmen der Intervention vermittelt wurde, ist zwar ungewöhnlich, spielt jedoch, wie in Kapitel 4 dargelegt, für die Beurteilung keine Rolle, da die Quelle eindeutig zugeordnet werden kann.

7 Fazit

In diesem Beitrag wurde ein theoretisch und empirisch fundiertes Ratinginstrument zur Beurteilung von Kontroversenreferaten in der Sekundarstufe II vorgestellt, das sowohl zur Erfassung wissenschaftlicher Textqualität als auch als Feedbackinstrument für Schüler:innen eingesetzt werden kann. Diese doppelte Ausrichtung hat sich bei der Entwicklung des Instruments als besonders fruchtbar erwiesen. Die Notwendigkeit, die Kriterien für Schüler:innen verständlich und transparent zu gestalten, etwa durch Tabellen oder farbliche Markierungen, hat zu einer kleinschrittigen und präzisen Operationalisierung beigetragen und damit die Reliabilität und Objektivität des Verfahrens erhöht. Gleichzeitig sind mit dem Instrument auch Einschränkungen verbunden. So ist es in seiner aktuellen Form eng an die im ElWisS-Projekt verwendeten Aufgabenstellungen und Bezugstexte gebunden. Für eine Anwendung auf andere Themen oder Aufgabenformate ist daher eine Anpassung erforderlich, insbesondere die Erstellung neuer Referenztabellen für die inhaltliche Vollständigkeit.

Aktuell wird das Rating- bzw. Feedbackinstrument im ElWisS-Projekt eingesetzt, um die Wirksamkeit genrebasierter Schreibfördermaßnahmen zur Entwicklung wissenschaftlicher Textkompetenz in der Sekundarstufe II longitudinal zu untersuchen. Für die zukünftige Forschung wären Erkenntnisse zur praktischen Implementierung des Feedbackinstruments im Unterricht von Bedeutung. Von Interesse wäre etwa, wie und unter welchen Bedingungen Lehrkräfte das Instrument einsetzen, welche Kompetenzen sie für einen zielführenden Einsatz benötigen und wie es in bestehende Schreib- und Beurteilungspraktiken integriert werden kann.

Vor dem Hintergrund des zunehmenden Einsatzes generativer KI-Systeme ergeben sich darüber hinaus neue Perspektiven für die Weiterentwicklung des Instruments. Zukünftige Studien sollten prüfen, inwieweit bestehende KI-gestützte Feedbacktools (z. B. FelloFish) geeignet sind, Teile des Ratingprozesses zu automatisieren. Dabei stellt sich insbesondere die Frage, in welchem Ausmaß KI-basierte Bewertungen mit menschlichen Ratings übereinstimmen und welche Aspekte wissenschaftlicher Textkompetenz sich automatisiert zuverlässig erfassen lassen. Gleichzeitig bietet das Instrument auch Potenzial als Reflexionsgrundlage im Unterricht, um mit Schüler:innen Unterschiede zwischen eigenständig verfassten und KI-gestützten Kontroversenreferaten sichtbar zu machen und den Einsatz von KI in wissenschaftlichen Kontexten kritisch zu reflektieren.

8 Literatur

- Akbulut, M., Ebner, C., & Schmölzer-Eibinger, S. (2021). Wissenschaftliche Textkompetenz von Schülerinnen und Schülern messen. Konstruktion und Evaluation eines holistisch-analytischen Ratingverfahrens. Knorr, Dagmar/Heine, Carmen (Hg.): *Schreibwissenschaft methodisch*. Frankfurt a. M.: Peter Lang, 165-182.
- Akbulut, M., & Schmölzer-Eibinger, S. (2021). Das Universum gleicht einem Fußball – oder doch nicht? Förderung wissenschaftlicher Textkompetenz in sprachlich heterogenen Klassen anhand naturwissenschaftlicher Kontroversen. *ÖDaF-Mitteilungen*, 37(2), 94–114. <https://doi.org/10.14220/odaf.2021.37.2.94>
- Akbulut, M., Schmölzer-Eibinger, S., & Rauter, E. (2022). „Zusammengefasst hat jede Hypothese einen wahren Kern“ – Herausforderungen der Synthese beim wissenschaftlichen Schreiben und wie diese von SchülerInnen gemeistert werden. *Didaktik Deutsch*, (52/53). <https://doi.org/10.21248/dideu.92>
- Akbulut, M., Reitbrecht, S., Gösweiler, A., Menditti, I., & Schmölzer-Eibinger, S. (2025). Developing genre awareness in novice academic writers in secondary education: A longitudinal intervention study on students' academic writing development. *Pedagogical Linguistics*. Advance online publication. <https://doi.org/10.1075/pl.24017.akb>
- Bachmann, T., & Feilke, H. (Hrsg.) (2014). *Werkzeuge des Schreibens. Beiträge zu einer Didaktik der Textprozeduren*. Fillibach bei Klett.
- Brommer, S. (2018). *Sprachliche Muster: eine induktive korpuslinguistische Analyse wissenschaftlicher Texte*. De Gruyter.
- Bushati, B., Ebner, C., Niederdorfer, L., & Schmölzer-Eibinger, S. (2018). *Wissenschaftlich schreiben lernen in der Schule*. Schneider.
- Clark, I. L., & Hernandez, A. (2011). Genre awareness, academic argument, and transferability. *The WAC Journal*, 22, 65–78.
- Cumming, A., Cho, Y., Burstein, J., Everson, P., & Kantor, R. (2021). Assessing academic writing. In X. Xi & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 107–151). Routledge. <https://doi.org/10.4324/9781351142403>
- Decker, L., Guschker, B., Hensel, S., & Schindler, K. (2021). *Wissenschaftliches Schreiben lernen in der Sekundarstufe II: Fachdebatte und Praxisprojekte*. wbv.
- Decker, L., & Kaplan, I. (2014). *Beurteilen mittels Ratingverfahren im Projekt AkaTex – Methoden* (AkaTex Working Papers, 5). 2., korrigierte Auflage. Siegen und Köln: Universität Siegen und Universität zu Köln.
- Eco, U. (1973/2020). *Wie man eine wissenschaftliche Abschlussarbeit schreibt: Doktor-, Diplom- und Magisterarbeit in den Geistes- und Sozialwissenschaften*. utb.
- Ehlich, K. (1984). Zum Textbegriff. In A. Rothkegel & B. Sandig (Hrsg.), *Text – Textsorten – Semantik* (S. 9–25). Buske.
- Emmrich, V. (2019). Kontroversen darstellen: Kontrastieren und Positionieren. In S. Steinseifer et al. (Hrsg.), *Eristische Literalität. Wissenschaftlich streiten – wissenschaftlich schreiben* (S. 209–244). Synchron.
- Feilke, H., & Lehnen, K. (2011). Wissenschaftlich Referieren: Positionen wiedergeben und konstruieren. *Der Deutschunterricht*, 5, 34–44.
- Feilke, H., & Rezat, S. (2025). *Textprozeduren: Werkzeuge für den Schreibunterricht–Grundlagen und unterrichtspraktische Anregungen*. Klett/Kallmeyer.
- Ghanbari, N., & Barati, H. (2020). Development and validation of a rating scale for Iranian EFL academic writing assessment: A mixed-methods study. *Language Testing in Asia*, 10(1), 17.
- Golebiowski, A. (2015). Zum Nominalstil der Wissenschaftssprache. In E. Blachut & A. Golebiowski (Hrsg.), *Kontroversen in der heutigen germanistischen Linguistik: Ansichten, Modelle, Theorien* (S. 49–64). Wrocław–Dresden.
- 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>
- Jansen, T., Vögelin, C., Machts, N., Keller, S. D., & Möller, J. (2021). Don't just judge the spelling! The influence of spelling on assessing second-language student essays. *Frontline Learning Research*, 9(1), 44-65. <https://doi.org/10.14786/flr.v9i1.541>
- Johns, A. M. (2008). Genre awareness for the novice academic student: An ongoing quest. *Language Teaching*, 41(2), 237–252. <https://doi.org/10.1017/s0261444807004892>

- Knight, S., Shibani, A., Abel, S., Gibson, A., & Ryan, P. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research*, 20, 1–20. <https://doi.org/10.17239/jowr-2020.12.01.06>
- Knorr, D. (2019). Sprachsensibles Kompetenzmodell wissenschaftlichen Schreibens. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 24 (1), 165–179.
- Kretzenbacher, H. L. (1991). Syntax des wissenschaftlichen Fachtextes. *Fachsprache: International Journal of LSP*, 13(1–2), 118–137.
- Kühtz, S. (2011). *Wissenschaftlich formulieren: Tipps und Textbausteine für Studium und Schule*. Schöningh.
- Lindauer, N., & Sommer, T. (2018). Verfahren der Textbeurteilung. Merkmale und Vorzüge eines holistischen Benchmarkratings. *Leseräume*, 5(5), 1–14.
- Margolin, B., Ram, D. & Mashias, R. (2013). Building and validating a tool for assessing academic writing skills. *International Journal of Linguistics*, 5(4), 188. <https://doi.org/10.5296/ijl.v5i4.3498>
- Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education*, 24(4), 435–451. <https://doi.org/10.1007/BF03178760>
- O'Connor, C., & Joffe, H. (2020). Intercode reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1–13. <https://doi.org/10.1177/1609406919899220>
- Philipp, M. (2020). Außer Schreiben nichts gewesen?! Materialgestütztes argumentatives Schreiben aus der Sicht der Leseforschung. *Mitteilungen des Deutschen Germanistenverbandes*, 67(2), 183–196.
- Pohl, T. (2007). *Ontogenese wissenschaftlicher Textkompetenz: Studien zur Ontogenese wissenschaftlichen Schreibens*. Niemeyer.
- Pohl, T. (2019). Eristisches Schreiben nach Bologna: Überlegungen zu Chancen und Grenzen einer veränderten Hochschulkultur. In M. Steinseifer, H. Feilke, & K. Lehnen (Hrsg.), *Eristische Literalität. Wissenschaftlich streiten – wissenschaftlich schreiben* (S. 327–345). Synchron.
- Popper, K. R. (1959/2002). *The logic of scientific discovery*. Routledge.
- Scheiter, K., Bauer, E., Omarchevska, Y., Schumacher, C., & Sailer, M. (2025). *Künstliche Intelligenz in der Schule. Eine Handreichung zum Stand in Wissenschaft und Praxis*. Online abrufbar unter https://www.empirischebildungsforschung-bmbfsfj.de/img/KI_Review_20250318_Veroeffentlichung.pdf
- Schüler, L. (2017). *Materialgestütztes Schreiben argumentierender Texte: Untersuchungen zu einem neuen wissenschaftspropädeutischen Aufgabentyp in der Oberstufe*. Schneider Verlag Hohengehren.
- Schüler, L., & Lehnen, K. (2014). Anbahnung wissenschaftlicher Schreib- und Textkompetenz in der Oberstufe: Textkonzeption und -komposition bei materialgestützten Schreibaufgaben. In T. Bachmann & H. Feilke (Hrsg.), *Werkzeuge des Schreibens: Beiträge zu einer Didaktik der Textprozeduren* (S. 223–246). Fillibach bei Klett.
- Segev-Miller, R. (2007). Cognitive processes in discourse synthesis: The case of intertextual processing strategies. In M. Torrance, L. Van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 231–250). https://doi.org/10.1163/9781849508223_016
- Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating information: An analysis of the processes involved and the products generated in a written synthesis Task. *Written Communication*, 30(1), 63–90. <https://doi.org/10.1177/0741088312466532>
- Steinhoff, T. (2007). *Wissenschaftliche Textkompetenz: Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Niemeyer.
- Steinhoff, T. (2008). Kontroversen erkennen, darstellen, kommentieren. In I. Bons, D. Kaltwasser, & T. Gloning (Eds.), *Fest-Platte für Gerd Fritz* (S. 1–13). http://festschrift-gerd-fritz.de/files/steinhoff_2008_kontroversen_erkennen_darstellen_und_kommentieren.pdf
- Steinhoff, T. (2024). Textprozeduren in Interventionsstudien: Methodische Verfahren und theoretische Fragen am Beispiel dreier Feldexperimente. In S. Rezat, E. Grundler, H. Feilke, & S. Schmölzer-Eibinger (Hrsg.), *Textprozeduren in Spannungsfeldern* (S. 95–114). Stauffenburg.
- Steinseifer, M. (2018). Referieren: Struktur und Didaktik einer intertextuellen Kompetenz. In S. Schmölzer-Eibinger et al. (Hrsg.), *Wissenschaftlich schreiben lehren und lernen* (S. 241–262). Waxmann.
- Steinseifer, M., Feilke, H., & Lehnen, K. (Hrsg.) (2019). *Eristische Literalität. Wissenschaftlich streiten – wissenschaftlich schreiben*. Synchron Verlag.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Hogrefe.

9 Korrespondenzangaben

Dr. Muhammed Akbulut, Universität Graz, Fachdidaktikzentrum Deutsch als Zweitsprache & Sprachliche Bildung

Prof. Dr. Sandra Reitbrecht, Otto-Friedrich-Universität Bamberg, Institut für Germanistik

Prof. Dr. Sabine Schmölder-Eibinger, Universität Graz, Fachdidaktikzentrum Deutsch als Zweitsprache & Sprachliche Bildung

Korrespondenz an: muhammed.akbulut@uni-graz.at

Danksagung

Einen besonderen Dank möchten wir allen Schüler:innen, Erziehungsberechtigten, Lehrpersonen und Schuldirektor:innen aussprechen, die die Umsetzung dieser Studie ermöglicht haben.

Funding

This research was funded in whole, or in part, by the Austrian Science Fund (FWF) [10.55776/ P36976]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.