

Lukas Mientus

---

# Reflexion und Reflexivität Befunde reflexionsbezogener Dispositionen

---

Dissertation zur Erlangung des Grades Doktor der Naturwissenschaften (Dr. rer.-nat.) in der  
Didaktik der Physik am Institut der Physik und Astronomie. Eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam im Juni 2023.

Soweit nicht anders gekennzeichnet, ist dieses Werk unter einem Creative-Commons-Lizenzvertrag Namensnennung 4.0 lizenziert.  
Dies gilt nicht für Zitate und Werke, die aufgrund einer anderen Erlaubnis genutzt werden.  
Um die Bedingungen der Lizenz einzusehen, folgen Sie bitte dem Hyperlink:  
<https://creativecommons.org/licenses/by/4.0/legalcode.de>

Disputation am 25.09.2023 unter Begutachtung durch

1. Prof. Dr. Andreas Borowski, Universität Potsdam,
2. Prof.in Dr.in Claudia von Aufschnaiter, Justus-Liebig-Universität Giesen und
3. Prof.in Dr.in Birte Friedrich, Universität Potsdam.

Weitere Mitglieder der Prüfungskommission:

- Prof. Dr. Helmut Prechtel (Vorsitz)
- Prof.in Dr.in Ulrike Lucke.

Online veröffentlicht auf dem

Publikationsserver der Universität Potsdam:

<https://doi.org/10.25932/publishup-61000>

<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-610003>



## **Zusammenfassung**

Reflexion gilt in der Lehrkräftebildung als eine Schlüsselkategorie der professionellen Entwicklung. Entsprechend wird auf vielfältige Weise die Qualität reflexionsbezogener Kompetenzen untersucht. Eine Herausforderung hierbei kann in der Annahme bestehen, von der Analyse schriftlicher Reflexionen unmittelbar auf die Reflexivität einer Person zu schließen, da Reflexion stets kontextspezifisch als Abbild reflexionsbezogener Argumentationsprozesse angesehen werden sollte und reflexionsbezogenen Dispositionen unterliegt. Auch kann die Qualität einer Reflexion auf mehreren Dimensionen bewertet werden, ohne quantifizierbare, absolute Aussagen treffen zu können.

Daher wurden im Rahmen einer Physik-Videovignette  $N = 134$  schriftliche Fremdre reflexionen verfasst und kontextspezifische reflexionsbezogene Dispositionen erhoben. Expert\*innen erstellten theoriegeleitet Qualitätsbewertungen zur Breite, Tiefe, Kohärenz und Spezifität eines jeden Reflexionstextes. Unter Verwendung computerbasierter Klassifikations- und Analyseverfahren wurden weitere Textmerkmale erhoben. Mittels explorativer Faktorenanalyse konnten die Faktoren Qualität, Quantität und Deskriptivität gefunden werden. Da alle konventionell eingeschätzten Qualitätsbewertungen durch einen Faktor repräsentiert wurden, konnte ein maximales Qualitätskorrelat kalkuliert werden, zu welchem jede schriftliche Fremdre reflexion im Rahmen der vorliegenden Vignette eine computerbasiert bestimmbare Distanz aufweist. Diese Distanz zum maximalen Qualitätskorrelat konnte validiert werden und kann die Qualität der schriftlichen Reflexionen unabhängig von menschlichen Ressourcen quantifiziert repräsentieren. Abschließend konnte identifiziert werden, dass ausgewählte Dispositionen in unterschiedlichem Maße mit der Reflexionsqualität zusammenhängen. So konnten beispielsweise bezogen auf das Physik-Fachwissen minimale Zusammenhänge identifiziert werden, wohingegen Werthaltung sowie wahrgenommene Unterrichtsqualität eng mit der Qualität einer schriftlichen Reflexion in Verbindung stehen können.

Es wird geschlussfolgert, dass reflexionsbezogene Dispositionen moderierenden Einfluss auf Reflexionen nehmen können. Es wird empfohlen bei der Erhebung von Reflexion mit dem Ziel der Kompetenzmessung ausgewählte Dispositionen mit zu erheben. Weiter verdeutlicht diese Arbeit die Möglichkeit, aussagekräftige Quantifizierungen auch in der Analyse komplexer Konstrukte vorzunehmen. Durch computerbasierte Qualitätsabschätzungen können objektive und individuelle Analysen und differenzierteres automatisiertes Feedback ermöglicht werden.

## **Abstract**

Reflection is considered as a key category of professional development in teacher education. Thus, the quality of reflection-related performance has been studied in a variety of ways. To derive teacher's reflection-related personal Pedagogical Content Knowledge (PCK) from the analysis of a written reflection (reflection-related enacted PCK) seems to be challenging. The enactment of reflection-related personal PCK is context-specific and should be seen as a manifestation under the influence of Amplifiers & Filters. Also, it is difficult to make quantifiable statements of reasoning quality in a written reflection without using stage models or categorical scoring.

Therefore,  $N = 134$  (preservice) physics teachers wrote a reflection text in the context of a video vignette and answered items related to context-specific reflection-related dispositions. Experts rated the quality of each reflection text according to the breadth, depth, coherence, and specificity. Using computer-based classification and analysis, additional text features were extracted. An exploratory factor analysis was used to reduce data to the factors quality, quantity, and descriptiveness of a written reflection. Cause experts' quality ratings were represented by just one factor, a maximum quality-correlate for the present vignette was calculated. Each written reflection was determined a distance to this maximum computer-based. This quality index was validated and can represent the quality of the written reflections in a quantified way without the need of human expertise. Finally, it could be identified that selected Amplifiers & Filters are related to the reflection quality. For example, minimal correlations could be identified with respect to physics content knowledge, whereas values and perceived teaching quality can be closely related to the quality of a written reflection.

It is concluded that reflection-related Amplifiers & Filters can have a measurable influence on reflection-related enacted PCK. It is recommended to include measurements of Amplifiers & Filters in each research of reflection with the aim of measuring competence. Further, this work illustrates the possibility of meaningful quantification even in the analysis of complex constructs. Computer-based quality assessments can enable objective and individualized analyses and more differentiated automated feedback.

## Inhaltsverzeichnis

<i>Motivation</i>	1
<b>1 Theorie</b>	<b>3</b>
<b>1.1 Professionswissen von Lehrkräften der Naturwissenschaften</b>	<b>4</b>
1.1.1 Pedagogical Content Knowledge & Refined Consensus Model	5
1.1.2 Reflexion im Refined Consensus Model	8
<b>1.2 Reflexion &amp; Reflexivität</b>	<b>9</b>
1.2.1 Operationalisierung von Reflexion	16
1.2.2 Qualitätsmerkmale in schriftlichen Abbildern von Reflexion	18
1.2.2.1 Methodischer Einschub zur explorativen Faktorenanalyse	20
1.2.3 Reflexionsbezogene Dispositionen / Amplifiers & Filters	21
1.2.4 Reflexionsanlässe in Lehre und Forschung	23
1.2.5 Computerbasierte Analyse schriftlicher Reflexionen	25
<b>1.3 Forschungsanliegen</b>	<b>26</b>
1.3.1 Forschungsfragen	27
1.3.2 Übersicht über die Studie	29
<b>2 Methode</b>	<b>30</b>
<b>2.1 Automatisierung der Analyse von Reflexionstexten</b>	<b>31</b>
2.1.1 Zugrundeliegender Klassifikationsalgorithmus	32
2.1.2 Entwicklung eines ML-basierten Qualitätsmaßes für die Struktur von Reflexionstexten	34
<b>2.2 Videovignette (Lernkontext)</b>	<b>38</b>
2.2.1 Analyse schriftlicher Reflexionen	39
2.2.1.1 ML-basierte Textanalyse nach dem Rahmenmodell für Reflexion	40
2.2.1.2 ML-basierte Modellierung von Qualität schriftlicher Fremdrelexionen	41
2.2.1.3 Qualitätsdimensionen schriftlicher Reflexionen	42
2.2.2 Skalen der reflexionsbezogenen Amplifiers & Filters	42
<b>3 Ergebnisse</b>	<b>46</b>
<b>3.1 Beschreibung der Stichprobe</b>	<b>46</b>
<b>3.2 Schriftliche Fremdrelexionen</b>	<b>47</b>
3.2.1 Deskriptive Textmaße	48

3.2.2	Strukturdimension: LOS als strukturbezogenes Textmaß . . . . .	50
3.2.3	Qualitätsbewertungen . . . . .	54
3.2.4	Explorative Faktorenanalyse . . . . .	57
3.2.4.1	Anwendung der Faktorenanalyse . . . . .	57
3.2.4.2	Untersuchung der Qualitätskorrelate . . . . .	63
3.2.4.3	Methodische Exploration eines absoluten Qualitätskorrelats . . . . .	67
3.2.5	Validierung eines absoluten Qualitätskorrelats . . . . .	69
<b>3.3</b>	<b><i>Skalen der Amplifiers &amp; Filters</i></b> . . . . .	<b>72</b>
3.3.1	Amplifiers & Filters - deskriptiv . . . . .	72
3.3.2	Amplifiers & Filters und Textmaße . . . . .	75
3.3.3	Amplifiers & Filters und Qualitätsabschätzung . . . . .	78
<b>4</b>	<b><i>Zusammenfassung, Limitation und Diskussion</i></b> . . . . .	<b>82</b>
4.1	Kontextspezifische computerbasierte Qualitätsabschätzung . . . . .	84
4.2	Zusammenhänge von Amplifiers & Filters zur Reflexionsqualität . . . . .	86
<b>5</b>	<b><i>Ausblick und Konklusion</i></b> . . . . .	<b>89</b>
5.1	Theoretischer Mehrwert . . . . .	89
5.2	Methodischer Mehrwert . . . . .	90
5.3	Empirischer Mehrwert . . . . .	91
	<i>Literaturverzeichnis</i> . . . . .	93
	Anhang . . . . .	103
A1	Vignette (Prompt, Video, RT & Skalen) . . . . .	104
A2	Definitionen für Qualitätsindikatoren nach 1.2.2 . . . . .	108
A3	Testung der erhobenen Skalen auf Normalverteilung . . . . .	109
A4	Skalen und Items . . . . .	110
A4.1	Skala der wahrgenommenen Unterrichtsqualität . . . . .	110
A4.2	Skala der reflexionsbezogenen Selbstwirksamkeit. . . . .	112
A4.3	Fachwissen – Skala . . . . .	113
A4.4	Wert – Skala. . . . .	118
A5	Tabellenverzeichnis . . . . .	119
A6	Abbildungsverzeichnis . . . . .	120



## *Motivation*

Im Zuge öffentlicher Debatten zur Bewältigung des akuten Lehrkräftemangels in Deutschland (Klemm, 2022) wird erneut bewusst, dass der Beruf der Lehrkraft (nicht nur aus ökonomischen Gründen) einen der wichtigsten Einzelfaktoren für erfolgreiches Lernen von Schüler\*innen darstellt (Hattie, 2003). Einen wesentlichen Baustein der professionellen Entwicklung von Lehrkräften kann Reflexion ausmachen (Korthagen & Kessels, 1999). Zusammenhänge von reflexionsbezogenen Kompetenzen, Professionswissen sowie affektiven und motivationalen Einstellungen erscheinen daher aktueller denn je und sind im besonderen Maße in Betracht zu ziehen. Dem Beruf der Lehrkraft können aktuell hohe Belastungen zugeschrieben werden. Reflexion kann sich jedoch als anerkanntes Mittel im Umgang mit diesen Unsicherheiten erweisen (vgl. Cramer et al., 2019). Wenngleich zahlreiche Forschende betonen, dass Reflexion von hoher Relevanz für die Professionalisierung von Lehrkräften ist, besteht in der Lehrkräftebildung Uneinigkeit über das, was sich hinter dem Begriff Reflexion verbirgt (Häcker, 2017; 2022; Leonhard, 2022; Collin et al., 2013; Reintjes & Kunze, 2022; Wyss & Mahler, 2021).

Da für Kompetenz generell nur indirekte Evidenz gewonnen werden kann (Schaper, 2009), muss auch Reflexionsfähigkeit operationalisiert werden. Eine Herausforderung hierbei ist es, dass in der reflexionsbezogenen Kompetenzmessung individuelle Ressourcen mit kontextspezifischen oder motivationalen Aspekten zusammengefasst erhoben werden (Keller-Schneider, 2020). Aus diesem Grund werden in der vorliegenden Arbeit reflexionsbezogene Kompetenzen als komplexes Konstrukt diskutiert, welches sich lediglich in Abbildern, wie schriftlichen Fremdreflexionen, zeigen kann. Durch Festlegung einer Operationalisierung von Reflexion sowie einen spezifischen Kontext soll im Rahmen einer Physik-Videovignette untersucht werden, auf welche Weise Reflexion und reflexionsbezogene Dispositionen verknüpft sind. Mit dieser Arbeit soll auf einflussreiche Dispositionen evidenzbasiert hingewiesen werden, welche im Kontext der reflexionsbezogenen Kompetenzforschung in zukünftigen Studien mit erhoben werden sollten.

In der vorliegenden Arbeit werden zwei viel zitierte Modellierungen miteinander verbunden, um Unterscheidungen zwischen reflexionsbezogenen Kompetenzen, reflexionsbezogenen Disposition und reflexionsbezogener Performanz einordnen zu können. Hierbei wird ein nationales Rahmenmodell für Reflexivität als Kompetenz in ein internationales Professionsmodell eingebettet. Für die Untersuchung reflexionsbezogener Performanz wurde eine online-Videovignette eingesetzt. In einem Prompt wurden die Teilnehmenden auf eine Operationalisierung für Reflexion aufmerksam gemacht. Im Anschluss

sahen sie eine authentische Videographie eines Physikunterrichtes im Kontext des Freien Falls und verfassten eine schriftliche Fremdreflexion. Der Videoausschnitt wurde hierbei explizit ausgewählt, da er augenscheinlich gelungenen Unterricht zeigt, welcher jedoch aus fachlicher, fachdidaktischer als auch pädagogischer Perspektive mehrere Reflexionsauslöser<sup>1</sup> beinhaltet. Die Qualität der schriftlichen Fremdreflexionen wurde zunächst durch geeignete Experten konventionell bewertet und im weiteren Verlauf auf Basis maschinellen Lernens (ML) und entsprechender Modellierungen abgeschätzt, um die Qualitätsbewertungen quantifizieren zu können. Zur Untersuchung der Frage nach den Zusammenhängen zwischen Reflexionsqualität und reflexionsbezogenen Dispositionen beantworteten die Teilnehmenden abschließend Items zu Skalen ausgewählter Dispositionen (z. B. Selbstwirksamkeit oder Fachwissen). Auf diese Weise konnten Befunde für Zusammenhänge reflexionsbezogener Dispositionen und schriftlichen Reflexionen gefunden werden. Implikationen dieser Zusammenhänge für die Untersuchung reflexionsbezogener Kompetenzen werden mit dieser Arbeit diskutiert.

Im Rahmen einer theoriegeleiteten Modellierung von Reflexivität als Kompetenz nach von Aufschnaiter et al. (2019) ist diese Dissertationsschrift auf der Suche nach Evidenz im Kontext dieses Modells. Hierzu werden Ansätze individualisierter Fallanalyse sowie quantitative Fragenbogen-Maße verwendet, um Reflexionen, Dispositionen und Reflexivität von (angehenden) Physik-Lehrkräften zu analysieren. Neben den Kernkonstrukten der Reflexion beziehungsweise Reflexivität werden in der vorliegenden Arbeit Qualitätskorrelate qualitativer Textdaten reflexionsbezogener Argumentationsprozesse unter Verwendung modellierender Methoden maschinellen Lernens quantitativ analysiert und individuellen grundlegenden Dispositionen der reflektierenden Lehrkräfte gegenübergestellt.

---

<sup>1</sup> *Reflexionsanlass* als Gelegenheit zu reflektieren (z. B. in einer Unterrichtshospitation oder im Seminar), *Reflexionsauslöser* meint in dieser Arbeit wiederum den wahrgenommenen Impuls aus einer Unterrichtsbeobachtung, von welcher ein reflexionsbezogener Argumentationsprozess ausgehen kann.

## **1 Theorie**

Lernen im Unterricht wird durch eine Vielzahl von Faktoren beeinflusst. Dennoch wird spätestens seit Hattie (2003) argumentiert, dass sich Forschung im Bereich der Lehrkräftebildung auf die größte Einzelquelle aller Einflussfaktoren konzentrieren sollte, wenn Forschungsergebnisse zu effektiveren Lernprozessen führen sollen. Obwohl sich mit rund 50% aller Varianz für erfolgreiche Bildungsbiographien die Leistungen der Schüler\*innen als relevant darstellen (hohe Korrelation zwischen Begabung und Leistung), entfallen mit etwa 30% aller Faktoren Aspekte, wie das Feedback, Unterrichtsqualität, Classroom Management oder Herausforderungen und Zielsetzung überwiegend auf die unterrichtende Lehrkraft (Hattie, 2003). Neben dem Hintergrund weiterer Faktoren steht somit die Lehrkräfteprofession, das Wissen und Können einer Lehrkraft, im Zentrum gelingenden Unterrichts.

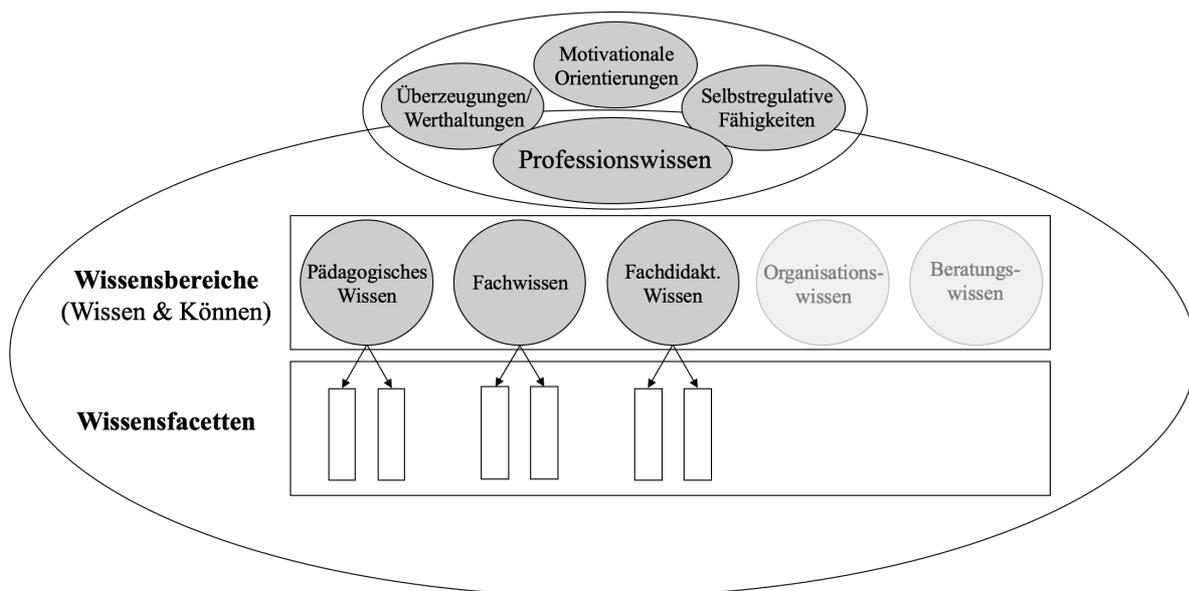
Ein solides Professionswissen wird hierbei als entscheidend für gelingenden Unterricht angesehen (Shulmann, 1987; Abell, 2007; Fischer et al., 2012). Auch konnten empirische Studien (z. B. COACTIV, Kunter et al., 2011) konkretere Zusammenhänge zwischen dem Fachwissen einer Lehrkraft sowie Wissensfacetten aus Pädagogik und Fachdidaktik und der Lernwirksamkeit von Unterricht aufzeigen (Sadler et al., 2013; Fischer et al., 2014; Liepertz & Borowski, 2019). Fachwissen, fachdidaktisches Wissen sowie weitere Wissensfacetten (wie vertieftes Schulwissen) entwickeln sich im Verlauf der ersten Phase der Lehrkräftebildung (Enkrott, 2021), können jedoch auch in Praxisphasen des Lehramtsstudiums oder der zweiten Phase der Lehrkräftebildung weiterentwickelt werden (Schubarth et al., 2009). Außerhalb extrinsischer Faktoren, wie Lehrveranstaltungen oder Hospitationsgelegenheiten, kann sich Reflexion als ein wesentliches Element zur professionellen Entwicklung darstellen.

Aus diesem Grund wird in den folgenden Kapiteln zunächst diskutiert, worin die Profession einer Lehrkraft in der Forschungsgemeinschaft der Naturwissenschaftsdidaktik besteht und wie deren Entwicklung aus theoretischer Perspektive mittels reflexionsbezogener Argumentationsprozesse erfolgen kann. In einem zweiten Schritt wird dieser reflexionsbezogener Argumentationsprozess sowie die Kompetenz selbigen zu durchschreiten konzeptualisiert, bevor Qualitätsmerkmale von Reflexion aufgezeigt werden. Abschließend wird ein Überblick über Erhebungsmöglichkeiten von Reflexion gegeben und diskutiert, inwieweit sich hierbei Herausforderungen der Kompetenzmessung zeigen.

### 1.1 Professionswissen von Lehrkräften der Naturwissenschaften

Im Kontext naturwissenschaftlicher Bildung erkennen Forschende die Bedeutung einer Lehrkraft bei der Unterstützung des Lernens der Schüler\*innen an, wobei viele hierbei ihre Aufmerksamkeit auf die zentrale Rolle des beruflichen Wissens der Lehrer für die Qualität des Unterrichts richten (z. B. Shulman, 2001; Berliner, 2001; Magnusson et al., 1999). Um die Qualität des Unterrichts zu charakterisieren, konzeptualisierten Forschende das notwendige professionelle Wissen und Können einer Lehrkraft als professionelle Handlungskompetenz.

Grundlegend fußen Systematisierungen professioneller Handlungskompetenz dennoch auf explizierten Wissensfacetten. Shulman (1986) separierte in einer ersten Konzeptualisierung allgemeines pädagogisches Wissen (general pedagogical knowledge), Fachwissen (subjectmatter content knowledge), fachdidaktisches Wissen (pedagogical content knowledge) und curriculares Wissen (curriculum knowledge) (Shulman, 1986). In weiteren Arbeiten wurden das Konstrukt eines Verständnisses für die Lernenden (knowledge of learners), für den Lernkontext (knowledge of educational context) sowie für Philosophie, Theorie und Historie der Pädagogik hinzugefügt (Shulman, 1987). Eine Reihe von Autor\*innen haben diese Kernkomponenten des professionellen Wissens von Lehrkräften weiter ausdifferenziert (z. B. Grossman, 1990), wobei sich in seiner allgemeinsten Form die Unterscheidung in pädagogisches Wissen, Fachwissen und fachdidaktisches Wissen durchgesetzt hat (vgl. Helmke, 2003). Entsprechend bezogen Baumert und Kunter (2006) diese als zentrale Kompetenzfacetten in ihre Konzeptualisierung mit ein (siehe Abbildung 1).



**Abb. 1** Modellierung der für den Lehrerberuf spezifizierten Wissenskomponenten des allgemeinen Modells professioneller Handlungskompetenz (Baumert & Kunter, 2006, S. 482)

International kann im deutschsprachigen Verständnis des Professionswissens der Terminus des Pedagogical Content Knowledge (PCK) verwendet werden, welches in diesem weiteren Sinne auch von Shulman (1999) begrifflich verwendet wird. Eine Gleichsetzung fachdidaktischen Wissens mit PCK sollte somit als unvollständig angesehen werden, da sich PCK als ganzheitliches Wissens- und Kompetenzgefüge verstehen lässt, kompetent unterrichten zu können. Seit Hattie (2003) stützt auch neuere Forschung die Behauptung, dass dieses PCK von Lehrkräften eine wesentliche Grundlage für effektiven Unterricht ist (Baumert et al., 2010; Sadler et al., 2013). Basierend auf der früheren Arbeit von Gess-Newsome (2015) wurde PCK mit der Einführung des Refined Consensus Modells of PCK in Science Teaching (RCM) neu konzeptualisiert (Carlson et al., 2019). Seit der Einführung des RCM verorten viele Forschende eigene Studien im Kontext Lehrkräfteprofessionalisierung im Modell, sodass der angestrebte Konsens in der Kommunikation von Forschungsergebnissen erreicht werden konnte (Mientus et al., 2022). Mientus et al. (2022) betonen, dass das RCM als theoretisches Gerüst für ein breites Spektrum von Forschungsschwerpunkten geeignet ist und eine solide Grundlage für die Kommunikation empirischer Ergebnisse innerhalb der Gemeinschaft und darüber hinaus bietet.

Aus diesem Grund werden im Folgenden das RCM detaillierter betrachtet und weiterführend Prozesse von Reflexion, Analyse und Förderung vor dem Hintergrund des RCM diskutiert. Erweiternd wird eine Konzeptualisierung von Reflexivität als Kompetenz in das RCM eingebettet und entsprechend des vorliegenden Forschungsanliegens spezifiziert.

### *1.1.1 Pedagogical Content Knowledge & Refined Consensus Model*

Pedagogical Content Knowledge (PCK) hat sich seit Shulman (1986) als besonders wichtiges Konstrukt für Lehrkräfte der Naturwissenschaften herausgestellt, da es das Potenzial hat, berufliches Wissen und die Fähigkeiten von Fachlehrkräften zu charakterisieren. Auch konnten Zusammenhänge zwischen PCK und effektiven Unterricht empirisch belegt werden (Baumert et al., 2010; Sadler et al., 2013; Kirschner et al., 2016). So wurde PCK in den naturwissenschaftlichen Fächern zu einem zentralen Bestandteil vieler universitärer Programme der Lehrkräftebildung (Gess-Newsome et al., 2019). Um angehenden Lehrkräften für naturwissenschaftliche Fächer die Möglichkeit zu geben, relevante PCK-Facetten zu entwickeln, wurden in Lehrveranstaltungen beispielsweise Fehlvorstellungen oder Alltagskonzepte von Schüler\*innen, Unterrichtsstrategien für den Umgang damit oder Bewertungskompetenzen explizit adressiert (Park & Oliver, 2008).

Ursprünglich betrachtete Shulman (1986; 1987) PCK als eine Verschmelzung von Inhalt und Pädagogik, so dass PCK als eine neue und eigenständige Form von Wissen angesehen

werden kann. Fachwissen (Content Knowledge [CK]), das sich auf den Lehrstoff und naturwissenschaftliche Phänomene bezieht und pädagogisches Wissen (Pedagogical Knowledge [PK]), das allgemeines Wissen über das Lehren und Lernen im Unterricht darstellt, überschneiden sich nach diesem Verständnis im PCK. Dieses einfache Grundmodell wurde für die mathematisch-naturwissenschaftlichen Fächer weiter ausdifferenziert (z. B. Grossman, 1990; Magnusson et al. (1999); Park & Oliver, 2008b). Anstatt die Forschungsrichtungen zu klären, ist es jedoch möglich, dass diese unterschiedlichen Sichtweisen über die Natur (und Entwicklung) von PCK in der naturwissenschaftlichen Bildung den Fortschritt in diesem Bereich zu behindern begannen und mit der Zeit zu Forderungen nach einem Konsens über die Konzeptualisierung von PCK führten (Abell, 2008).

*„Es besteht weitgehende Übereinstimmung darüber, dass Wissen und Können – also deklaratives, prozedurales und strategisches Wissen – zentrale Komponenten der professionellen Handlungskompetenz [im Sinne von PCK] von Lehrkräften darstellen. Weitaus weniger Übereinstimmung besteht in Bezug auf die Struktur und Topologie professionellen Wissens und Könnens, die unterschiedlichen Wissenstypen und ihren epistemischen Status, die mentale Repräsentation dieser Wissenstypen und die Genese professionellen Wissens und Könnens“*  
(Baumert & Kunter, 2006, S.481)

Eine Gemeinsamkeit der verschiedenen Modelle besteht darin, dass stärker ausgeprägte Wissensgrundlagen zu einem ausgeprägteren PCK führen können. Weiter besteht der Konsens, dass das PCK einer Lehrkraft auch vom Grad der Vernetzung zwischen den einzelnen Komponenten (Park & Suh, 2019) und Wissen als auch Fähigkeiten umfasst. Da PCK stets themenspezifisch, sehr persönlich und idiosynkratisch ist (Chan & Hume, 2019), kann es per se nur für einzelne Lehrkräfte beschrieben werden. Unter diesem Gesichtspunkt muss die Forschung über die Art und Entwicklung des PCK von Lehrkräften der Naturwissenschaften den Wechselwirkungen zwischen den PCK-Komponenten oder anderen Variablen (z. B.: Einstellungen und Überzeugungen) mehr Aufmerksamkeit schenken als den einzelnen Komponenten selbst (Park & Suh, 2019), um Hinweise auf Einflussfaktoren allgemeiner Art zu erhalten.

Basierend auf den früheren Arbeiten haben Carlson et al. (2019) im Jahr 2016 PCK für den naturwissenschaftlichen Unterricht im RCM neu positioniert (Abbildung 2). Aufbauend auf ausgewählten, bestehenden PCK-Komponenten früherer Modelle (Content Knowledge [CK], Pedagogical Knowledge [PK], Knowledge of Students [KS], Curricular Knowledge [CuK] und Assessment Knowledge [AK]) führt das RCM drei Ebenen von PCK ein, die das Verständnis potenzieller Quellen beziehungsweise Mechanismen für die PCK-Entwicklung

von Lehrkräften abzubilden versuchen. Diese Ebenen umfassen collective (gemeinschaftliches) PCK (cPCK), personal (persönliches) PCK (pPCK) und enacted (angewandtes) PCK (ePCK).

Dem RCM zufolge existiert cPCK innerhalb der größeren wissenschaftlichen Gemeinschaft, denen Forschende und Lehrkräfte angehören (Carlson et al., 2019). Lehrkräfte im Vorbereitungsdienst, in Fort- und Weiterbildung aber auch im kollegialen Austausch werden sich cPCK bewusst, indem sie während der Phasen der Lehrkräftebildung und darüber hinaus in die Gemeinschaft der Wissenschaftsbildung eintauchen. Hier haben sie die Möglichkeit verschiedene Ressourcen (z. B. Lehrpläne oder Unterrichtsmaterialien) kennenzulernen und sich mit anderen Akteur\*innen auszutauschen. Sie beginnen, neben anderen Wissensformen eine einzigartige und persönliche Form von PCK zu entwickeln, die als pPCK bezeichnet wird und im Laufe zunehmender Professionalisierung immer stärker durch die Erfahrungen der Lehrkräfte im Unterricht geprägt ist. pPCK wird zu einer wichtigen intellektuellen Ressource, die es Lehrkräften der Naturwissenschaften ermöglicht, sowohl auf expliziter als auch auf impliziter Ebene als Lehrkräfte zu denken und zu handeln (Alonzo et al., 2019; Eraut, 2000).

Im Zentrum des RCM (siehe Abbildung 2) wird deutlich, dass eine einzelne Lehrkraft während des Unterrichts einen pädagogischen Argumentationsprozess zur Entscheidungsfindung (Pedagogical Reasoning) nutzt, um Aspekte aus dem pPCK in das situationsspezifische ePCK während eines Zyklus aus Planen, Lehren und Reflektieren zu überführen (Alonzo et al., 2019). Daher kann die Unterrichtspraxis zu einem Forum für den Wissensaustausch und die Entwicklung von PCK werden. Der pädagogische Argumentationsprozess wird als wesentlicher Mechanismus dargestellt, durch den dieser Austausch und diese Entwicklung stattfinden. Allerdings gibt es derzeit kaum schlüssige Übersichten über die tatsächlichen Entwicklungsmechanismen zwischen pPCK und ePCK (Gess-Newsome et al., 2017; Aydin et al., 2013) sowie ein Defizit an empirisch-fundierter Forschung (Mientus et al., 2022), um z. B. die Größe von Effekten zu schätzen oder die qualitativen Zusammenhänge zu verbessern (Wilson et al., 2019).

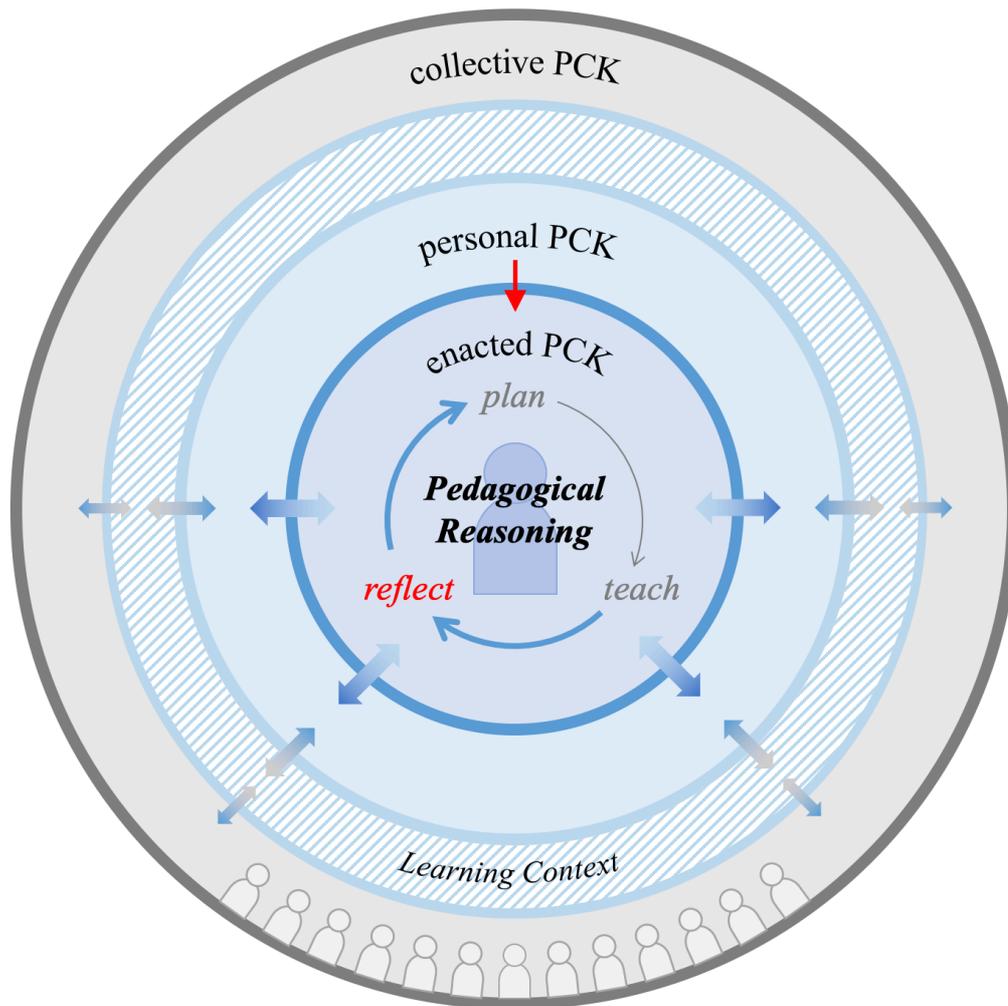


Abb. 2 RCM of PCK for teaching science ohne Wissensfacetten (nach Carlson et al., 2019, S. 83)

### 1.1.2 Reflexion im Refined Consensus Model

Alonzo et al. (2019) identifizieren zwei verschiedene Planen-Lehren-Reflektieren-Zyklen - einen Makrozyklus, der sich auf die Planung der gesamten Unterrichtsstunde, das Unterrichten und die retrospektive Reflexion konzentriert, und einen Mikrozyklus innerhalb der Unterrichtsphase des Makrozyklus, der sich mit aktuellen Unterrichtsentscheidungen während des Unterrichts befasst. Im Mikrozyklus leitet ePCK die Interpretation (Wahrnehmung und Reflexion) von Unterrichtssituationen, die während des Unterrichts auftreten, und die Planung von Entscheidungen und Handlungen als Reaktion auf diese Momente. Diese Unterscheidung ist in der Reflexionsforschung mit Donald Schöns Verständnis der reflection-on-action beziehungsweise der reflection-in-action (Schön, 1983) anschlussfähig. Unter Verwendung beider Zyklen argumentieren Alonzo et al. (2019) zusammenfassend, dass das ePCK einer Lehrkraft unmittelbar mit deren/dessen Performanz in Planung (ePCKp), Unterricht (ePCKt) und Reflexion (ePCKr) zusammenhängt, gleichzeitig wird jedoch die empirische Trennbarkeit von Planungs-, Unterrichts und Reflexionskompetenz diskutiert (z. B. Vogelsang et al., 2022).

Von den Wissensbasen im äußersten Kreis bis hin zum ePCK einer Lehrkraft im Inneren sind alle drei Ebenen sowie der Kontext der Betrachtung des PCK anschaulich mit Doppelpfeilen verbunden. Diese Doppelpfeile repräsentieren Dispositionen zwischen den jeweiligen Ebenen, welche von Carlson et al. (2019) als *Amplifiers & Filters*<sup>2</sup> benannt werden. Unter der Bedingung dieser Dispositionen lotet z. B. eine einzelne Lehrkraft aus, welche Argumente als belastbare Wissensfacette einer Gemeinschaft (cPCK) aufgenommen bzw. aus dem pPCK in eine konkrete Unterrichtssituation (ePCK) übernommen werden. Das RCM geht davon aus, dass vor allem ePCK auf komplexe Weise mit pPCK verknüpft ist und theoretisch auch differenzierte und testbare Rückschlüsse auf diese Zusammenhänge gezogen werden können. In ihrem systematischen Literaturreview stellen Mientus et al. (2022) jedoch fest, dass viele Forschende Wissensfacetten und Schritte (auch in Mikro- und Makrozyklus) nicht hinreichend trennscharf untersuchen, um einzelne Wirkmechanismen aufdecken zu können und empfehlen daher, sich in weiteren Studien auf konkrete und einzelne Zusammenhänge zu fokussieren sowie die Rolle von Amplifiers & Filters stärker in den Blick zu nehmen.

Um der selbst gesetzten Forderung nachzukommen und zu einem empirisch belastbarem Detailverständnis beizutragen, soll den reflexionsbezogenen Wechselwirkungen zwischen pPCK und ePCKr unter besonderer Berücksichtigung des Einflusses von Amplifiers & Filters auf der Makroebene nähergekommen werden. Aus diesem Grund werden im nachfolgenden Kapitel zunächst reflexionsbezogene Wechselwirkungen diskutiert. Im Weiteren werden diese Diskussionen in das RCM eingebettet und die diskutierten Konstrukte verortet. Das bestehende deutschsprachige Verständnis von Reflexion kann somit für eine international anschlussfähige Kommunikation von Forschungserkenntnissen erweitert werden.

## ***1.2 Reflexion & Reflexivität***

In Anlehnung an Blömeke et al. (2015) lässt sich PCK als Kontinuum deuten, auf dem situationsspezifische Fähigkeiten, (kognitive und affektiv-motivationale) Dispositionen und das tatsächliche Verhalten einer Lehrkraft im Unterricht zusammenspielen. Diese Deutung bleibt auch im RCM gültig, wenn Kompetenzen aus dem personal PCK in das enacted PCK überführt werden. Neben den Dispositionen, welche eine Lehrkraft in eine Unterrichtssituation einbringt, gelten also auch die professionelle Wahrnehmung unterrichtlicher Aspekte und dessen fundierte Analyse als wichtige Voraussetzungen für den Erwerb bzw. die Weiterentwicklung von professionellen Kompetenzen (van Es & Sherin, 2021; Gruber, 2021).

---

<sup>2</sup> Amplifiers & Filters wird in dieser Arbeit mit dem „&“-Zeichen als Eigenname verwendet um den Lesenden in Folge der Zweisprachigkeit der Begrifflichkeit die Lesart in „und“ oder „and“ selbst zu überlassen.

Als verbindendes Element von Wissen und Wahrnehmung gesehen, vermag Reflexion daher einen nachhaltigen intrinsischen Professionalisierungsprozess anzuregen (Roters, 2012). In der Lehrkräfteprofessionalisierung kann die Reflexionsfähigkeit von Lehrkräften somit als Kernkompetenz definiert werden (Combe & Colbe, 2008).

Im Professionsmodell von Baumert und Kunter (2006) kann Reflexion als selbstregulative Fähigkeit verstanden werden, wohingegen sie im RCM eher als Prozess der pädagogischen Argumentation verortet ist. In beiden Fällen wird jedoch Reflexion als zentraler Baustein für die professionelle Entwicklung von Lehrkräften angesehen (Schön, 1983; Korthagen & Kessels, 1999; Christof et al., 2018; Sorge et al., 2018) und scheint unabdingbar für die Entwicklung der eigenen Lehrpersönlichkeit (vgl. Rothland, 2021). Auch aus diesem Grund ist Reflexion in nationalen und internationalen Standards der Lehrkräftebildung fest verankert (z. B. KMK, 2004; 2019; NBPTS<sup>3</sup>, 2016) und wird in zahlreichen Programmen der Lehrkräftebildung verfolgt (Darling-Hammond, 2012). Die Förderung der Fähigkeit zur Reflexion legitimiert sich somit als ein Ziel der Lehrkräftebildung (vgl. Browning & Korthagen, 2021), obwohl die Wirksamkeit von Reflexion für den Unterricht empirisch nach wie vor unklar ist (Collin et al., 2013; Wyss, 2013, Häcker, 2019, Labott & Reintjes, 2022).

Reflexion kann auf der Makroebene als Verbindung von der Durchführung einer Unterrichtssituation und einer neuen Planung angesehen werden (vgl. Korthagen, 2001; Nordine et al., 2021), als Prozess aber gefahrlaufen eine bloße Analyse zu werden (von Aufschnaiter, im Druck). Wenngleich eine differenzierte Unterrichtsanalyse ein tiefes Verständnis von der reflektierten Situation zeigen kann, wird erst durch eine persönliche Bezugnahme (z. B. durch das ableiten von Konsequenzen zur Verbesserung der Lehrperson oder des Unterrichtes) in einer Reflexion eine professionelle Entwicklung angeregt (von Aufschnaiter et al., 2019). Inhalte, Ziele und Methoden für Reflexion sind vielfältig (Christof et al., 2018) und können zusätzlich von weiteren dispositiven Faktoren, wie dem situationsspezifischen Vorwissen, affektiven und motivationalen Prozessen als auch situativen Faktoren (z. B. einem Kontext) abhängen (vgl. Zlatkin-Troitschanskaia et al., 2019). Forschende empfehlen daher u.a. Reflexionsanlässe sorgfältig zu planen um beispielsweise Effekte von Inattentional Blindness<sup>4</sup> (Kellog, 2007) zu vermeiden. Mit geeigneten Fragestellungen, die den Reflexionsprozess strukturieren oder mit entsprechenden Prompts und Methoden die Reflexion unterstützt werden (Wyss, 2018). Auf einer Mikroebene (micro ePCKr bzw. reflection-in-action) wird jedoch die Anwendbarkeit geeigneter Vorgehensweisen

---

<sup>3</sup> National Board for Professional Teaching Standards

<sup>4</sup> Dt.: Unaufmerksamkeitsblindheit beschreibt den Effekt, dass bei fokussierter Aufmerksamkeit einzelne Geschehnisse oder Objekte nicht wahrgenommen werden.

fragwürdig und erfordert eine immanent, kompetent reflektierende Lehrkraft. Im Prozess zunehmender Professionalisierung sollen daher Lehrkräfte zu sog. reflective practitioners werden (Schön, 1983; Altrichter, 2000; Abels, 2011). Da jedoch einer Reflexion einer wahrgenommenen Unterrichtssituation immer eine Form der Analyse dieser Wahrnehmung vorausgeht (Meschede, 2014) und eine Lehrkraft kompetent in der Analyse von Unterricht sein kann, ohne kompetent reflektieren können zu müssen, kann eine Abgrenzung zu „analytical practitioners“ (von Aufschnaiter, im Druck) nützlich sein.

Während Reflexion anfangs eher als eine systematische Denkweise charakterisiert wurde (Dewey, 1933) wurde sie seit Aebli (1980; 1981) zunehmend als ein kognitiver Prozess klassifiziert, der eine (Neu-)Strukturierung von Erfahrungen, Wissen und Überzeugungen auf der Grundlage selbst erlebter oder stellvertretender, z. B. beobachteter, Erfahrungen beinhaltet (Korthagen, 2001). Differenzierter kann Reflexion aber auch als kriteriengeleitetes Nachdenken über eine Handlung sowie zugehörige Konsequenzen für weiteres Handeln verstanden werden (Wyss, 2013). Wird Reflexion als Fähigkeit verstanden, wird meist von Reflexionskompetenz gesprochen. Die Fähigkeit bzw. die Kompetenz zu reflektieren kann im Sinne einer „reflexiblen Eigenschaft oder Möglichkeit des [Sich]rückbeziehens“ (Dudenredaktion, o.D.) auch als Reflexivität benannt werden. Auf diese Weise können Kompetenz und Prozess verbal unterschieden werden.

Von Aufschnaiter et al. (2019) diskutieren Reflexion als einen „Prozess des strukturierten Analysierens, in dessen Rahmen zwischen den eigenen [Wissen], Fähigkeiten, Einstellungen/[Werthaltungen] und/oder [den Dispositionen einer Lehrkraft] und dem eigenen situationsspezifischen Denken und Verhalten [...] eine Beziehung hergestellt wird mit dem Ziel, [das eigene Wissen], die eigenen Einstellungen... und/oder das eigene Denken und Verhalten (weiter) zu entwickeln“ (von Aufschnaiter et al., 2019, S. 148). In einer Reflexion wird somit eine Vielzahl von Facetten professions- und persönlichkeitsbezogener Merkmale als komplexes System kognitiv verarbeitet. Die zugrundeliegende Reflexivität sollte folglich ebenso als komplexe Kompetenz konzeptualisiert werden. Eingebettet in den Kontext des RCM soll in Anlehnung an die Selbstbezüglichkeit und Situationsspezifität aus von Aufschnaiter et al. (2019) Reflexion (also der Prozess) für die vorliegende Arbeit wie folgt definiert werden:

*Reflexion ist ein strukturierter Prozess, in dessen Rahmen zwischen personal PCK und enacted PCK eine Beziehung hergestellt wird, mit dem Ziel personal PCK zu entwickeln.*

Wenngleich nicht ausgeschlossen werden soll, dass nicht auch im Übergang zwischen pPCK und cPCK ein reflexionsbezogener Argumentationsprozess inbegriffen ist, soll sich an dieser Stelle ausschließlich auf individuelle reflexionsbezogene Argumentationsprozesse innerhalb

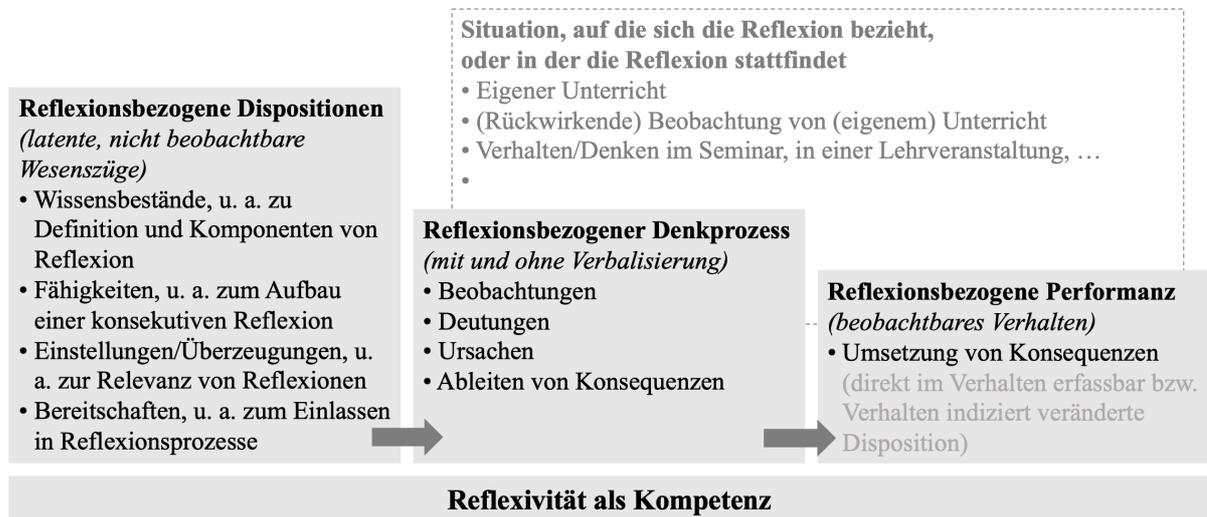
einer wahrgenommenen Unterrichtssituation bezogen werden. Auf diese Weise können detaillierte Wirkmechanismen im Kontext des RCM untersucht werden, um zu objektiven und validen Aussagen zu gelangen.

In ihrer Modellierung von Reflexivität als Kompetenz unterscheiden von Aufschnaiter et al. (2019) reflexionsbezogene Dispositionen, reflexionsbezogenen Argumentationsprozess und reflexionsbezogene Performanz (siehe Abbildung 3). In Anlehnung an Blömeke et al. (2015) wird jeweils das zugehörige Wissen und Können der einzelnen Komponenten als Kompetenz betrachtet. Disposition beschreibt in der Wissenschaftstheorie eine (zugrundeliegende) Eigenschaft, also das Vermögen eines Forschungsgegenstandes (z. B. einer Lehrkraft), sich in einer bestimmten Weise zu verhalten (Kanterian, 2004). Reflexionsbezogene Dispositionen nach von Aufschnaiter et al. (2019) entsprechen somit bei Blömeke et al. (2015) den Eigenschaften einer Person, einen reflexionsbezogenen Argumentationsprozess zu durchlaufen. Der reflexionsbezogene Argumentationsprozess kann als situationsspezifische Fähigkeit angesehen werden und die reflexionsbezogene Performanz analog als Anwendung.

Unter Berücksichtigung des Unvermögens, Kompetenzen direkt messen zu können (Labott & Reintjes, 2022), soll an dieser Stelle weiter limitiert werden, dass es schwierig ist etwas derartig komplexes wie Reflexivität zu messen und stattdessen lediglich Evidenzen generiert werden können. Diese können jedoch in der Lage sein, entsprechende Hinweise auf zugrundeliegende Kompetenzen zu liefern. So kann angenommen werden, dass die Qualität einer Reflexion (bzw. eines Reflexionstextes) als hinreichender Indikator für die Ausprägung der zugrundeliegenden Reflexivität anzusehen ist, da eine hohe Qualität auf eine entsprechende Kompetenz hinweisen kann (Leonhard, 2008, S. 55). Schlussfolgernd kann die Reflexion einer Lehrkraft (Denkprozess und Performanz) als Abbild der zugrundeliegenden Reflexivität in einem spezifischen Kontext interpretiert werden. Dispositionen, wie Befinden, Überzeugungen und Werthaltungen einer Lehrkraft fungieren somit eher als Vermittler zwischen Kompetenz und Performanz, zwischen Reflexivität und Reflexion, zwischen pPCK und ePCK, also als Amplifiers & Filters.

Zusammenfassend kann Reflexion eine referentielle Form des Denkens und somit einen Prozess darstellen (vgl. Häcker et al. 2008) und reflexionsbezogene Dispositionen die Möglichkeiten oder Bedingungen, unter welchen eine Lehrkraft Ressourcen des pPCK nutzt, um diese Reflexion zu durchschreiten. Reflexivität erscheint somit als Kompetenz, die eigenen kontextspezifischen Ressourcen unter persönlichen, dispositionellen Eigenschaften (pPCK, Amplifiers & Filters) in einen reflexionsbezogenen Argumentationsprozess (ePCKr) zu überführen. Diese Interpretation reiht sich ein in die Forderung an ein Rahmenmodell für

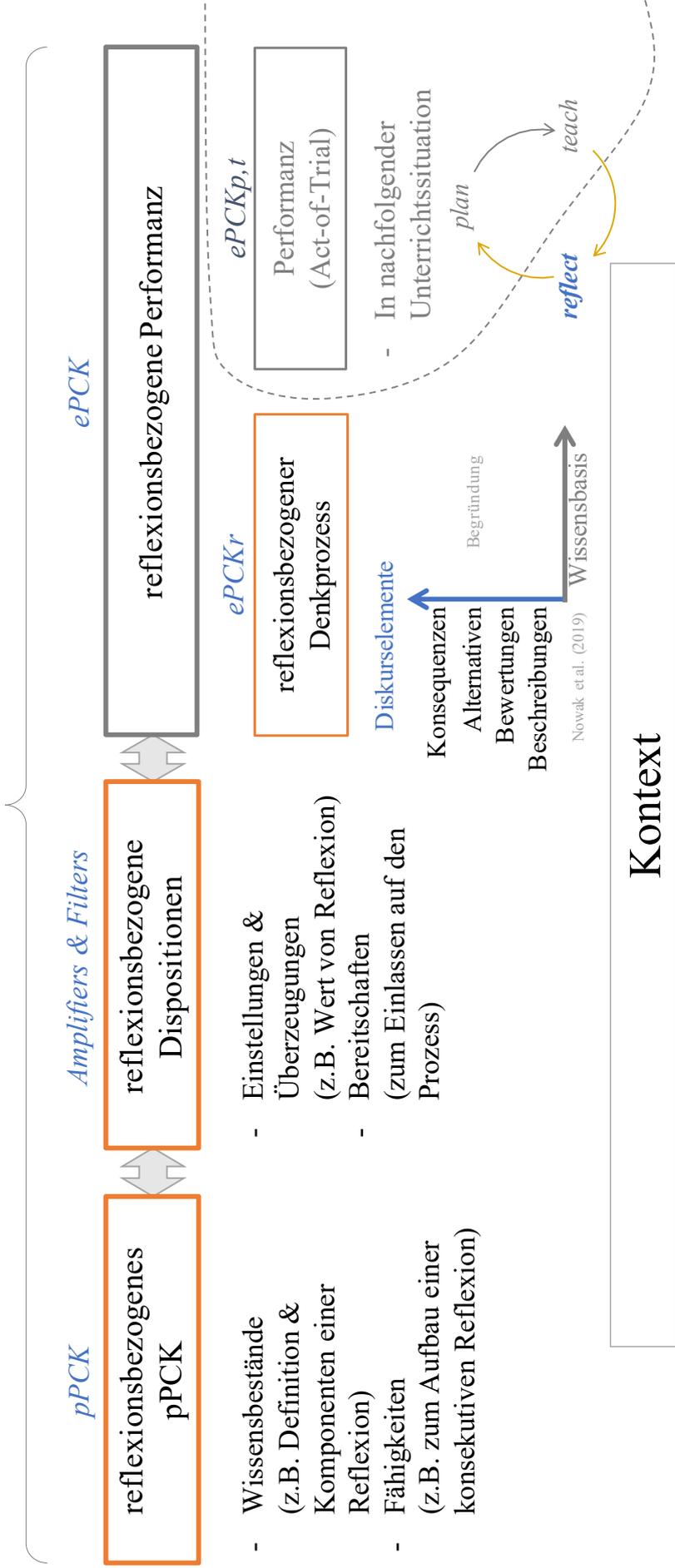
Reflexion und Reflexivität nach von Aufschnaiter (2019), dass „[unterschiedliche Perspektiven, sowie] bestimmte Schwerpunktsetzungen und Fokussierungen ermöglicht.“ (von Aufschnaiter et al., 2019, S. 151).



**Abb. 3** Reflexivität als Kompetenz (von Aufschnaiter et al., 2019, S. 152)

Mit der in Abbildung 3 visualisierten Konzeption der Reflexivität als Kompetenz sprechen sich von Aufschnaiter et al. (2019) dafür aus, dass in der Lehrkräftebildung „zunächst einmal Dispositionen mit Bezug auf Reflexion aufgebaut werden sollten“ (von Aufschnaiter, 2019, S. 155), bevor zugehörige Fähigkeiten und Fertigkeiten entwickelt werden können. Gleichzeitig wird nicht konkret diskutiert, inwieweit auch nicht-reflexionsbezogene Dispositionen Effekte auf einen reflexionsbezogenen Argumentationsprozess ausüben können. Im Sinne der Erfassung und Analyse von Reflexivität kann jedoch die Anwendung reflexionsbezogener Kompetenzen dem Einfluss individueller Reflexivität und situationsspezifischer Dispositionen unterliegen. Reflexivität beschreibt somit an dieser Stelle die Summe aller abrufbaren reflexionsbezogenen Kompetenzen unter dem Einfluss von Amplifiers & Filters. In Anlehnung an Weinert (2001) zeigen sich diese Kompetenzen in den Fähigkeiten, in variablen Reflexionsanlässen Wissen anzuwenden, um Reflexionsauslöser zu durchdenken um personal PCK oder den eigenen Unterricht weiterzuentwickeln. Aus diesem Grund kann es wertvoll sein, Reflexivität im Sinne der Kompetenz von einzelnen Dispositionen zu separieren und vor dem Hintergrund des RCM neu zu konzeptionieren. In Abbildung 4 wird daher ein Modell vorgeschlagen, welches die Rolle reflexionsbezogener Dispositionen im Übertrag von Reflexivität auf Reflexion repräsentiert.

*Reflexivität als Summe aller anwendbaren reflexionsbezogenen Kompetenzen*



**Abb. 4** Einbettung des Modells für Reflexivität als Kompetenz (nach von Aufschnaiter et al., 2019) in das Refined Consensus Modells of Pedagogical Content Knowledge (Carlson et al., 2019) unter Berücksichtigung der Operationalisierung von Reflexion nach Nowak et al. (2019)

Unter Berücksichtigung einer Operationalisierung von Reflexion<sup>5</sup> kann das Modell für Reflexivität als Kompetenz nach von Aufschnaiter et al. (2019) in das RCM eingebettet werden und skizziert auf diese Weise reflexionsbezogene Mechanismen zwischen pPCK und ePCK (siehe Abbildung 4), welche sich auf eine Situation beziehen. Reflexivität als Kompetenz zeigt sich hiermit kontextspezifisch im Übertrag von pPCK zu ePCK und umgekehrt. Allgemeine Wissensbestände und Fähigkeiten zu Definitionen, Komponenten oder zum Aufbau einer Reflexion werden vor dem Hintergrund persönlicher Wissensbestände in eine reflexionsbezogene Performanz überführt. Diese Performanz wird im Modell unterschieden in den eigentlichen reflexionsbezogenen Argumentationsprozess ePCK reflect (ePCKr) und die reflexionsbezogene Performanz im Sinne des Act-of-Trial ePCK plan, teach (ePCKp,t) in einer nachfolgenden Unterrichtssituation. Unter der Annahme, dass auch aus der reflexionsbezogenen Performanz im Sinne des Pedagogical Reasoning Cycles im ePCK nach Alonzo et al. (2019) nach der Konzipierung von Reflexivität nach von Aufschnaiter et al. (2019) eine Reflexion mit dem Ziel durchlaufen wird, pPCK (weiter-)zuentwickeln, beschreibt das Modell auch einen Übertrag aus ePCK zum pPCK. Im Gegensatz zur Konzeption nach von Aufschnaiter et al. (2019) wird dank der Einbettung in das RCM die Rolle reflexionsbezogener Dispositionen deutlich, so stehen Einstellungen, Überzeugungen, Motivationen und Bereitschaften als Amplifiers & Filters im Austausch zwischen beiden PCK-Ebenen.

Die Fähigkeit reflektieren zu können, bzw. reflektiert zu handeln wird mit diesem Modell und somit in dieser Arbeit grundlegend vom Prozess einer Reflexion selbst separiert. Begrifflich werden hier der *reflexionsbezogene Denk- bzw. Argumentationsprozess* mit dem reflexionsbezogenen *pedagogical Reasoning* gleichgesetzt. Auf diese Weise wird die Bedeutung des *pedagogical Reasoning Cycle* im Sinne des wiederkehrenden Zyklus aus planen, unterrichten und reflektieren herausgestellt. Es ergibt sich die Annahme, dass ein reflexionsbezogener Argumentationsprozess nicht erst mit der Anwendung gewonnener Erkenntnisse in einer neuen Unterrichtsplanung oder -situation beendet sein muss. Aspekte, welche Einfluss darauf haben, wie eine Lehrkraft ihre persönliche Kompetenz (Wissen und Können) zu reflektieren in einen expliziten reflexionsbezogenen Argumentationsprozess einbringen kann, werden in dieser Arbeit als reflexionsbezogene Dispositionen oder als Amplifiers & Filters zwischen dem personal PCK und dem enacted PCK interpretiert. Im Sinne des RCM wird somit der reflexionsbezogene Argumentationsprozess begrifflich dem reflexionsbezogenen Argumentationsprozess im enacted PCK gleichgesetzt und kann (etwas

---

<sup>5</sup> Im vorliegenden Fall wird die Operationalisierung nach Nowak et al. (2019) verwendet. Weitere Ausführungen folgen in Kapitel 1.2.1

verkürzt) als Reflexion bezeichnet werden. Reflexionsbezogene Performanz kann im Wesentlichen zwei Bedeutungen ausdrücken. Zum einen die Performanz im reflexionsbezogenen Argumentationsprozess (z. B. Reflexion nach Wyss, 2013<sup>6</sup> oder nach Nowak et al., 2019<sup>7</sup>) und zum anderen die Überführung gewonnener Erkenntnisse aus diesem Prozess in eine Handlung (z. B. ALACT-Reflexionsmodell<sup>8</sup> nach Korthagen, 2001 oder EDAMA<sup>9</sup> nach Aeppli & Lötcher, 2016). Der von von Aufschnaiter et al. (2019) adressierten Doppeldeutigkeit der reflexionsbezogenen Performanz und dessen Gehalt für die empirische Forschung kann somit Sorgegetragen werden.

Um reflexionsbezogenes pPCK oder relevante Amplifiers & Filters untersuchen zu können, sollten auch Operationalisierungen für reflexionsbezogene Argumentationsprozesse verwendet werden, um zugehörige Abbilder (z. B. Reflexionstexte) untersuchen zu können. Zur Operationalisierung der reflexionsbezogenen Performanz im Sinne des ePCK plan/teach können Beobachtungskriterien (z. B. nach Meyer, 2021) oder Merkmale guten Physikunterrichts (z. B. nach Piko-Brief Nr. 4, Duit & Wodzinski, 2004) verwendet werden.

### *1.2.1 Operationalisierung von Reflexion*

Wie bereits angeklungen wird Reflexion spätestens seit Korthagen (2001) auch als Prozessmodell operationalisiert. Wohingegen es sich bei ALACT um einen zirkulären Prozess handelt, bei welchem ausgehend von einer Handlung ein Prozess durchlaufen wird, welcher mit dem sogenannten act-of-trial in einer nächsten Handlung endet, öffneten Nowak et al. (2019) den Kreislauf für den spezifischen Kontext der schriftlichen Reflexion von Unterrichtspraktika in universitären Lehramtsstudiengängen wieder. In der obig dargestellten Modellierung reflexionsbezogener Kompetenzen erscheint diese Öffnung essentiell, da so Performanz in den eigentlichen reflexionsbezogenen Argumentationsprozess (Reflexion) und einer neuen Unterrichtsplanung in einem neuen Kontext ausdifferenziert werden kann. Hierbei lässt sich auch nach dem RCM sowohl auf Makro- als auch auf Mikroebene Reflexion von Planung und Unterricht trennschärfer definieren.

---

<sup>6</sup> Reflexion als gezieltes Nachdenken über Handlungen bzw. Geschehnisse mit (1) erweiterten Blickwinkel, (2) Einbezug eigener Werte, Erfahrungen, Überzeugungen und (3) Berücksichtigung des größeren Kontextes.

<sup>7</sup> Konsekutive Reflexion bestehend aus (1a) Rahmenbedingungen, nur bei Selbstreflexion, (1b) Beschreibung, (2) Bewertung, (3) Alternativen, (4) Konsequenzen

<sup>8</sup> ALACT: (1) Action, (2) Looking back on the action, (3) getting Awareness of essential aspects, (4) Creating alternative methods of action and (5) Starting a new Action with a Trial of alternatives.

<sup>9</sup> EDAMA: (1) Erleben, (2) Darstellen, (3) Aanalysieren, (4) Mmaßnahmen entwickeln und planen, (5) Anwenden.

Zur Operationalisierung von Reflexion bemerken Nowak et al. (2019), dass ein vollständiger reflexionsbezogener Argumentationsprozess im Wesentlichen aus vier Schritten besteht: (1) Beschreibungen<sup>10</sup> der konkreten Unterrichtssituation, in denen das Agieren von Lehrkraft und Schüleri\*nnen im Fokus steht, (2) Bewertung, wie eine Person (Lehrkraft selbst oder Beobachtende) die Unterrichtssituation wahrnimmt (und warum), (3) Handlungsalternativen für die erlebten Aktivitäten, bei denen die Lehrkraft bedenken sollte, was sie anders hätte machen können, um die wahrgenommenen Erlebnisse besser zu gestalten, und (4) Konsequenzen, die sie für ihre eigene berufliche Weiterentwicklung ziehen kann.

Mit diesen konsekutiv aufgebauten Diskurselementen intendierten Nowak et al. (2019) zwei Kernziele. Einerseits sollten Lehramtsstudierenden durch die Bereitstellung eines Rahmenmodells für Reflexion in Praxisphasen unterstützt werden, Unterrichtssituationen zu reflektieren. So kann es angehenden Lehrkräften besser gelingen, sich in einer Reflexion auf wesentliche Faktoren zu konzentrieren, wenn ihnen eine Strukturhilfe bereitgestellt wird (Lai & Calandra, 2010, Poldner et al., 2014). Diese Hilfen können sowohl ein fragestellungsgeleitetes Reflexionstool oder ein Interview als auch jeder andere anregende Impuls für reflektierendes Denken sein. Kost (2019) fand beispielsweise heraus, dass die Aufforderung an angehende Physiklehrkräfte, in ihren Reflexionen vertiefende Überlegungen anzustellen, die Lehrkräfte tatsächlich dazu veranlasste, dies zu tun. Häufig werden Prozessmodelle zur Förderung des reflektierenden Denkens eingesetzt, um angehenden Lehrkräften zu helfen, ihr Denken zu strukturieren (Lin et al., 1999). Andererseits sollte das Rahmenmodell für Reflexion Forschenden die Beurteilung der Qualität der schriftlichen Reflexion ermöglichen. So scheinen beispielsweise in eher basalen, schriftlichen Reflexionen deskriptive Diskurselemente (Rahmenbedingungen und Beschreibungen) einen großen Stellenwert einzunehmen, wohingegen diskursive Elemente (Bewertungen, und begründete Alternativen) auf höherwertige reflexionsbezogene Argumentationsprozesse hindeuten. Da es in der Lehrkräftebildung bei Reflexion (in Separierung von Analyse) um die persönliche berufliche Weiterentwicklung geht (von Aufschnaiter et al., 2019), ist das Entwickeln von Konsequenzen auch in dem Prozessmodell von Nowak et al. (2019) enthalten und kann ebenfalls als eine Form des Denkens höherer Ordnung angesehen werden.

Da jedoch die bloße Verwendung höherwertiger Diskurselemente in einer schriftlichen Reflexion nicht automatisch auf einen höherwertigen Argumentationsprozess hindeuten muss, ist auch die Betrachtung weiterer Qualitätsmerkmale nötig.

---

<sup>10</sup> Im Diskurselement der Beschreibungen sind für Selbstreflexionstexte zusätzlich Rahmenbedingungen essentiell und wichtig. Für Fremdrelexionen ist diese Unterscheidung nur bedingt nötig und möglich.

### 1.2.2 Qualitätsmerkmale in schriftlichen Abbildern von Reflexion

Die Qualität reflexionsbezogener Argumentationsprozesse wird häufig auf zwei Merkmale fokussiert: die sogenannte *Breite* einer Reflexion und deren *Tiefe* (Leonhard et al. 2010). Die Breite einer Reflexion wird hierbei als Anzahl der adressierten Reflexionsauslöser verstanden und in Anlehnung an Beauchamp (2006) als die Reichweite der Themen und Ideen, die in der Reflexion behandelt werden, definiert. Eine Reflexion mit großer Breite berührt eine Vielzahl von Themen und Ideen, während eine Reflexion mit geringer Breite sich auf ein spezifisches Thema oder eine spezifische Erfahrung konzentriert. Auch kann das Einnehmen der Mehrperspektivität (Cramer et al., 2019) eine größere Reflexionsbreite zur Folge haben. Eine Reflexion mit großer Breite kann somit in der Lage sein, ein umfassenderes Verständnis der wahrgenommenen Aspekte in der Unterrichtssituation zu adressieren. Eine Reflexion mit geringer Breite muss jedoch nicht immanent von geringerer Qualität sein, da beispielsweise ein tieferes Verständnis für ein bestimmtes Thema oder eine spezifische Erfahrung adressiert werden kann und somit ein wertvoller Detailfokus entstehen kann. Der Argumentation folgend, dass tiefere Reflexionen höherwertigere reflexionsbezogene Argumentationsprozesse repräsentieren, wurden Stufenmodelle, wie z. B. das von Hatton und Smith (1995) konzeptualisiert, welche schriftliche Reflexionen oft ganzheitlich in deskriptive, dialogische und kritische Reflexionen unterscheiden. Die Tiefe einer Reflexion kann sich somit generalisierter auch auf die Gründlichkeit bzw. Intensität der Analyse und des Verständnisses mit welchen im Mittel reflexionsbezogene Argumentationsprozesse dargestellt werden, beziehen und Aspekte der Selbstbezüglichkeit (Aeppli & Lötscher, 2016) sowie des Einbezugs theoretisch formalen Wissens aus dem personal PCK (Bain et al., 1999) repräsentieren. Eine tiefere Reflexion kann somit dazu beitragen, ein tieferes Verständnis der eigenen Wahrnehmung im Zusammenhang mit der Unterrichtssituation zu erreichen, wohingegen eine geringere Tiefe eine eher auf Oberflächenmerkmale fokussierende und weniger detaillierte bzw. analytische Reflexion zur Folge haben kann.

Da es sich bei einem Reflexionstext in erster Linie um ein schriftliches Sprachprodukt handelt, wird an dieser Stelle das Qualitätsmerkmal der *Kohärenz* einbezogen. Hierbei bezieht sich Kohärenz auf die logische Verbindung von Ideen und Informationen innerhalb des Textes. Ein kohärenter Reflexionstext folgt somit im reflexionsbezogenen Argumentationsprozess einer klaren Struktur, deutlich nachvollziehbaren Zusammenhängen sowie einer klaren und verständlichen Sprache. Hierbei kann auch der „Grad der inhaltlichen Bezugnahme in den Komponenten bzw. konsekutiver/argumentativer Aufbau der Reflexion“ (von Aufschnaiter et al., 2019, S. 154) berücksichtigt werden. Weiter kann in Anlehnung an von Aufschnaiter et al.

(2019) der „Grad der Differenzierung einzelner Komponenten, z. B. im Sinne der expliziten Ausformulierung und/oder Erläuterung der Überlegungen“ (von Aufschnaiter et al., 2019, S. 154) mit der *Spezifität* als Qualitätsmerkmal berücksichtigt werden. Die Genauigkeit beziehungsweise die Bezugnahme auf die wahrgenommene Unterrichtssituation wird hierbei adressiert. Anstatt allgemeine Aussagen zu tätigen (z. B. Meinungen zu gutem Unterricht) kann einen spezifischen Reflexionstext auszeichnen, dass er detaillierte Informationen zur wahrgenommenen Unterrichtssituation darlegt und Argumentationen konkret und weniger holistisch erfolgen. Ein spezifischer Reflexionstext kann so besser in der Lage sein, Lesende von der eigenen pädagogischen Argumentation zu überzeugen und weist eine gewisse Nähe zur Unterrichtssituation im reflexionsbezogenen Argumentationsprozess auf.

Nach dem Rahmenmodell für Reflexion nach Nowak et al. (2019) kann somit die Qualität einer schriftlichen Reflexion unter anderem davon geprägt sein, dass möglichst unter konkreter Bezugnahme auf die Unterrichtssituation, die Wahrnehmung der reflektierenden Person von dieser Situation transparent und möglichst wertungsfrei beschrieben wird, bevor in weiteren Schritten erst in Bewertungen, später in Alternativen und zuletzt in Konsequenzen Bezüge zum eigenen Wissen hergestellt werden. Ohne die Berücksichtigung inhaltlicher Aspekte konnte bereits gezeigt werden, dass eine rein strukturelle Analyse der Reflexionselemente (Beschreibung, Bewertung, Alternativen, Konsequenzen) in der Lage ist, die Qualität von schriftlichen Reflexionen zumindest in Extremgruppen abschätzen zu können (Mientus et al., 2023). Wenngleich schriftliche Reflexionen den Vorteil bieten, dass Lehrkräfte ein externalisiertes Produkt erstellen, das unabhängig diskutiert werden kann (Poldner et al., 2014), kann es bei der Verwendung von Freitextantworten unerwünschte Nebeneffekte geben. So hängt die Qualität des geschriebenen Materials meist signifikant mit der Textlänge zusammen (Chodorow & Burstein, 2004; Leonhard & Rihm, 2011). Mit der computerbasierten Strukturanalyse aus Mientus et al. (2023) konnte jedoch dieses typische Qualitätskorrelat extrahiert<sup>11</sup> werden.

Zusammenfassend scheinen Stufenmodelle für die Einschätzung von Reflexionsqualität zwar geeignet zu sein, um beispielsweise einem Reflexionstext einen fixen Wert (Stufe 1, 2, 3, ...) zuzuschreiben. Gleichzeitig wird deutlich, dass sich die Qualität in einer Vielzahl von Qualitätsmerkmalen manifestieren kann. Da diese wiederum nur kategoriale Qualitätsaussagen zulassen (z. B. die Reflexionstiefe ist sehr gut, obwohl die Breite der adressierten Themen sehr gering ist), schien es weniger möglich zu sein die Qualität von schriftlichen Reflexionen valide

---

<sup>11</sup> Signifikant positive Korrelationen zwischen dem Qualitätskorrelat „Level of Structure“ und den relativen Häufigkeiten an höherwertigen Diskurselementen im Text ohne Zusammenhang zur Textlänge (Mientus et al., 2023).

zu quantifizieren. Aus diesem Grund soll in der vorliegenden Arbeit die explorative Faktorenanalyse eingesetzt werden, um die Vielzahl an Qualitätsmerkmalen auf wenige modellierbare Faktoren zurückzuführen.

#### *1.2.2.1 Methodischer Einschub zur explorativen Faktorenanalyse*

Eine explorative Faktorenanalyse ist eine statistische Methode zur Identifikation von zugrunde liegenden Dimensionen maximaler Varianz von Items in einem Datensatz (Noack, 2007). Sie wird verwendet, um zu untersuchen, ob mehrere Variablen, die in verschiedenen Konstellationen miteinander korrelieren können, durch eine kleinere Anzahl von Faktoren erklärt werden können. Die Faktoren werden dabei so gewählt, dass sie eine maximale Varianz in den beobachteten Daten erklären. Die explorative Faktorenanalyse kann nützlich sein, um komplexe Datensätze zu reduzieren und Muster und Strukturen in den Daten zu entdecken. Sie kann beispielsweise in der Soziologie eingesetzt werden, um die zugrunde liegenden Faktoren von Persönlichkeitsmerkmalen zu identifizieren. Durch die explorative Faktorenanalyse können auch Hypothesen generiert werden, die dann mit weiteren Studien überprüft werden können.

Um zu ermitteln, wie viele Faktoren einem Datensatz zugrunde liegen, sind mehrere statistische Verfahren anwendbar. Bei der Maximum-A-Posteriori (MAP) handelt es sich um eine geläufige Methode der Bayes'schen Statistik, um die wahrscheinlichste zugrunde liegende Verteilung der Parameter eines Modells zu schätzen. Es basiert auf dem Bayes'schen Theorem, das besagt, dass die Wahrscheinlichkeit einer Hypothese (auch als Posteriori-Wahrscheinlichkeit bezeichnet) durch die Wahrscheinlichkeit der Daten unter der Hypothese (Likelihood) und die Prior-Wahrscheinlichkeit der Hypothese bestimmt wird. Das Ziel von MAP ist es, den Parameterwert zu finden, der die Posterior-Wahrscheinlichkeit maximiert. Der MAP-Schätzer ist der Parameterwert, der die höchste Wahrscheinlichkeit hat, basierend auf den beobachteten Daten und der a priori Annahme über die Verteilung der Parameter. Mit anderen Worten, es ist der Punkt in der Parameterverteilung, der am besten zu den beobachteten Daten passt, unter Berücksichtigung der a priori Annahme über die Verteilung der Parameter. Die Berechnung des MAP-Schätzers erfordert in der Regel die Wahl einer Prior-Verteilung, die die a priori Annahme über die Verteilung der Parameter ausdrückt. Diese Prior-Verteilung wird dann mit der Likelihood-Funktion multipliziert, um die Posterior-Verteilung zu erhalten, die die Wahrscheinlichkeit der Parameterwerte nach der Beobachtung der Daten darstellt. Die maximale Stelle der Posterior-Verteilung entspricht dann dem MAP-Schätzer.

MAP wird häufig in der statistischen Inferenz und dem maschinellen Lernen angewendet, um Modelle mit parametrischen Annahmen zu schätzen und Vorhersagen zu machen. Es bietet eine Möglichkeit, die Unsicherheit in den Schätzungen durch die Wahl einer a priori Annahme über die Verteilung der Parameter zu quantifizieren und zu berücksichtigen. Durch die Verwendung von Apriori-Informationen (wie in MAP) über die Verteilung der Parameter kann die Schätzung genauer und stabiler gemacht werden. Die MAP-Schätzung kann hilfreich sein, um einer Überanpassung (Overfitting) der Faktorenanalyse an die Daten zu vermeiden.

Die zwei gängigsten Ansätze zur Faktorenanalyse zur Untersuchung von Beziehungen zwischen Variablen sind die orthogonale und obliquen Verfahren. Orthogonale Verfahren wie die Hauptkomponentenanalyse und die Varimax-Rotation erzeugen Faktoren, die unabhängig voneinander sind und senkrecht aufeinander stehen. Das bedeutet, dass jeder Faktor nur einen Teil der Gesamtvarianz erklärt und keine Korrelation zwischen den Faktoren vorhanden ist. Oblimite Verfahren wie die Promax-Rotation erzeugen Faktoren, die korreliert sein können. Diese Art der Rotation erlaubt es, Faktoren zu extrahieren, die in Beziehung zu einander stehen und nicht unabhängig voneinander sind. Die Promax-Rotation ist besonders hilfreich, wenn man davon ausgeht, dass die Faktoren miteinander korrelieren und in der Lage sind, gemeinsam eine bestimmte Variable zu erklären. Beide Verfahren haben ihre Vor- und Nachteile und sollten in Abhängigkeit von der Forschungsfrage und den Daten, die untersucht werden, ausgewählt werden. Orthogonale Verfahren eignen sich besser, wenn die Faktoren unabhängig voneinander sind und keine Korrelation zwischen ihnen besteht. Oblimite Verfahren sind besser geeignet, wenn Korrelationen zwischen Faktoren erwartet werden oder wenn man annimmt, dass die Faktoren miteinander in Beziehung stehen.

Um mit einer einzigen Stichprobe die Exploration der Faktorenanalyse durchführen zu können und im Rahmen derselben Studie zu validieren, kann ein Datensatz in einen Trainings- und Testdatensatz aufgeteilt. Als geeignetes Verhältnis kann hierbei in einer Aufteilung von 80% Trainingsdaten und 20% Testdaten darstellen. Mit der Faktorenanalyse kann es möglich sein, identifizierte Textmerkmale oder Qualitätsbewertungen von Expert\*innen auf wenige Faktoren zu reduzieren, mit welchen mittels mathematischer Modellierung Reflexionsqualität in einem spezifischen Kontext valide quantifiziert werden kann.

### *1.2.3 Reflexionsbezogene Dispositionen / Amplifiers & Filters*

Für den Austausch zwischen den drei Ebenen des PCK (collective, personal & enacted) unterstreicht das RCM die Relevanz verschiedener Amplifiers & Filters. Diese wirken als begünstigende oder einschränkende Moderatoren in der Kompetenzperformanz bei Lehrkräften

der Naturwissenschaften (Carlson et al., 2019). Diese Moderatoren können ausschlaggebend dafür sein, welche Informationen und Ratschläge eine Lehrkraft aus der Gemeinschaft der Lehrkräftebildenden oder dem Kollegium für das personal PCK aufnimmt (oder einbringt) (Carlson et al., 2019). Analog können neben den Wissensbeständen und Fähigkeiten z. B. zum Aufbau einer konsekutiven Reflexion auch Amplifiers & Filters in reflexionsbezogenen Argumentationsprozessen einen Übertrag von Information von ePCK in das pPCK begünstigen oder hemmen (Carlson et al., 2019). Eine Qualitätsaussage über eine Reflexion kann somit unmittelbar im Zusammenhang mit diesen Amplifiers & Filters stehen.

In der Forschung der Lehrkräfteprofessionalisierung wird beispielsweise den Selbstwirksamkeitserwartungen eine zentrale Bedeutung zugeschrieben (Bach, 2022; Franken & Preisfeld, 2023). Selbstwirksamkeit kann in Anlehnung an Bandura (1997) als die Überzeugung angesehen werden, aufkommende Herausforderungen auf Grundlage der eigenen Kompetenzen bewältigen zu können. So steht in vielen Merkmalen die eigene Unterrichtsqualität mit Selbstwirksamkeit in Zusammenhang (Klassen & Tze, 2014; Kuusinnen, 2016) – ein Effekt, welcher sich potentiell auch auf die Reflexion von Unterricht übertragen lässt (Lohse-Bossenz et al., 2019). Vereinzelt Effekte von Reflexionsanlässen auf Selbstwirksamkeit konnten beispielsweise in inklusionsspezifischen Kontexten nachgewiesen werden (Gaßner-Hofmann, 2022). Da Selbstwirksamkeit aber als kontext- oder domänenspezifisch gilt (Bandura, 1997; 2006), sollten Aussagen studienbezogen geprüft werden.

Generell kann der Lernkontext selbst bereits eine große Rolle einnehmen (Carlson et al., 2019). Wenngleich einzelne Forschung darauf hindeutet, dass sich das PCK von Lehrkräften der Naturwissenschaften unabhängig vom Lernkontext entwickelt (Berry et al., 2016, Kersting et al., 2010) heben andere Forschende die Kontextspezifität von PCK sogar heraus (Park & Oliver, 2008b, Dunekacke et al., 2015). Hinreichend trennscharfe Erhebungen scheinen nicht trivial zu sein, weshalb es gewinnbringend sein kann, nicht nur einen Lernkontext mit einer Vignette zu standardisieren, sondern die wahrgenommene Unterrichtsqualität zu erheben (Schlesinger et al., 2018). Die individuelle Wahrnehmung von Lehrkräften kann maßgeblich vom bewussten Wissen (bzw. situationsspezifischen personal PCK) beeinflusst sein (Wahl, 2002). Dies gilt im Besonderen für subjektive Theorien über das, was guten Unterricht auszeichnet (Wiemann, 2007) und sollte daher berücksichtigt werden um der Kontextspezifität auch auf individueller Ebene gerecht zu werden.

Aus inhaltlicher Perspektive kann sich beispielsweise Fachwissen als ein wichtiger Moderator für unterrichtsbezogene Performanz erweisen (Abell, 2007; Sadler et al., 2013;

Fischer et al., 2014). So konnten beispielsweise Sorge et al. (2019) in der Untersuchung von Zusammenhängen zwischen collective PCK und Aspekten der Unterrichtsqualität im enacted PCK feststellen, dass sich das Fachwissen der untersuchten Lehrkräfte als relevanter Indikator für die PCK-Entwicklung erwies. Reflexionsbezogen scheinen zum aktuellen Zeitpunkt weniger Zusammenhänge zum Fachwissen aufgezeigt.

Aus methodischer Sicht kann für eine Reflexion Werthaltung eine wichtige Rolle einnehmen und so adressieren auch von Aufschnaiter et al. (2019) die Einstellungen zur Relevanz von Reflexionen als reflexionsbezogene Disposition (siehe Abbildung 3). Werthaltung ist eine stabile Disposition, welche im wissenschaftlichen, deutschen Sprachgebrauch unter anderem charakterisieren kann, wie wichtig ein Wert oder eine Handlung (wie eben das Reflektieren an sich) für ein Individuum ist (Dorsch, 2021). Werthaltungen können im Gegensatz zu einer Einstellung, Wahrnehmungen und Handlungen über viele spezifische Situationen, Personen und Gegenstände hinweg bestehen, da sie zum Teil schon sehr früh erworben beziehungsweise entwickelt werden und auch über lange Zeiträume stabil bleiben (Philipp, 2007). Doch scheint auch die Forschung von Zusammenhängen zwischen Werthaltung und Reflexion derzeit rar zu sein.

An den ausgewählten Beispielen für Amplifiers & Filters kann deutlich werden, dass in der Reflexionsforschung zum aktuellen Zeitpunkt in Teilen diffus bleibt, inwieweit statt der angestrebten Kompetenzmessung eher unerwünschte Nebeneffekte von Amplifiers & Filters untersucht werden. Um zugehörige Amplifiers & Filters eingrenzen zu können oder ggf. explizite Verbindungen zwischen Abbildern reflexionsbezogener Argumentationsprozesse und ausgewählten Amplifiers & Filters zu untersuchen, kann ein normierter Reflexionsanlass gewinnbringend sein. Dieser kann im Rahmen der eigenen Unterrichtserfahrung (siehe Nowak, eingereicht), als auch im Rahmen einer explizierten Lerngelegenheit durch z. B. eine Videovignette bestehen.

#### *1.2.4 Reflexionsanlässe in Lehre und Forschung*

Möglichkeiten zur Beobachtung realer Unterrichtssituationen für Studierende um Denk- und Lernprozesse von Schüler\*innen zu verstehen sind nach Fischer und Weinert (2021) rar. Daher können sich Videographien eigenen Unterrichts einzeln oder in der Gruppe besprochenen als eine wertvolle Lerngelegenheit erweisen. Auf diese Weise kann die Unterrichtssituation jederzeit wiederholt und unabhängig betrachtet werden. Dies kann die Betrachtung des Videos aus verschiedenen Perspektiven ermöglichen und kann somit dazu beitragen, die Reflexionsfähigkeit zu entwickeln und zu verbessern (Elsner et al., 2020). Die Kompetenz,

relevante Unterrichtssituationen zu erkennen, adäquat zu interpretieren und auf dieser Grundlage funktionale Schlussfolgerungen für das professionelle Handeln zu ziehen (Sherin, 2001), kann mit Hilfe von Unterrichtsvideos gefördert werden.

Unterrichtsvideos können einen Ausgangspunkt für Selbst- und Fremdrelexionen von angehenden Lehrkräften bilden (Gaudin & Chaliès, 2015) und können von angehenden Physiklehrkräften genutzt werden, um theoretisches Wissen und erworbene Konzepte in spezifischen und authentischen Unterrichtssituationen zu üben (Wyss, 2018). Um Reflexivität explizit zu fördern, wird der Einsatz von fremden Unterrichtsvideos in Erwägung gezogen (van Es & Sherin, 2021; Elsner et al., 2020a; Hoth et al., 2018). Unterrichtsvideographien stellen eine Quasi-Teilnahme an realen Unterrichtssituationen dar und eröffnen somit Lernmöglichkeiten zur Verknüpfung von Theorie und Praxis (Krammer & Reusser, 2005; Seidel et al., 2013) und können sich als Reflexionsanlass eignen.

Besonders in Videovignetten können (angehende) Lehrkräfte lernwirksamen Unterricht wahrnehmen und analysieren und weiter berufsrelevante Kompetenzen erwerben (Santagata & Angelici, 2010; Alonzo et al., 2012; Seidel & Stürmer, 2014). Aufgrund der Möglichkeit einer größeren Distanzierung können fremde Unterrichtsvideos (auch in Vignetten) eine kritische Haltung fördern (Kücholl & Lazarides, 2021). Unter anderem, da die Komplexität oder Unmittelbarkeit von Unterrichtssituationen sowie die Reaktion der videographierten Lehrkraft in einer Vignette abgebildet werden können (Lindmeier, 2011).

Zur Analyse von Reflexionskompetenzen werden häufig offene Bewertungsformate verwendet. Diese ermöglichen Lehrkräften, ihr personal PCK und ihre Dispositionen einzusetzen, um eine Unterrichtssituation zu reflektieren. Die Bewertung offener Antwortformate ist jedoch eine Herausforderung. In den meisten Fällen wird in der Lehrkräftebildung eine Form der schriftlichen Bewertung, wie z. B. Aufsätze oder Berichte, verwendet (Poldner et al., 2014). Ausgehend von der Annahme, dass reflexives Schreiben eine wichtige pädagogische Praxis ist, die durch die zur Verfügung stehende Unterrichtszeit begrenzt ist (Ullmann, 2017), bietet sich die Reflexion im Rahmen einer Videovignette als geeignete alternative Lernmöglichkeit an, um reflexives Schreiben in der universitären Lehrkräftebildung zu üben.

Schriftliche Reflexionen (auch in Vignetten) werden häufig manuell und qualitativ ausgewertet. Zur Analyse der schriftlichen Produkte wird die Inhaltsanalyse eingesetzt (Hume, 2009). Die manuelle Inhaltsanalyse ist jedoch mit Herausforderungen behaftet. Leonhard & Rihm (2011) berichten beispielsweise, dass sie die Kodierung von schriftlichen Reflexionen nicht skalieren können, weil sie zu ressourcenintensiv ist. Lehramtsstudiengängen ist daher

meist eine eher ganzheitliche (und weniger analytische) Kodierung von schriftlichen Reflexionen möglich (Ullmann, 2019; Poldner et al., 2014). Viele Forschende, die sich mit der Inhaltsanalyse schriftlicher Reflexionen befassen, mussten weitere Schwierigkeiten beim Erreichen von Interrater-Übereinstimmungen feststellen, die möglicherweise durch unklare Kodiereinheiten und hochinferentielle Kategorien verursacht werden (z. B. Kost, 2019, Abels, 2011). Außerdem impliziert die Inhaltsanalyse oft eine ganzheitlichere Bewertung, bei der eine Antwortkategorie mehrere konzeptionelle Elemente umfasst (Jescovitch et al., 2021) und abschließend diffus bleibt, welche Qualitätsmerkmale auf welche Weise in eine Bewertung einfließen.

Während eher kurze konstruierte Antworten meist zu einer akzeptablen Interrater-Übereinstimmung führen, sind längere schriftliche Reflexionen (wie sie in der Praxis anzutreffen sind), wesentlich schwieriger zu handhaben. Schriftliche Fremdre reflexionen stellen ein komplexes Argumentationssystem dar, in welchem es schwierig sein kann, valide Qualitätsbewertungen einer potentiell noch komplexeren Kompetenz vorzunehmen. Daher scheint auch eine Qualitätsabschätzung von Antwortformaten in einer Vignette durch den Vergleich mit einer Expertennorm eine bewährte Praxis zu sein (z. B. Oser & Heinzer, 2009; Oser et al., 2010).

Aus theoretischer Sicht können datengesteuerte und computergestützte Ansätze zur Analyse des reflexiven Schreibens neue Möglichkeiten bieten, da sie Sprachdaten auf eine systematische Weise analysieren wie dies für menschliche Rater kaum oder nur mit großem Aufwand möglich ist. Dadurch können neue Zusammenhänge in komplexen Daten erfasst werden und potenziell Rückschlüsse auf komplexe Kompetenzen identifiziert werden.

### *1.2.5 Computerbasierte Analyse schriftlicher Reflexionen*

Während Studien im Bereich der Reflective Writing Analytics (Buckingham Shum et al., 2017; Ullman, 2019; Wulff et al., 2022) darauf hindeuten, dass computerbasierte Methoden wie maschinelles Lernen (ML) gut geeignet sind, um schriftliche Reflexionen analytisch zu bewerten, ist eher unklar, inwieweit aus diesen ML-basierten Bewertungen Rückschlüsse auf die Qualität schriftlicher Reflexionen möglich sind. Datengesteuerte und computergestützte Ansätze für die Analyse reflektierten Schreibens können neue Möglichkeiten bieten, da sie (im Prinzip) nicht durch menschliche Schwächen wie begrenztes Kurzzeitgedächtnis, Tagesform oder unterschiedliches Sprachgefühl und Vorwissen eingeschränkt sind. In der Praxis ergeben sich im Zusammenhang mit ML-Methoden jedoch neuartige Probleme (wie Biases), da sie derzeit durch die Datenbasis, auf die sie zurückgreifen können, eingeschränkt sind. Dennoch

ist es erwähnenswert, dass die induktive Informations- und Wissensextraktion aus Textdaten (einschließlich der gesamten Wikipedia oder des Internets) mit generativen Sprachmodellen wie GPT4 neue Dimensionen erreicht hat. Diese haben medizinische, pädagogische oder mathematische Prüfungen bereits ohne spezielles Training auf dem Niveau menschlicher Leistungen bestanden (Katz et al., 2023). Auch für das Fach Physik ist dies bereits nachgewiesen, sowohl für Multiple-Choice Fragen als auch für kurze Essays (West, 2023a, 2023b; Kortemeyer, 2023; Yeadon et al., 2023). So bleibt weiter offen, inwieweit bereits aus den genannten Qualitätsmerkmalen, welche computerbasiert analysierbar sind ein Qualitätskorrelat modelliert werden kann, welches Expertenbewertungen entsprechend die Qualität schriftlicher Reflexionen abschätzt.

### ***1.3 Forschungsanliegen***

Aus den theoretischen Überlegungen dieser Arbeit ging hervor, dass Reflexion (im Sinne des Reflektierens oder der Performanz) und Reflexivität (als Summe aller anwendbaren reflexionsbezogenen Kompetenzen) zu unterscheiden sind. Zusammengefasst zeigen sich diese Kompetenzen (in Anlehnung an Blömeke, 2015) in den Fähigkeiten, in variablen Reflexionsanlässen Wissen anzuwenden, um Reflexionsauslöser zu durchdenken um personal PCK oder den eigenen Unterricht weiterzuentwickeln. Ausgehend von einem Reflexionsauslöser kann ein reflexionsbezogener Argumentationsprozess angeregt werden, welcher beispielsweise aufbauend auf eine neutrale Beschreibung und die Bewertung der wahrgenommenen Unterrichtssituation Alternativen und Konsequenzen diskutiert. Reflexion kann hierbei als allgemeiner Begriff für die Ausführung reflexionsbezogener Argumentationsprozesse in Handlungen (bspw. Schreiben eines Reflexionsberichts) angesehen werden und stellt somit immer nur ein Abbild von Reflexivität dar.

Da aktuell einer großen Zahl von normativen Überlegungen zu Reflexion und Reflexivität ein nur schwach ausgeprägter empirischer Forschungsstand gegenübersteht (von Aufschnaiter et al., 2019) soll das Ziel dieser Arbeit darin bestehen, Zusammenhänge von reflexionsbezogenen Argumentationsprozessen und Amplifiers & Filters zu identifizieren. Mit der vorliegenden Studie kann in der Reflexionsforschung darauf aufmerksam gemacht werden, welche Amplifiers & Filters auf welche Weise zu berücksichtigen sind, wenn Reflexivität untersucht werden soll, aber lediglich Abbilder reflexionsbezogener Argumentationsprozesse (wie eben Reflexionstexte) analysiert werden können. In der Arbeit soll daten-gestützt ein Qualitätsmaß für reflexionsbezogene Argumentationsprozesse entwickelt werden, um Zusammenhänge zwischen Reflexion und Amplifiers & Filters zu identifizieren.

### 1.3.1 Forschungsfragen

Wie aus der Literatur hervorgeht, sind Qualitätsbewertungen einer Reflexion durch Experten eine verbreitete Praxis. In dieser Arbeit soll anhand derartiger Expertenbewertungen empirisch untersucht werden, durch welche computerbasiert analysierbaren Textmerkmale die Qualität von Reflexionen modelliert werden kann, um Qualitätsbewertungen zuverlässig zu quantifizieren. Es sollen dabei Zusammenhänge mit allgemeinen und fachspezifischen Amplifiers & Filters aufgeklärt werden. Da nach den obig ausgeführten Annahmen Abbilder von reflexionsbezogenen Argumentationsprozessen nicht mit Reflexivität im Sinne von Kompetenz gleichgesetzt werden können, wird angenommen, dass es zu voreilig ist, von Reflexionen unmittelbar auf Reflexivität (angehender) Lehrkräfte zu schließen. Aus diesem Grund soll entlang der folgenden offenen Forschungsfragen dem Anliegen nachgegangen werden. Weil es sich bei den Forschungsfragen der vorliegenden Arbeit um ein exploratives Vorgehen handelt, werden keine gerichteten Hypothesen aufgestellt.

*Forschungsfrage 1:* Auf welche zentralen Faktoren sind die Qualitätsmerkmale aus 1.2.2 zurückzuführen und wie stehen diese zueinander?

*Exploration:* Im Rahmen der Analyse schriftlicher Reflexionstexte sind eine Vielzahl von Variablen identifizierbar. Dies sind die unter 1.2.2 diskutierten Qualitätsmerkmale der Breite, Tiefe, Kohärenz und Spezifität der schriftlichen Reflexion sowie die computerbasiert analysierbaren Textmerkmale der Textlänge oder Argumentationsstruktur (repräsentiert durch den LOS nach Mientus et al., 2023). Da wie diskutiert die Qualität von schriftlichen Reflexionen von verschiedenen Aspekten beeinflusst sein kann, soll mittels Faktorenanalyse exploriert werden, auf welche Weise Qualitätsmerkmale von Reflexion zu Aspekten vom Umfang der schriftlichen Reflexionen oder deren beschreibenden Anteil stehen.

*Forschungsfrage 2:* Inwieweit sind Qualitätskorrelate zwischen den computerbasiert analysierbaren Textmerkmalen und Qualitätsbewertungen von Experten in schriftlichen Fremdrelexionen auflösbar?

*Exploration 2:* Unter der Annahme, dass eine Qualitätsabschätzung in der vorliegenden Stichprobe valide ist, ergibt sich die Frage, wie eine schriftliche Fremdrelexion beschaffen sein muss, um mit einer möglichst hohen Qualität zu korrelieren. Mit dieser Forschungsfrage,

soll exploriert werden, inwieweit es möglich ist, ein Optimierungsproblem auf die bestehenden Analysen zu übertragen, um computerbasierte Qualitätsabschätzungen möglichst nah an menschlichen Qualitätsbewertungen heranführen zu können.

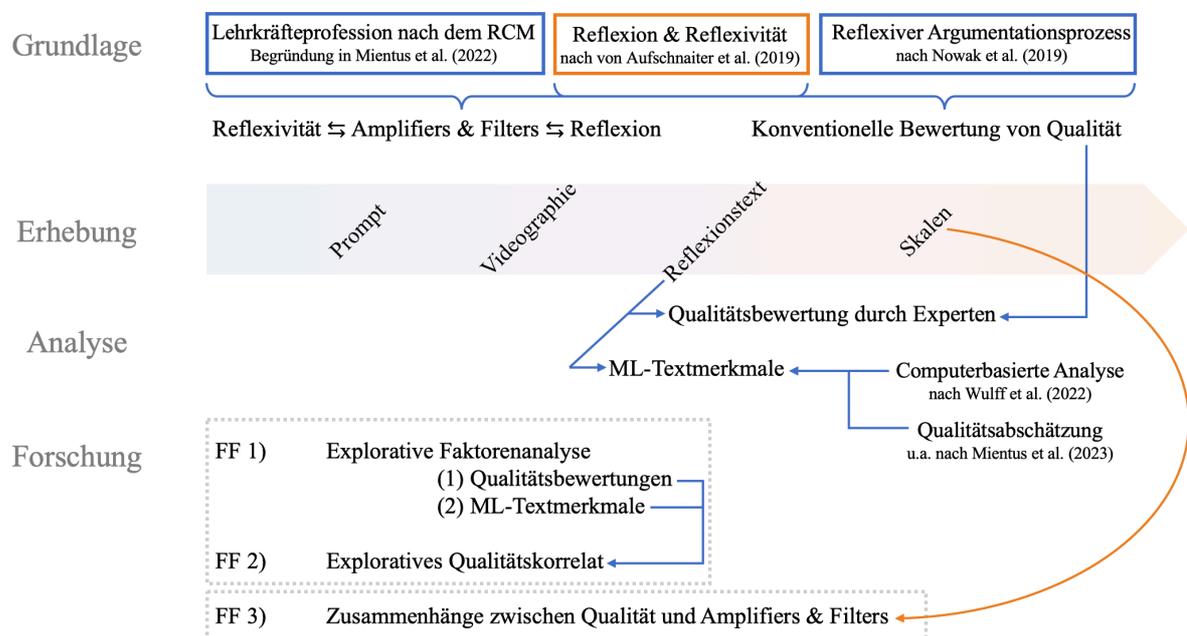
*Forschungsfrage 3:* Inwieweit stehen Amplifiers & Filters mit den Qualitätskorrelaten der Reflexionstexte in Zusammenhang?

*Exploration 3:* Nachdem computerbasiert abgeschätzt werden kann, worin die Qualität schriftlicher Reflexionen im Rahmen dieser Studie besteht, sollen Zusammenhänge der erhobenen Amplifiers & Filters zu den Qualitätskorrelaten analysiert werden. Auf diese Weise soll aufgedeckt werden können, welche weiteren Einflussfaktoren die Qualität einer schriftlichen Fremdrelexion beeinflussen könnten und somit für den Rückschluss vom Abbild des reflexionsbezogenen Argumentationsprozesses auf die Reflexivität einer Person ein Rauschen in der Kompetenzmessung erzeugen.

Zusammenfassend soll untersucht werden, inwieweit bereits computerbasierte Modellierungen die Qualität von Reflexionstexten abschätzen können und inwieweit diese mit reflexionsbezogenen Amplifiers & Filters in Zusammenhang stehen. Schlussfolgernd können Amplifiers & Filters aufgezeigt werden, welche im Bereich der Messung von Reflexivität in weiteren Studien besonders zu berücksichtigen sein könnten, wenn von der Qualität reflexionsbezogener Performanz (im Sinne der Abbilder reflexionsbezogener Argumentationsprozesse) auf Reflexivität geschlossen werden soll. Zwei Untersuchungsgegenstände sind somit für die vorliegende Studie von Relevanz: (1) die schriftliche Fremdrelexion sowie alle zugehörigen Textmerkmale und (2) die weiteren erhobenen Skalen, welche ausgewählte Amplifiers & Filters abbilden sollen.

### 1.3.2 Übersicht über die Studie

Für eine Erleichterung des Verständnisses der vorliegenden Untersuchung gibt Abbildung 5 einen Überblick über die zentralen Elemente aus der Theorie, die methodischen Modellierungen der computerbasierten Textanalyse, der Vignettenstudie sowie der Analysewege nach den Forschungsfragen. Nachdem die theoretischen Vorüberlegungen (Grundlagen) bereits diskutiert wurden, wird im Folgenden zunächst die Erhebung im Rahmen einer Physik-Videovignette (Erhebung) vorgestellt und in zugehörige Analysemethoden (Analyse) eingeführt. Gewonnene Ergebnisse werden entlang der formulierten explorativen Forschungsfragen vorgestellt und abschließend diskutiert.



**Abb. 5** Visualisierung der Studienstruktur in Grundlagen, Erhebung und Analyse

## 2 Methode

Um der aufgezeigten Breite der Reflexionsforschung anschlussfähig zu sein, wurde auch für die vorliegende Studie eine online Videovignette implementiert. Die Teilnehmenden (1) erhielten einen Prompt mit einer Einführung in das Reflexionsmodell nach Nowak et al. (2019), (2) sahen einen 17-minütigen Ausschnitt einer authentischen Unterrichtssituation zur Einführung des Freien Falls, (3) verfassten einen schriftlichen Fremdrelexionstext und (4) beantworteten Fragen zu den erhobenen Skalen.

Infolge des Einsatzes der Videovignette in unterschiedlichen Kontexten (Praxissemester als Pre-Post-Erhebung, Lehrkräftefortbildung in der zweiten Ausbildungsphase, Vergleich Bachelor-Masterstudierende verschiedener Hochschulstandorte mit und ohne Aufwandsentschädigung) wurden die erhobenen Daten aus den verschiedenen Vignettenumgebungen in einem ersten Schritt in einen Datensatz zusammengefasst, welcher alle Items aus allen Versionen erhält und zugehörige Bearbeitungszeiten der Teilnehmenden erfasst. Auf diese Weise bleiben im ersten Schritt der Datenaufbereitung alle Informationen erhalten. In einem zweiten Schritt werden erhobene Variablen vereinheitlicht. Beispielsweise wurde bei den Potsdamer Studierenden der Studienstandort nicht erhoben und muss folglich im Prozess der Datenaufbereitung unter dem Item ‚Standort‘ entsprechend ergänzt werden. Ebenso wurde die Vignette an anderen Hochschulstandorten in beispielsweise zwei Seminaren eingesetzt, welche in der Datenaufbereitung eindeutig den Bachelor- oder Masterstudierenden zugeordnet werden können. Bis zu diesem Punkt bleiben alle Originaldaten erhalten.

Im Prozess der Datenaufbereitung wird in einem ersten Schritt allen Daten, bei welchen die TN kein persönliches Pseudonym erstellt haben, da sie beispielsweise die Befragung vorher Abbrechen ein neutrales Pseudonym (z. B. ‚ZZZZZZ01‘) zugeordnet um den Datensatz weiterhin einheitlich nach Pseudonymen auswerten zu können. Im zweiten Schritt des Cleanings werden alle Datensätze entfernt, welche keinen oder einen stichpunktartigen Reflexionstext beinhalten. Dieser Prozess wird durchgeführt, da die Verschriftlichung des reflexionsbezogenen Argumentationsprozesses als reflexionsbezogene Performanz die zentrale Datenquelle dieser Studie darstellt. Im dritten Schritt werden auf Grundlage der Pseudonyme unter Berücksichtigung der Zugriffszeiten der Teilnehmenden auf die Vignette alle Erstteilnahmen identifiziert und übrige Datensätze zusätzlich entfernt. Wenngleich eine doppelte Teilnahme insbesondere unter der Beibehaltung der neutral generierten Pseudonyme nicht ausgeschlossen werden kann, wird somit Einflüssen eines Wiederholungseffektes entgegengewirkt und die Objektivität der durchgeführten Untersuchung erhöht.

## ***2.1 Automatisierung der Analyse von Reflexionstexten***

Da für die Beantwortung der explorativen Forschungsfragen die ML-basierten Textanalysen von zentraler Relevanz sind, werden diese im Folgenden zunächst losgelöst vorgestellt. Zusätzlich kann so für den weiteren Verlauf dieser Arbeit ein besseres Verständnis über die zugrundeliegenden Algorithmen ermöglicht werden.

Dank des Fortschritts in der Modellierung und Automatisierung der Sprachverarbeitung können heute Methoden des maschinellen Lernens (ML) und der Verarbeitung natürlicher Sprache (NLP – Natural Language Processing) genutzt werden, um Abbilder komplexer Kompetenzen wie schriftliche Reflexionen zu analysieren und zu bewerten. Die Anwendungen von ML und MLP sind vielfältig und umfassen unter anderem die Bewertung von offenen Antwortformaten oder gar die Generierung von Feedback (Shermis et al., 2019; Zhai et al., 2020). ML ermöglicht es Computerprogrammen, aus neuen Daten zu lernen und ihre Leistung bei einer bestimmten Aufgabe zu verbessern, während sich NLP auf die effektive Darstellung und automatische Verarbeitung von Sprache bezieht (Mitchell, 1997).

Es gibt verschiedene Ansätze des maschinellen Lernens, darunter überwachtes (supervised) und unüberwachtes (unsupervised) Lernen. Beim überwachten Lernen werden annotierte Daten vorgelegt, um einen Algorithmus zu trainieren eine gewisse Zuordnung trefflich leisten zu können. Beim unüberwachten Lernen können Algorithmen in der Lage sein auch ohne gezielteres Training Cluster in den Daten zu entdecken. Beim Einsatz von unüberwachtem Lernen zur Textanalyse können beispielsweise Themen, welche in Texten behandelt werden, extrahiert werden. Überwachtes Lernen ist besonders nützlich, wenn menschliche Rater reliabel auf der Grundlage eines Kodiermanuals beispielsweise Segmente eines Textes nach einem Modell kodieren. Um in dieser Weise eine bestehende Kodierungsaufgabe fortzusetzen, müssen folglich nicht immer neue Experten ausgebildet werden, da ML-basierte Methoden vergleichbar zuverlässige Ergebnisse erzielen können. Das Training eines derartigen Algorithmus mit den bestehenden menschlichen Zuordnungen ermöglicht die Durchführung einer automatischen Bewertung von neuen, unbekanntem Daten, welches mit vergleichbaren Reliabilitäten zuverlässig für die spezifische Kodieraufgabe eingesetzt werden kann.

Im Kontext schriftlicher Reflexionen wurden beide Ansätze bereits erfolgreich umgesetzt (z. B. Ullmann, 2017; Wulff et al., 2023). So verwendete beispielsweise Ullmann (2019) als theoretische Rahmung acht Kategorien<sup>12</sup> zur Kodierung, welche aus 24 Modellen des

---

<sup>12</sup> reflection (depth of reflection), experience, feelings, personal belief, awareness of difficulties, perspective, lessons learned, and future intention (breadth of reflection) (Ullmann, 2019, p. 218)

reflektierenden Schreibens abgeleitet wurden und konnte vergleichbare Beurteilerübereinstimmungen zwischen Mensch und ML-Modell erreichen, wie zwei menschliche Kodierer untereinander.

### *2.1.1 Zugrundeliegender Klassifikationsalgorithmus*

Nach ähnlichem Vorgehen trainierten auch Wulff et al. (2020) einen Klassifikationsalgorithmus für das Rahmenmodell für Reflexion von Nowak et al. (2019). Das ML-Modell wurde anhand der Analyse von Physiktexten trainiert, da Physik als ein schwieriges und wissensintensives Fach gilt, in dem Rahmenmodelle zur Unterstützung der Reflexion für angehende Lehrkräfte von Vorteil sein könnten, um zu verstehen, wie schwierig Physik für Lernende ist (Hume, 2009; Sorge et al., 2018). Um das ML-Modell zu erstellen, wurden die schriftlichen Reflexionen in Segmente (in unserem Fall Sätze) unterteilt, die in Merkmale (Prädiktorvariablen) umgewandelt werden, die es dem Klassifikator ermöglichen, Wahrscheinlichkeiten für jedes der Reflexionselemente für jedes einzelne Segment vorherzusagen. In der Studie von Wulff et al. (2020) wurde erwartet, dass die in einem Segment verwendeten Wörter ein wichtiges Merkmal für die Repräsentation der Segmente sind, da Beschreibung wahrscheinlich Prozessverben wie „ein (Experiment) durchführen“ hervorruft, Bewertung durch Wörter wie „gut“ oder „schlecht“ charakterisiert sein könnte und Alternativen das Schreiben im konditionalen Modus erfordern (nach Ullmann, 2019). Der Klassifikator basiert also primär auf der Analyse von Wörtern, da diese häufig als Ausgangspunkt für die Modellierung verwendet wird (Ullmann 2019).

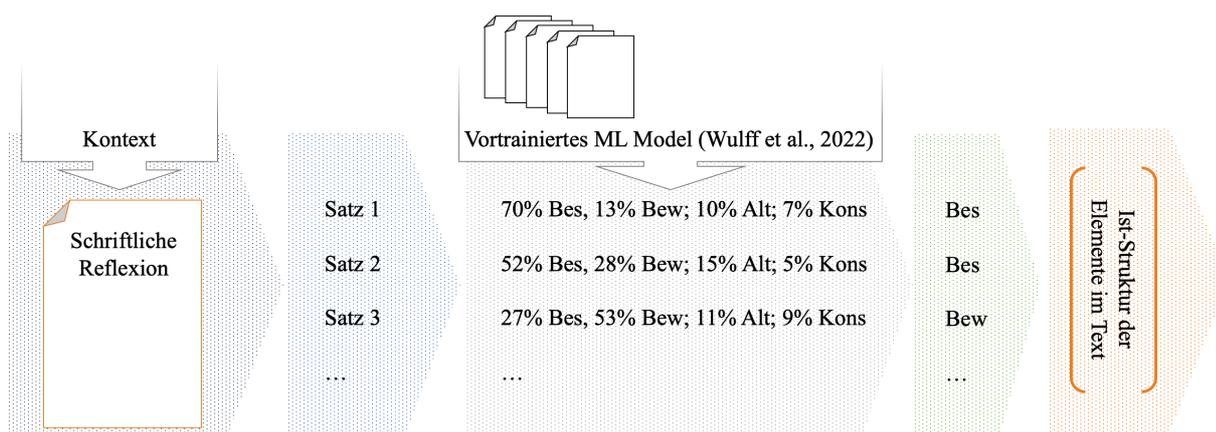
Wulff et al. (2020) trainierten ihr Klassifikationsmodell zur Analyse der reflexionsbezogenen Argumentation in schriftlichen Reflexionen von Physiklehrkräften unter Verwendung des Sprachmodells BERT. BERT steht für Bidirectional Encoder Representations from Transformers. Es ist ein vortrainiertes Deep-Learning-Modell, das eine Transformator-Architektur verwendet, um authentische Texte zu analysieren und zu verstehen. Das Modell wurde von Google entwickelt und 2018 veröffentlicht. Es wurde auf einem großen Korpus von Textdaten trainiert, so dass es kontextuelle Beziehungen zwischen Wörtern in einem Satz lernen kann. BERT muss mit satzbasierten Kodierungen für eine bestimmte Aufgabe trainiert werden und ist dann in der Lage, diese Kodierungen auf unbekannte Texte anzuwenden. Die Grundlage für das satzbasierte Klassifikationsmodell von Wulff et al. (2020) ist das Rahmenmodell für Reflexion nach Nowak et al. (2019). Das Modell wurde zudem im spezifischen Kontext von schriftlichen Reflexionen in Physik-Lehramtspraktika in universitären Lehramtsstudiengängen

entwickelt und konzeptualisiert die Elemente (Rahmenbedingungen, Beschreibungen, Bewertungen, Alternativen und Konsequenzen) als konstitutiv.

Das ML-Modell wurde nach dem Training im Hinblick auf seine Klassifikationsleistung durch Mensch-Maschine-Abgleich bewertet. Mit dem so genannten F1-Score wurde hier eine häufig verwendete Metrik in der ML-Klassifikation verwendet, welche ein harmonisches Mittel aus Precision und Recall wiedergibt. Die Precision gibt hierbei an, wie viele der vorhergesagten positiven Ergebnisse tatsächlich korrekt sind, während der Recall angibt, wie viele der tatsächlich positiven Ergebnisse korrekt vorhergesagt wurden. Beide Metriken sind wichtig, da sie unterschiedliche Aspekte der Leistung eines Modells berücksichtigen. Ein hoher Precision-Score zeigt an, dass das Modell in der Lage ist, falsch positive Vorhersagen zu minimieren, während ein hoher Recall-Score darauf hinweist, dass das Modell in der Lage ist, falsch negative Vorhersagen zu minimieren. Folglich gibt es keine absoluten Schwellwerte für einen guten F1-Score, da diese Bewertung von verschiedenen Faktoren abhängt, in der Regel gilt jedoch, dass ein höherer F1-Score besser ist, da dies darauf hinweist, dass das Modell eine höhere Precision und einen höheren Recall hat. Eine Möglichkeit, um zu beurteilen, ob ein F1-Score gut ist oder nicht, besteht darin, ihn mit dem F1-Score eines Baseline-Modells oder anderen Modellen zu vergleichen, die für das gleiche Problem trainiert wurden.

In einem ersten Verfahren erwies sich eine regularisierte multinomiale logistische Regression mit einem F1-Wert von 0,56 auf dem Testdatensatz als der für die vorliegenden Zwecke am besten geeignete Klassifikator (Wulff et al. 2020). Dabei kritisieren Wulff et al. (2020), dass etwa 11 % der Mensch-Mensch-Vergleiche und 40 % der Computer-Mensch-Vergleiche in den Bereichen Rahmenbedingungen, Beschreibung und Bewertung unklar sind. Es wird weiter darauf hingewiesen, dass ein Grund für die Unklarheit in diesen Elementen auf Ungenauigkeiten in den schriftlichen Reflexionen der Studierenden zurückzuführen sein könnte. So stellte beispielsweise Kost (2019) im Rahmen der Analyse der schriftlichen Reflexionen von angehenden Physiklehrkräften auch ohne ML fest, dass die Studierenden Schwierigkeiten hatten, beschreibende und bewertende Texte in den Reflexionen zu trennen. Mit einem Deep-Learning-basierten Ansatz verbesserten Wulff et al. (2022) die Klassifikationsgenauigkeit für die Kodierung der Diskurselemente in den schriftlichen Reflexionen deutlich auf einen F1-Score von .81, was als erhebliche Mensch-Maschine-Übereinstimmung angesehen werden kann (Wulff et al., 2022). Daher wird dieses ML-Modell als ein valides Instrument für die spezifische Kodieraufgabe anerkannt, in schriftlichen Reflexionen die Diskurselemente aus Nowak et al. (2019) zu identifizieren.

Der auf dem ML-Modell basierende Klassifikator segmentiert zunächst einen Text in dessen Sätze. Für jeden Satz wird eine Zuordnungswahrscheinlichkeit für jedes Diskurselement anhand des vortrainierten Wortkorpus ermittelt. Das Diskurselement mit der größten Übereinstimmung hat die höchste Wahrscheinlichkeit, dass auch ein menschlicher Kodierer dieses Element wählen würde. Auf diese Weise wird jedem Satz ein Diskurselement mit der höchsten Wahrscheinlichkeit zugeordnet. Die ursprünglichen Informationen werden in einen Datensatz umgewandelt, der nur die zugeordneten Diskurselemente pro Satz enthält. Dieser Datensatz gibt Auskunft darüber, welcher Satz des Reflexionstextes einer Rahmenbedingung, einer Beschreibung, einer Bewertung, einer Alternative oder einer Konsequenz entspricht. Die komplexen Textdaten einer schriftlichen Reflexion, die den Prozess der reflexionsbezogenen Argumentation abbilden, können auf diese Weise modelliert und zu einer Ist-Struktur der verwendeten Reflexionselemente vereinfacht werden. Abbildung 6 gibt einen Überblick über diese Datenreduktion.



**Abb. 6** Schema der satzweisen Klassifizierung einer schriftlichen Reflexion unter Verwendung eines vortrainierten ML-Modells

Mit diesen Datensätzen kann der Anteil der formulierten Rahmenbedingungen, Beschreibungen, Bewertungen, Alternativen und Konsequenzen in jedem Text berechnet werden. Darüber hinaus kann die Gesamtheit der Positionierung der Diskurselemente im Text selbst skaliert werden, was nachfolgend erläutert ist.

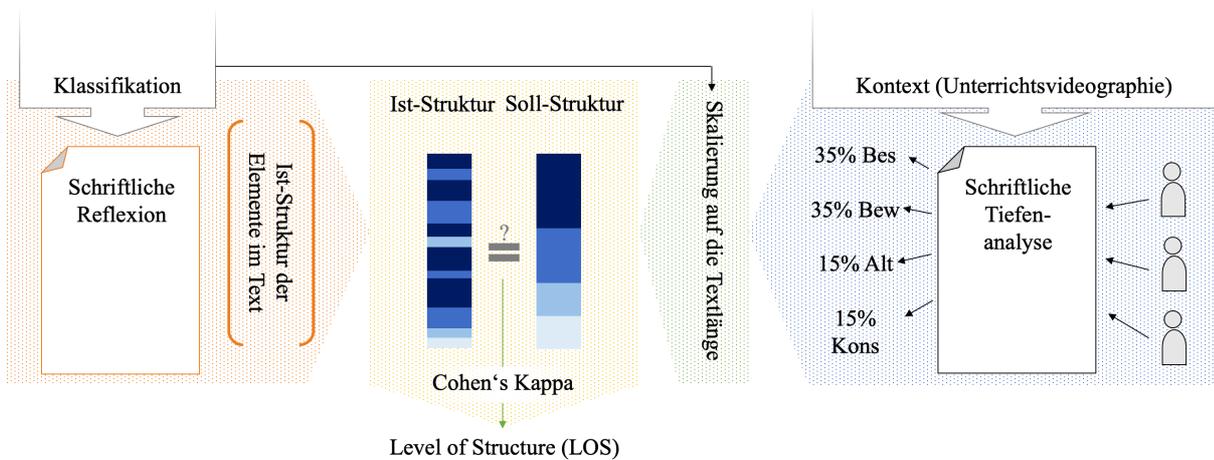
### 2.1.2 Entwicklung eines ML-basierten Qualitätsmaßes für die Struktur von Reflexionstexten

Inwieweit bereits die ML-basierte Positionierung der Diskurselemente des Reflexionsmodells nach Nowak et al. (2019) im Text eine Aussage über die Qualität des zugrundeliegenden Reflexionstextes erlaubt, wurde in einer ersten Modellierung von Mientus et al. (2023) erprobt. Hierfür wurde eine Online-Videovignette implementiert, um angehenden Physiklehrkräften die Möglichkeit zu geben, stellvertretende Erfahrungen zu machen.

Nowak et al. (2019) berücksichtigten in den analysierten Selbstreflexionstexten die formulierten Rahmenbedingungen zwar konzeptuell als Teil der Beschreibungen, jedoch zeigte sich, dass für Selbstreflexionen die Berücksichtigung als eigenständiges Element gewinnbringend sein kann. Entsprechend trainierten auch Wulff et al. (2020) auf die Kodierung aller fünf Elemente hin. Mientus et al. (2023) betrachteten Rahmenbedingungen und Beschreibungen im Sinne der Fremdreﬂexionen erneut zusammen. Zum einen, da die Konzeptualisierung von Nowak et al. (2019) dies erlaubt und zum anderen, da die Anzahl an formulierten Rahmenbedingungen in der Fremdreﬂexion im Rahmen einer Videovignette deutlich weniger umfangreich ausfällt als in Selbstreflexionen.

Um zu untersuchen, welche Diskurselemente zu welchen Anteilen in einer standardisierten Unterrichtsvideographie aufkommen sollten, wurden in Mientus et al. (2023) drei Professor\*innen der Physikdidaktik gewonnen, die eine eingehende Analyse der videographierten Unterrichtssituation vornahmen. Die Professor\*innen erstellten eine schriftliche Analyse, in der Möglichkeiten für pädagogische und fachdidaktische Überlegungen aufgezeigt wurden, welche nach dem Rahmenmodell für Reflexion von Nowak et al. (2019) strukturiert wurden. Auf der Grundlage dieser kollaborativen schriftlichen Reflexion als normativer Verteilung für eine schriftliche Reflexion wurden 35% Beschreibungen, 35% Bewertungen, 15% Alternativen und 15% Konsequenzen gewählt. Auch wenn sich diese Anteile nicht aus der Theorie ableiten lassen, argumentieren Mientus et al. (2023), dass sie eine vernünftige Verteilung darstellen, bei der kein einzelnes Element überproportional vertreten oder ausgelassen ist. Weiter wird davon ausgegangen, dass die Elemente in Blöcken auftreten und dass es in qualitativ hochwertigen Reflexionen keine Vertauschung von Elementen gibt. Ausgehend von der Wortzahl einer schriftlichen Reflexion kann für jeden Text eine Soll-Struktur bestimmt werden, die eine Reflexionsstruktur gemäß der genannten, normativen Verteilung repräsentiert. In einem weiteren Schritt kann diese Soll-Struktur mit der durch das ML-Modell ermittelten Ist-Struktur eines Textes verglichen werden. Dabei wird die Übereinstimmung der Ist-Struktur einer schriftlichen Reflexion mit der Soll-Struktur unter Verwendung von Cohen's Kappa bestimmt. Abbildung 7 stellt dieses Verfahren schematisch dar. Auf diese Weise kann jedem Text ein so genannter *Level of Structure* (LOS) zugewiesen werden, welcher dem Kappakoeffizient entspricht. Ein LOS von 1 würde einer schriftlichen Reflexion entsprechen, die eine Sequenzierung der Diskurselemente gemäß dem Rahmenmodell für Reflexion aufweist und mit der normativen Verteilung der Elemente übereinstimmt. Ein Strukturierungsgrad von 0 bedeutet eine maximale Unordnung der

Diskurselemente im Text; Grad -1 bedeutet eine Sequenzierung der Reflexionselemente in umgekehrter Reihenfolge.



**Abb. 7** Schema zur Berechnung des Level of Structure (LOS)

Zur Validierung dieses Vorgehens wurde eine Stichprobe von  $N = 110$  Reflexionstexten nach dem jeweils ermittelten LOS in Extremgruppen aufgeteilt (weniger, moderat und gelungen strukturierte Texte). Um zu überprüfen, inwieweit Expert\*innen für Beobachtung, Reflexion und Feedback von Physikunterricht die schriftlichen Reflexionstexte in ähnlicher Weise unterscheiden würden, wurden zufällig ausgewählte Reflexionstexte aus den drei Extremgruppen ohne Kennzeichnung der Zugehörigkeit bewertet. Vier Physik-Fachseminarleitende verglichen je eine zufällige Auswahl von 18 Reflexionstexten (neun Texte, drei aus jeder Extremgruppe) in insgesamt 108 Vergleichen und bewerteten die Schwierigkeit der einzelnen Vergleiche. Hierfür wurde (1) die Übereinstimmung der Expert\*innen untereinander und (2) die Übereinstimmung der Expert\*innenbewertungen mit der ML-basierten Bewertung berechnet (Mientus et al., 2023). Zusammenfassend rangiert die Mensch-Mensch-Übereinstimmung und die Mensch-Maschine-Übereinstimmung auf vergleichbarem Niveau, wobei Vergleiche, welche von den Fachseminarleitenden als anspruchsvoller bewertet wurden, auch geringere Übereinstimmungen aufwiesen. Mientus et al. (2023) schlussfolgerten, dass der LOS den Bewertungen der Expert\*innen entspricht und folglich als Qualitätseinschätzung einer schriftlichen Reflexion (zumindest in Extremgruppen) angesehen werden kann. Die Genauigkeit der ML-Modellierung scheint in der Lage zu sein, eine klare Qualitätsaussage insbesondere für weniger strukturierte Reflexionstexte zu treffen. Die ML-basierte Unterscheidung zwischen weniger strukturierten und moderat strukturierten Reflexionstexten kann als Teil einer Qualitätsbeurteilung der schriftlichen Reflexionen verwendet werden, da sie den Bewertungen der Expert\*innen entspricht. Bei der ML-basierten

Analyse der schriftlichen Reflexionen konnten keine Zusammenhänge zwischen der Textlänge und Elementen höherer Ordnung festgestellt werden. Daher argumentierten Mientus et al. (2023), dass im LOS ein Maß gefunden wurde, in dem die Länge als typischer Qualitätsindikator extrahiert wird, so dass andere Indikatoren in den Fokus rücken können. Gemessen an der vorgeschlagenen Bestimmung eines LOS konnte bestätigt werden, dass strukturierte Texte auch mehr höherwertige Reflexionselemente enthalten. Dennoch scheint ein Mindestmaß an formulierten Sätzen für einen erfolgreich strukturierten und höherwertigen Reflexionstext relevant zu sein.

Für eine zusätzliche Validierung des LOS stellten Mientus et al. (2023b) den auf Fremdrelexionen entwickelten und validierten computer-basierten LOS einem auf Selbstreflexionen entwickelt und validierten, qualitativen Instrument gegenüber. Hierzu verfassten  $N = 30$  Teilnehmende je eine Fremdrelexion zu einem standardisierten Unterrichtsvideo mit Demonstrationsexperimenten und eine Selbstreflexion zu einer selbst gehaltenen Experimentiersituation im Unterricht. Beide Instrumente wurden auf alle 60 Reflexionstexte (30 Selbst- und 30 Fremdrelexionen) angewandt, um die jeweilige Qualitätseinschätzung vergleichen zu können. Beide Verfahren zur Beurteilung von Qualität schriftlicher Reflexionen sollten so extern validiert werden. Gemäß den jeweiligen Modellierungen konnten zwar keine signifikanten Korrelationen beider Qualitätsabschätzungen identifiziert werden, beide Instrumente ergaben jedoch in Rangsummentests vergleichbare Platzierungen von Reflexionstexten, sodass geschlussfolgert werden kann, dass beide Testinstrumente die Qualität der Selbst- und Fremdrelexionen vergleichbar abbilden können. Neben den jeweiligen Validierungen der Instrumente selbst stellt die Studie von Mientus et al. (2023b) ein zusätzliches Validitätsargument für den modellierten LOS dar.

Der LOS ermöglicht eine Qualitätseinschätzung in schriftlichen Fremdrelexionen unabhängig von der Textlänge und ist in der Lage Reflexionstexte nach ihrer Qualität vergleichbar mit menschlichen Bewertungen vorzunehmen. Dennoch weisen Mientus et al. (2023) darauf hin, dass nicht nur die normative Verteilung alleinstehend eine Aussage über die Qualität von schriftlichen Reflexionen treffen kann, sondern lediglich eine Näherung darstellen soll. Neben Untersuchungen der Qualität schriftlicher Reflexionen bleibt somit weiterhin offen, in welchen Zusammenhängen Qualitätsmerkmale (ggf. als Indizien für ausgeprägte Reflexivität) und reflexionsbezogenen Dispositionen (Amplifiers & Filters) verknüpft sind. Um hier aussagekräftige Befunde generieren zu können, kann eine Konkretisierung von Rahmenbedingungen hilfreich sein.

## **2.2 Videovignette (Lernkontext)**

Da PCK stets persönlich und themenspezifisch ist (Chan & Hume, 2019), berücksichtigt auch das RCM den Kontext einer Unterrichtssituation. In dessen Rahmen können auch erlernte Kompetenzen (wie Reflexivität) als auch deren Anwendung (in einer Reflexion) im spezifischen Kontextbezug entwickelt oder angewandt werden. Entscheidend beim Kompetenzerwerb ist daher, dass Kompetenzen aktiv unter Berücksichtigung entsprechender Kontextbedingungen entwickelt bzw. gezeigt werden können (Schaper, 2009). Wenngleich eine Einschränkung des Kontextes eine Standardisierung und Vergleichbarkeit identifizierter Effekte erleichtert und dem Einsatz ML-basierter Analyseverfahren dienlich sein kann, kann der hohe Aufwand für die Analyse von Videos Stichprobengrößen einschränken (siehe Kulgemeyer & Riese, 2018). Aus diesem Grund sollen die grundlegenden Analysen aus Mientus et al. (2023) beibehalten werden. Auf Basis der bestehenden Videovignette aus Mientus et al. (2023) soll die Analyse der schriftlichen Fremdreflexionen erfolgen.

Kontext der Videovignette ist die Mechanik (im konkreten der Freie Fall, als Spezialfall der geradlinig beschleunigten Bewegung). Dieser Kontext wurde ausgewählt, da im Bereich der Mechanik sowohl auf die Lehrenden bezogen als auch aus Lernendenperspektive bereits eine Vielzahl physikdidaktischer Forschung verfolgt wurde. Die vorliegende Studie kann somit mit größeren Projekten und Erkenntnissen in Bezug gesetzt werden. Der Videoausschnitt bildet die ersten 17 Minuten einer Einführungsstunde in die Thematik des Freien Falls als Ausgangspunkt für beschleunigte Bewegungen ab. Die Unterrichtssituation ist lehrerzentriert und beinhaltet eine Vielzahl von experimentellen Anwendungen. Zu Beginn der Unterrichtssituation greift die Lehrkraft auf eine vorangegangene Stunde zurück, in welcher ein Wettrennen zwischen Hase und Igel thematisiert wurde. Im Unterrichtsgespräch werden verschiedene Fallversuche durchgeführt. Nach wiederkehrendem Schema fordert die Lehrkraft die Schüler\*innen auf Hypothesen für den Ausgang der Experimente zu nennen. Hierzu stehen der Lehrkraft eine Schraube und eine Feder, verschiedene Maße Stücke sowie eine evakuierte Glasröhre zur Verfügung. Im weiteren Verlauf der Unterrichtsstunde abstrahiert die Lehrkraft im Unterrichtsgespräch die Versuche von der Realsituation des Fallenlassens von Schraube und Feder über die unterschiedlich großen Massstücke bis hin zum Fallexperiment in der Glasröhre, welches einmal evakuiert und einmal mit Luftwiderstand durchgeführt wird. Abschließend zeigt die Lehrkraft eine vorbereitete Definition des Freien Falls am Beamer und bittet die Schüler\*innen um das Abschreiben. Währenddessen kommt es zu Nachfragen einzelner Schüler\*innen und schließlich zum Ende des Videoausschnitts.

Die eigentliche Videovignette verlief in den vier Blöcken Prompt, Video, Reflexionstext und Skalen. Im Sinne der Variablenkontrolle wurden einzelne reflexionsbezogene Dispositionen (in unserem Fall die Wissensbestände zum Aufbau einer konsekutiven Reflexion nach dem Rahmenmodell für Reflexion nach Nowak et al. (2019)) im Prompt der Videovignette eingegrenzt. Weiter wurde die konkrete Aufgabe formuliert, dass im Folgenden Unterricht zu beobachten und im Anschluss ein Reflexionstext zu verfassen sei.

Das Unterrichtsvideo im zweiten Block der Vignette konnten die Teilnehmenden nicht spulen, lediglich pausieren, um einer Authentizität der Unterrichtshospitation möglichst nahe zu kommen. Währenddessen und auch für das Schreiben der Fremdrelexion war der Prompt zum Reflexionsmodell einsehbar. Darüber hinaus wurden die Teilnehmenden motiviert während der Unterrichtsbeobachtung Mitschriften anzufertigen. Für das Verfassen der schriftlichen fremd Reflektionen stand den Teilnehmenden ein unbegrenztes Textfeld zur Verfügung. Auch wenn der Prompt zum Reflexionsmodell weiterhin oberhalb des Textfeldes einsichtig blieb, war es den Teilnehmenden ab diesem Punkt der Vignette nicht mehr möglich, die Unterrichtsvideographie noch einmal zu sehen. Eine Zeitvorgabe für das Verfassen des Reflexionstextes gab es nicht.

Nach Verfassen des Reflexionstextes beurteilten die Teilnehmenden verschiedene Aussagen, um eine unmittelbare Einstellung zu ausgewählten Amplifiers & Filters zu erhalten. Diese waren wie folgt gruppiert: (1) Aussagen zur wahrgenommenen Unterrichtsqualität, (2) Fragen zur Unterrichtsanalyse, welche die erhobenen Dimensionen der reflexionsbezogenen Selbstwirksamkeit sowie der Werthaltung von Reflexion beinhalteten und (3) Fachwissensitems, welche den Teilnehmenden als Quizfragen kommuniziert wurden. Detailliertere Ausführungen zu den Skalen der reflexionsbezogenen Dispositionen / Amplifiers & Filters findet sich im Abschnitt 2.2.2. Zuletzt beantworteten die Teilnehmenden Fragen zu personenbezogenen Daten, welche Aufschluss über den Grad der Professionalisierung geben, bevor abschließend ein Pseudonym erstellt wurde. Für einen genaueren Einblick in die Erhebungsstrukturen, ist im Anhang A1 sämtliches Schriftwerk (Prompt, Aufgaben, Items, etc.) der online Vignette einzusehen.

### *2.2.1 Analyse schriftlicher Reflexionen (Untersuchungsgegenstand 1)*

Als Untersuchungsgegenstand 1 stehen in der vorliegenden Studie die Reflexionstexte der Teilnehmenden zur Verfügung. Unabhängig vom Untersuchungsgegenstand 2, den weiteren erhobenen Skalen, wird die Textanalyse lediglich über das hinterlegte Pseudonym in der vorliegenden Arbeit mit dem Untersuchungsgegenstand 2 in Verbindung gebracht. Die Analyse

der Texte beruht im Wesentlichen auf zwei Säulen: (1) der computerbasierten Analyse mittels des vor trainierten ML Modells nach Wulff et al. (2022) sowie der modellierten Strukturanalyse nach Mientus et al. (2023) und (2) Einschätzungen von Qualitätsindikatoren der Reflexionstexte durch zwei unabhängige Experten. Als Qualitätsindikatoren wurden die in der Reflexionsforschung viel verwendeten Dimensionen *Breite* und *Tiefe* ausgewählt. Unter der Breite eines Reflexionstextes ist in der vorliegenden Arbeit die qualitative Einschätzung der Anzahl diskutierter Reflexionsauslöser zu verstehen, wobei unter der Tiefe die Intensität der jeweiligen Diskussion im subjektiven Mittel über den gesamten Text verstanden wird. Da es sich bei schriftlichen Reflexionen wie eingangs beschrieben nicht um reflexionsbezogene Argumentationsprozesse, sondern lediglich deren Abbilder handelt, wurden geläufige Dimensionen der qualitativen Textanalyse, wie *Kohärenz* und *Spezifität*, als Qualitätsindikatoren angefügt. Unter Kohärenz wird im vorliegenden Kontext der Zusammenhang des verschriftlichten reflexionsbezogenen Argumentationsprozesses verstanden, wobei die Spezifität eines Textes dessen Bezugnahme auf den Gegenstand (also die Unterrichtsvideographie) repräsentiert.

Die schriftlichen Reflektionen der Teilnehmenden, welche als qualitative Daten angesehen werden können, fließen somit nicht direkt in die vorliegende Studie ein, sondern werden über zwei Verfahren analysiert und als quantitative Daten berücksichtigt. Diese in Teilen computerbasiert-erhobenen quantitativen Daten beinhalten Aussagen über die Verwendung von Elementen nach dem Reflexionsmodell von Nowak et al. (2019) sowie dessen Positionen im Text. Die Einschätzungen bezüglich der Qualitätsindikatoren werden durch Expertenbewertungen abgebildet und im Folgenden methodisch diskutiert.

#### 2.2.1.1 ML-basierte Textanalyse nach dem Rahmenmodell für Reflexion

Wie unter 1.2.1 bereits genannt, ist der Klassifikator von Wulff et al. (2022) in der Lage jeden Reflexionstext in die einzelnen Sätze des Textes zu segmentieren und den Text satzweise den Diskurselementen nach Nowak et al. (2019) zuzuordnen. Auf diese Weise ergibt sich für jede schriftliche Fremdrelexion ein Vektor der Länge  $n$ , wobei  $n$  der Anzahl an Sätzen des jeweiligen Reflexionstextes entspricht. Dieser  $n$ -dimensionale Vektor wird *Ist-Struktur* genannt und wird im Weiteren als das erste Textmerkmal weiterverarbeitet. Aus dieser *Ist-Struktur* konnten mit der Analysesoftware R-Studio durch Aufsummierung die folgenden Textmerkmale errechnet werden:

- Anzahl der *Sätze* im Text,
- Anzahl der Sätze, welche *Beschreibungen* zugeordnet werden konnten,

- Anzahl der Sätze, welche *Bewertungen* zugeordnet werden konnten,
- Anzahl der Sätze, welche *Alternativen* zugeordnet werden konnten,
- Anzahl der Sätze, welche *Konsequenzen* zugeordnet werden konnten.

Mit diesen Informationen lassen sich sowohl absolute Häufigkeiten (im Weiteren jeweils mit *Beschreibungen*, *Bewertung*, *Alternativen* oder *Konsequenzen* benannt) als auch die relativen Häufigkeiten (im Weiteren jeweils mit *Beschreibungen\_%*, *Bewertung\_%*, *Alternativen\_%* oder *Konsequenzen\_%* benannt) bestimmen. Während die absoluten Häufigkeiten ganzzahlige Ordinale annehmen, werden die relativen Häufigkeiten durch Werte zwischen 0 und 1 bzw. 0% und 100% angegeben. Gemäß den ganzzahligen absoluten Häufigkeiten, werden die relativen Häufigkeiten in dieser Arbeit mit bis zu zwei zuverlässigen Ziffern angegeben.

#### 2.2.1.2 ML-basierte Modellierung von *Qualität schriftlicher Fremdrelexionen*

Im Rahmen der Modellierung des LOS (Kap. 2.1.2) wurde neben der Ist-Struktur beschrieben, wie jedem Text eine *Soll-Struktur* zugeordnet werden kann, welche aus der prozentualen Verteilung der normativen Reflexion (35% Beschreibung, 35% Bewertung, 15% Alternativen, 15% Konsequenzen) errechnet wurde. Hierbei werden die folgenden Größen repräsentiert:

- *Textlänge* als Anzahl der Worte in einem Text,
- *Level of Structure (LOS)* aus Mientus et al. (2023) als Repräsentant der Textstruktur gemäß der normativen Verteilung in Anlehnung an das Rahmenmodell für Reflexion von Nowak et al. (2019).

Wenngleich die Variable der Textlänge lediglich Werte aus den Natürlichen Zahlen annehmen kann, kann der LOS jegliche Dezimalwerte zwischen -1 und 1 annehmen. Da auch für den LOS die satzweisen Klassifikationen nach Wulff et al. (2022) grundlegend sind und ebenso satzweise ein Abgleich zwischen Ist-Struktur und Soll-Struktur erfolgt, sollte auch der LOS mit nur bis zu zwei zuverlässigen Ziffern angegeben werden. Mathematisch verbirgt sich hinter dem LOS Cohens‘ Kappa und somit eine Division aus absoluten Häufigkeiten an (nicht-) Übereinstimmungen. Mathematisch basiert Cohens Kappa auf der Konzeptualisierung eines sogenannten beobachteten Übereinstimmungsindex und eines erwarteten Übereinstimmungsindex. Der beobachtete Übereinstimmungsindex gibt an, wie oft die Beobachter tatsächlich übereinstimmen, während der erwartete Übereinstimmungsindex angibt, wie oft sie übereinstimmen würden, wenn ihre Entscheidungen zufällig wären. Der Cohens Kappa-Koeffizient wird dann berechnet, indem der beobachtete Übereinstimmungsindex um den erwarteten Übereinstimmungsindex korrigiert wird, um die zufällige Übereinstimmung zu berücksichtigen. Die Werte von Cohens Kappa liegen zwischen

-1 und 1. Ein Wert von 1 zeigt eine perfekte Übereinstimmung zwischen den Beobachtern an, während ein Wert von -1 auf eine vollständige Diskrepanz hindeutet. Ein Wert nahe bei 0 deutet darauf hin, dass die Beobachter nicht besser abschneiden als bei rein zufälligen Entscheidungen.

### *2.2.1.3 Qualitätsdimensionen schriftlicher Reflexionen*

Um der Forschungsfrage 2 nachkommen zu können, explorativ ein Qualitätskorrelat auf Basis der computerbasierten Textmerkmale zu entwickeln, werden nach Forschungsfrage 1 (Exploration von Hintergrundfaktoren) die computerbasierten Textmerkmale Qualitätsbewertungen von Expert\*innen gegenübergestellt.

Zu Beurteilung der Qualitätsindikatoren der schriftlichen Reflexionen werden von zwei unabhängigen Gutachtenden die Ausprägung des jeweiligen Textes in Breite, Tiefe, Kohärenz und Spezifität eingeschätzt, wobei jeweils vierstufige Bewertungen vorgenommen werden („Nicht ausgeprägt“ = 1; „Gering ausgeprägt“ = 2; „Eher ausgeprägt“ = 3; „Sehr ausgeprägt“ = 4). Die vereinbarten Definitionen sind angelehnt an die theoretischen Überlegungen aus 1.2.2 und sind im Anhang A2 aufgelistet. Bei den Experten handelte es sich um zwei Experten für Analyse und Feedback zu fachdidaktischen schriftlichen Fremd- und Selbstreflexionen im Rahmen des Praxissemesters Physik. Ohne weitere Abstimmung konnte auf Basis der vereinbarten Definitionen eine substantielle Übereinstimmung (Breite: .77<sup>\*\*\*</sup>, Tiefe: .67<sup>\*\*\*</sup>, Kohärenz: .71<sup>\*\*\*</sup>, Spezifität: .74<sup>\*\*\*</sup>) erreicht werden.

### *2.2.2 Skalen der reflexionsbezogenen Amplifiers & Filters (Untersuchungsgegenstand 2)*

Zur Beantwortung der Forschungsfrage 3 nach den Zusammenhängen von Qualitätsbewertungen in schriftlichen Fremdrelexionen und reflexionsbezogenen Amplifiers & Filters werden für den Untersuchungsgegenstand 2 zunächst geeignete Skalen für die Erhebung der ausgewählten Amplifiers & Filters diskutiert. Ausgewählte Skalen der Selbstwirksamkeitserwartung, der wahrgenommenen Unterrichtsqualität, des Fachwissens sowie der Werthaltung der Teilnehmenden, nach der in der Videovignette eingeforderten Weise eine Fremdrelexion durchzuführen, werden im Fragebogenformat mit Multiple Choice Items erhoben. Wie bereits unter 1.2.2, der theoriegeleiteten Herleitung der Qualitätsmerkmale schriftlicher Reflexionen, wird der Lernkontext als ebenfalls einflussreicher Amplifier & Filter mit dem Vignettenvideo selbst standardisiert und unter Verwendung der Skala der wahrgenommenen Unterrichtsqualität berücksichtigt.

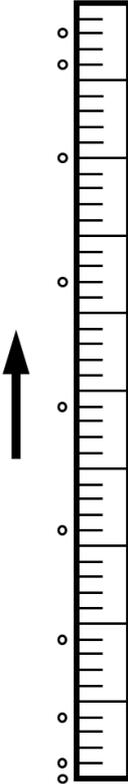
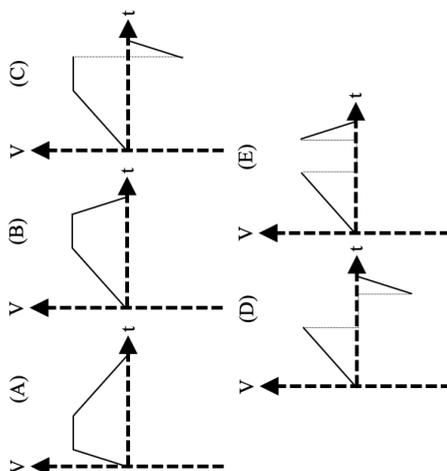
Analog zum Untersuchungsgegenstand 1, den Reflexionstexten, werden zunächst die hier aufgeführten Skalen gesondert betrachtet, bevor eine Verbindung zu den Qualitätsbewertungen hergestellt wird. Der Untersuchungsgegenstand 2 setzt sich aus (1) 12 Items zur reflexionsbezogenen Selbstwirksamkeit (aus Lohse-Bossenz et al., 2018), (2) 60 Items zur wahrgenommenen Unterrichtsqualität (aus Schlesinger et al., 2018), (3) 8 Fachwissensitems (6 Items aus dem Mechanics Baseline Test (MBT) aus Cardamone (2011) sowie 2 Eigenentwicklungen) und (4) 4 Items zur Werthaltung von Reflexion (Eigenentwicklung) zusammen. Die Eigenentwicklungen der Items der Fachwissensskala beziehen sich explizit auf den Lernkontext des Freien Falls und wurden ergänzt, da der MBT als Testinstrument nicht für einen spezifischen Kontext konstruiert und somit nicht für den spezifischen Kontext validiert ist. Die Berechnung des skalenbezogenen Cronbachs'  $\alpha$  wird abschließend Aufschluss über die interne Konsistenz der Fachwissensskala geben und stellt somit dar, inwieweit die 8 Items der Fachwissensskala dasselbe Konstrukt zu messen scheinen. Ebenso wird mit den vier Items der Wert-Skala verfasst, für welche keine geeignete Literatur gefunden werden konnte. Auch die Skala der wahrgenommenen Unterrichtsqualität soll einer zusätzlichen Prüfung auf interne Konsistenz unterzogen werden, da zum einen alle Items aus dem englischen ins Deutsche übersetzt werden mussten und zum anderen nicht alle Subskalen oder Items übernommen wurden, da es sich im Original um die Analyse von ganzen Mathematikunterrichtsstunden handelt. Im Rahmen unseres physikspezifischen Videoausschnittes einer authentischen Unterrichtssituation erscheinen Itemgruppen zu *Effective use of lesson time* oder *Self-directed learning* ungeeignet.

Tabelle 1 gibt einen Einblick in die verwendeten Skalen. Für jede Skala sowie bestehender Subskalen ist hier eine ausgewählte Abkürzung, die Itemanzahl, ein Beispielitem, zugehörige Antwortmöglichkeiten sowie Antwortkodierungen und dem statistischen Mittelwert der Skala repräsentiert. Die Skala der reflexionsbezogenen Selbstwirksamkeit (rSWK) bietet Subskalen, welche sich auf die Beschreibung (rSWK\_Beschreiben), die Bewertung (rSWK\_Bewerten), das Formulieren von Alternativen (rSWK\_Alternativen), Konsequenzen (rSWK\_Konsequenz) sowie Begründungen (rSWK\_Begründung) bezieht. Die wahrgenommene Unterrichtsqualität (wUQ) bezieht sich auf Subskalen zum Classroom Management (wUQ\_Classroom Management), zur Unterstützung von Schüler\*innen durch die Lehrkraft (wUQ\_Student Support), zur kognitiven Aktivierung (wUQ\_Cognitive Activation), zur inhaltlich-fachlichen Unterrichtsqualität (wUQ\_subject-related Quality), sowie zur Unterrichtsqualität, welche sich auf die Performanz der Lehrkraft im Unterrichtsausschnitt bezieht (wUQ\_teaching-related Quality).

Tab. 1 Skalen, Itemanzahl, Beispielimitem und Antworten

<i>Skala</i>	<i>#Item</i>	<i>Beispielimitem</i>	<i>Antwort (Kodierung)</i>	<i>Stat. <math>\sigma</math></i>
rSWK	12			
rSWK_Beschreiben	3	Obwohl die Situation schon vergangen ist, kann ich mich an die Handlungen gut erinnern.		
rSWK_Bewerten	3	Ich konnte gut einschätzen, inwiefern die Handlungen geeignet waren, die Unterrichtssituation positiv zu beeinflussen.		
rSWK_Alternativen	2	Mir ist es gut gelungen, Alternativen für die Gestaltung der Unterrichtssituation zu entwickeln.	Trifft gar nicht zu (1); ...; Trifft voll zu (5)	2.50
rSWK_Konsequenz	2	Obwohl es viele Möglichkeiten gab, seine Handlungen in der Unterrichtssituation zu ändern, gelang es mir gut, mich auf ein spezifisches Veränderungsziel festzulegen.		
rSWK_Begründung	2	Trotz der Komplexität der Unterrichtssituation, konnte ich einzelne Gründe für die Handlungen identifizieren.		
wUQ	60			
wUQ_Classroom Management	13	Die Unterrichtssituation besaß einen „roten Faden“ .		
wUQ_Student Support	16	Es gab anforderungsbezogene, differenzierende Aufgabenstellungen.	Trifft gar nicht zu (1); ...; Trifft voll zu	
wUQ_Cognitive Activation	15	Der Lehrer fragte nach den Vorstellungen der Lernenden zum Thema.	(5); Kann ich für die beobachtete Unterrichtssituation nicht beantworten	2.50
wUQ_subject-related Quality	12	Der Lehrer beschränkte sich auf das Wesentliche.		
wUQ_teaching-related Quality	4	Schülerinnen und Schüler konnten eigene Erfahrungen und Interessen einbringen.		

Tab. 1 Skalen, Itemanzahl, Beispielimem und Antworten

Skala	#Item	Beispielimem	Antwort (Kodierung)	Stat. $\theta$
Fachwissen	8	<p>Beziehen Sie sich auf folgende Abbildung, wenn Sie die ersten beiden Fragen beantworten. Das Diagramm stellt eine Stroboskop-Aufnahme eines Körpers dar, welcher sich entlang einer horizontalen Ebene bewegt. Die Positionen, wie sie im Diagramm eingezeichnet sind, sind durch jeweils gleichgroße Zeitabschnitte voneinander getrennt. Die erste Aufnahme ist zu dem Zeitpunkt gemacht, an dem der Körper mit seiner Bewegung beginnt. Die letzte Aufnahme zeigt den Körper, als er gerade wieder zur Ruhe gekommen ist.</p>  <p>1) Welcher der folgenden Graphen stellt am besten die Geschwindigkeit des Körpers als Funktion der Zeit dar?</p>	 <p>„Weiß ich nicht“ (Richtig (1); Falsch (0); NA)</p>	4
Wert	4	Ich denke, dass es sinnvoll ist, so wie eben gesehen über Unterrichtssituationen nachzudenken.	Trifft gar nicht zu (1); ...; Trifft voll zu (5)	2.50

### 3 Ergebnisse

Der in Kapitel 2 ausgeführten Datenaufbereitung lagen als Rohdaten insgesamt 17 Erhebungen mit dem Vignetteninstrument zugrunde, welche in einem ersten Schritt zu einem Gesamtdatensatz additiv verbunden wurden. Von den  $N = 161$  Teilnehmenden konnte  $N = 13$  kein Pseudonym zugeordnet werden.  $N = 6$  Freitextantworten beinhalteten keinen eindeutig identifizierbaren Reflexionstext (keine Bearbeitung, „test“, „Text“ oder Stichpunkte). Auf Grundlage der Pseudonyme sowie der Zugriffszeiten mussten  $N = 21$  Datensätze entfernt werden, da es sich um eine Doppelteilnahme handelte, sodass eine Gesamtstichprobengröße von  $N = 134$  für diese Arbeit vorliegt (nur Erstteilnahmen wurden berücksichtigt). Obwohl lediglich sechs Teilnehmende offensichtlich die Befragung abbrachen oder beispielsweise zehn Personen nicht ihre Abiturnote angaben (kein Item war ein Pflichtfeld), sind in den Analysen alle 134 Teilnehmende berücksichtigt, da je ein Reflexionstext vorliegt.

Wenngleich in Bezug auf Qualitätsaspekte von Reflexionsprozessen in der Reflexionsforschung Aussagen generiert werden konnten, welche sich zu großen Teilen auf Professionalisierungsprozesse fokussieren, fällt bei Zusammenhangsanalysen auf, dass schriftliche Unterrichtsanalysen und -reflexionen überwiegend deskriptiv sind. Erweisen sich auch die Texte der vorliegenden Studie als überwiegend deskriptiv, wird angenommen, dass sich die Stichprobe in die empirische Studienlage einreihen kann. Ebenso soll geprüft werden, inwieweit die ausgewählten Methodiken der Analyse schriftlicher Fremdre reflexionen auf die vorliegende Studie übertragbar sind. Mientus et al. (2023) modellierten mit ihrem *Level of Structure* (LOS) einen Qualitätsindikator, welcher auf die Qualität der Reflexionstexte hindeutet und hierbei das typische Qualitätskorrelat der Textlänge vernachlässigen kann. LOS repräsentiert den Anteil höherwertiger Diskurselemente nach dem Rahmenmodell für Reflexion von Nowak et al. (2019) und kann als computerbasierter Indikator für gelungene Reflexionen angesehen werden. Es soll daher geprüft werden, inwieweit sich der LOS als valide erweist. Da es sich bei dieser Stichprobe um denselben Erhebungskontext handelt (z. B. in Stichprobe, Vignettenvideo) und sich die Stichprobe in Teilen mit der Stichprobe aus Mientus et al. (2023) bzw. Mientus et al. (2023b) überschneidet, kann erwartet werden, dass sich die Zusammenhänge zwischen Qualitätsindikatoren und den weiteren Textmerkmalen reproduzierbar zeigen.

#### 3.1 Beschreibung der Stichprobe

Von den  $N = 134$  Teilnehmenden ordneten sich  $N = 35$  einem weiblichen,  $N = 90$  einem männlichen und  $N = 1$  weiteren Geschlechtern zu. Das Alter der Teilnehmenden reichte von 19

bis 49 ( $\sigma = 26$ ,  $sd = 5$ ).  $N = 48$  Bachelorstudierende,  $N = 60$  Masterstudierende,  $N = 11$  Referendar\*innen und  $N = 8$  Noviz\*innen (Lehrkräfte in den ersten beiden Jahren nach dem zweiten Staatsexamen) konnten für eine Teilnahme gewonnen werden. Der Abiturdurchschnitt reichte von 1.0 bis 3.4 ( $\sigma = 2.0$ ,  $sd = 0.6$ ), wobei die einzelnen Notenpunkte der für diese Erhebung relevantesten Schulfächer sich als diverser darstellte. Die Physiknoten reichten von 7 bis 15 Notenpunkten ( $\sigma = 12$ ,  $sd = 1$ ) und stellten sich somit als besser dar, als die Mathematiknoten 2 bis 15 Notenpunkte ( $\sigma = 10$ ,  $sd = 3$ ) oder die Notenpunkte in Deutsch, welche zwischen 5 und 15 ( $\sigma = 10$ ,  $sd = 2$ ) lagen. Insgesamt nahmen  $N = 61$  Personen aus Potsdam, vier aus Cottbus, zehn weitere aus dem Bundesland Brandenburg,  $N = 33$  Personen aus Gießen, je acht aus Kiel und Heidelberg, zwei aus Berlin, eine Person aus Aachen sowie eine Person sonstiger Angabe teil.

Die vorliegende Stichprobe stellt eine repräsentative Auswahl an (angehenden) Physiklehrkräften aus mehreren deutschen Bundesländern dar, welche zwar alle Phasen der Lehrkräftebildung bedient, gemäß der Rahmenbedingungen jedoch einerseits zugunsten der ersten Phase und andererseits Brandenburg und die zugehörigen Standorte der Lehrkräftebildung gewichtet ist. Besonders die fachspezifische Verteilung der Geschlechter von ca. 28% zu 71% deutet auf eine repräsentative Stichprobe hin (Faulstich-Wieland, & Nyssen, 1998; Klika, 2007). Der Effekt einer Positivauswahl an Teilnehmenden kann ausgeschlossen werden, da die meisten Daten im Rahmen von Seminaren oder Workshops eingesetzt wurden, um im Sinne einer reflexionsbezogenen Lerngelegenheit eine Seminareinheit zu gestalten.

Im Weiteren soll zunächst rein deskriptiv auf den Untersuchungsgegenstand 1 (die Reflexionstexte) fokussiert werden, bevor vertiefende Analysen aufgezeigt werden und eine Bezugnahme auf den Untersuchungsgegenstand 2 (die Skalen der Amplifiers & Filters) erfolgen kann.

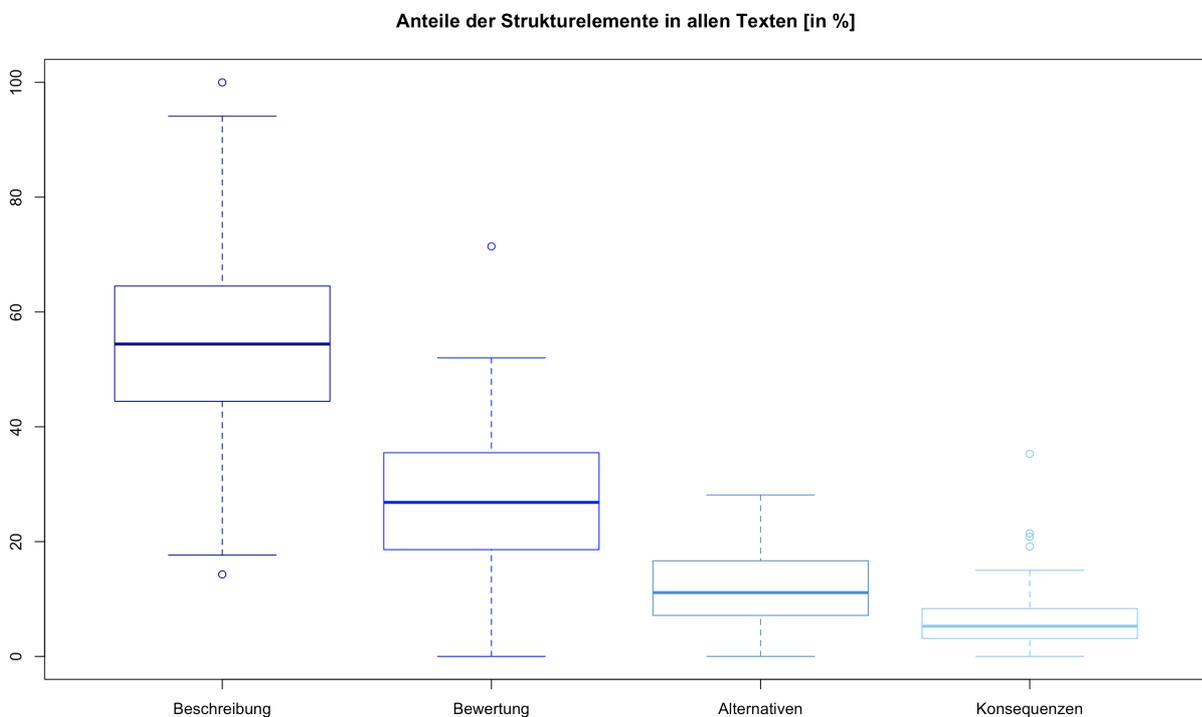
### **3.2 Schriftliche Fremdreflexionen (Untersuchungsgegenstand 1)**

Zur Beantwortung der explorativen Forschungsfragen 1 und 2 werden im Folgenden die Ergebnisse der Qualitätsbewertungen als auch der computerbasiert analysierten Textmerkmale vorgestellt und im weiteren Verlauf dieses Kapitels bis hin zu einem Qualitätskorrelat modelliert. Die Analyse der schriftlichen Fremdreflexionen erfolgt, wie im Methodenteil dieser Arbeit ausgeführt, auf Basis des Klassifikationsalgorithmus von Wulff et al. (2022) sowie der mit dem Level of Structure entworfenen Metrik für die Abschätzung von Qualitätskorrelaten nach Mientus et al. (2023), welche zumindest in Extremgruppen in der Lage ist, Qualität

abzubilden. Gegenübergestellt werden im weiteren Verlauf dieses Abschnittes die Qualitätsdimensionen Breite, Tiefe, Kohärenz und Spezifität.

### 3.2.1 Deskriptive Textmaße

Die Länge der Reflexionstexte beträgt im Mittel 721 Worte bei einer Standardabweichung von 320 Worten. Der kürzeste Text ist 102 Worte lang, wobei die längste schriftliche Reflexion mit 1755 Worten um das 17-fache länger ist. Im Mittel sind die Reflexionstexte überwiegend deskriptiv. Die Textlänge der schriftlichen Fremdrelexionen in Worten und Sätzen (welche nach den Diskurselementen kodiert werden) korrelieren nach einer Pearsonschen Korrelation mit  $\rho = .938$  hoch signifikant ( $p < .001$ ). Mit einem Anteil von 55% ( $sd = 16\%$ ) ist der Anteil an formulierten Beschreibungen signifikant<sup>13</sup> größer ( $p < .001$ ) als alle übrigen Diskurselemente. Bewertungen, Alternativen und Konsequenzen wurden im Mittel zu 27% ( $sd = 12\%$ ), 12% ( $sd = 6\%$ ) beziehungsweise 6% ( $sd = 5\%$ ) formuliert. Genau einem Text wurden ausschließlich Beschreibungen zugeordnet, wengleich auch in weiteren Texten keine Alternativen (fünf Texte insgesamt) oder Konsequenzen (insgesamt 16 Texte) identifiziert werden konnten. Abbildung 8 illustriert die Verteilung der Diskurselemente.



**Abb. 8** Relativer Anteil der Diskurselemente über alle Texte

<sup>13</sup> Zweiseitige t-Tests

Die mittleren Unterschiede zur normativen Verteilung aus der Verteilungsvorgabe nach Mientus et al. (2023) (35% Beschreibungen, 35% Bewertungen, 15% Alternativen, 15% Konsequenzen) betragen +20% (Beschreibungen), -8% (Bewertungen), -3 % (Alternativen) und -9% (Konsequenzen). Alle Unterschiede stellen sich ebenfalls als signifikant dar<sup>14</sup>. Die Teilnehmenden stimmen also am meisten mit dem Anteil der Alternativen überein, die mit unserer gemeinsamen schriftlichen Reflexion formuliert wurden.

Neben der Stichprobe der Teilnehmenden stellen sich somit die Texte auf ähnliche Weise deskriptiv dar, wie die Texte der Stichprobe von Mientus et al. (2023), unter welcher der Level of Structure als strukturbezogenes Qualitätskorrelat modelliert und getestet wurde. Um erneut belastbare Aussagen treffen zu können, sollen im Weiteren die Aussagen zum Level of Structure aus Mientus et al. (2023) für die erweiterte Stichprobe überprüft werden.

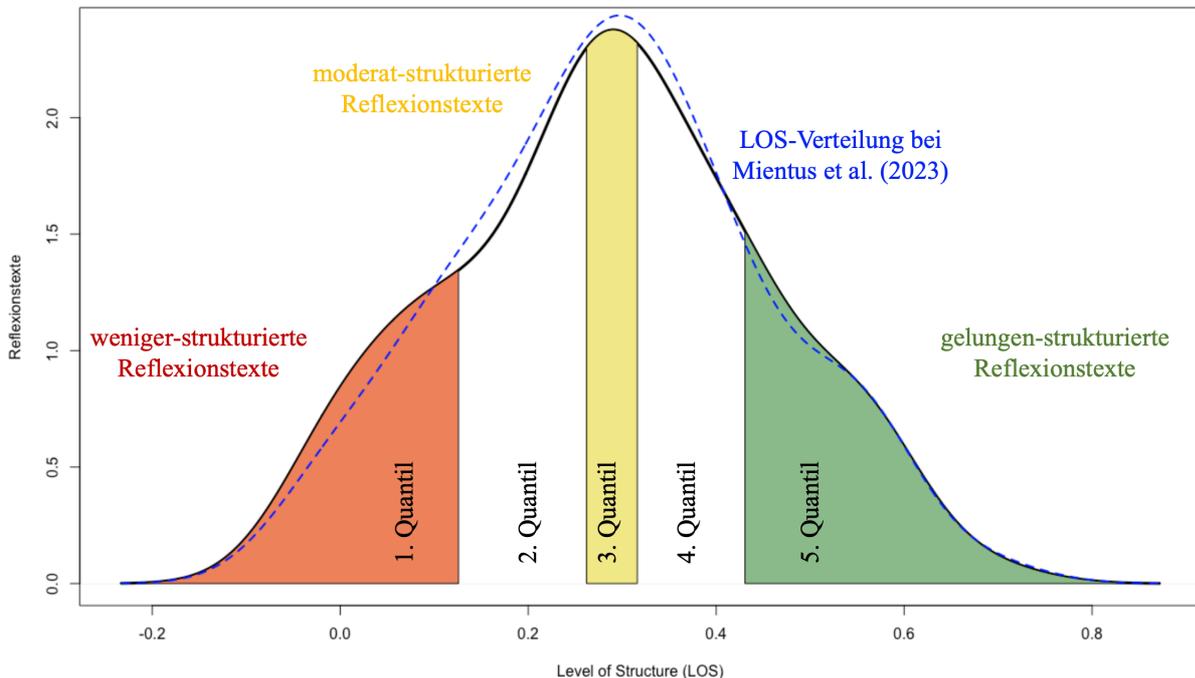
### 3.2.2 Strukturdimension: LOS als strukturbezogenes Textmaß

In Anlehnung an Mientus et al. (2023) wird der LOS zunächst deskriptiv diskutiert. Der ermittelte LOS der vorliegenden  $N = 134$  Reflexionstexte beträgt im Durchschnitt  $.285$  ( $sd = .168$ ), wobei die Werte von  $-.064$  bis  $.703$  reichen, und stellt nur bedingt eine Normalverteilung dar ( $W = .99$ ,  $p = .293$ ). Im direkten Vergleich zu Mientus et al. (2023) stellen sich somit alle Werte als nicht normalverteilt dar<sup>15</sup>. Besonders deutlich wird dies in Abbildung 9, in welcher die Verteilung des LOS über die Texte graphisch vergleichbar sind. Der LOS von fünf Texten ist kleiner als Null. Die Intervallgrenzen der LOS-Extremgruppen verhalten sich gemäß der größeren Standardabweichung erwartbar weiter als bei Mientus et al. (2023) und liegen bei  $[-.064 | .127]$  für das erste Quantil,  $[.259 | .317]$  für das dritte Quantil und  $[.427 | .703]$  für das fünfte Quantil. Wenngleich sich die Extremgruppen gemäß des LOS in ihrer Struktur unterscheiden, sind keine signifikanten Unterschiede in den Textlängen identifizierbar (siehe Abbildung 10), was den LOS als längenunabhängiges Textmaß (vgl. Mientus et al., 2023) unterstreicht.

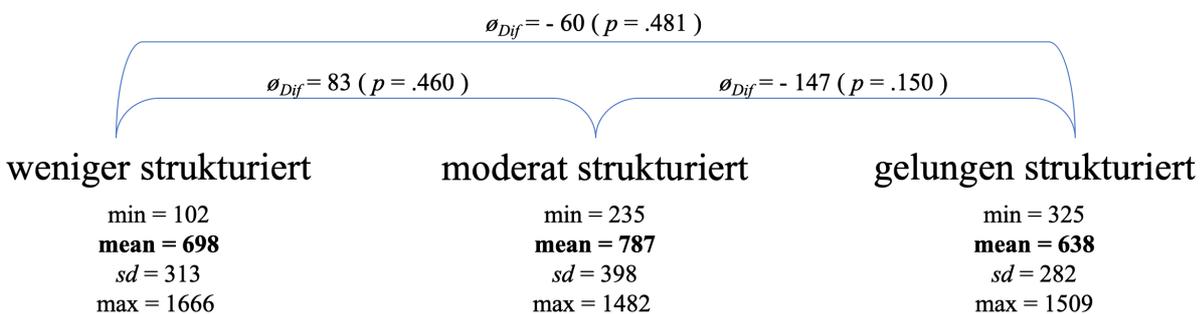
---

<sup>14</sup> Zu beachten ist, dass für alle deskriptiven Textmaße (Länge sowie absolute und relative Häufigkeit der Diskurselemente) von einer Normalverteilung ausgegangen werden kann (siehe Anhang A3) und somit zweiseitige t-Tests angewandt werden dürfen.

<sup>15</sup> Mientus et al. (2023) berichten einen LOS im Bereich von  $-.064$  bis  $.703$  bei einem Mittelwert von  $.289$  und einer Standardabweichung von  $0.161$  bei Shapiro-Wilk-Testergebnissen von  $W = .991$  und  $p = .716$ .



**Abb. 9** Verteilung der LOS über alle 134 schriftlichen Reflexionen und die drei Qualitätsgruppen (rot: weniger strukturiert, gelb: moderat strukturiert, grün: gelungen strukturiert) in Gegenüberstellung der LOS-Verteilung aus Mientus et al. (2023)



**Abb. 10** Mittlere Unterschiede der Textlänge der Extremgruppen

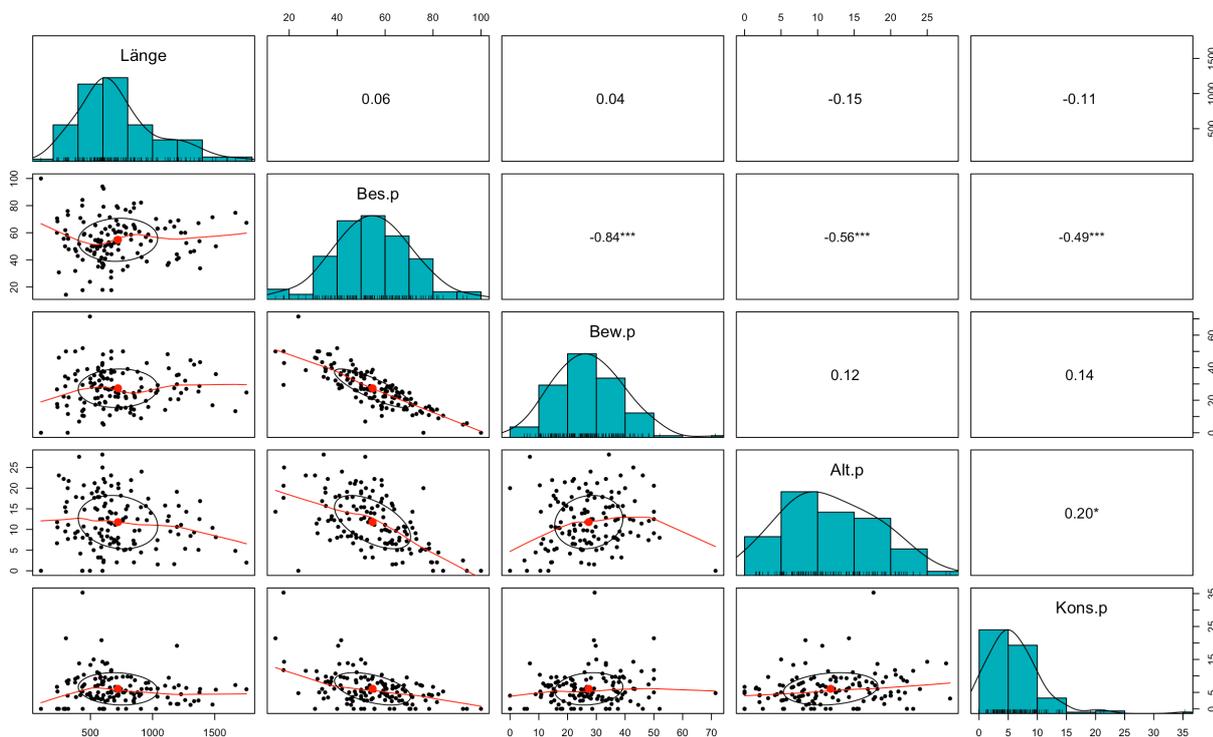
Wenngleich die Vergleichbarkeit der drei Extremgruppen unter anderen Aspekten möglich erscheint, da die drei Extremgruppen von der Textlänge unabhängig sind, gibt Abbildung 11 eine Vorstellung über den LOS indem die drei am wenigsten, am ehesten moderat und am gelungensten Reflexionstexte graphisch verdeutlicht sind.

ID LOS Sent	Weniger strukturierte Texte			Moderat strukturierte Texte			Gelungen strukturierte Texte		
	KARAH07 -.0643 37	ZZZZZ07 -.0377 22	INRUBE10 -.0372 49	DAMIBE04 .2794 77	UTPEPA20 .2802 62	CHMIKI05 .2822 54	DOHALU08 .6070 36	JETHWE28 .6260 23	HEDIRO19 .7030 24
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung		Bewertung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung		Beschreibung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Beschreibung		Beschreibung
Bewertung	Bewertung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Alternativen	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung		Beschreibung
Bewertung	Bewertung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung		Beschreibung
Bewertung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung		Beschreibung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung		Beschreibung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Beschreibung	Beschreibung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung		Bewertung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Beschreibung	Bewertung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung		Bewertung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Bewertung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung		Bewertung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Bewertung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung		Bewertung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Bewertung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung		Bewertung
Bewertung		Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Bewertung	Bewertung	Bewertung
Alternativen	Alternativen	Beschreibung	Beschreibung	Konsequenzen	Bewertung	Bewertung	Bewertung	Bewertung	Bewertung
Beschreibung	Bewertung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Bewertung	Beschreibung	Bewertung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Bewertung	Beschreibung	Alternativen	Alternativen	Bewertung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Konsequenzen	Alternativen	Bewertung
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Alternativen	Alternativen	Bewertung
Beschreibung	Beschreibung	Bewertung	Beschreibung	Beschreibung	Alternativen	Beschreibung	Bewertung	Alternativen	Alternativen
Beschreibung		Bewertung	Beschreibung	Beschreibung	Konsequenzen	Beschreibung	Beschreibung	Beschreibung	Alternativen
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Bewertung	Alternativen	Alternativen
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Alternativen	Bewertung	Bewertung
Beschreibung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Alternativen	Alternativen	Alternativen	Konsequenzen	Alternativen
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Bewertung	Bewertung	Alternativen	Alternativen	Alternativen
Beschreibung	Beschreibung	Bewertung	Beschreibung	Beschreibung	Bewertung	Alternativen	Alternativen	Alternativen	Konsequenzen
Beschreibung		Beschreibung	Beschreibung	Beschreibung	Bewertung	Alternativen	Bewertung	Alternativen	Konsequenzen
Bewertung	Bewertung	Beschreibung	Beschreibung	Beschreibung	Beschreibung	Bewertung	Konsequenzen	Konsequenzen	Konsequenzen
Bewertung		Beschreibung	Beschreibung	Beschreibung	Bewertung	Konsequenzen	Konsequenzen		Konsequenzen
Alternativen	Alternativen	Bewertung	Beschreibung	Beschreibung	Alternativen	Konsequenzen	Alternativen	Beschreibung	Bewertung
		Konsequenzen	Beschreibung	Beschreibung	Bewertung	Beschreibung			

Abb. 11 Exemplarische Visualisierung der Textstruktur

Nach der deskriptiven Prüfung, nach welcher die vorliegende Stichprobe die Stichprobe der Studie von Mientus et al. (2023) nur in Teilen widerspiegelt, sollen die Kernergebnisse von Mientus et al. (2023) geprüft werden, um zu bestimmen, inwieweit Aussagen zu Qualitätskorrelaten unter Verwendung des LOS dennoch möglich sind.

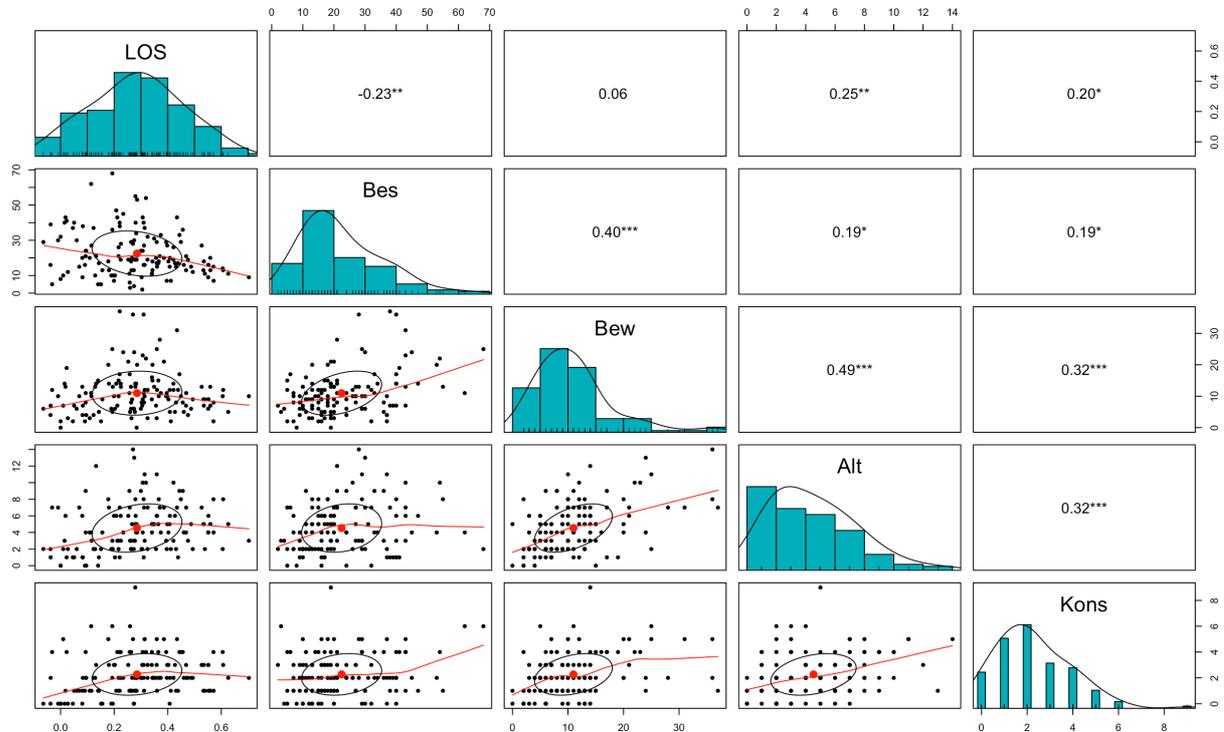
Entsprechend soll der Zusammenhang der Länge einer schriftlichen Reflexion mit der relativen Verteilung der Diskurselemente betrachtet werden, da die absoluten Anteile ohnehin mit der Textlänge zunehmen. Es zeigt sich, dass die Textlänge auch in der erweiterten Stichprobe weniger mit den relativen Anteilen der formulierten Beschreibungen, Bewertungen, Alternativen und Konsequenzen zusammenhängt. Es sind jedoch signifikante, negative Korrelationen zwischen dem Anteil der Beschreibungen und den anderen drei Diskurselementen identifizierbar. Weiter besteht eine signifikant positive Korrelation zwischen dem Anteil der Alternativen und der Konsequenzen. In Texten mit einem höheren Anteil an Alternativen konnten mehr Konsequenzen identifiziert werden und umgekehrt. Abbildung 12 veranschaulicht diese Zusammenhänge in Form eines Korrelationspanels. Die Verteilung der Variablen (Textlänge & Anteil der Beschreibungen, Bewertungen, Alternativen und Konsequenzen) sind als Histogramme auf der Diagonalen dargestellt. Oberhalb der Diagonalen sind die Cohen's Kappa-Koeffizienten mit den Signifikanzsternen<sup>16</sup> nach Landis und Koch (1977) dargestellt. Unterhalb der Diagonale sind die Korrelationen als Streudiagramme dargestellt, um einen detaillierteren Eindruck zu vermitteln.



**Abb. 12** Korrelationspanel der Textlänge und des prozentualen Anteils der Reflexionselemente

<sup>16</sup> Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$

Da der LOS von den relativen Häufigkeiten der Diskurselemente abhängt, wurde eine zweite Korrelationsmatrix berechnet, in der der LOS mit den absoluten Anteilen der Diskurselemente korreliert ist (Abbildung 13). Da der LOS u.a. aus den prozentualen Anteilen der Reflexionselemente abgeleitet wird, wird der LOS in Verbindung mit den absoluten Anteilen der formulierten Sätze analysiert, die als Beschreibungen, Bewertungen, Alternativen und Konsequenzen klassifiziert wurden. Die festgestellte positive Korrelation zwischen Alternativen und Konsequenzen wird auch in dieser Gegenüberstellung deutlich. Darüber hinaus gibt es positive Korrelationen zwischen den absoluten Anteilen der Bewertungen und den absoluten Anteilen der formulierten Alternativen und Konsequenzen, also allen diskursiven Elementen. Der Eindruck aus Abbildung 12 mit den signifikanten negativen Korrelationen zwischen Beschreibungen und allen anderen Diskurselementen findet sich im Korrelationspanel in Abbildung 13 nicht wieder. Stattdessen korrelieren die absoluten Anteile der Reflexionselemente Beschreibung und Bewertung signifikant positiv. Ein möglicher Einfluss der Struktur zeigt sich in einer signifikant negativen Korrelation des LOS mit dem absoluten Anteil der Beschreibungen bzw. den positiven Korrelationen mit dem Anteil der formulierten Alternativen oder Konsequenzen.



**Abb. 13** Korrelationspanel der LOS und der absoluten Anzahl der formulierten Reflexionselemente

Diese Daten deuten darauf hin, dass die konstruierte Metrik des LOS weiterhin als Indikator für die Qualität einer schriftlichen Reflexion angesehen werden kann, da weiterhin

ein höherer LOS im Zusammenhang mit einem weniger beschreibenden Text und einem reflexionsbezogenen Argumentationsprozess mit einem größeren Anteil an formulierten Alternativen und Konsequenzen steht. Weniger Bewertungen (absolute Häufigkeit) scheinen mit weniger Beschreibungen verbunden zu sein, welche wiederum für die Formulierung von Alternativen und Konsequenzen notwendig zu sein scheinen. Es bestätigt sich, dass die Länge einer schriftlichen Reflexion nicht universell mit dessen Qualität in Zusammenhang zu stehen scheint. Nichtsdestotrotz bleibt zu hinterfragen, inwieweit Häufigkeitsanalyse von identifizierten Diskurselementen oder deren Position im Text eine belastbare Qualitätsaussage über eine schriftliche Reflexion zulassen kann.

### 3.2.3 Qualitätsbewertungen

Neben den computerbasiert analysierten Strukturdimensionen der schriftlichen Reflexionen wurden von zwei unabhängigen Gutachtenden die vier Qualitätsindikatoren Breite, Tiefe, Kohärenz und Spezifität des Reflexionstextes auf einer vierstufigen Skalierung bewertet. Die Breite bildet hierbei die Anzahl der genannten Reflexionsauslöser ab, die Tiefe adressiert, wie differenziert ein Reflexionsauslöser diskutiert wurde, die Kohärenz bezieht sich auf die Nachvollziehbarkeit der reflexionsbezogenen Argumentation und die Spezifität eines Reflexionstextes repräsentiert, wie nah an den Geschehnissen der Videographie eine Argumentation stattfindet. Tabelle 2 zeigt die Ergebnisse der Interraterreliabilität zwischen beiden Gutachtenden sowie die Korrelationen zwischen den Qualitätsindikatoren. Alle Korrelationen konnten mit einem *p*-Wert von deutlich unter .001 als hoch signifikant eingeordnet werden. Die Beurteilerübereinstimmung mittels Cohens' Kappa erlangte mit mindestens .67 eine substantielle Übereinstimmung (Landis & Koch, 1977). Mit W-Werten von mindestens .842 und ebenfalls kleinen *p*-Werten (siehe Anhang A3) kann für alle Skalen eine Normalverteilung angenommen werden.

**Tab. 2** Korrelationen der Qualitätsdimensionen

	<i>Rater 1</i>				<i>Rater 2</i>			
	<i>Breite</i>	<i>Tiefe</i>	<i>Kohärenz</i>	<i>Spezifität</i>	<i>Breite</i>	<i>Tiefe</i>	<i>Kohärenz</i>	<i>Spezifität</i>
Breite	1	.58	.45	.47	1	.68	.75	.64
Tiefe		1	.46	.51		1	.65	.48
Kohärenz			1	.41			1	.57
Spezifität				1				1

Auf Grundlage der substantiellen Übereinstimmung der beiden Gutachtenden und der positiven Korrelationen aller Qualitätsbewertungen werden für die weiteren Analysen die Bewertungen von Rater 1 verwendet, da die vier Skalen bei Rater 2 stärker mit einander korrelieren und somit ein stärkerer Einfluss eines Bias, was einen qualitativ hochwertigen Reflexionstext kennzeichnet weniger ausgeschlossen werden kann.

Um analog zum vorherigen Kapitel um die deskriptiven Textmaße nach der deskriptiven Statistik auf korrelative Zusammenhänge zu stoßen, werden in den Tabellen 3 und 4 die Qualitätsbewertungen zu den computerbasiert analysierten Textmerkmalen in Verbindung gebracht. Tabelle 3 stellt die Bewertungen der Qualitätsindikatoren Breite, Tiefe, Kohärenz und Spezifität den absoluten und relativen Anteilen der Diskurselemente gegenüber. Eindrücklich unterscheiden sich die Zusammenhänge zu den Qualitätsbewertungen zwischen den absoluten und den relativen Anteilen. Fast alle Korrelationen zu absoluten Häufigkeiten werden signifikant und deuten auf positive Zusammenhänge hin. Lediglich mit der absoluten Häufigkeit formulierter Konsequenzen scheinen die Qualitätsbewertungen geringer ausgeprägt zu sein, was sich möglicherweise mit den in der Theorie identifizierten Diskurs von Unterschieden und Gemeinsamkeiten von Analyse und Reflexion erklären lässt, da eine gelungene Unterrichtsanalyse bereits einige Teile einer gelungenen Unterrichtsreflexion beinhalten kann (siehe Kap. 1.2.). In Bezug auf die relativen Häufigkeiten der Diskurselemente im Text scheinen (wenn auch weder signifikant noch stark ausgeprägt) die Qualitätsbewertungen negativ mit dem Anteil formulierter Konsequenzen zu sein, was im gängigen Verständnis einer gelungenen Reflexion (Stichwort Selbstbezug) als eher weniger erwartbar gelten kann. Um ein Verständnis der bewerteten Qualitätsbewertungen weiter aufzulösen, kann weiter aus Tabelle 3 gelesen werden, dass ein hohes Scoring in der Qualitätsdimension Kohärenz mit weniger Beschreibungen aber mehr Bewertungen und Alternativen einhergeht. Mit dieser Beobachtung werden die Befunde aus 3.2.2 unterstrichen, nach welchen deskriptive Texte im Allgemeinen weniger hochwertig sind. Abschließend geht aus den Korrelationen hervor, dass offenbar der relative Anteil an Beschreibungen positiv mit der Spezifität verbunden ist, da mutmaßlich für eine spezifische Gedankenführung mehr Beschreibungen vorgenommen werden müssen. Die Breite der in einer schriftlichen Reflexion adressierten Themen scheint wiederum maßgeblicher von den vorgenommenen Bewertungen verbunden zu sein.

**Tab. 3** Korrelationsmatrix zwischen Qualitätsdimensionen und Elementen  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$ )

	<i>Elemente (absolut)</i>				<i>Element (relativ)</i>			
	<i>Bes</i>	<i>Bew</i>	<i>Alt</i>	<i>Kons</i>	<i>Bes %</i>	<i>Bew %</i>	<i>Alt %</i>	<i>Kons %</i>
Breite	.34***	.47***	.42***	.16 $\cdot$	-.08	.13	.06	-.15 $\cdot$
Tiefe	.31***	.37***	.41***	.21*	-.04	.03	.08	-.07
Kohärenz	.31***	.45***	.41***	.24*	-.12	.11	.14	-.05
Spezifität	.45***	.37***	.29***	.18*	.11	-.05	-.08	-.12

In Tabelle 4 kommt diese Lesart in vereinfachter Form erneut zum Vorschein. Hier sind die beiden grundlegenden Repräsentanten für die absoluten und relativen Anteile an Diskurselementen den Qualitätsbewertungen gegenübergestellt. Zusammenfassend wird der aus der Literatur bekannte Zusammenhang zwischen Textlänge und Qualität signifikant deutlich. Der LOS scheint weniger verknüpft zu sein, wenngleich die Tiefe eines reflexionsbezogenen Argumentationsprozesses in der vom LOS repräsentierten Struktur identifizierbar ist, beziehungsweise eine modellgemäße Struktur die Tiefe eines Textes begünstigen könnte. Spezifität und LOS sind negativ korreliert. Vergegenwärtigt man sich die Visualisierung der Struktur nach dem LOS aus Abbildung 10, kann interpretiert werden, dass eine hohe Spezifität mit wiederkehrender Bezugnahme auf die Unterrichtssituation (in Form von Bewertungen) einhergeht und derartige Strukturen weniger vom LOS abgebildet werden.

**Tab. 4** Korrelationen zwischen LOS und Länge zu den Qualitätsdimensionen  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$ )

	<i>Breite</i>	<i>Tiefe</i>	<i>Kohärenz</i>	<i>Spezifität</i>
LOS	.07	.10	.03	-.19*
Textlänge	.54***	.47***	.49***	.53***

Da in Mientus et al. (2023) bemerkt wurde, dass der LOS in der Lage ist, die Qualität schriftlicher Reflexionen zumindest in Extremgruppen abzubilden, werden die Qualitätsbewertungen in Tabelle 5 den Texten nach Extremgruppen des LOS gegenübergestellt. Zur Bildung der Extremgruppen wurde die Stichprobe in Quantile unterteilt. Das erste, dritte bzw. fünfte Quantil beinhaltet jeweils die nach dem LOS 27 schwach, moderat oder gelungen strukturierten Reflexionstexte. Rein deskriptiv ist zu bemerken, dass die Qualitätsbewertungen für die moderat strukturierten Texte tendenziell höher bewertet wird, als für die schwach oder gelungen strukturierten Texte. Auffällig ist dennoch, dass unter den schwach strukturierten Texten wie auch unter den gelungen strukturierten Texten überwiegend negative Korrelate zu allen vier Qualitätsbewertungen bestehen, was die Aussagekraft des LOS für die Qualität schriftlicher Reflexionen relativiert.

**Tab. 5** Qualitätsdimensionen nach Extremgruppen (bei  $N = 134$  ist pro Extremgruppe  $n = 27$ )

	<i>Breite</i>	<i>Tiefe</i>	<i>Kohärenz</i>	<i>Spezifität</i>
Über alle Texte	$\sigma = 2.34$ $sd = .75$	$\sigma = 2.44$ $sd = .83$	$\sigma = 2.46$ $sd = .79$	$\sigma = 2.10$ $sd = .82$
schwach strukturiert LOS: [ -.064   .127 ]	$\sigma = 2.11$ $sd = .64$ $\kappa = .00$	$\sigma = 2.19$ $sd = .79$ $\kappa = -.13$	$\sigma = 2.41$ $sd = .84$ $\kappa = -.25$	$\sigma = 2.26$ $sd = .90$ $\kappa = -.08$
moderat strukturiert LOS: [ .259   .318 ]	$\sigma = 2.50$ $sd = .81$ $\kappa = -.05$	$\sigma = 2.54$ $sd = .81$ $\kappa = .16$	$\sigma = 2.42$ $sd = .81$ $\kappa = .11$	$\sigma = 2.15$ $sd = .88$ $\kappa = .35^*$
gelingen strukturiert LOS: [ .428   .703 ]	$\sigma = 2.30$ $sd = .78$ $\kappa = -.19$	$\sigma = 2.48$ $sd = .94$ $\kappa = -.44^*$	$\sigma = 2.56$ $sd = .85$ $\kappa = -.22$	$\sigma = 1.85$ $sd = .82$ $\kappa = -.38^*$

Trotz der Vorteile der computerbasierten und entsprechend ökonomischen Abschätzung der Qualität schriftlicher Reflexionen, welche der LOS nach Mientus et al. (2023) bieten kann, bleibt der LOS alleinig nur bedingt aussagekräftig. In einer vertiefenden Faktorenanalyse sollen daher explorativ zugrundeliegende Konstrukte identifiziert werden, welche die Skalen der Textanalyse repräsentieren.

### 3.2.4 Explorative Faktorenanalyse

Mit den bis hierher vorgestellten Methoden ist die Erhebung einer Fülle von Qualitätsbewertungen und Textmerkmalen schriftlicher Reflexionen möglich. Aus diesem Grund soll im Folgenden der Forschungsfrage 1 nachgegangen werden. Es soll untersucht werden auf welche zentralen Faktoren die erhobenen Variablen der Texte hindeuten. Eine geeignete Methode hierzu ist die explorative Faktorenanalyse, welche im Weiteren angewandt werden soll.

Um mit der bestehenden Stichprobe die Exploration der Faktorenanalyse durchführen zu können und im Rahmen dieser Studie zu validieren wurde an dieser Stelle der Datensatz in einen Trainings- und Testdatensatz aufgeteilt. Als geeignetes Verhältnis wurden 80:20 gewählt, sodass 107 Reflexionstexte für die Faktorenanalyse verwendet und 27 Texte für eine Validierung der entwickelten Verfahren zurückgehalten wurden.

#### 3.2.4.1 Anwendung der Faktorenanalyse

Da die bereits erhobenen Textmerkmale keine Ausreißer aufweisen und nach dem Bartlett-Test<sup>17</sup> eine hinreichende Signifikanz für die Durchführung einer explorativen Faktorenanalyse

<sup>17</sup> Der Bartlett-Test ist ein statistischer Test, der verwendet wird, um zu überprüfen, ob die Varianzen mehrerer Stichproben gleich sind. Er wird häufig in der multivariaten Statistik verwendet, um die Homogenität der Varianzen in einer Gruppe von Stichproben zu testen. Insbesondere kann der Bartlett-Test verwendet werden,

gegeben ist ( $\text{Chi}^2 = 850.59$ ,  $p = .001$ ), werden alle Variablen, die für die schriftlichen Reflexionen vorliegen und eine Aussage über deren Qualität haben könnten in der Faktorenanalyse berücksichtigt. Im Einzelnen sind dies:

- die Länge des Reflexionstextes,
- die absoluten Häufigkeiten aller Diskurselemente nach Nowak et al. (2019) (Beschreibung, Bewertung, Alternativen, Konsequenzen),
- die prozentuale Häufigkeit der Beschreibungen innerhalb der Reflexionstexte,
- der LOS, sowie
- die vier Qualitätsdimensionen (Breite, Tiefe, Kohärenz & Spezifität).

Die relativen Häufigkeiten der Diskurselemente Bewertungen, Alternativen und Konsequenzen werden nicht berücksichtigt, da sich alle vier Diskurselemente gegenseitig beeinflussen. Nach dem Kaiser-Meyer-Olkin-Test<sup>18</sup> sollten für eine Faktorenanalyse ausschließlich Variablen in eine Faktorenanalyse einfließen, welche Korrelationen über .50 annehmen, was keine Variable unter den relativen Häufigkeiten der Diskurselemente erreicht. Die relative Häufigkeit der Beschreibungen wird mit einem KMO<sup>19</sup> von .45 als wichtiger Repräsentant der relativen Häufigkeiten der Diskurselemente dennoch mit in die Faktorenanalyse einbezogen, da (wie genannt) alle vier relativen Häufigkeiten direkt voneinander abhängig sind und die KMO-Faktoren unter Berücksichtigung der übrigen drei Elemente schwächer ausfallen ( $\text{KMO}_{\text{Bew}_\%} = .29$ ,  $\text{KMO}_{\text{Alt}_\%} = .39$  und  $\text{KMO}_{\text{Kons}_\%} = .44$ ).

Für die vorliegenden Daten wird nach MAP ein Faktor vorgeschlagen. Nach weiteren Methoden für die Faktorextraktion, wie der Hauptkomponentenanalyse oder der Maximum-Likelihood-Methode werden zwei bis drei Faktoren vorgeschlagen. Aus diesem Grund ist in Abbildung 14 ein Scree-Plot<sup>20</sup> dargestellt, nach welchem drei Faktoren sinnvoll erscheinen.

---

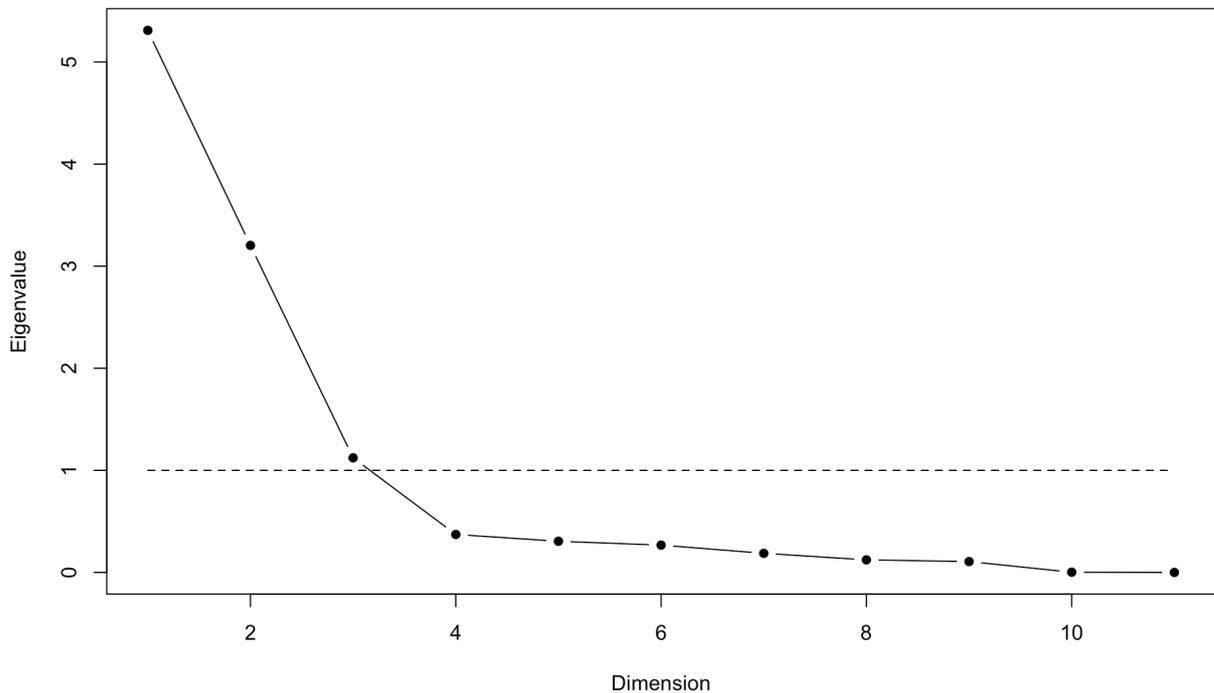
um zu prüfen, ob die Varianzen in verschiedenen Gruppen gleich sind, beispielsweise um zu untersuchen, ob die Varianz der Textlänge in verschiedenen Studiengruppen gleich ist. Der Test berücksichtigt dabei sowohl die Größe der Stichprobe als auch die Anzahl der Gruppen.

<sup>18</sup> Der Kaiser-Meyer-Olkin-Test ist ein statistischer Test, der in der multivariaten Statistik verwendet wird, um die Eignung der Daten für eine Faktorenanalyse zu überprüfen. Er prüft, ob die Korrelationen zwischen den Variablen ausreichend sind, um eine Faktorenanalyse durchzuführen. Konkret wird der Kaiser-Meyer-Olkin-Test verwendet, um die Stärke der Korrelationen zwischen den Variablen zu messen und zu prüfen, ob diese ausreichend sind, um eine Faktorenanalyse durchzuführen. Wenn die Korrelationen zwischen den Variablen hoch genug sind, kann die Faktorenanalyse verwendet werden, um die zugrunde liegenden Faktoren zu identifizieren, die die Variation in den Daten erklären. Der Kaiser-Meyer-Olkin-Test berechnet einen Wert zwischen 0 und 1, der als Maß für die Abhängigkeit der Variablen voneinander dient. Ein Wert von 1 bedeutet, dass alle Variablen perfekt miteinander korrelieren und somit ideal für eine Faktorenanalyse geeignet sind. Ein Wert von 0 hingegen zeigt an, dass die Variablen unabhängig voneinander sind und daher keine Faktorenanalyse durchgeführt werden sollte.

<sup>19</sup> Kaiser-Meyer-Olkin-Faktor

<sup>20</sup> Ein Scree Plot ist ein grafisches Werkzeug, das in der multivariaten Statistik verwendet wird, um zu bestimmen, wie viele Faktoren oder Dimensionen in einem Datensatz signifikant sind. Er wird oft in der Faktorenanalyse eingesetzt, um zu entscheiden, wie viele Faktoren extrahiert werden sollten. Ein Scree Plot ist ein Streudiagramm, das die Eigenwerte der Faktoren (auf der y-Achse) gegen ihre Rangordnung (auf der x-Achse)

Zum einen zeigt sich dies auf Grund der Anzahl an Dimensionen, welche noch Eigenwerte von  $\geq 1$  aufweisen und zum zweiten an einem interpretierbaren „Knick“ in der abflachenden Kurve.



**Abb. 14** Scree Plot – Faktorenanalyse

Da wir für die vorliegenden Daten eher annehmen können, dass Korrelationen zwischen den Faktoren bestehen, da beispielsweise nach der MAP-Methode bereits nur ein Faktor (mutmaßlich der Reflexionstext selbst) vorgeschlagen ist, wird eine obimite Faktorenanalyse in der vorliegenden Studie bevorzugt. Gleichzeitig ergeben orthogonale Verfahren, wie die Varimax-Rotation immer unspezifischere Faktoren, als oblique Verfahren (Venkatesh et al., 2016).

Aus diesem Grund sind in der Tabelle 6<sup>21</sup> die Faktorladungen und die zugehörige Faktoreninterpretation für drei Faktoren abgebildet. Tabelle 7 zeigt eine Übersicht der Varianzen, die die einzelnen Variablen in der zugehörigen Faktorenanalyse erklären können.

---

darstellt. Ein Eigenwert ist ein Maß für die Varianz, die durch den jeweiligen Faktor erklärt wird. Ein Faktor mit einem hohen Eigenwert erklärt also eine größere Menge an Varianz in den Daten als ein Faktor mit einem niedrigeren Eigenwert. Im Scree Plot werden die Eigenwerte absteigend sortiert und in abnehmender Größe auf der y-Achse dargestellt. Die Rangordnung der Faktoren wird auf der x-Achse dargestellt. Der Scree Plot zeigt einen Knick, an dem sich die Steigung der Kurve ändert, von steil zu flach. Die Anzahl der Faktoren, die vor diesem Knick liegen, werden als signifikante Faktoren betrachtet, da sie einen wesentlichen Beitrag zur Varianz in den Daten leisten. Die Entscheidung darüber, wie viele Faktoren extrahiert werden sollen, kann auf der Grundlage des Scree Plots getroffen werden. Die Anzahl der signifikanten Faktoren kann durch eine visuelle Inspektion des Scree Plots bestimmt werden..

<sup>21</sup> Aus Gründen der Übersichtlichkeit sind in den Tabellen alle Werte zwischen ] - .10 | .10 [ ausgeblendet und für jede Variable der gewichtigste Faktor fett gedruckt.

**Tab. 6** Faktorenanalyse nach Promax-Verfahren (oblimin)

<i>Variablen</i>	<i>Drei Faktoren</i>		
	<i>F1</i>	<i>F2</i>	<i>F3</i>
Kumulative Varianz	.22	.42	.59
Länge	<b>.98</b>		.29
Beschreibung	<b>.69</b>		.69
Bewertung	<b>.69</b>	.10	-.16
Alternativen	<b>.34</b>	.30	-.30
Konsequenzen	<b>.50</b>		-.20
Beschreibung_%	-.23	.12	<b>.95</b>
LOS			<b>-.43</b>
Breite		<b>.73</b>	
Tiefe	-.16	<b>.89</b>	
Kohärenz	.17	<b>.50</b>	
Spezifität		<b>.61</b>	.22
Interpretation: (Latente Variablen)	<i>Quantität</i>	<i>Qualität</i>	<i>Deskriptivität</i>

**Tab. 7** Aufklärung der Varianz nach Variablen

<i>Variable</i>	<i>Bes %</i>	<i>Länge</i>	<i>Bes</i>	<i>Bew</i>	<i>Tiefe</i>	<i>Breite</i>	<i>Spezifität</i>	<i>Alt</i>	<i>Kohärenz</i>	<i>Kons</i>	<i>LOS</i>
Anteil an Varianz	.966	.959	.883	.634	.624	.554	.467	.462	.400	.269	.185

Mit Ausnahme der absoluten Häufigkeit der Beschreibungen in Tabelle 6 laden alle Variablen eindeutig auf jeweils einen der drei Faktoren, sodass eine Interpretation der Faktoren nach Qualität (Variablen der Qualitätsdimensionen), Quantität (Textlänge und absolute Häufigkeiten der Diskurselemente) und Deskriptivität (relative Häufigkeit der Beschreibungen und der LOS) möglich ist. Diese Faktorladungen spiegeln einerseits die Qualitätsindikatoren bzw. Textmerkmale der Textanalyse in Qualität und Quantität wider, ergänzen diese jedoch um den Faktor der Deskriptivität. Mit einem Anteil von .966 ist allein die relative Häufigkeit der formulierten Beschreibungen ausschlaggebend für die Auflösung der Varianz der Stichprobe. Vereinfacht gesagt scheint es für die Faktorenanalyse ausschlaggebend zu sein, wie deskriptiv eine schriftliche Reflexion ist. Zu beachten ist, dass größere Werte der Faktoren Qualität und Quantität eher positiv bewertet werden können, da sie qualitativ höherwertige bzw. längere Texte repräsentieren, wobei der Faktor Deskriptivität invers zu betrachten ist. Von einem höherwertigen Reflexionstext kann die Rede sein, wenn dieser Text einen geringen Wert in der Deskriptivität aufweist und somit weniger beschreibend ist.

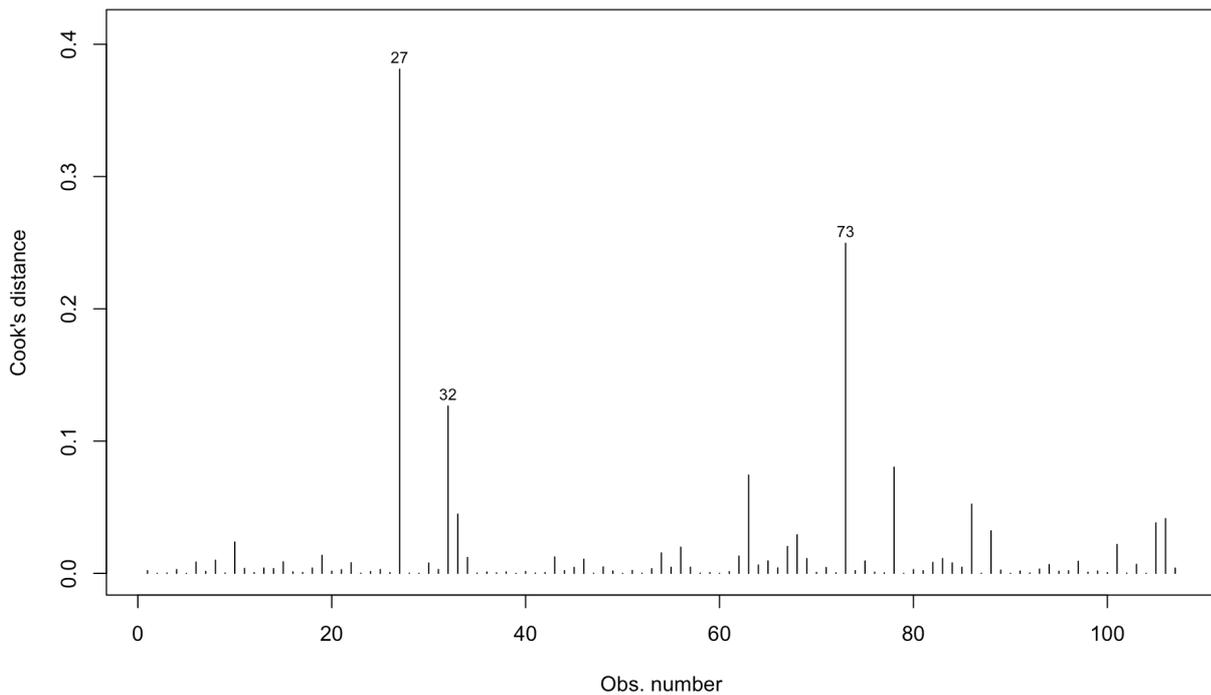
Zum besseren Verständnis der Faktoren sind in Tabelle 8 die Korrelationen zwischen den Variablen und den zugehörigen Faktoren, auf welche die Variablen stark laden dargestellt.

**Tab. 8** Korrelationen zwischen Faktoren und zugeordneten Variablen ( $p = .001$ )

<b>Faktoren</b>	<b>Variablen</b>				
<b>Qualität</b>	<i>Breite</i>	<i>Tiefe</i>	<i>Kohärenz</i>	<i>Spezifität</i>	
	.86	.86	.69	.76	
<b>Quantität</b>	<i>Länge</i>	<i>Beschreibung</i>	<i>Bewertung</i>	<i>Alternativen</i>	<i>Konsequenzen</i>
	.90	.55	.77	.57	.55
<b>Deskriptivität</b>	<i>LOS</i>		<i>Beschreibung %</i>		
	-.43		.92		

Die Faktoren Qualität ( $W = .983, p = .084$ ) und Quantität ( $W = .973, p = .010$ ) können als normalverteilt angenommen werden. Der Faktor Deskriptivität nach dem Shapiro-Wilk-Test mit  $W = .99$  bei  $p = .800$  ist nicht normalverteilt, wobei dies auf die bereits nicht oder nur bedingt normalverteilten Variablen LOS und Beschreibungen\_% zurückzuführen ist (siehe Anhang A3). Alle Faktoren einer Faktorenanalyse sind um den Mittelwert Null verteilt, wobei sich die Verteilung der Texte über die Faktoren unterscheidet ( $sd_{\text{Qualität}} = .92$  bei einer Range von -1.79 bis 2.47,  $sd_{\text{Quantität}} = 1.04$  bei einer Range von -3.32 bis 2.98 und  $sd_{\text{Deskriptivität}} = 0.99$  bei einer Range von -2.53 bis 2.73). Nach einer Probe auf Ausreißer unter den drei Faktoren (Qualität in Abhängigkeit von Quantität und Deskriptivität) konnten via Cook-Distanzen<sup>22</sup> drei Texte identifiziert werden, welche in Abbildung 15 zu erkennen sind. Aus Tabelle 9 geht hierbei hervor, dass alle drei Reflexionstexte an den Skalenenden der Textlänge liegen, wobei den Texten bestenfalls eine moderate Strukturierung attestiert werden kann. Die zuvor Vorgestellten Ergebnisse der explorativen Faktorenanalyse bleiben jedoch auch bei Ausschluss der Ausreißertexte stabil, weshalb ein Ausschluss aus dem Datensatz nicht nötig ist.

<sup>22</sup> Cook-Distanzen beschreiben den Abstand von geschätzten und tatsächlichen Werten der abhängigen Variablen. Im Rahmen einer Faktorenanalyse können so Daten identifiziert werden, welche der Faktorisierung im Verhältnis zu den übrigen Daten tendenziell widersprechen.



**Abb. 15** Cook-Distanzen innerhalb des Trainingsdatensatzes mit Kennzeichnung der Ausreißertexte

**Tab. 9** Ausreißertexte und zugehörige Faktoren bzw. Variablen

<i>ID</i>	<i>Cook-Distance</i>	<i>Qualität</i>	<i>Quantität</i>	<i>Deskriptivität</i>	<i>Länge</i>	<i>LOS</i>
CLMARÜ20	.4517	-0.1874	2.3392	2.4374	1666	.1139
CONOGI30	.3906	1.2598	2.4832	2.1296	1755	.1930
JALABE05	.2676	-1.5204	-3.4382	1.8159	102	.0000

Betrachtet man die linearen Zusammenhänge der Faktoren Qualität, Quantität und Deskriptivität, so sind deutlich die erwarteten Korrelationen zwischen Qualität und Quantität ersichtlich ( $\kappa = .73, p = .001$  bzw.  $\kappa = .78, p = .001$  ohne Ausreißer). Qualität und Deskriptivität weisen einen gering ausgeprägten Zusammenhang auf ( $\kappa = -.02, p = .79$  und  $\kappa = -.10, p = .265$  ohne Ausreißer), wobei Quantität und Deskriptivität erneut in größerem Zusammenhang stehen ( $\kappa = -.10, p = .23$  und  $\kappa = -.16, p = .074$  ohne Ausreißer). Analog zu den graphischen Verdeutlichungen der Qualitätskorrelate in der Diskussion des LOS (Kap 3.2.2) ist in Abbildung 16 ein Korrelationspanel der drei Faktoren in der restriktiven Form inklusive der Ausreißer abgebildet.

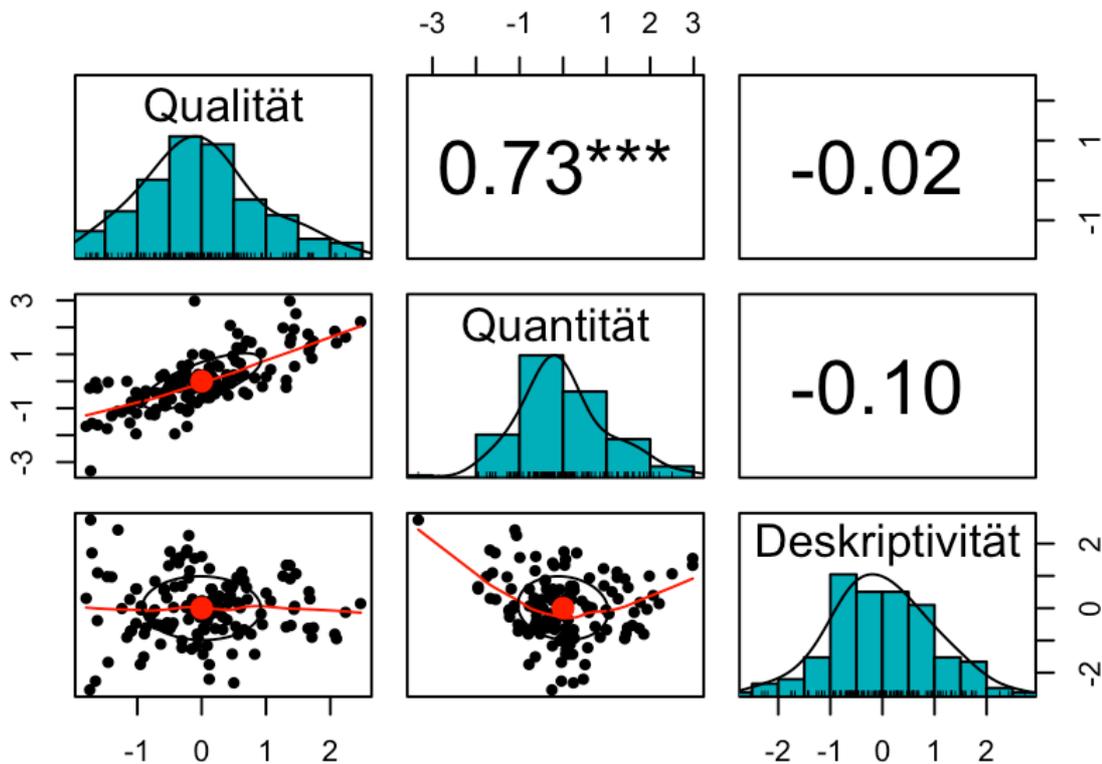
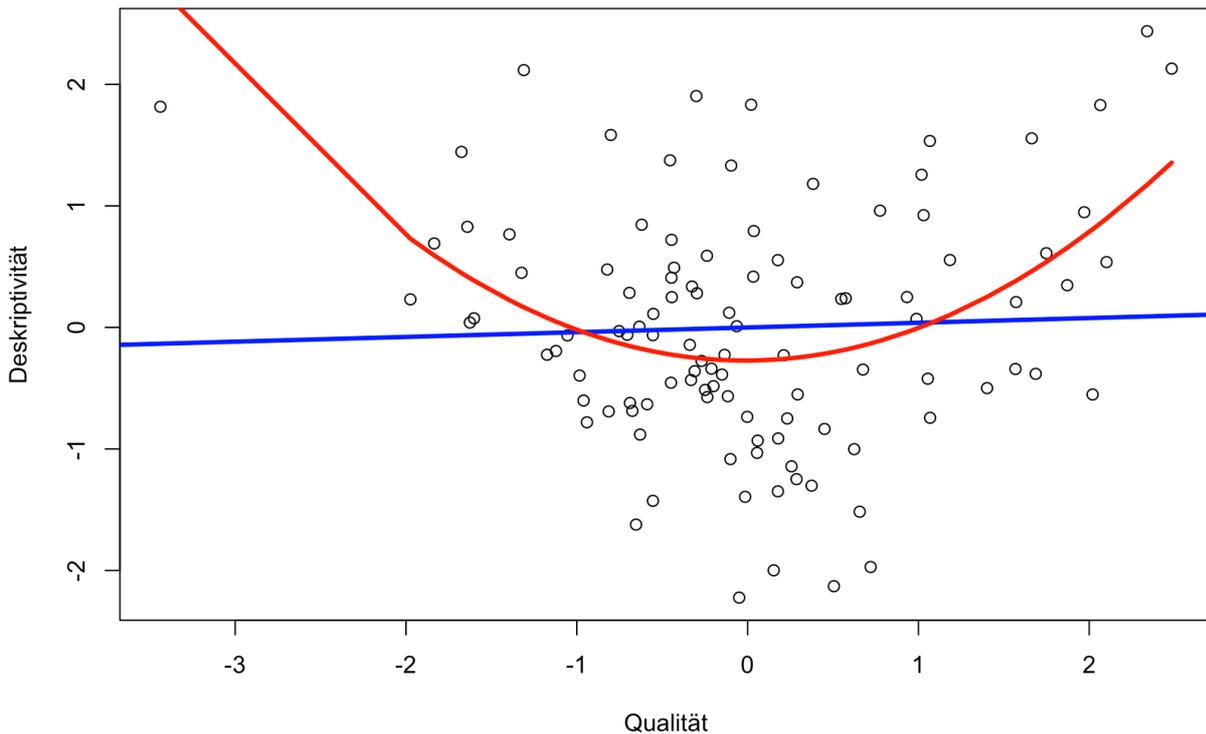


Abb. 16 Korrelationsmatrix der drei Faktoren unter Einbezug der Ausreißer

### 3.2.4.2 Untersuchung der Qualitätskorrelate

Cohens' Kappa geht von linearen Regressionen aus, welche aus dem Streudiagramm von Quantität und Deskriptivität in Abbildung 16 nicht hervorgeht. Vielmehr kann auch rechnerisch getestet werden, dass eine quadratische Regression mit einem  $R^2 = .203$  und einem Standardfehler von .883 eine geeignetere Näherung aufweist, als eine lineare Regression mit einem  $R^2 = .002$  bei einem Standardfehler von .983. Der höhere  $R^2$ -Wert sowie der etwas geringere Fehler bedeuten, dass quadratisch eine bessere Näherung erreicht werden kann. Zuletzt kann mit einem Partial F-test die Verteilung der Stichprobe im quadratischen Zusammenhang gegen die Nullhypothese, dass keine signifikanten Unterschiede zu einem quadratischen Polynom vorliegen, getestet werden. Eine ANOVA<sup>23</sup> ergab einen p-Wert von deutlich unter .001, sodass die Nullhypothese abgelehnt werden muss und ein quadratischer Zusammenhang angenommen werden kann. Abbildung 17 verdeutlicht die angenommenen linearen und quadratischen Regressionen und unterstreicht die Passung eines quadratischen Zusammenhangs.

<sup>23</sup> Die ANOVA oder Varianzanalyse ist ein multivariantes Analyseverfahren, das die Mittelwerte von verschiedenen, normalverteilten Stichproben vergleicht. In begründeten Fällen kann eine ANOVA auch mit nicht-normalverteilten Variablen durchgeführt werden. Dies ist beispielsweise bei hinreichend großen Stichproben oder (wie in diesem Fall) in der inhaltlichen Relevanz einer Variable erlaubt.

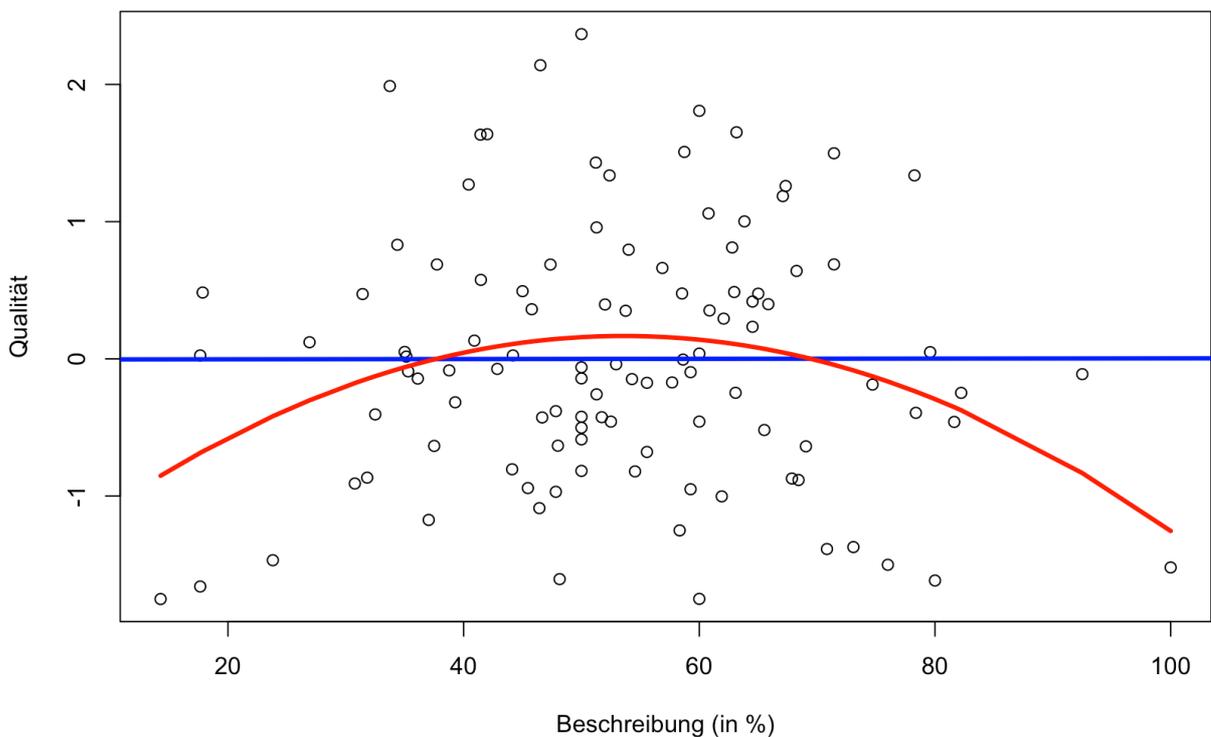


**Abb. 17** Lineare (blau) versus quadratische (rot) Regression im Quantität-Qualität-Streudiagramm

Wo polynome Zusammenhänge identifizierbar sind, sind Extrema identifizierbar, sodass Aussagen zum absoluten Zusammenhang zwischen Quantität und Qualität einer schriftlichen Reflexion relativiert werden können. Aus den vorliegenden Daten geht hervor, dass ein längerer Reflexionstext nicht per se eine höhere Qualität aufweist. Auf die evidenzbasierte Aussage, dass Qualität und Quantität positiv korrelieren, kann dank des quadratischen Zusammenhangs zwischen Quantität und Deskriptivität nun bestimmt werden, wie eine maximale Qualität erreicht werden kann, ohne die Quantität der schriftlichen Reflexion inflationär zu erhöhen. Mittels Extremwertberechnung kann ermittelt werden, dass circa bei einer Quantität von -0.27 die minimale Deskriptivität mit -0.01 für die vorliegende Videovignette erreicht werden kann. Hieraus sind jedoch keine ad hoc interpretierbaren Schlussfolgerungen zu ziehen.

Im Weiteren wurden daher Variablen näher untersucht, welche am stärksten mit einem Faktor assoziiert sind. Da die Variablen Beschreibung\_% und Textlänge mit .92\*\*\* bzw. .90\*\*\* am stärksten mit dem Faktor Deskriptivität bzw. Quantität korrelieren (siehe Tabelle 8) und beide Variablen erheblichen Anteil der Varianz der Faktorenanalyse erklären (siehe Tabelle 7) kann geschlussfolgert werden, dass beide Variablen als Repräsentanten für den jeweiligen Faktor bestehen. Gemäß der in Mientus et al. (2023) dargestellten Ergebnisse soll als weiteres Textmerkmal der LOS als weiterer Repräsentant des Faktors Deskriptivität in die weiteren Betrachtungen einbezogen werden.

Aus diesem Grund wird der identifizierte quadratische Zusammenhang zur Qualität auch zu den beiden Variablen, welche am höchsten mit den Faktoren Quantität und Deskriptivität korrelieren, untersucht. Für den Faktor Quantität stellt sich die Textlänge mit einer Korrelation von  $.90^{***}$  und für den Faktor der Deskriptivität die prozentualen Beschreibungen mit einem Korrelationskoeffizient von  $.92^{***}$  als repräsentativ dar. In Abbildung 18 ist daher der Zusammenhang zwischen dem Faktor Qualität und der Variablen Beschreibung\_% (als Repräsentant für Deskriptivität) abgebildet. Eine quadratische Näherung ist mit einem  $R^2 = .071$  und einem Standardfehler von  $.909$  geringfügig geeigneter, als eine lineare Näherung ( $R^2 = .001$ ) bei einem Standardfehler von  $.939$ ), wenngleich die quadratische Näherung signifikant besser ist ( $p = .006$ ).

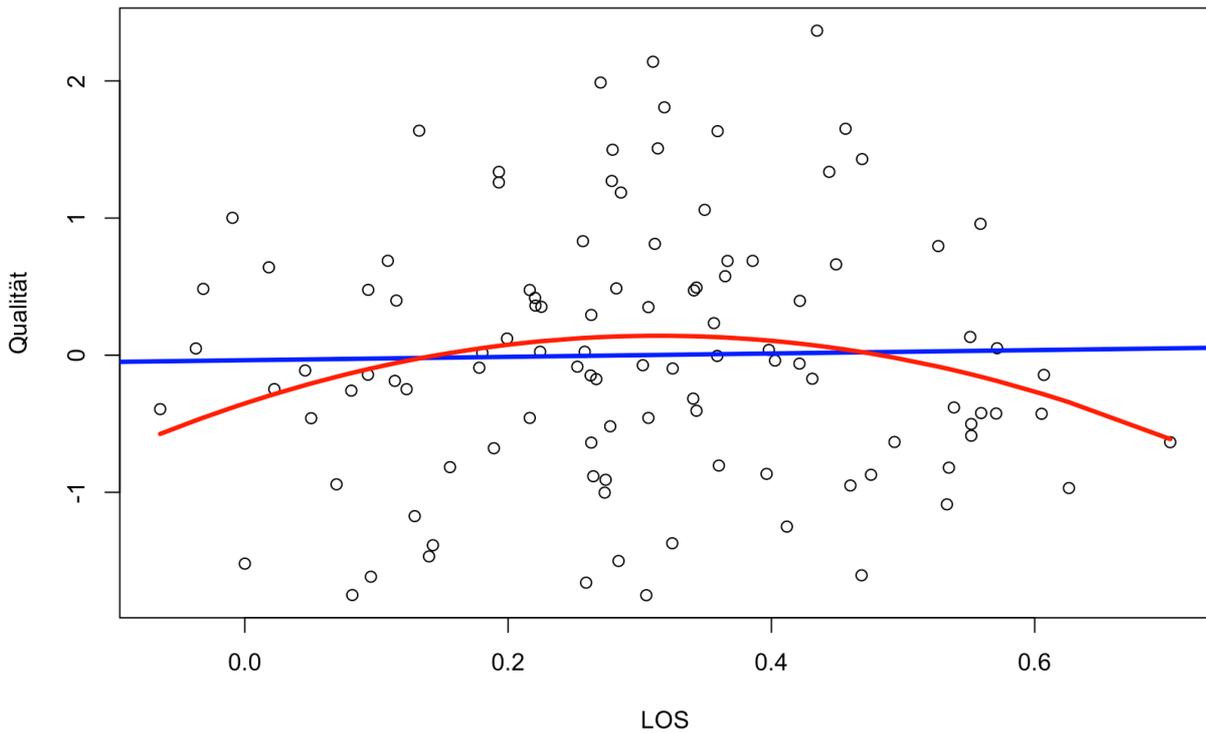


**Abb. 18** Qualität nach Beschreibungen\_% mit linearer (blau) und quadratischer (rot) Regression

Nach der quadratischen Regression kann eine maximale Qualität in einer schriftlichen Reflexion bei einem Anteil an formulierten Beschreibungen von  $Bes\_ \% = 53.58\%$  erreicht werden. Trotz des schwächeren Zusammenhangs, ermöglicht selbiges Vorgehen bei der Gegenüberstellung von Qualität und LOS die Erkenntnis, dass ein LOS von  $.3144$  mit einer maximalen Qualität zusammenhängt<sup>24</sup> (siehe Abbildung 19). Eine lineare Näherung ( $R^2 = .001$ , Error =  $.938$ ) stellt sich hierbei vernachlässigbar schlechter dar ( $p = .060$ ), als eine quadratische Näherung ( $R^2 = .034$ , Error =  $.927$ ).

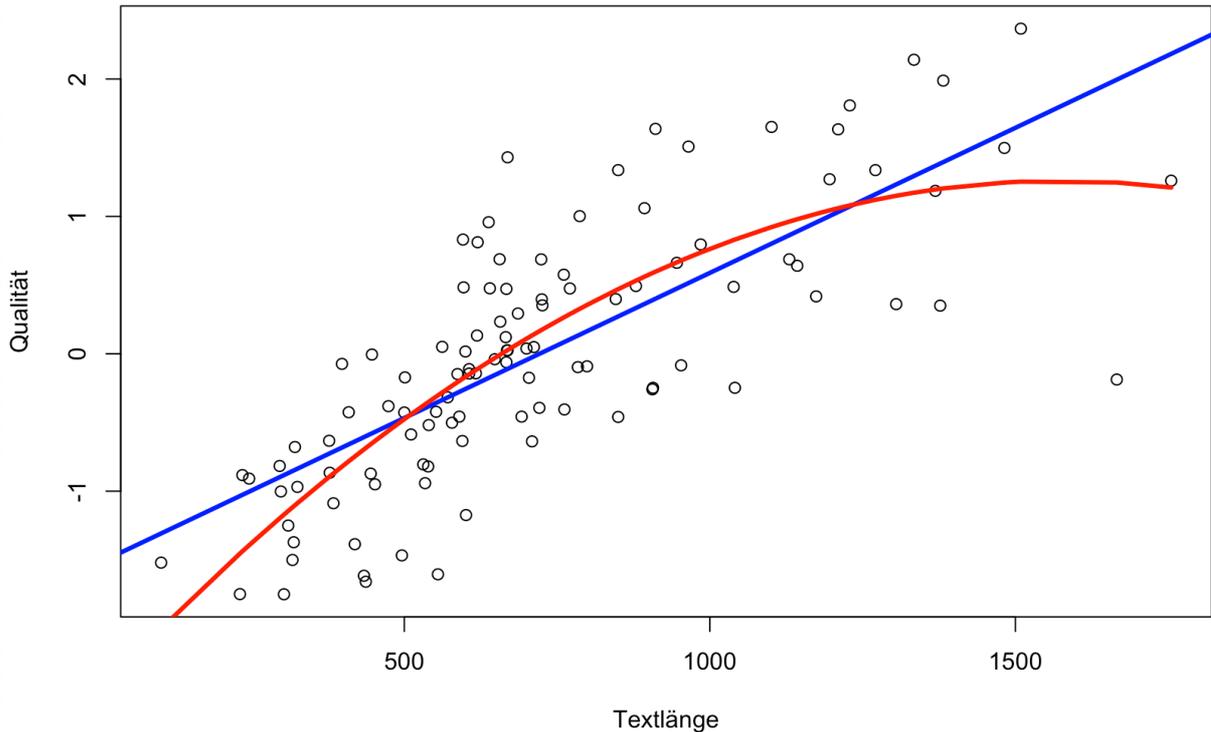
<sup>24</sup> Lineare Regression:  $R^2 = .003177$ , Error =  $.925$

Quadratische Regression:  $R^2 = .04448$ , Error =  $.9091$  und  $p = .001877$



**Abb. 19** Qualität nach LOS mit linearer (blau) und quadratischer (rot) Regression

In der Gegenüberstellung von Qualität und Textlänge stellt sich die quadratische Näherung ebenfalls als lediglich geringfügig geeigneter dar ( $p = .001$ ). Die linearere Regression weist hierbei ein  $R^2 = .566$  bei einem Standardfehler von  $.618$  auf, wobei die quadratische Regression Werte von  $R^2 = .620$  und  $\text{Error} = .582$  aufweist. Das theoretische Maximum der Qualität liegt in diesem Fall bei einer Textlänge von ca. 1574 Worten und damit weit oberhalb der meisten Textlängen (siehe Abbildung 20).



**Abb. 20** Qualität nach Textlänge mit linearer (blau) und quadratischer (rot) Regression

Unter der Annahme, dass quadratische Zusammenhänge bestehen, ergibt sich ein dreidimensionaler Vektor, an welchem Länge, prozentuale Beschreibung und LOS so sind, dass die Qualität maximal wird. Jedem Reflexionstext kann nach diesen Überlegungen ein dreidimensionaler Vektor aus Textlänge, Beschreibung\_% und LOS zugeordnet werden:

$$\overrightarrow{QK} = \begin{pmatrix} \text{Textlänge} \\ \text{Beschreibung\_}\% \\ \text{LOS} \end{pmatrix}$$

### 3.2.4.3 Methodische Exploration eines absoluten Qualitätskorrelats

Dank der Betrachtungen aus dem vorherigen Kapitel ergibt sich ein dreidimensionaler Ortsvektor aus Textlänge, Beschreibungen\_% und LOS, zu welchem die entsprechenden Vektoren eines jeden Reflexionstextes eine bestimmbare Distanz haben. Inwieweit mit dieser Distanz zum maximalen Qualitätskorrelat ein computerbasiert analysierbares Qualitätskorrelat gefunden wurde, welches die Qualitätsbewertungen der Expert\*innen abschätzen kann (Forschungsfrage 2) soll im Folgenden untersucht werden.

Hierzu ist es jedoch zunächst nötig, die bestehenden computerbasierten Textmerkmale einer z-Standardisierung zu unterziehen. Die z-Standardisierung, auch bekannt als die z-Score-Transformation, ist eine statistische Methode, die verwendet wird, um Variablen zu transformieren, damit sie auf eine Standardnormalverteilung mit einem Mittelwert von 0 und

einer Standardabweichung von 1 gebracht werden. Die z-Standardisierung wird verwendet, um Daten zu vergleichen und zu analysieren, die auf unterschiedlichen Skalen gemessen wurden, da sie es ermöglicht, Daten auf eine gemeinsame Skala zu bringen. Die z-Standardisierung wird durchgeführt, indem man den Mittelwert von jedem Datenpunkt in einer Variable subtrahiert und das Ergebnis durch die Standardabweichung der Variable teilt. Das Ergebnis ist ein neuer Wert, der angibt, wie viele Standardabweichungen ein Datenpunkt von der durchschnittlichen Position entfernt ist. Ein z-Wert von 0 bedeutet, dass der Datenpunkt genau im Durchschnitt liegt, während positive z-Werte angeben, dass der Datenpunkt über dem Durchschnitt liegt, und negative z-Werte, dass der Datenpunkt unter dem Durchschnitt liegt. Die z-Standardisierung wird in vielen Bereichen wie der Finanzanalyse, der medizinischen Forschung, der Psychologie und anderen Bereichen verwendet, in denen es wichtig ist, Daten auf eine gemeinsame Skala zu bringen, um sie vergleichen oder analysieren zu können.

Um nun die unter 3.2.4.3 ermittelten maximalen Qualitätskorrelate von Textlänge (1574 Worte), prozentuaalem Anteil an Beschreibung (53%) sowie Grad der Strukturiertheit (LOS von .3144) ebenfalls einer z-Score-Transformation zu unterziehen, werden auch diese normativen Werte durch die Standardabweichung der jeweiligen Textvariable (Länge, Beschreibung\_% bzw. LOS) dividiert. Es ergibt sich ein Ortsvektor des maximalen Qualitätskorrelates ( $\overrightarrow{QK_{max}}$ ) aus:

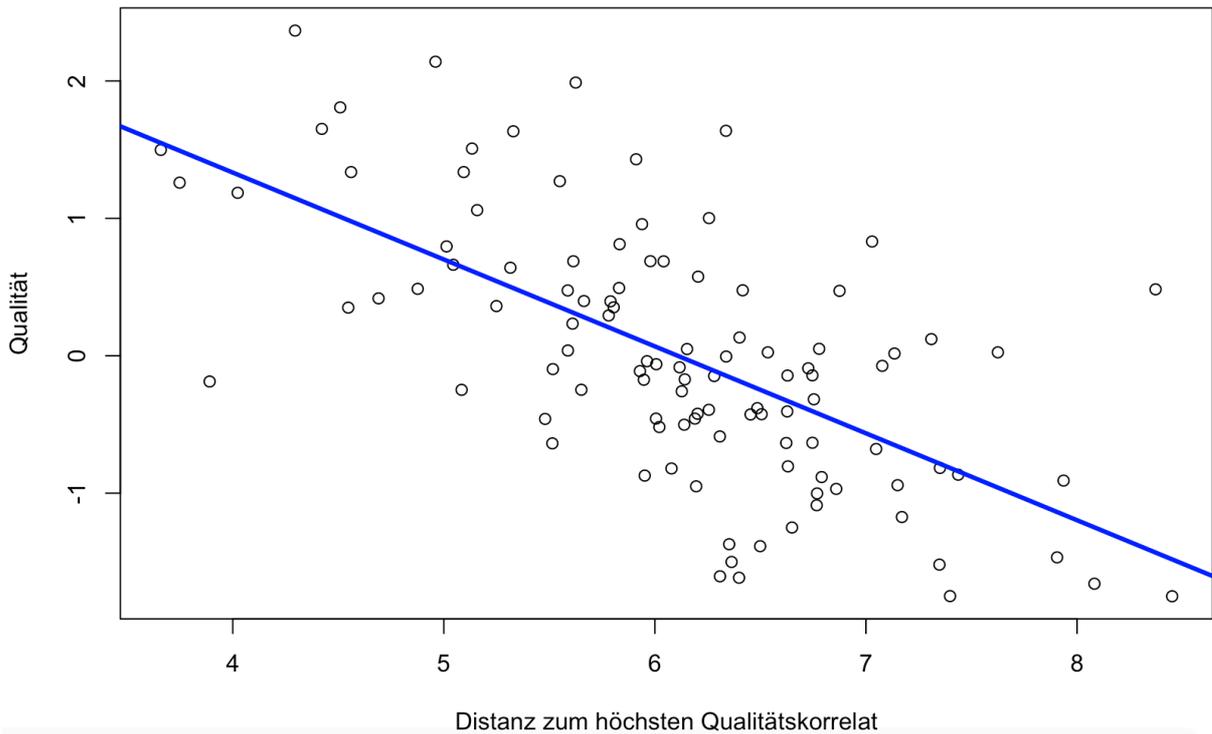
$$\overrightarrow{QK_{max}} = \begin{pmatrix} Länge_{max} \\ Beschreibung\_ \%_{max} \\ LOS_{max} \end{pmatrix} = \begin{pmatrix} 1574 : sd(Textlänge) \\ 53\% : sd(Beschreibung\_ \%) \\ .3144 : sd(LOS) \end{pmatrix}$$

Zuletzt kann für jede schriftliche Reflexion die Distanz im dreidimensionalen Vektorraum zwischen dem computerbasiert-bestimmten Ortsvektor aus Textlänge, Beschreibung\_% und LOS zum maximalen Qualitätskorrelat bestimmt werden:

$$Distanz = |\overrightarrow{QK_{max}} - \overrightarrow{QK}| = \left| \begin{pmatrix} Länge_{max} \\ Beschreibung\_ \%_{max} \\ LOS_{max} \end{pmatrix} - \begin{pmatrix} Textlänge \\ Beschreibung\_ \% \\ LOS \end{pmatrix} \right|$$

$$= \sqrt{(Länge_{max} - Textlänge)^2 + (Beschreibung\_ \% - Beschreibung\_ \%)^2 + (Struktur_{max} - LOS)^2}$$

In Abbildung 21 ist die durch menschliche Rater ermittelte Qualität aller schriftlichen Reflexionstexte nach dessen computerbasiert berechneten Distanz zum maximalen Qualitätskorrelat aufgetragen. Deutlich zu erkennen ist die signifikant negative Korrelation (siehe Tabelle 10). Ein negativer Zusammenhang ist an dieser Stelle als positiv zu bewerten, da ein Text als höherwertiger eingeschätzt werden kann, wenn dessen Distanz zum maximalen Qualitätskorrelat möglichst gering ist.



**Abb. 21** Korrelationsplot zwischen Qualität und modellierter Distanz zum höchsten Qualitätskorrelat (inkl. Regressionsgerade)

**Tab. 10** Korrelationen zwischen der Distanz zum höchsten Qualitätskorrelat und Beschreibungen\_%, LOS, Länge und Qualität  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und · bei  $p < .1$ )

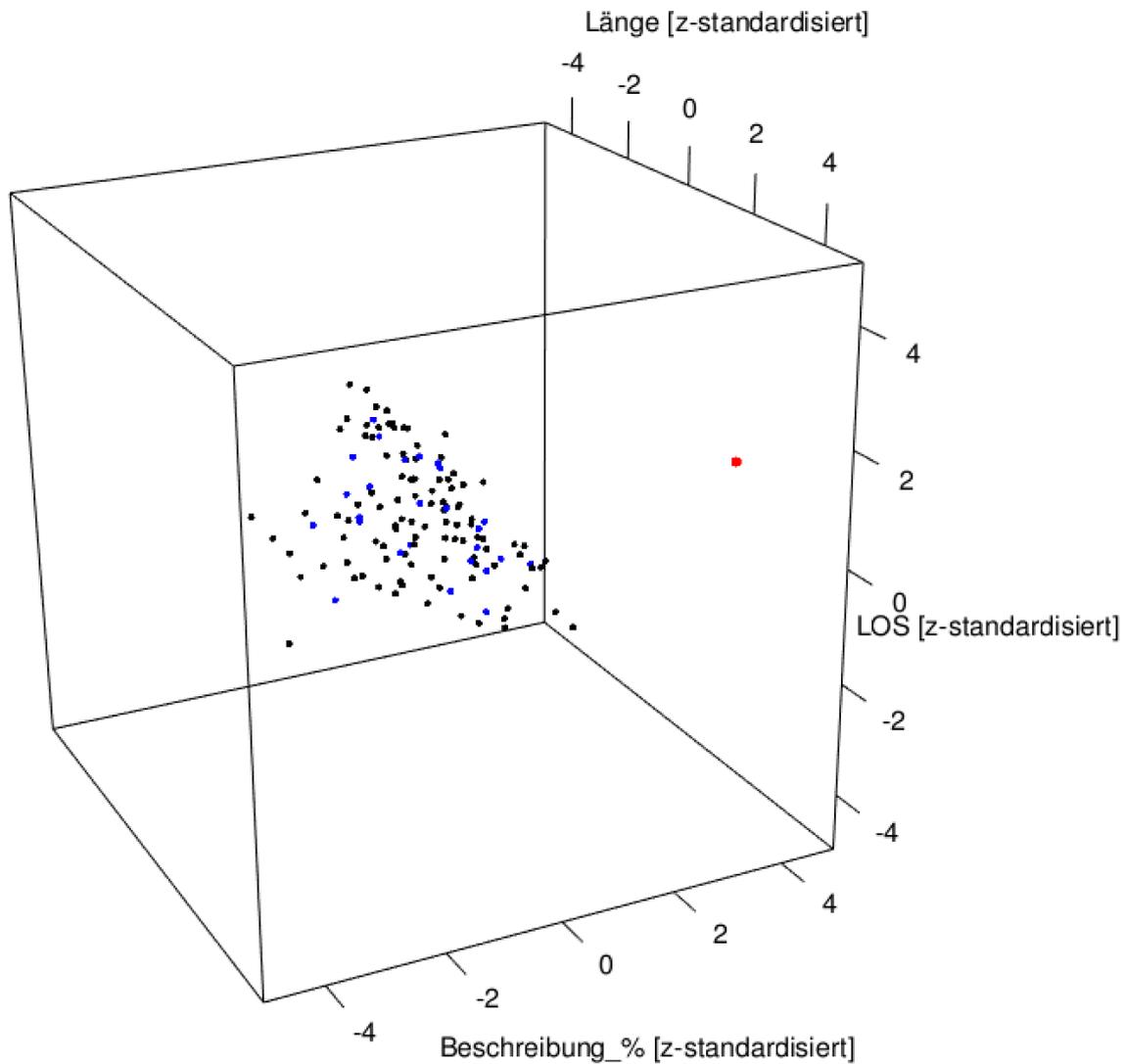
	<i>Beschreibung %</i>	<i>LOS</i>	<i>Länge</i>	<i>Qualität</i> <sup>25</sup>
Distanz zum höchsten Qualitätskorrelat	-0.35***	-0.18·	-0.77***	-0.64***

### 3.2.5 Validierung eines absoluten Qualitätskorrelats

Zur Validierung der modellierten Distanz zum maximalen Qualitätskorrelat und somit zur Beantwortung der explorativen Forschungsfrage 2, wurden nun die Qualitätskorrelate aller Reflexionstexte des Testdatensatzes berechnet. Abbildung 22 gibt einen weiteren detaillierteren Einblick in den konkreten Zusammenhang der Textmaße Länge, Beschreibung\_% und LOS zueinander und ermöglicht ein Verständnis für die Distanz zum maximalen Qualitätskorrelat. Aus der Abbildung kann deutlich hervorgehen, dass keine der einbezogenen Textmaße einen wesentlich größeren Einfluss auf die Distanz ausübt, als ein anderes Textmaß. Die minimale Distanz liegt bei 3.26, wohingegen die maximale Distanz bei 7.95 liegt. Im Mittel liegen die Distanzen der Texte zum maximalen Qualitätskorrelat bei 5.57, bei einer Standardabweichung

<sup>25</sup> Zu beachten ist, dass die Korrelation zwischen der Distanz zum höchsten Qualitätskorrelat und der Qualität ausschließlich auf dem Trainingsdatensatz beruht, da die Qualität nach der Faktorenanalyse nicht mit dem Testdatensatz berechnet wurde.

von .895. In Folge der z-Standardisierung ist auch das Qualitätskorrelat der Distanz normalverteilt.



**Abb. 22** Z-standardisierte Qualitätskorrelate Länge, Beschreibung\_% und LOS aller Fremdreﬂexionen zum höchsten Qualitätskorrelat (roter Punkt), unter Einbezug aller Reﬂexionstexte des Testdatensatzes (blau)

Für eine erste Validierung der Distanz zum maximalen Qualitätskorrelat wurden die, für die in Mientus et al. (2023) für den Expertenvergleich verwendeten Texte, mit der Distanz verglichen. Mit einem Agreement von 70.1% ( $\kappa = .670^{***}$ ) konnten abermals Mensch-Maschine-Übereinstimmungen identifiziert werden, welche den Mensch-Mensch-Übereinstimmungen gleichkommen. Konnte in Mientus et al. (2023b) in der Übereinstimmung von Qualitätseinschätzungen lediglich berichtet werden, dass „Rangsummentests im Instrumentenvergleich [zwischen Mientus et al. (2023) und der Metrik aus Ergebnissen nach Nowak (2023)] vergleichbare Platzierungen [ergaben]“, so können nun im direkten Vergleich zwischen Distanz und der modellierten Metrik aus Mientus et al. (2023) an dieser Stelle Korrelationen zwischen beiden Instrumenten von  $-.18'$  berichtet werden. Zu berücksichtigen

ist, dass Nowak (2023) keine absolute Qualitätsbewertung verfolgt, sondern eine differenzierte und mehrdimensionale Einschätzung in internal und external diskutierte reflexionsbezogene Argumentationsprozesse, welche auf positiven oder negativen Reflexionsauslösern beruhen können.

Um die Aussagekraft der mathematischen Modellierung auf Basis der explorativen Faktorenanalyse zu stützen sind in Tabelle 11 die Qualitätsbewertung sowie deren computerbasiert-bestimmte Distanz zum maximalen Qualitätskorrelat von neun schriftlichen Reflexionen abgedruckt. Ausgewählt wurden je die drei Texte, welche die geringste, mittelste oder größte Distanz aufweisen und somit Texte repräsentieren, welche auf Grundlage der computerbasierten Analyse die höchsten, moderatesten oder geringsten Qualitäten aufweisen.

**Tab. 11** Gegenüberstellung der Qualitätsbewertung zur modellierten Distanz

	<i>Breite</i>	<i>Tiefe</i>	<i>Kohärenz</i>	<i>Spezifität</i>	<i>Distanz</i>
Texte geringster Distanz	4	3	3	3	5.175
	3	3	2	2	5.390
	4	4	4	4	6.031
Texte moderatesten Distanz	2	3	3	3	6.727
	2	1	2	2	6.900
	2	2	3	3	6.913
Texte größter Distanz	2	1	1	1	8.441
	2	2	2	2	8.442
	2	1	1	2	8.501

Tabelle 11 gibt deutliche Indizien dafür, dass auch für die schriftlichen Fremdrelexionen des Testdatensatzes die Distanz zum maximalen Qualitätskorrelat aussagekräftig die konventionell bestimmten Qualitätsbewertungen abschätzen kann. Wenn auch keine signifikante Korrelation zwischen der Distanz zum maximalen Qualitätskorrelat und der Aufsummierung der Qualitätsbewertungen (als ad hoc Näherung), konnte in Rangsummentests die Nullhypothese verworfen werden ( $p = .134$ ), sodass anzunehmen ist, dass beide Verfahren vergleichbare Bewertungen darstellen.

Aus diesem Grund können für die weiteren Betrachtungen die Distanz zum maximalen Qualitätskorrelat, welches auf der Grundlage validierter Analyse- und Bewertungsverfahren entwickelt wurde, als repräsentierend angesehen werden, wengleich das vorgeschlagene Konstrukt perspektivisch zusätzlich einer externen Validierung bedarf.

### 3.3 *Skalen der Amplifiers & Filters (Untersuchungsgegenstand 2)*

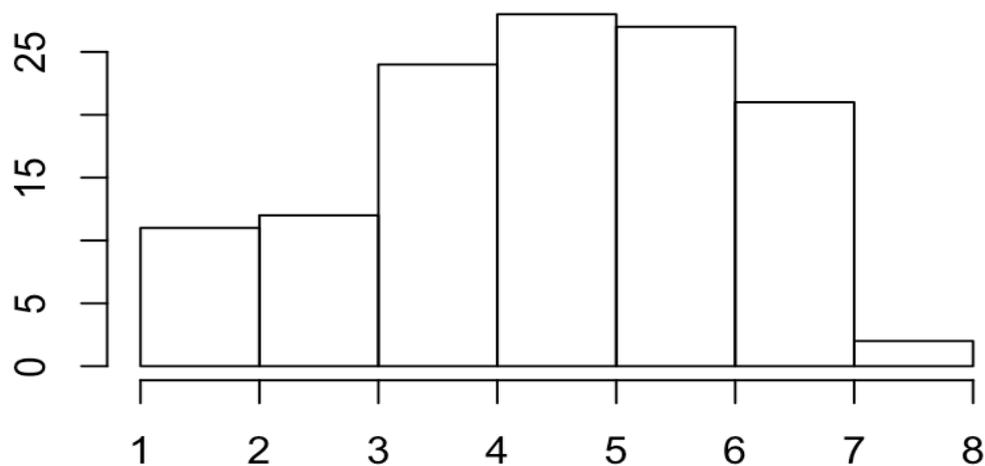
Nachdem im vorherigen Abschnitt Ergebnisse präsentiert wurden, welche auf verschiedene Merkmale des Abbildes eines reflexionsbezogenen Argumentationsprozesses hindeuten, werden im Folgenden ausgewählte Dispositionen untersucht, welche im Wechselspiel zwischen Kompetenz und Performanz als Amplifiers & Filters zwischen einem reflexionsbezogenen personal PCK (der Reflexivität) und einem reflexionsbezogenen enacted PCK (der Reflexion) abzielen. Analog zum Untersuchungsgegenstand 1, den Reflexionstexten, werden erst nach der rein deskriptiven Statistik Zusammenhänge und Bezüge zu bisherigen Ergebnissen hergestellt. Die Items aller Skalen sind im Anhang A4 einsehbar. Mit diesem Kapitel wird explorativ der Forschungsfrage 3 nachgegangen, welche im weiteren Verlauf dieses Kapitels diskutiert wird.

#### 3.3.1 *Amplifiers & Filters - deskriptiv*

Wie bereits unter 2.2.2 ausgeführt, wurden als Repräsentation der Amplifiers & Filters vier Skalen aufgenommen. Die 60 Items zur wahrgenommenen Unterrichtsqualität (wUQ), welche nach Schlesinger et al. (2018) in übersetzter Form in die Untersuchung eingeflossen sind bilden mit akzeptablen Konsistenzen vergleichbar mit der Originalquelle die Subskalen ab. Bei Schlesinger et al. (2018) werden für das Classroom Management  $\alpha = .86$ , für den Student Support  $\alpha = .71$ , für die Cognitive Activation  $\alpha = .82$ , bei der Subject-related Quality  $\alpha = .77$  und bei der Teaching-related Quality  $\alpha = .69$  angegeben. Aus Lohse-Bossenz et al. (2018) wurden alle 13 Items zur Erhebung der reflexionsbezogenen Selbstwirksamkeit (rSWK) aufgenommen, aus welchen sich mit einem Cronbachs'  $\alpha$  von  $.70$  und  $N = 143$  Teilnehmenden eine etwas weniger konsistente Skala ergibt, als in der Originalerhebung (mit  $\alpha = .87$  bei  $N = 746$ ). Aus testtheoretischen Gründen fallen die zugehörigen Werte der Subskalen der reflexionsbezogenen Selbstwirksamkeit schwächer aus (siehe Tabelle 12). Die drei selbst-entwickelten Items zur Werthaltung von Reflexion weisen eine gute interne Konsistenz auf und sind somit in der Lage die Werthaltung der Teilnehmenden gegenüber der durchgeführten Art der Reflexion abzubilden. Das Fachwissen wurde aus sechs Items aus dem Mechanics Baseline Test (MBT) aus Cardamone (2011) zusammengestellt und um zwei selbst entwickelte Aufgaben ergänzt. Wenngleich sich diese Skala mit einem Cronbachs'  $\alpha$  von  $.51$  als weniger konsistent erweist, wird die Skala weiterhin berücksichtigt, da sich Fachwissen häufig als Prädiktor für professionsbezogene Leistungen darstellt. Zusätzlich konnte jedes Item von mindestens einer Person erfolgreich bearbeitet werden und neben der gegebenen statistischen Normalverteilung (siehe Anhang A3) sind lediglich latente Bodeneffekte identifizierbar (siehe Abbildung 23).

**Tab. 12** Deskriptive Statistik der erhobenen Skalen (inklusive der Abweichung des Mean vom Statistischen Mittelwert einer Skala bzw. zur Skalenmitte bei Subskalen)

<i>Skala</i>	<i>#Item</i>	<i>Alpha</i>	<i>min</i>	<i>Mean</i> <i>(Abweichung)</i>	<i>sd</i>	<i>max</i>	<i>NA</i>
wUQ	60	.94	2.16	3.24 (+0.74)	0.50	4.54	3
wUQ_Classroom Management	13	.79	2.20	2.73 (-0.51)	0.58	4.67	4
wUQ_Student Support	16	.84	1.71	2.93 (-0.31)	0.64	4.80	3
wUQ_Cognitive Activation	15	.79	1.86	3.38 (+0.14)	0.60	4.60	6
wUQ_subject-related Quality	12	.77	2.00	3.19 (-0.05)	0.52	4.63	7
wUQ_teaching-related Quality	4	.61	1.00	2.59 (-0.35)	0.83	5.00	7
rSWK	13	.70	2.00	3.44 (+0.94)	0.43	4.33	5
rSWK_Beschreiben	3	.45	1.67	3.55 (+0.11)	0.62	5.00	5
rSWK_Bewerten	3	.49	1.67	3.41 (-0.03)	0.57	4.67	5
rSWK_Alternativen	2	.65	1.00	3.61 (+0.17)	0.75	5.00	5
rSWK_Konsequenz	2	.30	1.50	3.38 (-0.06)	0.68	5.00	5
rSWK_Begründung	3	.60	1.50	3.28 (-0.16)	0.73	4.50	5
Wert	4	.80	2.75	4.34 (+1.84)	0.55	5.00	6
Fachwissen	8	.51	1.00	4.92 (+0.92)	1.62	8.00	9



**Abb. 23** Histogramm der Fachwissensskala mit Indiz für Ausreißeritem

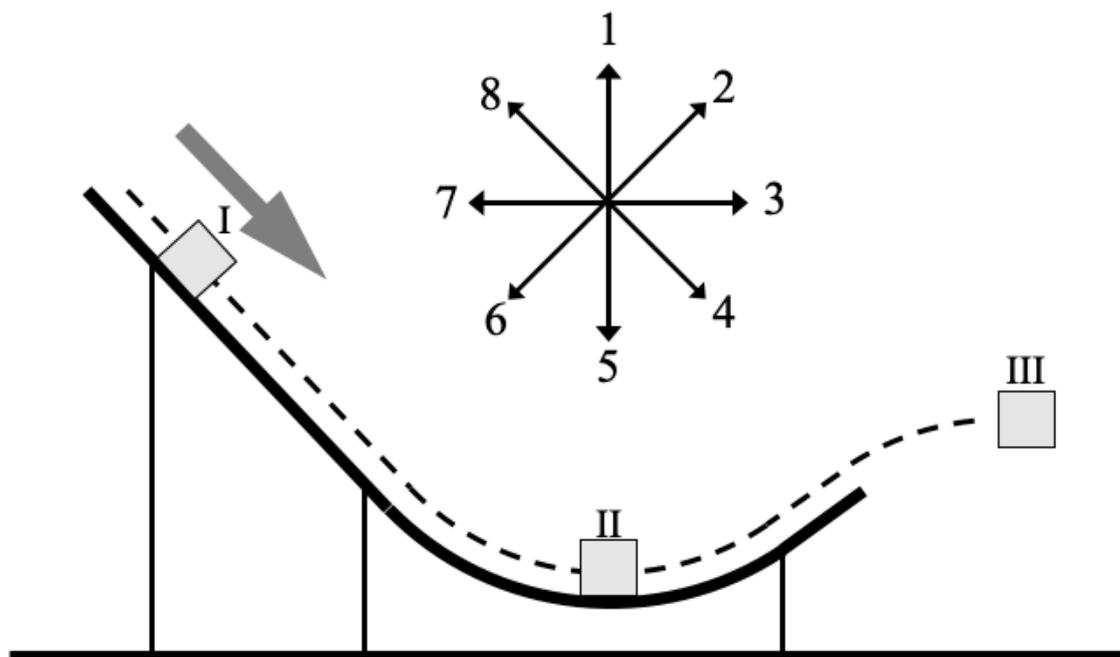
Um eine mögliche Inkonsistenz der Fachwissensitems besser nachvollziehen zu können, zeigt Tabelle 13 die Lösungshäufigkeiten der einzelnen Aufgaben gegenübergestellt der absoluten Bearbeitungshäufigkeit.

**Tab. 13** Lösungen einzelner Fachwissensitems  
( $\circ$  mindestens zwei Personen gaben durchgehend „weiß nicht an“)

Item	N =	„Weiß ich nicht“	Falsche Lösung	Richtige Lösung
Stroboskop 1	127	3 $^\circ$	34	90 (70.87%)
Stroboskop 2	127	4 $^\circ$	44	79 (62.20%)
Klotz 1	127	5 $^\circ$	47	75 (59.06%)
<b>Klotz 2</b>	<b>127</b>	<b>8<math>^\circ</math></b>	<b>117</b>	<b>2 (1,57%)</b>
Klotz 3	127	5 $^\circ$	21	101 (79.53%)
SenkWurf	126	8 $^\circ$	56	62 (49.21%)
FreiFall 1	126	3 $^\circ$	11	112 (88.89%)
FreiFall 2	126	16 $^\circ$	16	94 (74.60%)

Beziehen Sie sich auf die untere Grafik, wenn Sie die nächsten drei Fragen beantworten.

Die Grafik stellt einen Klotz dar, der auf einer reibungsfreien Rampe gleitet. Die acht nummerierten Pfeile repräsentieren die Richtungen, auf die in den Antworten zu den Fragen Bezug genommen wird.



Durch welchen Pfeil wird die Richtung der Beschleunigung des Klotzes in Position II am besten repräsentiert?

„1“	„3“	„5“	„7“	„Keiner der Pfeile, die Beschleunigung ist Null“	„Weiß ich nicht“
2	18	42	0	56	8

**Abb. 24** Fachwissensitem „Klotz 3“ und zugehöriges Antwortverhalten (Richtig ist Antwort „1“)

Deutlich zu erkennen ist, dass die Aufgabe „Klotz 2“ mit einem Antwortverhalten von 2 richtigen Lösungen zu 117 falschen Lösungen ein Ausreißeritem darstellen kann. Zum besseren Verständnis ist dieses Item in Abbildung 24 abgedruckt. Die häufigsten Antworten entfallen mit 56 von 126 auf die Aussagen, dass der Klotz in reibungsfreier Bewegung an Position II

keine Beschleunigung erfährt. Hierbei werden bekannte Fehlvorstellungen oder -Konzepte abgebildet. Wenngleich die Antwort „Pfeil 1“ die richtige ist, wird lediglich die falsche Antwort „Pfeil 7“, welche auf einer nicht-reibungsfreien Rampe eine Folge der Gleitreibung repräsentieren kann, seltener adressiert. Die falsche Antwort „Pfeil 3“ adressiert die Verwechslung von Beschleunigungs- und Bewegungsrichtung, Antwort „Pfeil 5“ impliziert eine Gravitationskraft, bzw. unter Vernachlässigung der Rampe eine Erdbeschleunigung. Am häufigsten wählen die Teilnehmenden die Antwort, dass die Beschleunigung gleich Null sei und offenbaren damit die Fehlvorstellung, dass eine Beschleunigung eine „Tempoänderung“ sei (Schecker & Wilhelm, 2018, S.66) und vernachlässigen hiermit vektorielle Effekte der Beschleunigung.

Wenngleich sich das Item „Klotz 2“ eher als Item darstellt, welches offenbar in der Lage ist auch bei angehenden Physiklehrkräften bekannte Fehlvorstellungen hervorzurufen, würde ein Auslassen dieses Items die Konsistenz der Skala Fachwissen nicht markant ändern ( $\alpha = .50$ ). Aus diesem Grund werden alle acht Items für die weiteren Analysen der reflexionsbezogenen Dispositionen in der Fachwissensskala berücksichtigt. Da in den übrigen Skalen keine Ausreißer identifiziert werden konnten, können die Ergebnisse der validen Skalen zueinander und zu den Textmaßen in Verbindung gesetzt werden.

### *3.3.2 Amplifiers & Filters und Textmaße*

In der Gegenüberstellung der Amplifiers & Filters Fachwissen, Werthaltung und wahrgenommene Unterrichtsqualität wird deutlich, dass die genannten drei Dispositionen nicht korrelieren und so in der Tat als zueinander unabhängige Konstrukte diskutiert werden können (siehe Tabelle 14). Die reflexionsbezogene Selbstwirksamkeit kann in der vorliegenden Untersuchung in einen stärkeren Zusammenhang zu anderen Skalen gesetzt werden. So korrelieren Werthaltung und Selbstwirksamkeit signifikant positiv ( $p = .001$ ). Für die wahrgenommene Unterrichtsqualität und die reflexionsbezogene Selbstwirksamkeit bestehen ebenfalls signifikante, positive Zusammenhänge ( $p = .002$ ). Mit der Fachwissensskala sind keine ausgeprägten Korrelationen identifizierbar. Auch bleibt der LOS nach Mientus et al. (2023) weiterhin unabhängig von den Amplifiers & Filters. Erwartungsgemäß korreliert die Werthaltung, die dem Reflektieren zugeschrieben wird mit der Textlänge der schriftlichen Reflexion.

**Tab. 14** Korrelationen reflexionsbezogener Amplifiers & Filters mit ausgewählten Qualitätskorrelaten  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$ )

	<i>Fachwissen</i>	<i>Wert</i>	<i>wUQ</i>	<i>rSWK</i>
LOS	.10	-.03	.11	.05
Länge	.09	.22**	-.13	.08
Distanz	-.09	-.12	.15 $\cdot$	-.01
Fachwissen	1	-.00	.01	-.09
Wert		1	-.08	.32***
wUQ			1	.28**
rSWK				1

Da unter den ausgewählten Amplifiers & Filters in verschiedenen Erhebungskontexten Fachwissen und Selbstwirksamkeit meisthin in Verbindung zu anderen Skalen stehen und an dieser Stelle keine Zusammenhänge zur Textlänge oder zum LOS identifiziert werden konnten, kann es gewinnbringend sein, die Subskalen der reflexionsbezogenen Selbstwirksamkeit nach Lohse-Bossenz et al. (2018) zur (1) Fähigkeit wahrgenommene Situationen zu beschreiben, (2) Bewertungen vorzunehmen, (3) Alternativen zu formulieren, (4) Konsequenzen für die eigene Professionalisierung abzuleiten sowie (5) zugehörige reflexionsbezogene Argumentationsprozesse zu begründen, beiden Textmerkmalen gegenüberzustellen. Diese fünf Subskalen werden in Tabelle 15 den Textlängen und den absoluten Häufigkeiten der Diskurselemente nach Nowak et al. (2019) sowie dem LOS und den relativen Häufigkeiten der Elemente gegenübergestellt. Auffällig ist, dass kaum starke Korrelationen und keine berichtenswerten Signifikanzen erreicht werden. Lediglich die Selbstwirksamkeit in Bezug auf das Formulieren von Alternativen korreliert signifikant positiv mit der absoluten Häufigkeit an formulierten Alternativen. Weiter wurde die Gegenüberstellung der formulierten Diskurselemente den Amplifiers & Filters gegenübergestellt. Die ermittelten Korrelationen sind in Tabelle 16 dargestellt.

**Tab. 15** Subskalen der rSWK nach Lohse-Bossenz et al. (2019) bezogen auf Elemente (Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$ )

	<i>Beschreiben</i>	<i>Bewerten</i>	<i>Alternativen</i>	<i>Konsequenzen</i>	<i>Begründen</i>
Länge	.06 ( $p = .513$ )	.15 ( $p = .104$ )	.14 ( $p = .118$ )	.03 ( $p = .717$ )	.06 ( $p = .489$ )
Beschreibung	-.01 ( $p = .892$ )	.06 ( $p = .538$ )	.05 ( $p = .592$ )	-.09 ( $p = .336$ )	.03 ( $p = .717$ )
Bewertung	.00 ( $p = .987$ )	.14 ( $p = .115$ )	.13 ( $p = .135$ )	.10 ( $p = .251$ )	.11 ( $p = .235$ )
Alternativen	-.03 ( $p = .777$ )	.17 $\cdot$ ( $p = .051$ )	<b>.21*</b> ( $p = .020$ )	.14 ( $p = .121$ )	-.01 ( $p = .925$ )
Konsequenzen	.13 ( $p = .156$ )	.10 ( $p = .242$ )	.05 ( $p = .575$ )	<b>.05</b> ( $p = .616$ )	.10 ( $p = .243$ )
LOS	.01 ( $p = .930$ )	.03 ( $p = .721$ )	.02 ( $p = .814$ )	.14 ( $p = .109$ )	.01 ( $p = .953$ )
Beschreibung_%	-.08 ( $p = .377$ )	-.11 ( $p = .210$ )	-.12 ( $p = .191$ )	-.16 $\cdot$ ( $p = .074$ )	-.08 ( $p = .353$ )
Bewertung_%	.06 ( $p = .526$ )	<b>.05</b> ( $p = .611$ )	.08 ( $p = .389$ )	.11 ( $p = .238$ )	.11 ( $p = .229$ )
Alternativen_%	-.03 ( $p = .713$ )	.15 $\cdot$ ( $p = .086$ )	.13 ( $p = .133$ )	.17 $\cdot$ ( $p = .064$ )	-.07 ( $p = .454$ )
Konsequenzen_%	.16 $\cdot$ ( $p = .074$ )	.05 ( $p = .597$ )	.01 ( $p = .886$ )	<b>.04</b> ( $p = .644$ )	.10 ( $p = .279$ )

**Tab. 16** Korrelationen zwischen absoluten bzw. relative Elementhäufigkeiten und den Amplifiers & Filters (Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$ )

<i>Absolut:</i>	<i>Beschreibung</i>	<i>Bewertung</i>	<i>Alternativen</i>	<i>Konsequenzen</i>
<i>Fachwissen</i>	.06	.05	.04	.03
<i>Werthaltung</i>	.12	<b>.28**</b>	.13	<b>.21*</b>
<i>wUQ</i>	<b>-.19*</b>	-.10	.00	.05
<i>rSWK</i>	-.02	.04	.10	.14
<i>Relativ:</i>	<i>Beschreibung</i>	<i>Bewertung</i>	<i>Alternativen</i>	<i>Konsequenzen</i>
<i>Fachwissen</i>	-.01	.04	-.04	-.00
<i>Werthaltung</i>	-.07	.13	-.11	.05
<i>wUQ</i>	-.17 $\cdot$	.05	<b>.17*</b>	<b>.19*</b>
<i>rSWK</i>	-.15 $\cdot$	.08	.11	.13

Wie aus Tabelle 16 hervor geht, steht eine hohe Werthaltung von Reflexion erwartbar mit dem Umfang einer Reflexion und dem Umfang der einzelnen Elemente und auch der Länge eines Reflexionstextes in Verbindung (teils signifikant positive Korrelationen zwischen Werthaltung und allen absoluten Häufigkeiten der Diskurselemente). Mit hoher Werthaltung für die erlebte Methode der strukturierten Reflexion scheinen Studierende gut in der Lage zu sein, signifikant mehr Bewertungen zu formulieren und mutmaßlich Bezüge zwischen Beobachteten (enacted PCK) und dem eigenen Wissen bzw. den eigenen Erfahrungen (personal PCK) herzustellen. Ebenso scheint der gezielte Rückbezug auf die eigene Professionalisierung in Form des signifikant positiven Zusammenhangs zur absoluten Häufigkeit an Konsequenzen

identifizierbar zu sein. Wird der beobachtete Unterricht als qualitativ höherwertiger bewertet, fällt nicht nur die Beschreibung im Text selbst kürzer aus, sondern auch der prozentuale Anteil an formulierten Beschreibungen. Entgegengesetzt werden Alternativen und Konsequenzen mit höherer wahrgenommener Unterrichtsqualität prozentual häufiger im Text adressiert. In den Reflexionstexten kann sich dieses Ergebnis beispielsweise in extremen und weniger differenzierten Bewertungen der Gesamtsituation zeigen. Wird in einem Reflexionstext die Situation durchgehend positiv bewertet, so kann die Beschreibung kurz und die formulierten Konsequenzen umfangreich ausfallen, warum es z. B. gut so ist, zu unterrichten und dass als Konsequenz derartig positiv wahrgenommenen Handlungsweisen in die eigene Profession übernommen werden sollen. Gleichzeitig kann die Unterrichtssituation unter einer kurzen Beschreibung konsequent negativ bewertet werden und anschließend eine Vielzahl von Alternativen diskutiert werden. Die Skala der reflexionsbezogenen Selbstwirksamkeit deutet auf einen lediglich latenten Zusammenhang hin, wenngleich mit hoher Selbstwirksamkeit prozentual und absolut mehr Konsequenzen formuliert werden. Nach dem derzeitigen Einblick scheint eine hohe reflexionsbezogene Selbstwirksamkeit mit einem weniger deskriptiven Reflexionstext in Verbindung zu stehen (siehe Korrelation von  $-.15'$  bei Gegenüberstellung von  $r_{\text{SWK}}$  und  $\text{Beschreibung}_{\%}$ ). Das Fachwissen weist auf eher untergeordnete Einflüsse hin.

### *3.3.3 Amplifiers & Filters und Qualitätsabschätzung*

Neben den erhobenen Textmerkmalen und Qualitätsbewertungen (aus 3.2.3) wurde unter 3.2.4 diskutiert, aus welchen Gründen für den vorliegenden Datensatz eine explorative Faktorenanalyse gewinnbringend sein kann. Weiter konnten die drei Faktoren Qualität, Quantität und Deskriptivität bestimmt werden, welche die erhobenen Textmaße repräsentieren (Forschungsfrage 1). Ergänzend konnten ausgewählte Textmerkmale, welche bereits computerbasiert analysierbar sind und die Faktoren Quantität und Deskriptivität repräsentieren können, zu einem maximalen Qualitätskorrelat modelliert werden. Dieses maximale Qualitätskorrelat entspricht einem fiktiven Reflexionstext mit einer Textlänge von 1574 Worten, einem relativen Anteil von 53% Beschreibungen der Unterrichtssituation und einem Level of Structure von .3144. Ferner kann für jeden Reflexionstext eine Distanz zum maximalen Qualitätskorrelat berechnet werden, welche die Qualität einer schriftlichen Fremdrelexion in der vorliegenden Videovignette abschätzen kann (Forschungsfrage 2). Abschließend soll nun mit der Forschungsfrage 3 untersucht werden, in welchen Zusammenhängen die computerbasierte Qualitätsabschätzung zu den Amplifiers & Filters sowie zugehörige Subskalen steht. Aus diesem Grund sind in Tabelle 18 die Faktoren der

Qualität, Quantität und Deskriptivität sowie die ermittelte Distanz zum maximalen Qualitätskorrelat den Amplifiers & Filters gegenübergestellt.

**Tab. 17** Faktorenbestimmung nach ortho und obli in Korr. zu Amplifiers & Filters  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und · bei  $p < .1$ )

	<i>Fachwissen</i>	<i>Wert</i>	<i>wUQ</i>	<i>rSWK</i>
Qualität	-.02 ( $p = .803$ )	<b>.31</b> <sup>***</sup> ( $p = .0003$ )	-.07 ( $p = .435$ )	.05 ( $p = .592$ )
Quantität	.11 ( $p = .237$ )	<b>.18</b> <sup>*</sup> ( $p = .043$ )	-.08 ( $p = .384$ )	.13 ( $p = .129$ )
Deskriptivität	.01 ( $p = .956$ )	-.03 ( $p = .712$ )	<b>-.20</b> <sup>*</sup> ( $p = .020$ )	-.14 ( $p = .103$ )
Distanz	-.09 ( $p = .303$ )	-.12 ( $p = .179$ )	<b>.15</b> <sup>·</sup> ( $p = .092$ )	-.01 ( $p = .903$ )

In Tabelle 17 spiegelt sich der geringe Zusammenhang des Fachwissens mit den Faktoren wider. Das Fachwissen der Teilnehmenden scheint in keinem Zusammenhang mit der Qualität, Quantität, Deskriptivität oder gar Distanz zu stehen. Texte von Teilnehmenden, die reflexionsbezogene Argumentationsprozesse für wertvoll erachten, stellen sich als qualitativ hochwertiger dar. Hierbei zeigt sich der in Abbildung 16 identifizierte stark positive Zusammenhang zwischen Qualität und Quantität der Texte, da eine hohe Werthaltung weiterhin (siehe Tabelle 17) positiv mit beiden Faktoren korrelieren. Nehmen Teilnehmende die Unterrichtsqualität des Vignettenvideos als eher hoch wahr, fallen schriftliche Reflexionen signifikant weniger deskriptiv aus. Das kann bedeuten, dass gemäß der ermittelten Textstruktur eher diskursiv mit dem Reflexionsgegenstand umgegangen wurde. Gleichzeitig scheint eine stärker ausgeprägte Deskriptivität von Nöten zu sein, wenn die Qualität des Unterrichts schlechter wahrgenommen wird. Um zugehörige Effekte potentiell noch detaillierter auflösen zu können, werden in Tabelle 19 die Korrelationen zwischen den Faktoren und den Subskalen der wahrgenommenen Unterrichtsqualität gegenübergestellt. Ebenso werden in Tabelle 18 Zusammenhänge der Subskalen der reflexionsbezogenen Selbstwirksamkeit aufgelöst, da an dieser Stelle keine signifikanten Zusammenhänge identifizierbar sind. Auf diese Weise soll der wechselhafte Einfluss der reflexionsbezogenen Selbstwirksamkeit (in Tabelle 14 eher starke Zusammenhänge zu den übrigen Amplifiers & Filters und in den Tabellen 14 und 16 eher geringe Zusammenhänge zu den erhobenen Textmaßen) überprüft werden.

**Tab. 18** Subskalen der wahrgenommenen Unterrichtsqualität  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und  $\cdot$  bei  $p < .1$ )

	<i>Classroom Management</i>	<i>Student Support</i>	<i>Cognitive Activation</i>	<i>Subject-related Quality</i>	<i>Teaching-related Quality</i>
<i>Qualität</i>	.01 ( $p = .950$ )	-.19 ( $p = .279$ )	-.09 ( $p = .569$ )	-.05 ( $p = .781$ )	.08 ( $p = .443$ )
<i>Quantität</i>	-.10 ( $p = .578$ )	-.18 ( $p = .300$ )	.08 ( $p = .715$ )	.07 ( $p = .701$ )	.11 ( $p = .289$ )
<i>Deskriptivität</i>	-.26 ( $p = .143$ )	<b>-.40*</b> ( $p = .021$ )	-.29 ( $p = .059$ )	-.19 ( $p = .280$ )	<b>-.26**</b> ( $p = .009$ )
<i>Distanz</i>	.21 ( $p = .226$ )	<b>.38*</b> ( $p = .025$ )	.26 ( $p = .087$ )	.24 ( $p = .164$ )	.07 ( $p = .510$ )

Die signifikant negativen Korrelationen zwischen Deskriptivität und wahrgenommener Unterrichtsqualität stellen sich am deutlichsten auf der Subskala des „Student Support“ und der „Teaching-related Quality“ dar. Vergegenwärtigt man sich die zugrundeliegende Unterrichtsvideographie, welche maßgeblich ein lehrergeleitetes Unterrichtsgespräch repräsentierte, erscheint der direkte Zusammenhang zu den eher auf die Lehrkraft fokussierenden Subskalen gerechtfertigt. Mutmaßlich infolge des Fragend-entwickelnden Unterrichtes, scheint ein geringes Maß an Deskriptivität tendenziell mit positiv bewerteten Wahrnehmungen von der Interaktion zwischen Lehrkraft und Schüler\*innen geprägt zu sein. Eine negativer wahrgenommene Lehrkraft-Schüler\*innen-Interaktion steht mit einer erhöhten Deskriptivität in Verbindung, da für diese Bewertung mutmaßlich umfangreichere Beschreibungen oder eine weniger konsistente Reflexionsstruktur (repräsentiert durch den LOS) verbunden ist. Weiter hervorgehoben werden soll an dieser Stelle die Subskala der „Subject-related Quality“, welche gegenüber den anderen Subskalen zu allen drei Faktoren den geringsten Zusammenhang aufweist. Wenngleich die subject-related Quality eher als neutral bewertet wurde (siehe Tabelle 12), stellt sie sich als am wenigsten relevant für das Formulieren der schriftlichen Reflexion dar. Dies kann verschiedene Ursachen haben. Neben der bereits angesprochenen Fokussierung der Teilnehmenden auf die Lehrkraft und dessen Interaktion mit den Schüler\*innen, kann beispielsweise die videographierte Unterrichtssituation an sich weniger Anlässe für fachbezogene Reflexion bieten. Der nur schwach ausgeprägte Zusammenhang zwischen allen diskutierten Dimensionen und dem Fachwissen, stellt sich mit dieser Erkenntnis mutmaßlich als Testeffekt dieser Videovignette dar und bliebe mit einer anderen Unterrichtssituation zu prüfen.

**Tab. 19** Subskalen der reflexionsbezogenen Selbstwirksamkeit  
(Kennzeichnung der Signifikanz: \*\*\* bei  $p < .001$ , \*\* bei  $p < .01$ , \* bei  $p < .05$  und · bei  $p < .1$ )

	<i>Beschreiben</i>	<i>Bewerten</i>	<i>Alternativen</i>	<i>Konsequenzen</i>	<i>Begründen</i>
<i>Qualität</i>	-.04 ( $p = .701$ )	.16· ( $p = .067$ )	.10 ( $p = .265$ )	.11 ( $p = .224$ )	.06 ( $p = .489$ )
<i>Quantität</i>	.09 ( $p = .334$ )	<b>.18*</b> ( $p = .046$ )	<b>.18*</b> ( $p = .043$ )	.07 ( $p = .443$ )	.07 ( $p = .408$ )
<i>Deskriptivität</i>	-.07 ( $p = .465$ )	-.11 ( $p = .238$ )	-.10 ( $p = .243$ )	<b>-.16*</b> ( $p = .047$ )	-.07 ( $p = .451$ )
<i>Distanz</i>	-.01 ( $p = .905$ )	-.02 ( $p = .782$ )	-.05 ( $p = .613$ )	.01 ( $p = .995$ )	-.03 ( $p = .736$ )

Nachdem in der groben Gegenüberstellung der ermittelten Faktoren mit der reflexionsbezogenen Selbstwirksamkeit nur geringfügige Zusammenhänge identifiziert werden konnten (Tabelle 18), ist es unter der genaueren Betrachtung der Subskalen doch leistbar, einzelne Zusammenhänge trennschärfer aufzulösen. Zugehörige Korrelationen sind in Tabelle 19 abgedruckt. Dank der Subskalen kann gesagt werden, dass Selbstwirksamkeit und die Faktoren Qualität und Quantität insbesondere in den Subskalen der Bewertungen und Alternativen im signifikant positiven Zusammenhang stehen. Der Umfang, sowie die Qualitätsdimensionen Breite, Tiefe, Kohärenz und Spezifität fallen also tendenziell besser aus, wenn Teilnehmende erwarten, Bewertungen zu einer Unterrichtswahrnehmung festzuhalten und alternative Vorgehensweisen aufzuzeigen. Diese Befunde sind interessant, da die Subskalen der reflexionsbezogenen Selbstwirksamkeit in unserer Stichprobe nicht direkt mit den zugehörigen Diskurselementen des Rahmenmodells für Reflexion nach Nowak et al. (2019) zu korrelieren scheinen (siehe Tabelle 15). Weiter geht aus Tabelle 19 hervor, dass Teilnehmende, die eher erwarten Konsequenzen für die eigene Professionalisierung ableiten zu können, weniger deskriptive Reflexionstexte verfassen oder allgemeiner eine deskriptivere Reflexion mit einer eher geringeren Selbstwirksamkeitserwartung in Bezug auf Konsequenzen in Verbindung steht.

#### ***4 Zusammenfassung, Limitation und Diskussion***

Die professionelle Entwicklung von Lehrkräften unterliegt einer Vielzahl von Einflussfaktoren. Ebenso divers, wie der Forschungsdiskurs im Bereich der professionellen Entwicklung, stellt sich auch die Reflexionsforschung als vielschichtig dar. Im Bereich der naturwissenschaftlichen Lehrkräftebildung konnte 2019 das Pedagogical Content Knowledge (PCK) einer Lehrkraft im internationalen Konsens mit dem Refined Consensus Modell (RCM) neu konzeptualisiert werden (Hume et al., 2019). Weiter konnte gezeigt werden, dass mit dem RCM das Ziel erreicht werden konnte, Forschungsergebnisse der Professions- bzw. Professionalisierungsforschung zu vereinheitlichen (Mientus et al., 2022). Zwar wurde es mit dem RCM als Rahmenmodell möglich, empirische Befunde einheitlich zu kommunizieren, dennoch bestehen aktuell wenige belastbare Aussagen zu konkreten Wirkmechanismen. Aus diesem Grund identifizierten Mientus et al. (2022) mit ihrem systematischen Literatur Review die Relevanz, den Forschungsfokus mehr auf ausgewählte Wechselwirkungen zwischen den im RCM beschriebenen Aspekten zu richten.

Als ausgewählter Forschungsfokus, welcher mit dem RCM kommuniziert werden kann, konzentrierte sich diese Arbeit auf den Aspekt der Reflexion im Rahmen des Pedagogical Reasoning unter der Annahme, dass eine Form der grundlegenden Kompetenz zu reflektieren einem reflexionsbezogenen Argumentationsprozess zugrunde liegt. In Anlehnung an von Aufschnaiter et al. (2019) kann die Summe aller abrufbaren reflexionsbezogenen Kompetenzen als Reflexivität bezeichnet werden und im RCM im personal PCK verortet werden (siehe Abbildung 4 im Kapitel 1.2 auf Seite 14). Eine reflexionsbezogene Performanz wird dem enacted PCK zugeordnet und entgegen vielzitatierter Prozessmodelle für Reflexion (z. B. dem ALACT-Modell) vom Act-of-Trial im Sinne einer neuen Unterrichtsplanung in einer neuen Unterrichtssituation separiert. Als ein Teil der reflexionsbezogenen Performanz wurde in dieser Arbeit das Durchlaufen eines reflexionsbezogenen Argumentationsprozesses selbst angesehen. Dieser basiert zwar in Anlehnung an von Aufschnaiter et al. (2019) auf der Reflexivität einer Lehrkraft, wird jedoch in dieser Arbeit als beeinflusst durch reflexionsbezogene Dispositionen diskutiert, welche somit nicht mehr als Teil der Reflexivität betrachtet werden können. Im RCM entsprechen diese Dispositionen den Amplifiers & Filters, welche kontextspezifisch Reflexionen beeinflussen können und somit miterhoben werden sollten, wenn Reflexivität untersucht werden soll. Aus der Verbindung der Konzeptmodelle konnten Argumentationen zum Zusammenhang von Reflexion und Reflexivität aus von Aufschnaiter et al. (2019) in das RCM und somit die internationale Professionalisierungsforschung eingebettet und weiterentwickelt werden. In dieser Arbeit wurden zudem Abbilder reflexionsbezogener

Argumentationsprozesse analysiert und mit ausgewählten reflexionsbezogenen Dispositionen in Verbindung gebracht, um erste empirische Befunde im Kontext des Kompetenzmodells für Reflexivität zu bieten.

Zur Eingrenzung von Amplifiers & Filters und zur Vergleichbarkeit innerhalb der Studie wurden daher schriftliche Fremdrelexionen im Rahmen einer standardisierten Videovignette erhoben. Die Teilnehmenden lasen (1) einen Prompt zum Rahmenmodell für Reflexion nach Nowak et al. (2019), (2) sahen einen 17-minütigen Videoausschnitt einer Physik-Unterrichtsstunde zum Freien Fall, (3) verfassten eine schriftliche Fremdrelexion und (4) beantworteten Fragen zu ausgewählten Amplifiers & Filters. Reflexionstexte als Abbilder reflexionsbezogener Argumentationsprozesse wurden ausgewählt, da sie im Gegensatz zu Interviews geringen extrinsischen Effekten unterliegen, als Freitextantworten Kompetenzen handlungsnah abbilden und unabhängig analysiert werden können (Poldner et al., 2014). Da die Qualität schriftlicher Reflexionen jedoch signifikant mit der Aufsatzlänge zusammenhängen kann (Chodorow & Burstein, 2004; Leonhard & Rihm, 2011), wurde zunächst eine validierte Methode für die vorliegende Stichprobe geprüft.

Zur Beantwortung der Forschungsfrage 1 nach den zugrundeliegenden Faktoren der erhobenen Textmerkmale wurde eine explorative Faktorenanalyse angewandt. Berücksichtigt wurden hierbei die konventionell durch Expert\*innen für schriftliche Unterrichtsreflexionen eingeschätzten Qualitätsbewertungen der Breite, Tiefe, Kohärenz und Spezifität der Fremdrelexionen, sowie ausgewählte computerbasiert analysierbare Textmerkmale. Unter den Textmerkmalen wurden neben der Textlänge (als typisches Qualitätskorrelat) auch der Level of Structure (LOS) aus Mientus et al. (2023) sowie ausgewählte absolute und relative Häufigkeiten der Diskurselemente aus Nowak et al. (2019) berücksichtigt. Diese Textmerkmale wurden unter Verwendung eines vortrainierten Klassifikationsalgorithmus nach Wulff et al. (2022) erhoben. Um Qualitätsaussagen valide quantifizieren zu können wurden die zugrundeliegenden Faktoren Qualität, Quantität und Deskriptivität exploriert. Der Faktor Qualität wird hierbei repräsentiert von den konventionell generierten Qualitätsbewertungen. Der Faktor der Quantität kann durch die Textlänge sowie die absoluten Häufigkeiten der Diskurselemente in einem Reflexionstext repräsentiert werden, wobei der relative Anteil an formulierten Beschreibungen auf den Faktor Deskriptivität entfällt. Ebenfalls auf den Faktor der Deskriptivität lädt der LOS, wobei Deskriptivität und LOS negativ korrelieren. Die Befunde aus Mientus et al. (2023), dass der LOS mit der Qualität einer schriftlichen Reflexion linear zusammenhängt, konnte in dieser Arbeit nicht bestätigt werden.

Da der Faktor Qualität durch alle konventionell eingeschätzten Qualitätsbewertungen repräsentiert wird und die computerbasiert analysierbaren Textmerkmale über die Faktoren Quantität und Deskriptivität indirekt mit dem Faktor Qualität zusammenhängen, wurde unter Forschungsfrage 2 ein maximales Qualitätskorrelat exploriert. Hierbei konnten die signifikant bessere quadratische Näherung der Regression zwischen Qualität und Deskriptivität verwendet werden. Textlänge, LOS sowie der relative Anteil an Beschreibungen im Text können die Qualitätskorrelat der schriftlichen Reflexion ergeben. Mit dessen Distanz zum maximalen Qualitätskorrelat kann die Qualität der schriftlichen Reflexion abgeschätzt werden (Weiteres in Kap. 4.1). Abschließend wurde zur Beantwortung der Forschungsfrage 3 diese Distanz zum maximalen Qualitätskorrelat (sowie die explorierten Faktoren) den Skalen ausgewählter Amplifiers & Filters gegenübergestellt, um Einflussfaktoren zwischen Reflexion und Reflexivität zu untersuchen (genauer unter 4.2).

#### *4.1 Kontextspezifische computerbasierte Qualitätsabschätzung*

Interpretierend kann festgehalten werden, dass der Level of Structure (LOS) aus Mientus et al. (2023) zwar weiterhin brauchbare Qualitätsabschätzungen alleinig auf Basis computerbasierter Strukturanalysen leisten kann, diese jedoch ausschließlich relative Aussagen zur Qualität eines reflexionsbezogenen Argumentationsprozesses treffen können. So konnten beispielsweise signifikant positive Korrelationen zwischen dem LOS und den relativen Häufigkeiten formulierter Alternativen und Konsequenzen (also den höherwertigen Diskurselementen) gefunden werden ohne dass LOS und Textlänge korrelieren. Eine Herausforderung kann hierbei darin liegen, dass der LOS nach Mientus et al. (2023) eine blockweise Formulierung für die Diskurselemente nach Nowak et al. (2019) unterstützt. Nowak (2023) fand jedoch in einer Untersuchung von 132 Selbstreflexionstexten heraus, dass eine blockweise Formulierung der Diskurselemente keinen signifikanten Einfluss auf die Qualitätsmerkmale Breite und Tiefe hat.

In dieser Arbeit wurde daher unter Verwendung einer explorativen Faktorenanalyse (FF 1) eine Metrik entwickelt, welche eine Vielzahl von computerbasiert analysierbaren Textmerkmalen zusammenführt und in Zusammenhang mit Qualitätsmerkmalen steht (FF 2). Die Faktorenanalyse wurde an einer zufälligen Teilstichprobe von 80% der vorliegenden Daten durchgeführt und zugehörige Ergebnisse an einem Testdatensatz von 20% der Reflexionstexte validiert. Konkret wurden die Qualitätsmerkmale Breite, Tiefe, Kohärenz und Spezifität von Experten für die Begutachtung von Reflexionstexten vierstufig bewertet und mit den Textmerkmalen Länge, als typisches Qualitätskorrelat, absolute Häufigkeit aller

Diskurselemente, dem prozentualen Anteil an formulierten Beschreibungen und dem LOS als validierten Repräsentant einer hochwertigen Argumentationsstruktur zusammengeführt.

Wenngleich die konventionell eingeschätzten Qualitätsbewertungen der schriftlichen Fremdreflexionen, wie im Theorieteil dieser Arbeit diskutiert, Gefahr laufen können zu holistischen Gesamtschätzungen zu verfallen, erwies sich die hohe Beurteilerübereinstimmung als Indiz gegen individuelle und subjektive Qualitätseinschätzungen. Um etwaigen Effekten von Gesamtschätzungen entgegenzuwirken wurde sich bei der methodischen Verwendung der Qualitätsbewertungen für Rater 1 entschieden, da hier die Korrelationen zwischen den Qualitätsmerkmalen weniger stark ausgeprägt waren. Auf diese Weise kann davon ausgegangen werden, dass die einzelnen Qualitätsmerkmale differenzierter begutachtet und bewertet werden konnten. Nichtsdestotrotz können subjektive Einstellungen und Überzeugungen der Expert\*innen nicht ausgeschlossen werden. Aus diesem Grund könnte es wertvoll sein, eine gewisse Teilmenge an Reflexionstexten von unabhängigen Expert\*innen anderer Domänen (wie z. B. Erziehungswissenschaften) bewerten zu lassen und beide Bewertungen abzugleichen.

Mittels obliminer Faktorenanalyse wurden die genannten Textmaße drei Faktoren zugeordnet. Diesen wurden die Bedeutung der Qualität, der Quantität und der Deskriptivität zugeschrieben. Unter Verwendung mathematischer Optimierung konnte für die vorliegende Vignettenstudie ein maximales Qualitätskorrelat bestimmt werden. Nach diesem maximalen Qualitätskorrelat müsste im Rahmen dieser Videovignette ein Reflexionstext höchster inhaltlicher Qualität eine Textlänge von 1428 Worten, einen Anteil an Beschreibungen von 54% sowie einen LOS von .3144 aufweisen. Durch die computerbasiert analysierbaren Textmaße der Textlänge, der relativen Häufigkeit an Beschreibungen sowie des LOS, konnte jedem Reflexionstext eine Distanz zum maximalen Qualitätskorrelat automatisiert zugeordnet werden, welches zu Vergleichbaren Qualitätsabschätzungen gelangt, wie weitere validierte Instrumente (z. B. Mientus et al., 2023; Nowak, 2023). Auf diese Weise konnte exemplarisch gezeigt werden, dass es für den vorliegenden Kontext ein maximales Qualitätskorrelat gibt. Da die Qualitätsbewertungen konventionell durch Expert\*innen durchgeführt wurden und diese mit computerbasiert analysierten Textmerkmalen gemeinsam untersucht wurden, sind artifizielle Korrelationen nicht ausschließbar. Auch wurden für die Modellierung des maximalen Qualitätskorrelates in den Zusammenhängen zwischen Qualität und Textlänge, Anteilen an Beschreibungen und LOS quadratische Zusammenhänge angenommen. Dies ist fragwürdig, da eine quadratische Näherung meist eine bessere Regression darstellt, als eine lineare und insbesondere in der Regression zwischen Qualität und LOS große Standardfehler

identifiziert wurden. Gleichzeitig wurden die vorliegenden Analysen inklusive der zugrundeliegenden explorativen Faktorenanalyse lediglich an einer Teilstichprobe von 80% (107 Texte) kalkuliert. Für den Testdatensatz von 27 Texten (20%) konnten positive Zusammenhänge zwischen der Distanz zum maximalen Qualitätskorrelat und den Qualitätsbewertungen der Expert\*innen identifiziert werden.

Zu beachten bleibt, dass ein Fremdreflexionstext, welcher die obig genannten Werte des maximalen Qualitätskorrelates aufweist, keinesfalls in den Qualitätsbewertungen der Expert\*innen höchste Bewertungen erreichen muss, da in der vorgestellten Metrik inhaltliche Aspekte der schriftlichen Reflexionen nur indirekt berücksichtigt wurden. Gleichzeitig erwies sich die Distanz zum maximalen Qualitätskorrelat auch für den Testdatensatz als nützlich. Perspektivisch ist es daher im Rahmen dieser Videovignette nicht mehr zwingend nötig, konventionell durch Expert\*innen oder geschulte Mitarbeitende Qualitätsbewertungen vorzunehmen um eine Qualitätsabschätzung der schriftlichen Fremdreflexionen zu erreichen.

#### *4.2 Zusammenhänge von Amplifiers & Filters zur Reflexionsqualität*

Unter Verwendung der in dieser Arbeit vorgestellten Metrik konnte die Qualität der schriftlichen Reflexionen abgeschätzt und in Beziehung zu ausgewählten Amplifiers & Filters (reflexionsbezogene Dispositionen) gesetzt werden. Als eindeutige Befundlage konnte festgestellt werden, dass Fachwissen keine mit dieser Studie messbaren Korrelation zu einer Reflexion hat. Diese Beobachtung steht den Befunden aus Sorge et al. (2019) in Teilen entgegen und stellt sich auch als Widerspruch zur Befundlage des Zusammenhangs zwischen Professionswissen und Unterrichtsqualität (Fischer et al., 2014, Liepertz & Borowski, 2019) dar. Wenngleich nicht ausgeschlossen werden kann, dass mit der Kontextspezifität dieser Studie ein nur bedingt übertragbarer Befund besteht, scheint eine qualitativ hochwertige Fremdreflexion von Physikunterricht eher unabhängig vom physikalischen Vorwissen zu sein. Für einen Zusammenhang zwischen Reflexion und Fachwissenskontext konnte somit keine Evidenz gefunden werden, was darauf hindeutet, dass Reflexivität als eher kontextunabhängig betrachtet werden kann. Dies unterstützt eher die Forschung, dass sich PCK von Lehrkräften der Naturwissenschaften unabhängig vom Lernkontext entwickeln kann (Berry et al., 2016, Kersting et al., 2010). Offen bleibt an dieser Stelle, inwieweit fachliche Reflexionsauslöser ggf. nicht hinreichend präsent im Vignettenvideo identifizierbar sind. Eine inhaltliche Untersuchung der kollaborativ erstellten schriftlichen Reflexion der drei Professor\*innen (auf welcher Grundlage die Berechnung des LOS ruht) könnte hierbei aufschlussreich sein. Ebenso könnte

es von Wert sein, in der bestehenden Vignette eine fachdidaktische Skala zu ergänzen, um fachliche Inhalte näher am Schulkontext erheben zu können.

Auch scheint die Gesamtskala der reflexionsbezogenen Selbstwirksamkeit keine hier auflösbaren Effekte zu haben, sodass die Zuschreibung positiver Effekte auf eine Vielzahl von unterrichtsrelevanten Kompetenzen (Bach, 2022) nicht zwangsläufig auf Reflexivität bzw. reflexive Argumentation übertragen werden kann. Auf den Subskalen wird jedoch deutlich, dass Teilnehmende, welche nach Selbsteinschätzungen, erfolgreich Bewertungen, Alternativen und Konsequenzen formulieren konnten, längere und weniger deskriptive Texte verfassen. Diese Selbsteinschätzungen stehen dennoch in keinem Zusammenhang zur Qualität der schriftlichen Reflexionen oder zur Distanz zum maximalen Qualitätskorrelat. Alternativ kann an dieser Stelle ebenfalls geschlossen werden, dass es möglich sein kann, durch hinreichende Eingrenzung des Lernkontextes bzw. des Reflexionsanlasses Effekte der reflexionsbezogenen Selbstwirksamkeit ausblenden zu können.

Lediglich der Werthaltung für Reflexion sowie der wahrgenommenen Unterrichtsqualität konnten schwache Korrelationen zur Metrik zugeschrieben werden. So scheint Werthaltung von Reflexion (wie in dieser Studie durchgeführt) in einem positiven Zusammenhang zur Qualität der schriftlichen Reflexion zu stehen, während sich bei der wahrgenommenen Unterrichtsqualität ein entgegengesetzter Effekt zeigt. Texte von Personen, die den beobachteten Unterricht positiv bewerten, formulieren einen tendenziell weniger hochwertigen Reflexionstext. Nowak (2023) konnte in ihrer Untersuchung zu Selbstreflexionen identifizieren, dass positive Reflexionsauslöser tendenziell nur deskriptiv reflektiert werden. Es könnte also erkenntnisreich sein, die entwickelte Werthaltungsskala auch im Kontext von Selbstreflexionen zu erproben und zugehörige Ergebnisse zusammenzuführen.

Zusammenfassend muss also festgehalten werden, dass einzelne Amplifiers & Filters mit entscheidend bei der Untersuchung von Reflexivität als Kompetenz sein können, da reflexionsbezogene Argumentationsprozesse stets nur als Abbild einer zu untersuchenden Reflexivität betrachtet werden dürfen. Unter der Annahme, dass eine Reflexivität im Sinne von Kompetenz im personal PCK besteht, sich ein reflexionsbezogener Argumentationsprozess aber nur enacted in einer Reflexion zeigen kann, konnten mit dieser Arbeit Belege dafür identifiziert werden, dass bei der Untersuchung von reflexionsbezogenen Kompetenzen immer auch Amplifiers & Filters untersucht werden sollten. Umgekehrt können nur unter der Berücksichtigung dieser Amplifiers & Filters auch Rückschlüsse auf Reflexivität gezogen werden. In unserem Fall erweisen sich beispielsweise die Werthaltung sowie die wahrgenommene Unterrichtsqualität als beeinflussend, sodass diese beiden Skalen in weiteren

Untersuchungen von Reflexivität miterhoben werden sollten. An dieser Stelle ist jedoch erneut auf den bewusst ausgewählten Lernkontext und die spezifische Unterrichtsvideographie hinzuweisen. Inwieweit identifizierte Effekte auch auf andere Vignettenstudien übertragbar sind, bleibt an dieser Stelle offen.

Für die Qualität einer schriftlichen Reflexion kann im Kontext dieser Vignettenstudie festhalten werden, dass die Qualität der Reflexion nicht automatisch zunimmt, wenn ein längerer Text formuliert wird oder der Text weniger Deskriptiv erscheint. Die vorliegende Studie kann exemplarisch darstellen, dass für einen konkreten Reflexionsanlass ein optimales Verhältnis aus Quantität und Deskriptivität ermittelt werden kann – und dies computerbasiert.

## ***5 Ausblick und Konklusion***

Auf Basis dieser Erkenntnisse kann ein Mehrwert in drei Bereichen abgeleitet und Potentiale für Lehre und Forschung diskutiert werden. Bereicherungen für die Naturwissenschaftsdidaktik können auf die theoriegeleitete Diskussion, die methodischen Errungenschaften sowie die empirischen Erkenntnisse der vorgestellten Studie entfallen und werden an dieser Stelle für weitere Forschung zur Diskussion gestellt

### *5.1 Theoretischer Mehrwert*

Im Theorieteil dieser Arbeit vorbereitend wurde die Modellierung von Reflexivität als Kompetenz nach von Aufschnaiter et al. (2019) in das Refined Consensus Modell of PCK (Carlson et al., 2019) eingebettet. Dies hatte zur Folge, dass Reflexivität als potentiell kontextunspezifische Kompetenz konkreter von reflexionsbezogenen Dispositionen separiert betrachtet werden kann. Im eingebetteten Modell von Reflexivität als Kompetenz können somit Reflexivität (als Summe aller anwendbaren reflexionsbezogenen Kompetenzen im personal PCK), Reflexion (als reflexionsbezogener Argumentationsprozess im enacted PCK) sowie reflexionsbezogene Dispositionen (als Amplifiers & Filters zwischen ePCK und pPCK) beschrieben werden (siehe Abbildung 4 im Kapitel 1.2 auf Seite 14).

Da ausschließlich Abbilder reflexionsbezogener Argumentationsprozesse analysiert werden können und in bestehender Forschung teilweise von diesen Analysen auf Reflexivität geschlossen wird, können ungewollte Nebeneffekte der Messung von Amplifiers & Filters nicht ausgeschlossen werden. Aus der Theorie heraus erscheint es daher wertvoll, im Zuge einer Erhebung von Reflexion ausgewählte Skalen von Amplifiers & Filters parallel zu erheben. Auf diese Weise könnten mittels Triangulation von Daten valide Rückschlüsse auf die zugrundeliegende Reflexivität von Lehrkräften geschlossen werden. Wichtig zu beachten ist an dieser Stelle, dass in der vorliegenden Studie lediglich einzelne Amplifiers & Filters ausgewählt wurden. Ergänzend könnten beispielsweise auch Einstellungen zur Reflexion (nach Neuber und Göbel, 2020) oder das individuelle Reflexionsverständnis (vgl. Meier et al., 2021) erhoben und berücksichtigt werden.

Wie eingangs skizziert, stellt sich die deutschsprachige Reflexionsforschung begriffsdefinitiv als heterogen und dynamisch dar. Unter der einheitlichen Verwendung der in dieser Arbeit vorgestellten Modellierung von Reflexivität als Kompetenz in Einbettung in das RCM kann eine internationale Kommunikation von Forschungserkenntnissen erleichtert werden. Aus diesem Grund wird an diesem Punkt die Einbettung der eigenen (auch reflexionsbezogenen) empirischen Studien in das RCM eindrücklich empfohlen. Die

entwickelte Konzeption aus Abbildung 4 kann hierbei im Bereich der Reflexion und Reflexivität ein geeignetes Gerüst bieten.

### *5.2 Methodischer Mehrwert*

Im Rahmen der Forschungsfrage 1 nach den zugrundeliegenden Faktoren, welche Aussagen darüber ermöglichen, welche Textmerkmale auf welche Weise zusammenhängen konnten alle erhobenen Informationen über die schriftlichen Reflexionen auf drei Hintergrundfaktoren zurückgeführt werden. Mittels explorativer Faktorenanalyse konnten alle konventionell bewerteten Textmerkmale dem Faktor Qualität zugeordnet werden. Die übrigen computerbasiert bestimmbar Textmerkmale konnten den Faktoren Quantität und Deskriptivität zugeordnet werden. Auf der Grundlage der nicht-linearen Zusammenhänge dieser Größen konnte für die vorliegende Videovignette ein maximales Qualitätskorrelat ermittelt werden. Nach diesem scheint ein Reflexionstext maximaler Qualität mit einer bestimmten Kombination aus den drei Textmerkmalen Textlänge, relativen Anteil an Beschreibungen und dem Level of Structure (LOS) nach Mientus et al. (2023) zu korrelieren. Trotz in Teilen hoher Standardfehler konnte das entworfene Qualitätskorrelat mit einer Teilstichprobe getestet und validiert werden

Da auf Basis der computerbasierten Textanalysen von Wulff et al. (2022) und Mientus et al. (2023) für jeden Reflexionstext der verwendeten Videovignette die drei genannten Textmaße analysiert werden können, kann eine Distanz zum maximalen Qualitätskorrelat ermittelt werden. Diese Distanz stellt entgegen dem LOS aus Mientus et al. (2023) ein absolutes Qualitätskorrelat für diese Vignette dar. Auf dieser Analyse ist die Umsetzung automatisierten Feedbacks möglich. Hierbei könnte nicht nur zurückgemeldet werden, ob ein gelungener Reflexionstext länger, weniger beschreibend oder näher an dem Rahmenmodell für Reflexion nach Nowak et al. (2019) gestaltet sein sollte, sondern auch in welchem konkreten Maße. Konkrete Formulierungen und die praktische Umsetzbarkeit gilt es aktuell zu überprüfen

Verallgemeinernd konnte mit dem modellierten maximalen Qualitätskorrelat ein Beleg dafür gefunden werden, dass eine längere, weniger deskriptive oder eben strukturiertere Reflexion nicht immer einen höherwertigen reflexionsbezogenen Argumentationsprozess repräsentieren muss. Schlussfolgernd soll daher empfohlen werden auch in anderen Anwendungsgebieten von operationalisierten Analysen und ggf. Feedback nach Wegen zu suchen, empirisch fundierte Quantifizierungen vorzunehmen. Wenngleich nicht alle Herausforderungen von Bewertung und Analyse auf der Grundlage maschinellen Lernens oder modellierender Algorithmen gelöst werden können, offenbaren sich ressourcenschonende

Potentiale und Dimensionen für Forschung und Lehre, welche konventionell nicht bestehen könnten.

### *5.3 Empirischer Mehrwert*

Empirisch bestätigt werden konnte der generelle Zusammenhang von Amplifiers & Filters mit Abbildern reflexionsbezogener Argumentationsprozesse. In der vorliegenden Studie konnte so Evidenz dafür gewonnen werden, dass Reflexivität potentiell fachwissensunabhängig erscheint. Zur Diskussion kann daher gestellt werden, inwieweit in der Lehrkräftebildung explizite Reflexionsanlässe fachdidaktisch begleitet werden müssen. An dieser Stelle soll aber eher hinterfragt werden, inwieweit Teilnehmende überhaupt fachliche oder fachdidaktische Aspekte der Unterrichtsvideographie als Reflexionsauslöser oder als Argumente für ihren reflexionsbezogenen Argumentationsprozess verwendeten. Vertiefende inhaltliche Analysen der erhobenen Reflexionstexte stehen zum derzeitigen Zeitpunkt noch aus und sollten in weiteren Studien überprüft werden. Auch wurden zum derzeitigen Zeitpunkt keine Kohortenvergleiche auf Grundlage der erhobenen demographischen Daten durchgeführt. Variierende maximale Qualitätskorrelate zwischen Kohorten aus Bachelorstudierenden, Masterstudierenden oder Referendar\*innen können an dieser Stelle nicht ausgeschlossen werden. Weiter könnten Kohortenvergleiche zwischen den Qualitätsbewertungen oder Extremgruppen des LOS vorgenommen werden, um detailliertere Unterschiede der hier aufgezeigten Korrelationen aufzulösen.

Unter den ausgewählten Amplifiers & Filters der Werthaltung, wahrgenommenen Unterrichtsqualität sowie der reflexionsbezogenen Selbstwirksamkeit wurden differenzierte Zusammenhänge identifiziert. Die zentrale Kernerkenntnis ergibt sich jedoch auf der Metaebene. Der Forschungsfrage 3 nachgehend konnten mit dieser Arbeit deutliche Befunde reflexionsbezogener Dispositionen (bzw. Amplifiers & Filters) gefunden werden. Soll also Reflexivität untersucht werden erscheint es unzureichend, ausschließlich Reflexionen als Abbilder reflexionsbezogener Argumentationsprozesse zu untersuchen, da nicht hinreichend eindeutig ist, welche empirischen Evidenzen gemessen werden. Von großer Relevanz für die Reflexionsforschung scheint es zu sein, weitere Amplifiers & Filters auf deren Zusammenhang zu Qualitätseinschätzungen von Reflexionen hin zu untersuchen. Die in dieser Studie untersuchten Skalen stellen sich als divers und in Teilen (siehe Selbstwirksamkeit) nicht vollständig theoriegemäß zusammenhängend dar. Aus diesem Grund bleibt zum aktuellen Zeitpunkt diffus, welche Amplifiers & Filters den reflexionsbezogenen Übertrag zwischen

personal und enacted PCK zu beeinflussen scheinen. Erste Indizien konnten mit dieser Arbeit gefunden werden.

Abschließend gilt es weiterhin den Übertrag in die schulische Praxis im Bereich der Reflexion zu untersuchen. Hierbei steht eine kategoriale Verschiedenheit zwischen Theorie und Praxis einer linearen Übertragbarkeit von Forschungsergebnissen grundlegend entgegen Neuweg (2021). Wenngleich Reflexion als ein integraler Bestandteil von Professionalität in der Lehrkräftebildung hervorgehoben wird (Roters, 2012) warnen Forschende vor einer zu eingeschränkten Professionsförderung hin zur Reflexion (Neuweg, 2017; Häcker, 2017). Diese Arbeit erlaubte die fokussierte Betrachtung reflexionsbezogener Profession. Es wird sich erhofft, dass auch über die Reflexionsforschung hinaus weitere professionsbezogene Zusammenhänge derart fokussiert untersucht werden, um ein zunehmendes Detailverständnis für Pedagogical Content Knowledge zu fördern.

## Literaturverzeichnis

- Abell, S. K. (2007). Research on science teachers' knowledge. In N. G. Lederman & S. K. Abell (Hrsg.), *Handbook of Research on Science Education* (S. 1105–1149). Lawrence Erlbaum Associates
- Abell, S. K. (2008). Twenty Years Later: Does pedagogical content knowledge remain a useful idea?, *International Journal of Science Education*, 30:10, 1405-1416, <https://doi.org/10.1080/09500690802187041>
- Abels, S. (2011). *LehrerInnen als „Reflective Practitioner“*. Reflexionskompetenz für einen demokratieförderlichen Naturwissenschaftsunterricht. VS Verlag.
- Aebli, H. (1980). *Denken: das Ordnen des Tuns. Band I: Kognitive Aspekte der Handlungstheorie*. Klett-Cotta.
- Aebli, H. (1981). *Denken: das Ordnen des Tuns. Band II: Denkprozesse*. Klett-Cotta.
- Aeppli, J. & Lötscher, H. (2016). EDAMA - ein Rahmenmodell für Reflexion. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 34(1), 78–97.
- Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher-student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, 49(10), 1211–1239. <https://doi.org/10.1002/tea.21055>
- Alonzo, A., Berry, A., Nilsson, P. (2019). Unpacking the Complexity of Science teachers' PCK in Action. Enacted and Personal PCK. In A. Hume, R. Cooper & A. Borowski (Hrsg.), *Repositioning Pedagogical Content Knowledge in Teachers' Professional Knowledge* (271–286). Springer.
- Altrichter, H. (2000). Handlung und Reflexion bei Donald Schön. In G.H. Neuweg (Hrsg.), *Wissen – Können – Reflexion* (S. 201–221). Studien-Verlag.
- Aydin, S., Demirdogen, B., Tarkin, A., Kutucu, S., Ekiz, B., Akin, F. N., Tuysuz, M., & Uzuntiryaki, E. (2013). Providing a Set of Research-Based Practices to Support Preservice Teachers' Long-Term Professional Development as Learners of Science Teaching: Providing a set of research-based practices. *Science Education*, 97(6), 903–935. <https://doi.org/10.1002/sce.21080>
- Bach, A. (2022). Selbstwirksamkeit im Lehrberuf. Entstehung und Veränderung sowie Effekte auf Gesundheit und Unterricht. *Pädagogische Psychologie und Entwicklungspsychologie* (101). Waxmann. <https://doi.org/10.25656/01:24604>
- Bain, J. D., Ballantyne, R., Packer, J., & Mills, C. (1999). Using Journal Writing to Enhance Student Teachers' Reflectivity During Field Experience Placements. *Teachers and Teaching*, 5(1), 51–73. <https://doi.org/10.1080/1354060990050104>
- Bandura, A. (1997). *Self-efficacy. The Exercise of Control*. W. H. Freeman and Company.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Hrsg.), *Self-Efficacy Beliefs of Adolescents*. Information Age Publishing, 307-337.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Beauchamp, C. (2006). *Understanding Reflection in Teaching: A Framework for Analyzing the Literature*.
- Berliner, D.C. (2001). Learning about and learning from expert teachers. *International Journal of Education Research*, 35, 463–482. [http://doi.org/10.1016/S0883-0355\(02\)00004-6](http://doi.org/10.1016/S0883-0355(02)00004-6)

- Berry, A., Depaepe, F., & van Driel, J. (2016). Pedagogical content knowledge in teacher education. In J. Loughran & M. L. Hamilton (Eds.), *International handbook of teacher education* (Vol. 1, pp. 347–386). Springer.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie / Journal of Psychology*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/A000194>
- Browning, T. D., & Korthagen, F. A. J. (2021). The winding road of student teaching: Addressing uncertainty with core reflection. *European Journal of Teacher Education*, 1–18. <https://doi.org/10.1080/02619768.2021.1973421>
- Buckingham Shum, S., Sándor, Á., Goldsmith, R., Bass, R., & McWilliams, M. (2017). Towards Reflective Writing Analytics: Rationale, Methodology and Preliminary Results. *Journal of Learning Analytics*, 4(1), 58–84. <https://doi.org/10.18608/jla.2017.41.5>
- Cardamone, C. N. et al. (2011). "Item Response Theory Analysis of the Mechanics Baseline Test." *2011 Physics Education Research Conference: Omaha, Nebraska, USA : 3-4 August 2011*" editors, N. Sanjay Rebello, Paula V. Engelhardt, Chandralekha Singh., American Institute of Physics, Melville, N.Y.2012. (AIP Conf. Proc. p. 135–138).
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., ... Wilson, C. D. (2019). The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In A. Hume, R. Cooper, & A. Borowski (Hrsg.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (S. 77–94). Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
- Chan, K.K.; Hume, A. *Repositioning Pedagogical Content Knowledge in Teachers' Professional Knowledge*; Springer: Singapore, 2019.
- Chodorow, M. & Burstein, J. (2004). Beyond essay length. Evaluating e-rater's performance on Toefl essays. *ETS. Research Report Service*, i–38. <https://doi.org/10.1002/j.2333-8504.2004.tb01931.x>
- Christof, E., Köhler, J., Rosenberger, K., & Wyss, C. (2018). *Mündliche, schriftliche und theatrale Wege der Praxisreflexion: Beiträge zur Professionalisierung pädagogischen Handelns* (1. Auflage). hep der bildungsverlag.
- Collin, S., Karsenti, T., & Komis, V. (2013). Reflective practice in initial teacher training: Critiques and perspectives. *Reflective Practice*, 14(1), 104–117. <https://doi.org/10.1080/14623943.2012.732935>
- Combe, A. & Kolbe, F.-U. (2008). Lehrerprofessionalität: Wissen, Können, Handeln. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (S. 857-875). Springer.
- Cramer, C., Harant, M., Merk, S., Drahmman, M., & Emmerich, M. (2019). Meta-Reflexivität und Professionalität im Lehrerinnen- und Lehrerberuf. *Zeitschrift für Pädagogik*, 3, 401–423.
- Darling-Hammond, L. (2012). *Powerful teacher education: Lessons from exemplary programs*. John Wiley & Sons.
- Dewey, J. (1933). *How we think. A restatement of the relation of reflective thinking to the educative process*. D.C. Heath and Company. <https://archive.org/details/howwethinkrestat00dewerich/page/16/mode/2up>
- Dorsch, Lexikon der Psychologie (o.D.). Werthaltung. In Wirtz, M., A. (Hrsg.) *Dorsch, Lexikon der Psychologie* online. Hogrefe AG. Abgerufen am 19.05.2023, von [https://dorsch.hogrefe.com/stichwort/werthaltung#:~:text=%5Bengl.,wünschenswert%20erachtet%20werden%20\(Werte\)](https://dorsch.hogrefe.com/stichwort/werthaltung#:~:text=%5Bengl.,wünschenswert%20erachtet%20werden%20(Werte))

- Dudenredaktion. (o. D.). Reflexivität. In *Duden* online. Abgerufen am 11. Mai 2023, von <https://www.duden.de/node/119547/revision/1275000>
- Duit, R. & Wodzinski, C. T. (2004). Merkmale „guten“ Physikunterrichts. *PIKO-BRIEF NR. 4*.
- Dunekacke, S., Jenßen, L., & Blömeke, S. (2015). Mathematikdidaktische Kompetenz von Erzieherinnen und Erziehern: Validierung des KomMa-Leistungstests durch die videogestützte Erhebung von Performanz. *Zeitschrift für Pädagogik*, 61. Beiheft, 80 – 99.
- Elsner, D., Kreft, A., Niesen, H. & Viebrock, B. (2020a). Unterrichtsvideos als Reflexionsanlässe im Englischlehramtsstudium. Herausforderung Lehrer\*innenbildung - *Zeitschrift Zur Konzeption, Gestaltung Und Diskussion*, 3(2), 279–299. <https://doi.org/10.4119/hlz-2501>
- Elsner, D., Kreft, A., Niesen, H. & Viebrock, A. (2020). Unterrichtsvideos als Reflexionsanlässe im Englischlehramtsstudium: Verbindung von Theorie und Praxis am Beispiel der Heterogenitätsdimensionen Mehrsprachigkeit und Transkulturalität. *HZL*, 3(2), 279-299.
- Enkrott, P. (2021). *Entwicklung des fachlichen Wissens angehender Physiklehrkräfte*. <https://doi.org/10.25932/publishup-50040>
- Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, 70, 113-136. <https://doi.org/10.1348/000709900158001>
- Faulstich-Wieland, H. & Nyssen, E. (1998). Geschlechterverhältnisse im Bildungssystem – Eine Zwischenbilanz. In Rolff, Bauer, Klemm & Pfeiffer (Hrsg.). *Jahrbuch der Schulentwicklung*, Bd. 10. Weinheim, S.163-199.
- Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (Hrsg.). (2014). *Quality of instruction in physics: Comparing Finland, Germany, and Switzerland*. Waxmann.
- Fischer, H., Borowski, A., & Tepner, O. (2012). Professional Knowledge of Science Teachers. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 435–448). Springer. [https://doi.org/10.1007/978-1-4020-9041-7\\_30](https://doi.org/10.1007/978-1-4020-9041-7_30)
- Fischer, J. & Weinert, M. (2021). Reflexionsförderung bei Lehramtsstudierenden durch den Einsatz von videobasierten Aufgaben. In L. Humbert (Hrsg.). *INFOS 2021 – 19. GI-Fachtagung Informatik und Schule. Gesellschaft für Informatik* (S. 261-270).
- Franken, N. & Preisfeld, A. (2023). Das Praxissemester als Triebkraft für die Entwicklung von Lehrer\*innen- Selbstwirksamkeitserwartungen in den Naturwissenschaften. Herausforderung Lehrer\*innenbildung. *Zeitschrift Zur Konzeption, Gestaltung Und Diskussion*, 6 (1), 36–55. <https://doi.org/10.11576/hlz-3274>
- Gaßner-Hofmann, H. (2022). Die Bedeutung angeleiteter tiefer Reflexionsphasen für die Entwicklung der inklusionsspezifischen Selbstwirksamkeit aus Sicht der Studierenden. In E. Gläser, J. Poschmann, P. Büker, S. Miller (Hrsg.): *Reflexion und Reflexivität im Kontext Grundschule. Perspektiven für Forschung, Lehrer:innenbildung und Praxis*. Klinkhardt, S. 58-63. <https://doi.org/10.25656/01:25551>
- Gaudin, C. & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41-67.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. Results of the thinking from the PCK summit. In A. Berry, P. Friedrichsen, J. Loughran (Hrsg.), *Re-Examining Pedagogical Content Knowledge in Science Education*, Routledge.
- Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., & Stuhlsatz, M. A. M. (2017). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 1-20.

- Gess-Newsome, J.; Taylor, J. A.; Carlson, J.; Gardner, A.L.; Wilson, C. D. & Stuhlsatz, M.A.M. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 41:7, 944-963, <https://doi.org/10.1080/09500693.2016.1265158>
- Grossman, P.L. *The Making of a Teacher: Teacher Knowledge and Teacher Education; Professional Development and Practice*; Teachers College Press: New York, NY, USA, 1990.
- Gruber, H. (2021). Reflexion. Der Königsweg zur Expertise-Entwicklung. *Journal für LehrerInnenbildung*, 21(1), 108–117. <https://doi.org/10.35468/jlb-01-2021-10>
- Häcker, T., Hilzensauer W. & Reinmann, G. (2008). Reflexives Lernen. *Bildungsforschung*, 5(2), 1-4.
- Häcker, T. (2017). Grundlagen und Implikationen der Forderung nach Förderung von Reflexivität in der Lehrerinnen- und Lehrerbildung. In C. Berndt, T. Häcker, & T. Leonhard (Hrsg.), *Reflexive Lehrerbildung revisited. Traditionen - Zugänge - Perspektiven* (S. 21–45). Klinkhardt.
- Häcker, T. H. (2019). Reflexive Professionalisierung. Anmerkungen zu dem ambitionierten Anspruch, die Reflexionskompetenz angehender Lehrkräfte umfassend zu fördern. In M. Degeling, N. Franken, S. Freund, S. Greiten, D. Neuhaus, & J. Schellenbach-Zell (Hrsg.), *Herausforderung Kohärenz: Praxisphasen in der universitären Lehrerbildung: Bildungswissenschaftliche und fachdidaktische Perspektiven* (S. 81–96). Verlag Julius Klinkhardt.
- Häcker, T. (2022). Reflexive Lehrer\*innenbildung Versuch einer Lokalisierung in pragmatischer Absicht. In C. Reintjes & I. Kunze (Hrsg.), *Reflexion und Reflexivität in Unterricht, Schule und Lehrer:innenbildung* (S. 94–116). Klinkhardt.
- Hattie, J. (2003). Teachers Make a Difference. What Is the Research Evidence? (pp. 1-17) Australian Council for Educational Research Annual Conference on Building Teacher Quality. Auckland: University of Auckland.
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49. [https://doi.org/10.1016/0742-051X\(94\)00012-U](https://doi.org/10.1016/0742-051X(94)00012-U)
- Helmke, A. (2003). Unterrichtsqualität erfassen, bewerten, verbessern. Seelze.
- Hoth, J., Kaiser, G., Döhrmann, M., König, J., & Blömeke, S. (2018). A Situated Approach to Assess Teachers' Professional Competencies Using Classroom Video. In O. Buchbinder & S. Kuntze (Eds.), *Mathematics Teachers Engaging with Representations of Practice* (pp. 23–46). Springer International Publishing.
- Hume, A., Cooper, R., & Borowski, A. (2019). *In Repositioning Pedagogical Content Knowledge in Teachers' Professional Knowledge*. Springer.
- Hume, A. (2009). Promoting higher levels of reflective writing in student journals. *Higher Education Research & Development*, 28(3), 247–260.
- Jescovitch, L.N., Scott, E.E., Cerchiara, J.A., Merrill, J., Urban-Lurain, M., Doherty, J. H., & Haudek, K. C. (2021). Comparison of Machine Learning Performance Using Analytic and Holistic Coding Approaches Across Constructed Response Assessments Aligned to a Science Learning Progression. *Journal of Science Education and Technology*, 30, 150–167 (2021). <https://doi.org/10.1007/s10956-020-09858-0>
- Kanterian, E. (2004). *Analytische Philosophie*. Campus Verlag.
- Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). GPT-4 Passes the Bar Exam. *SSRN*. <http://dx.doi.org/10.2139/ssrn.4389233>
- Keller-Schneider, M. (2020). Was beschäftigt Studierende in der Rolle als Lehrpersonen? Mehrperspektivische Reflexion subjektiv bedeutsamer Situationen. In M. Keller-Schneider, K. Krammer, M. Trautmann, & K. Zierer (Hrsg.), *Jahrbuch für Allgemeine*

- Didaktik. Thementeil Allgemeine Didaktik und Reflexion von Lehr- und Lernprozessen* (S. 40–58). Schneider Verlag. [www.paedagogik.de](http://www.paedagogik.de)
- Kellog, R. T. (2007). *Fundamentals of Cognitive Psychology*, 1st Ed. Thousand Oaks, Sage.
- Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teacher's analysis of classroom video predicts student learning of mathematics: Further exploration of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61, 172 – 181.
- Klika, D. (2007). Fächerwahl im Lehramtsstudium - Zementierung der Geschlechtersegregation? In R. Casale & E. Borst (Hrsg.), *Ökonomien der Geschlechter*. Budrich, S. 103-115. <https://doi.org/10.25656/01:5346>
- Kirschner, S.; Borowski, A.; Fischer, H. E.; Gess-Newsome, J. & von Aufschnaiter, C. (2016). Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge, *International Journal of Science Education*, 38:8, 1343-1372, <https://doi.org/10.1080/09500693.2016.1190479>
- Klassen, R. M. & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59–76. <https://doi.org/10.1016/j.edurev.2014.06.001>
- Klemm, K. im Auftrag des Verbandes Bildung und Erziehung [VBE] (2022). Entwicklung von Lehrkräftebedarf und -angebot in Deutschland bis 2030. *Verband Bildung und Erziehung* (VBE).
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2004). Standards für die Lehrerbildung: Bildungswissenschaften.
- KMK. (2019). *Standards für die Lehrerbildung: Bildungswissenschaften* (Beschluss der Kultusministerkonferenz vom 16.12.2004 i .d. F. vom 16.05.2019). [www.kmk.org](http://www.kmk.org).
- Kortmeyer, G. (2023). *Could an Artificial-Intelligence agent pass an introductory physics course?*. <https://arxiv.org/pdf/2301.12127.pdf>
- Korthagen, F.A. & Kessels, J. (1999). Linking theory and practice: changing the pedagogy of teacher education. *Educational Research*, 28(4), 4–17.
- Korthagen, Fred. A. J. (2001). *Linking Practice and Theory. The Pedagogy of Realistic Teacher Education*. Lawrence Erlbaum Associates.
- Kost, D. (2019). Reflexionsprozesse von Studierenden des Physiklehramts. Dissertation at Justus-Liebig-University in Gießen.
- Krammer, K. & Reusser, K. (2005). Unterrichtsvideos als Medium der Aus- und Weiterbildung von Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 23(1), 35–50.
- Kücholl, D., & Lazarides, R. (2021). Video- und protokollbasierte Reflexionen eigener praktischer Unterrichtserfahrungen im Lehramtsstudium. *Zeitschrift für Erziehungswissenschaft*, 24(4), 985-1006.
- Kulgemeyer, C. & Riese, J. (2018). From Professional Knowledge to Professional Performance: The Impact of CK and PCK on Teaching Quality in Explaining Situations. *Journal of Research in Science Teaching*. <https://doi.org/10.1002/tea.21457>
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.
- Kuusinen, C. M. (2016). *The meaning and measure of teacher self-efficacy for effective classroom teaching practices*. Doctoral dissertation. [Letzter Zugriff: 12.06.2023]. [https://deepblue.lib.umich.edu/bitstream/handle/2027.42/133219/ckuus\\_1.pdf?sequence=1](https://deepblue.lib.umich.edu/bitstream/handle/2027.42/133219/ckuus_1.pdf?sequence=1)
- Labott, D., & Reintjes, C. (2022). Unvereinbarkeit von Bewertung und Reflexionsaufgaben in der Lehrer\*innenbildung. In C. Reintjes & I. Kunze (Hrsg.), *Reflexion und Reflexivität*

- in *Unterricht, Schule und Lehrer:innenbildung* (S. 170–184). Klinkhardt.  
<https://doi.org/10.35468/5969-10>
- Lai, G., & Calandra, B. (2010). Examining the effects of computer-based scaffolds on novice teachers' reflective journal writing. *Educational Technology Research and Development*, 58(4), 421–437. <https://doi.org/10.1007/s11423-009-9112-2>
- Landis, J. R., & Koch, G. G. (1977). *The Measurement of Observer Agreement for Categorical Data* (Bd. 33, Nummer 1). <https://about.jstor.org/terms>
- Leonhard, T., & Rihm, T. (2011). Erhöhung der Reflexionskompetenz durch Begleitveranstaltungen zum Schulpraktikum? Konzeption und Ergebnisse eines Pilotprojekts mit Lehramtsstudierenden. *Lehrerbildung auf dem Prüfstand*, 4(2), 240–270.
- Leonhard, T., Nagel, N., Rihm, T., Strittmatter-Haubold, V., & Wengert-Richter, P. (2010). Zur Entwicklung von Reflexionskompetenz bei Lehramtsstudierenden. In A. Gehrman, U. Hericks, & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 111–128). Klinkhardt.
- Leonhard, T. (2008). *Professionalisierung in der Lehrerbildung. Eine explorative Studie zur Entwicklung professioneller Kompetenzen in der Lehrererstausbildung*. Logos.
- Leonhard, T. (2022). Reflexionsregime in Schule und Lehrerbildung – Zwischen guter Absicht und transintentionalen Folgen. In C. Reintjes & I. Kunze (Hrsg.), *Reflexion und Reflexivität in Unterricht, Schule und Lehrer:innenbildung* (S. 77–93). Klinkhardt.  
<https://doi.org/10.35468/5969-05>
- Liepert, S., & Borowski, A. (2019). Testing the Consensus Model: Relationships among physics teachers' professional knowledge, interconnectedness of content structure and student achievement. *International Journal of Science Education*, 41(7), 890–910.  
<https://doi.org/10.1080/09500693.2018.1478165>
- Lin, X., Hmelo, C., Kinzer, C.K. et al. Designing technology to support reflection. *ETR&D* 47, 43–62 (1999). <https://doi.org/10.1007/BF02299633>
- Lindmeier A. (2011). *Modeling and Measuring Knowledge and Competencies of Teachers. A Threefold Domain- Specific Structure Model for Mathematics*, Waxmann.
- Lohse-Bossenz, H., Schönknecht, L. & Brandtner, M. (2019). Entwicklung und Validierung eines Fragebogens zur Erfassung Reflexionsbezogener Selbstwirksamkeit von Lehrkräften im Vorbereitungsdienst. *Empirische Pädagogik*, 33(2), 164–179.
- Magnusson, S.; Krajcik, J.S.; Borko, H. Nature, sources, and development of pedagogical content knowledge for science teaching. In *Examining Pedagogical Content Knowledge. The Construct and Its Implication for Science Education*; Gess-Newsome, J., Lederman, N.G., Eds.; Kluwer Academic: Dordrecht, The Netherlands, 1999; pp. 95–132.
- Meier, J., Vogelsang, C., Watson, C. & Schaper, N. (2021). „Reflexion ist erzwungenes Nachdenken“. Zusammenhänge zwischen dem Reflexionsverständnis Lehramtsstudierender & Facetten ihrer Reflexionsperformanz. In U. Stadler-Altmann (Hrsg.), *Digitale Lehrerbildung (Lehrerbildung auf dem Prüfstand, 14. Jahrgang, Heft 1 (2021))*. Verlag Empirische Pädagogik.
- Meschede, N. (2014). Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. Theoretische Beschreibung und empirische Forschung. In M. Hopf, H. Niedderer, M. Ropohl, & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* (Bd. 163). Logos-Verlag.
- Meyer, H. (2021). Was ist guter Unterricht?. 15. Auflage, Cornelsen.
- Mientus, L., Hume, A., Wulff, P., Meiners, A., & Borowski, A. (2022). Modelling STEM Teachers' Pedagogical Content Knowledge in the Framework of the Refined Consensus

- Model: A Systematic Literature Review. *Education Sciences*, 12, 385.  
<https://doi.org/10.3390/educsci12060385>
- Mientus, L., Wulff, P., Nowak, A., & Borowski, A. (2023). Fast-and-frugal means to assess reflection-related reasoning processes in teacher training – Development and evaluation of a scalable machine learning-based metric. *Zeitschrift für Erziehungswissenschaften*.  
<https://doi.org/10.1007/s11618-023-01166-8>
- Mientus, L., Borowski, A., Nowak, A., & Wulff, P. (2023b). Pädagogische Argumentation in Fremd- und Selbstreflexionstexten - Gegenüberstellung zweier Analyseverfahren zur Qualitätseinschätzung. In van Vorst, H. (Hrsg.). *Lernen, Lehren und Forschen in einer digital geprägten Welt*. GDCP-Jahrestagung 2022. Tagungsband 2023 - Band 43. S. 302 – 305. <https://doi.org/10.25656/01:27632>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math, Book News.
- NBPTS (National Board for Professional Teaching Standards) (2016). What Teachers Should Know and Be Able to Do.
- Neuber, K. & Göbel, K. (2020). Einstellungen zur Reflexion von angehenden und praktizierenden Lehrkräften. *Empirische Pädagogik*, 34(1), 64–78.
- Neuweg, G. H. (2017). Herrlich unreflektiert. Warum Könner weniger denken, als man denkt. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Reflexive Lehrerbildung revisited. Traditionen – Zugänge – Perspektiven* (S. 89–101). Bad Heilbrunn: Klinkhardt.
- Neuweg, G.H. Reflexivität. *Zeitschrift für Bildungsforschung*, 11, 459–474 (2021).  
<https://doi.org/10.1007/s35834-021-00320-8>
- Noack, M. (2007). *Faktorenanalyse*. Universität Duisburg-Essen. <https://www.uni-due.de/imperia/md/content/soziologie/stein/faktorenanalyse.pdf>. [Zuletzt abgerufen am 03.10.2023 11:14]
- Nordine, J., Sorge, S., Delen, I., Evans, R., Juuti, K., Lavonen, J., Nilsson, P., Ropohl, M., & Stadler, M. (2021). Promoting Coherent Science Instruction through Coherent Science Teacher Education: A Model Framework for Program Design. *Journal of Science Teacher Education*, 32(8), 911–933. <https://doi.org/10.1080/1046560X.2021.1902631>
- Nowak, A., Kempin, M., Kulgemeyer, C., & Borowski, A. (2019). Reflexion von Physikunterricht. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018* (Bd. 11, Nummer 1, S. 838–841). Universität Regensburg. [https://doi.org/10.1016/0742-051X\(94\)00012-U](https://doi.org/10.1016/0742-051X(94)00012-U)
- Nowak, A. (2023). Untersuchung der Qualität von Selbstreflexionstexten zum Physikunterricht. Entwicklung des Reflexionsmodells REIZ. In: M. Hopf & M. Ropohl (Hrsg.). *Studien zum Physik- und Chemielernen*. Logos.
- Oser, F. K. & Heinzer, S. (2009). Die Entwicklung eines Qualitätskonstrukts zur advokatorischen Erfassung der Professionalität. In: Zlatkin-Troitschanskaia, O., Beck, K., Sembill, D., Nickolaus, R., & Mulder, R. [Hrsgs.]. *Lehrerprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung*. Beltz Verlag, 167-180.
- Oser, F., Heinzer, S., & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. Chancen und Grenzen des advokatorischen Ansatzes. *Unterrichtswissenschaft*, 38/1, 5-28.
- Park, S. & Oliver, J.S. (2008). National Board Certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching*, 45: 812-834.  
<https://doi.org/10.1002/tea.20234>
- Park, S. & Oliver, J.S (2008b). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals.

- Research in Science Education*, 38, 261–284. <https://doi.org/10.1007/s11165-007-9049-6>
- Park, S. & Suh, J.K. (2019). The PCK Map Approach to Capturing the Complexity of Enacted PCK (ePCK) and Pedagogical Reasoning in Science Teaching. In: Hume, A., Cooper, R., Borowski, A. (Hrsg.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*. Springer. [https://doi.org/10.1007/978-981-13-5898-2\\_8](https://doi.org/10.1007/978-981-13-5898-2_8)
- Philipp, R. (2007). Mathematics Teachers' Beliefs and Affect. In F. K. Lester (Hrsg.), *Second Handbook of Research on Mathematics Teaching and Learning. A Project of the National Council of Teachers of Mathematics* (S. 257–315). Information Age Publishing.
- Poldner, E., Van Der Schaaf, M., Simons, R.-J., Van Tartwijk, J., & Wijngaards, G. (2014). Assessing student teachers' reflective writing through quantitative content analysis. *European Journal of Teacher Education*, 37(3), 348–373. <https://doi.org/10.1080/02619768.2014.892479>
- Reintjes, C., & Kunze, I. (2022). Editorial: Reflexion und Reflexivität in Unterricht, Schule und Lehrer\*innenbildung. In C. Reintjes & I. Kunze (Hrsg.), *Reflexion und Reflexivität in Unterricht, Schule und Lehrer\*innenbildung* (S. 9–16). Klinkhardt. <https://doi.org/10.35468/5969-01>
- Roters, B. (2012). *Professionalisierung durch Reflexion in der Lehrerbildung*. Waxmann.
- Rothland, M. (2021). Die „Lehrerpersönlichkeit“: das Geheimnis des Lehrberufs?. *Die Deutsche Schule*, 113 (2021) 2, 188-198.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms. *American Educational Research Journal*, 50(5), 1020–1049. <https://doi.org/10.3102/0002831213477680>
- Santagata, R. & Angelici, G. (2010). Studying the impact of the lesson analysis framework on pre-service teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education*, 61(4), 339–349.
- Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und -messung in der Lehrerbildung - In: *Lehrerbildung auf dem Prüfstand 2*, S. 166-199. <https://doi.org/10.25656/01:14697>
- Schecker, H. & Wilhelm, T. (2018). Schülervorstellungen in der Mechanik. In H. Schecker, T. Wilhelm, M. Hopf & R. Duit, *Schülervorstellungen und Physikunterricht - Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*, S. 63 – 88. Springer. <https://doi.org/10.1007/978-3-662-57270-2>
- Schlesinger, L., Jentsch, A. & Kaiser, G. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM: The International Journal on Mathematics Education*. 50(3), pp. 475 - 490. <https://doi.org/10.1007/s11858-018-0917-5>
- Schön, D. A. (1983). *The reflective practitioner: how professionals think in action*. Basic Books.
- Schubarth, W., Speck, K., Seidel, A., & Wendland, M. (2009). Unterrichtskompetenzen bei Referendaren und Studierenden. Empirische Befunde der Potsdamer Studien zur ersten und zweiten Phase der Lehrerbildung. In: *Lehrerbildung auf dem Prüfstand 2*, S. 304-323. <https://doi.org/10.25656/01:14705>
- Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, 51(4), 739–771. <https://doi.org/10.3102/0002831214531321>
- Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether

- teachers observe their own teaching or that of others? *Teaching and Teacher Education*, 27(2), 259–267. <https://doi.org/10.1016/j.tate.2010.08.009>
- Sherin, M. G. (2001). Developing a professional vision of classroom events. In T. L. Wood, B. S. Nelson & J. Warfield (Hrsg.), *Beyond classical pedagogy. Teaching elementary school mathematics* (S. 75–93). Erlbaum Associates.
- Shermis, M.D., Burstein, J., Higgins, D., & Zechner, K. (2019). Automated essay scoring: writing assessment and instruction. In P. Peterson, E. Baker, B. McGaw (Hrsg.), *International Encyclopedia of Education* (Third Edition), Elsevier, 2010, Pages 20-26. <https://doi.org/10.1016/B978-0-08-044894-7.00233-5>
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Shulman, L. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*. 1 April 1987; 57 (1): 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Shulman Lee S. (1999). *Taking Learning Seriously, Change: The Magazine of Higher Learning*, 31:4, 10-17, <https://doi.org/10.1080/00091389909602695>
- Shulman, L.S. Appreciating good teaching. A conversation with Lee Shulman by Carol Tell. *Education Leadership*. 2001, 58, 6–11.
- Sorge, S., Neumann, I., Knut Neumann, Parchmann, I., & Schwanewedel, J. (2018). Was ist denn da passiert? Ein Protokollbogen zur Reflexion von Praxisphasen im Lehr-Lern-Labor. *MNU Journal*, 6, 420–426.
- Sorge, S., Stender, A., & Neumann, K. (2019). The Development of Science Teachers' Professional Competence. In A. Hume et al. (eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*. (149 - 164). [https://doi.org/10.1007/978-981-13-5898-2\\_6](https://doi.org/10.1007/978-981-13-5898-2_6)
- Ullmann, T.D. (2017). Reflective writing analytics: empirically determined keywords of written reflection: LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference. In: *ACM International Conference Proceeding Series*, pp. 163–167.
- Ullmann, T.D. (2019). Automated analysis of reflection in writing: validating machine learning approaches. *International Journal Artificial Intelligence in Education*, 29(2), 217–257.
- van Es, E. A. & Sherin, M. G. (2021). Expanding on prior conceptualizations of teacher noticing. *ZDM Mathematics Education*, 53(1), 17–27. <https://doi.org/10.1007/s11858-020-01211-4>
- Venkatesh, V.; Thong, J. Y. L.; Chan, F. K. Y.; & Hu, P. J. H. (2016) Managing Citizens' Uncertainty in E-Government Services: The Mediating and Moderating Roles of Transparency and Trust. *Information Systems Research* 27(1):87-111. <https://doi.org/10.1287/isre.2015.0612>
- Vogelsang, C., Kulgemeyer, C. & Riese, J. (2022). Learning to plan by learning to reflect? Exploring relations between professional knowledge, reflection skills, and planning skills of preservice physics teachers in a one-semester field experience. *Education Sciences*, 12(7), 479. <http://dx.doi.org/10.3390/educsci12070479>
- von Aufschnaiter, C., Fraij, A. & Kost, D. (2019). Reflexion und Reflexivität in der Lehrerbildung. Herausforderung Lehrer\*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion, Bd. 2 Nr. 1. *Herausforderung Lehrer\_innenbildung*, (144-159). <https://doi.org/10.4119/HLZ-2439>
- von Aufschnaiter, C. (in Press): Reflexive Professionalisierung: Zentral – Vielschichtig – Herausfordernd. In L. Mientus, A. Nowak, & C. Klempin (Hrsg.), *Reflexion in der Lehrkräftebildung – empirisch, phasenübergreifend, interdisziplinär*. Potsdamer Universitätsverlag.

- Wahl, D. (2002). Mit Training vom trägen Wissen zum kompetenten Handeln? *Zeitschrift für Pädagogik*, 48(2), 227-241.
- Weinert, F.E. (2001). Vergleichende Leistungsmessungen in Schulen – eine umstrittene Selbstverständlichkeit. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-31). Beltz.
- West, C. G. (2023a). *Advances in apparent conceptual physics reasoning in GPT-4*. <https://arxiv.org/pdf/2303.17012.pdf>
- West, C. G. (2023b). *AI and the FCI: Can ChatGPT project an understanding of introductory physics?*. <https://arxiv.org/pdf/2303.01067.pdf>
- Wieman, C. E. (2007). The "Curse of Knowledge" or Why Intuition About Teaching Often Fails. *American Physical Society News*, 16(10).
- Wilson, C.D.; Borowski, A. & van Driel, J.H. (2019). Perspectives on the Future of PCK Research in Science Education and Beyond. In A. Hume, R., Cooper, A., Borowski (Hrsg.), *Repositioning Pedagogical Content Knowledge in Teachers' Professional Knowledge*, Springer. pp. 289–300.
- Wulff, P., Westphal, A., Mientus, L., Nowak, A., & Borowski, A. (2023). Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers' written reflections. *Frontiers in Education*, 7. <https://doi.org/10.3389/educ.2022.1061461>
- Wulff, P., Buschhüter, D., Westphal, A., Nowak, A., Becker, L., Robalino, H., Stede, M., & Borowski, A. (2020). Computer-Based Classification of Preservice Physics Teachers' Written Reflections. *Journal of Science Education and Technology*, 30(1), 1–15. <https://doi.org/10.1007/s10956-020-09865-1>
- Wulff, P., Buschhüter, D., Westphal, A., Mientus, L., Nowak, A., & Borowski, A. (2022). Bridging the Gap Between Qualitative and Quantitative Assessment in Science Education Research with Machine Learning—A Case for Pretrained Language Models-Based Clustering. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-022-09969-w>
- Wyss, C. & Mahler, S. (2021). Mythos Reflexion. Theoretische und praxisbezogene Erkenntnisse in der Lehrer\*innenbildung. *Journal für LehrerInnenbildung*, 21(1), 16-25. <https://doi.org/10.35468/jlb-01-2021-01>
- Wyss, C. (2013). *Unterricht und Reflexion. Eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften*. Waxmann.
- Wyss, C. (2018). Mündliche, kollegiale Reflexion von videografiertem Unterricht. In E. Christof, J. Köhler, K. Rosenberger, & C. Wyss (Hrsg.), *Mündliche, schriftliche und theatrale Wege der Praxisreflexion: Beiträge zur Professionalisierung pädagogischen Handelns*, 1. Auflage, S. 15–49. hep Verlag.
- Yeadon, W., Inyang, O.-O., Mizouri, A., Peach, A. & Testrow, C. P. (2023). The death of the short-form physics essay in the coming AI revolution. *Physics Education*, 58(3), 35027. <https://doi.org/10.1088/1361-6552/acc5cf>
- Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education*, 56(1), 111–151. <https://doi.org/10.1080/03057267.2020.1735757>
- Zlatkin-Troitschanskaia, O., Kuhn, C., Brückner, S., & Leighton, J. P. (2019). Evaluating a Technology-Based Assessment (TBA) to Measure Teachers' Action-Related and Reflective Skills. *International Journal of Testing*, 19(2), 148–171. <https://doi.org/10.1080/15305058.2019.1586377>

# ***Anhang***

## ***Verzeichnis***

A1	Vignette (Prompt, Video, RT & Skalen)	104
A2	Definitionen für Qualitätsindikatoren nach 1.2.2	108
A3	Testung der erhobenen Skalen auf Normalverteilung	109
A4	Skalen und Items	110
A4.1	Skala der wahrgenommenen Unterrichtsqualität	110
A4.2	Skala der reflexionsbezogenen Selbstwirksamkeit	112
A4.3	Fachwissen – Skala	113
A4.4	Wert – Skala	118
A5	Tabellenverzeichnis	119
A6	Abbildungsverzeichnis	120

## ***A1 Vignette (Prompt, Video, RT & Skalen)***

### *Erste Seite:*

Liebe angehende Physik-Lehrkräfte,

zur Qualitätsentwicklung in der Lehre untersuchen wir im Rahmen eines Projektes (PSI Potsdam) im Zusammenhang mit der „Qualitätsoffensive Lehrerbildung“ Fragen dazu, wie Studierende eine echte Unterrichtssituation in der Physik analysieren und in diesem Zusammenhang ihre Fähigkeiten der Unterrichtswahrnehmung selbst einschätzen.

Was Sie erwartet:

1. Anschauen eines Videoausschnittes aus einer Unterrichtsstunde (ca. 15 Minuten)
2. Verfassen eines Analysetextes zum Videoausschnitt
3. Multiple-Choice Fragen zum Videoausschnitt und dem Analysetext

Die Umfrage dauert durchschnittlich 60 bis 90 Minuten. Nehmen Sie sich diese Zeit bitte am Stück, sodass Sie die Umfrage nicht für längere Zeit unterbrechen müssen, da dies zu Datenverlust führen kann.

Hinweise zur Verwendung Ihrer Daten:

In der Umfrage verfassen Sie unter anderem einen Analysetext und geben Selbsteinschätzungen zu Ihrer Wahrnehmung und Ihren Fähigkeiten ab. Diese Daten möchten wir für die Weiterentwicklung der universitären physikdidaktischen Lehre verwenden, indem wir Fragen dazu klären, auf welche Weise unsere Studierenden analysieren und sich einschätzen. Wir versichern Ihnen, dass wir Ihre Antworten vollkommen anonym auswerten. Zu keiner Zeit werden wir Ihre Antworten mit Ihrer Person verbinden. Die Teilnahme am Fragebogen ist vollkommen freiwillig und Sie können Ihre Teilnahme jederzeit widerrufen.

Für den Text sowie den Fragen zur Selbsteinschätzung gibt es kein richtig oder falsch. Wir sind einzig an Ihrer persönlichen Meinung interessiert.

Beachten Sie, dass die Bearbeitung des Fragebogens keinerlei Auswirkungen auf Ihr Studium hat.

Hinweis zur Bearbeitung der Umfrage:

Wenn Sie an irgendeiner Stelle Probleme mit dem Fragebogen haben, melden Sie sich ganz einfach bei uns per E-Mail ([lukas.mientus@uni-potsdam.de](mailto:lukas.mientus@uni-potsdam.de)). Wir hoffen, dass wir auftretende Probleme einfach beheben können. Um Datenverlust zu vermeiden, klicken Sie bitte nicht auf den "Zurück"-Button Ihres Browsers.

Wenn Sie mehr Information über die Verarbeitung Ihrer personenbezogenen Daten wünschen, bitte auf folgenden Link klicken.

Ich stimme zu, dass meine personenbezogenen Daten gemäß den hier aufgeführten Angaben verarbeitet werden.

### *Zweite Seite:*

Auswertung einer Unterrichtssituation

Zunächst möchten wir uns bei Ihnen bedanken, dass Sie am Fragebogen teilnehmen! Füllen Sie den Fragebogen bitte in einem Durchlauf aus. Nehmen Sie sich also ca. 90 Minuten Zeit. Sollten Sie den Fragebogen zwischendurch schließen, kann es zum Datenverlust kommen.

Ihre Aufgabe:

In ihrem potentiellen Beruf als Lehrkraft werden Sie häufig Unterrichtssituationen erleben, über die Sie auch im Nachhinein noch nachdenken. Doch wie genau funktioniert dieses Nachdenken über Unterricht und wie kann man es strukturieren? Genau diese Fragen wollen wir auf der Basis des nachfolgenden Fragebogens weiter untersuchen.

Dazu sehen Sie im Folgenden einen kurzen Videoausschnitt einer echten Unterrichtsstunde, die wir aufgezeichnet haben.

Schauen Sie sich zunächst auf der nächsten Seite das Video der Unterrichtssituation bitte nur einmal an. Sie sollten sich für Ihren Analysetext während des Videoausschnittes Notizen machen.

Im Anschluss an den Videoausschnitt gelangen Sie auf die nächste Seite. Dort finden Sie ein (unbegrenztes) Textfeld, in welchem Sie Ihren Analysetext der Unterrichtssituation verfassen.

Ihre Aufgabe besteht darin einen Text zu verfassen, der Ihnen (nicht Ihrem Kollegen, den Sie im Video sehen) als Hilfe dienen kann sich persönlich professionell weiterzuentwickeln. Bitte verfassen Sie Ihren Text als Fließtext ohne Stichpunkte oder Teilüberschriften. Stellen Sie sich vor, Sie sitzen als Kollege oder Kollegin in diesem Unterricht. Sie verfassen Ihren Text mit dem Ziel, eine Bewertung der Unterrichtssituation vorzunehmen und für sich selbst persönliche Konsequenzen für Ihr Handeln als Lehrkraft abzuleiten.

Hinweis: Die im Folgenden aufgeführten Punkte 1), 2), 3) und 4) bekommen Sie während des Videos und im Texteditor oben eingeblendet.

Was schreiben Sie?

Beschreibung: Erinnern Sie sich zunächst daran, was in der Unterrichtssituation passiert ist. Beschreiben Sie die Situation in den aus Ihrer Sicht wichtigen Aspekten. Was hat die Lehrperson gemacht? Was haben die SuS gemacht?

Bewertung: Im Anschluss bewerten und begründen Sie die verschiedenen Aspekte der Unterrichtssituation. Was war an der Stunde gelungen/weniger gelungen und warum?

Alternativen: Finden Sie dann alternative Vorgehensweisen zu den Aspekten der Unterrichtssituation und vermuten Sie, warum diese zielführend wären. Was hätte die Lehrperson wie anders machen können und warum wäre dieses Handeln zielführend gewesen? Oder warum sollte sie es genauso wieder machen?

Konsequenzen: Überlegen Sie abschließend, welche Konsequenzen Sie aus dieser Unterrichtssituation ziehen: Welche Konsequenzen ziehe ich für die Weiterentwicklung der beobachteten Lehrperson und warum? Was schließe ich aus dem beobachteten Unterricht für meine persönliche Entwicklung als Lehrperson und warum?

Sind Sie nicht verunsichert, wenn Sie nicht genau wissen, was Ihre Aufgabe ist. Interpretieren Sie die Punkte 1), 2), 3) und 4) so, wie Sie es für richtig halten und verfassen auf dieser Basis Ihren Text.

Im Anschluss an den Text haben wir dann noch einige weitere Fragen für Sie.

Der Videoausschnitt:

Im Videoausschnitt sehen Sie eine Unterrichtssituation im Fach Physik einer 9. Klasse. Das Thema der gesamten Unterrichtsstunde ist die gleichförmig beschleunigte Bewegung. Sie sehen im Folgenden die ersten 15 Minuten dieser Unterrichtsstunde. Ziel der Unterrichtssequenz ist es, die Gesetzmäßigkeiten der gleichmäßig beschleunigten Bewegung zu erarbeiten.

*Dritte Seite:*

Die Unterrichtssituation - Videoausschnitt

Sehen Sie sich im Folgenden das Video eines Ausschnittes aus dem Unterricht an.

Zur Erinnerung: Was schreiben Sie?

Beschreibung: Erinnern Sie sich zunächst daran, was in der Unterrichtssituation passiert ist. Beschreiben Sie die Situation in den aus Ihrer Sicht wichtigen Aspekten. Was hat die Lehrperson gemacht? Was haben die SuS gemacht?

Bewertung: Im Anschluss bewerten und begründen Sie die verschiedenen Aspekte der Unterrichtssituation. Was war an der Stunde gelungen/weniger gelungen und warum?

Alternativen: Finden Sie dann alternative Vorgehensweisen zu den Aspekten der Unterrichtssituation und vermuten Sie, warum diese zielführend wären. Was hätte die Lehrperson wie anders machen können und warum wäre dieses Handeln zielführend gewesen? Oder warum sollte sie es genauso wieder machen

Konsequenzen: Überlegen Sie abschließend, welche Konsequenzen Sie aus dieser Unterrichtssituation ziehen: Welche Konsequenzen ziehe ich für die Weiterentwicklung der beobachteten Lehrperson und warum? Was schließe ich aus dem beobachteten Unterricht für meine persönliche Entwicklung als Lehrperson und warum?

### Unterrichtsvideo ###

*Vierte Seite:*

Bitte schreiben Sie im folgenden Textfeld Ihre Analyse der Unterrichtssituation

Bitte verfassen Sie die Reflexion als Fließtext in vollständigen Sätzen ohne Stichpunkte oder Teilüberschriften.

Hinweis: Schreiben Sie Ihren Text am besten auf Ihrem lokalen Rechner und speichern diesen dort. Kopieren Sie den Text dann in den Editor und gehen Sie zur nächsten Seite. Das stellt sicher, dass Ihre Arbeit nicht durch versehentliches Klicken auf "Aktualisieren" oder ähnliches verloren geht.

Zur Erinnerung: Was schreiben Sie?

Beschreibung: Erinnern Sie sich zunächst daran, was in der Unterrichtssituation passiert ist. Beschreiben Sie die Situation in den aus Ihrer Sicht wichtigen Aspekten. Was hat die Lehrperson gemacht? Was haben die SuS gemacht?

Bewertung: Im Anschluss bewerten und begründen Sie die verschiedenen Aspekte der Unterrichtssituation. Was war an der Stunde gelungen/weniger gelungen und warum?

Alternativen: Finden Sie dann alternative Vorgehensweisen zu den Aspekten der Unterrichtssituation und vermuten Sie, warum diese zielführend wären. Was hätte die Lehrperson wie anders machen können und warum wäre dieses Handeln zielführend gewesen? Oder warum sollte sie es genauso wieder machen

Konsequenzen: Überlegen Sie abschließend, welche Konsequenzen Sie aus dieser Unterrichtssituation ziehen: Welche Konsequenzen ziehe ich für die Weiterentwicklung der beobachteten Lehrperson und warum? Was schließe ich aus dem beobachteten Unterricht für meine persönliche Entwicklung als Lehrperson und warum?

### unbegrenztes Eingabefeld für Reflexionstext ###

*Weitere Seiten:*

### Items zur Erhebung der Amplifiers & Filters ###

## *A2 Definitionen für Qualitätsindikatoren nach 1.2.2*

Qualitätsbewertungen der schriftlichen Fremdreflexionen erfolgen in erster Linie auf Basis der Erfahrungswerte der Expert\*innen für schriftliche Reflexionen. Die Expert\*innen dieser Studie befassen sich intensiv mit schriftlichen Reflexionen in Forschung und Lehre. Ihnen kann ein weitreichendes Verständnis für die Diversität der verschriftlichten Reflexionen attestiert werden. Um dennoch Qualitätsbewertungen auf einer objektiven Grundlage durchzuführen sind im Folgenden die vereinbarten Definitionen der Qualitätsindikatoren festgehalten. Nach diesen Definitionen wurde jeder Reflexionstext vierstufig bewertet, wobei 1 = „Nicht ausgeprägt“; 2 = „Gering ausgeprägt“; 3 = „Eher ausgeprägt“ und 4 = „Sehr ausgeprägt“ repräsentiert.

Die Breite einer Reflexion klassifiziert die Anzahl der betrachteten Reflexionsauslöser und die adressierte Reichweite der Themen und Ideen, die in der Reflexion behandelt werden. Eine Reflexion mit großer Breite berührt eine Vielzahl von Themen und Ideen, während eine Reflexion mit geringer Breite sich auf ein spezifisches Thema oder eine spezifische Erfahrung konzentriert.

Die Tiefe einer Reflexion klassifiziert die Gründlichkeit bzw. Intensität der Analyse und des Verständnisses, mit welchen im Mittel reflexionsbezogene Argumentationsprozesse dargestellt werden. Der Einbezug theoretisch formalen Wissens kann ebenso auf eine ausgeprägtere Reflexionstiefe hindeuten.

Die Kohärenz eines Reflexionstext bezieht sich auf die logische Verbindung von Ideen und Informationen innerhalb des Textes. Ein kohärenter Reflexionstext folgt somit im reflexionsbezogenen Argumentationsprozess einer klaren Struktur, deutlich nachvollziehbaren Zusammenhängen sowie einer klaren und verständlichen Sprache.

Die Spezifität einer schriftlichen Reflexion repräsentiert die Genauigkeit beziehungsweise die Bezugnahme auf die wahrgenommene Unterrichtssituation. Anstatt allgemeine Aussagen zu tätigen kann einen spezifischen Reflexionstext auszeichnen, dass er detaillierte Informationen zur wahrgenommenen Unterrichtssituation darlegt und Argumentationen konkret und weniger holistisch erfolgen.

### A3 *Testung der erhobenen Skalen auf Normalverteilung*

In der folgenden Tabelle sind die Testergebnisse des Shapiro-Wilk-Testes dargestellt und zugehörig bewertet, ob von einer Normalverteilung ausgegangen werden darf oder nicht.

<i>Skala</i>	<i>W-Wert</i>	<i>p-Wert</i>	<i>Skalenniveau</i>	<i>Normalverteilung</i>
Textlänge	0.94457	3.392e-05	metrisch	annehmbar
Sätze	0.93383	5.859e-06	metrisch	annehmbar
Beschreibung	0.92051	3.507e-06	metrisch	annehmbar
Bewertung	0.88928	1.488e-08	metrisch	annehmbar
Alternativen	0.94395	3.051e-05	metrisch	annehmbar
Konsequenzen	0.91830	5.931e-07	metrisch	annehmbar
LOS	0.98795	0.2933	metrisch	bedingt annehmbar
Alter	0.85021	7.212e-10	metrisch	annehmbar
Abinote	0.97318	0.02527	ordinal	annehmbar
Physiknote	0.91672	7.156e-06	ordinal	annehmbar
Mathenote	0.90665	6.317e-07	ordinal	annehmbar
Deutschnote	0.97686	0.02785	ordinal	annehmbar
Expertise	0.26688	2.2e-16	ordinal	annehmbar
wUQ	0.98822	0.3263	metrisch	bedingt annehmbar
wUQ_CM	0.92926	0.02659	metrisch	annehmbar
wUQ_SS	0.93594	0.04671	metrisch	annehmbar
wUQ_CA	0.96999	0.3025	metrisch	bedingt annehmbar
wUQ_SQ	0.96277	0.2764	metrisch	bedingt annehmbar
wUQ_TQ	0.97266	0.03246	metrisch	annehmbar
SWK	0.97629	0.02313	metrisch	annehmbar
SWK_Bes	0.96424	0.002063	metrisch	annehmbar
SWK_Beg	0.93680	1.448e-05	metrisch	annehmbar
SWK_Bew	0.95819	0.0005717	metrisch	annehmbar
SWK_Alt	0.94168	3.159e-05	metrisch	annehmbar
SWK_Kons	0.94756	9.013e-05	metrisch	annehmbar
Wert	0.92160	1.537e-06	metrisch	annehmbar
CK	0.94161	3.858e-05	ordinal	annehmbar
Breite	0.84229	1.148e-10	ordinal	annehmbar
Tiefe	0.86605	1.171e-09	ordinal	annehmbar
Kohärenz	0.85993	6.285e-10	ordinal	annehmbar
Spezifität	0.85439	3.631e-10	ordinal	annehmbar
Quantität	0.97336	0.009843	metrisch	annehmbar
Qualität	0.98255	0.08389	metrisch	annehmbar
Deskriptivität	0.99344	0.7958	metrisch	nicht annehmbar

## A4 Erhebungsinstrumente und Skalen

### A4.1 Skala der wahrgenommenen Unterrichtsqualität (aus Schlesinger et al., 2018)

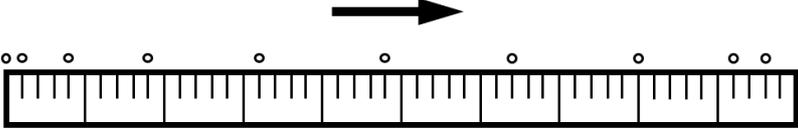
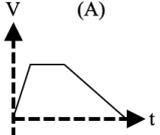
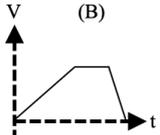
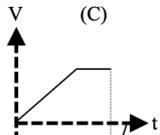
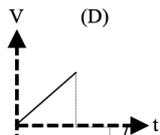
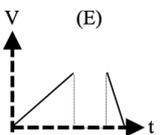
<i>Item-ID</i>	<i>Original</i>	<i>Übersetzung</i>	<i>Subscale</i>
wUQ1	Transitions between lesson phases run smoothly	Der Übergang zwischen den Unterrichtsphasen erfolgte reibungslos.	
wUQ2	The lesson time is used for content-related instruction	Die Unterrichtszeit wurde für unterrichtsbezogene Inhalte genutzt.	
wUQ5	The teacher is aware of everything that happens in the classroom	Der Lehrer hatte den Überblick über das Unterrichtsgeschehen.	
wUQ6	The teacher informs the students about the lesson objectives	Der Lehrer informierte die Schülerinnen und Schüler über Lernziele / Unterrichtsablauf.	<b><i>Classroom management</i></b>
wUQ8	Tasks are given in precise language	Arbeitsaufträge waren präzise formuliert.	
wUQ9	The volume level is appropriate	Die Lautstärke war angemessen.	
wUQ11	The students and the teacher do not interrupt each other	Die Schülerinnen und Schüler und der Lehrer haben sich nicht gegenseitig unterbrochen.	
wUQ12	A common thread is apparent in the lesson	Die Unterrichtssituation besaß einen „roten Faden“	
wUQ14	The teacher takes time for individuals	Der Lehrer nahm sich Zeit für einzelne Lernende.	
wUQ16	Tasks address different types of students	Es gab anforderungsbezogene, differenzierende Aufgabenstellungen.	
wUQ18	The teacher encourages students to work independently	Der Lehrer forderte eigenständiges Arbeiten in hohem Maße ein.	
wUQ19	The teacher's feedback is constructive	Die Rückmeldungen des Lehrers waren sachlich und konstruktiv.	<b><i>student support</i></b>
wUQ21	The teacher is patient	Der Lehrer hatte Geduld.	
wUQ23	The teacher encourages the students	Der Lehrer ermutigte die Lernenden.	
wUQ25	The teacher reacts to students' feedback	Der Lehrer ist auf Anregungen der SuS eingegangen.	
wUQ26	The teacher initiates collaborative learning processes between students	Der Lehrer initiierte gemeinsame Lernprozesse zwischen den Lernenden.	

<i>Item-ID</i>	<i>Original</i>	<i>Übersetzung</i>	<i>Subscale</i>
wUQ30	The teacher asks cognitively appropriate questions	Der Lehrer zeigte ein angemessenes Frageverhalten (nicht zu kleinschrittig).	<i>cognitive activation</i>
wUQ31	The teacher presents challenging tasks	Der Lehrer stellte kognitiv herausfordernde Aufgaben.	
wUQ32	The teacher spontaneously encourages cognitive challenges	Der Lehrer gab geeignete Impulse, z.B. ein Problem, offene Frage.	
wUQ34	The teacher asks for students' beliefs concerning the topic	Der Lehrer fragte nach den Vorstellungen der Lernenden zum Thema.	
wUQ35	The students explain the task in their own words	Die Schülerinnen und Schüler erklärten das Thema in eigenen Worten.	
wUQ39	Knowledge is developed coconstructively in class	Es fand eine gemeinsame Konstruktion neuen Wissens statt.	
wUQ40	Cognitively challenging teaching methods are used	Es wurden kognitiv aktivierende Methoden verwendet.	
wUQ41	The teacher provides enough time to think about the tasks	Der Lehrer gab angemessene Zeit zum Nachdenken im Plenum	
wUQ42	The teaching methods correspond to the content and the class	Die Methodenwahl war dem Inhalt und den Lernenden angemessen.	
wUQ46	The teacher analyzes students' errors and misconceptions	Der Lehrer analysierte Schülerfehler und fehlerhafte Denkweisen.	
wUQ48	The teacher does not make any content-related or formal mistakes	Der Lehrer machte keine inhaltlichen oder formalen Fehler.	
wUQ52	The teacher is precise concerning mathematical language and notation	Der Lehrer initiierte die angemessene Verwendung von Fachsprache.	
wUQ49	Teacher's explanations are wellstructured and precise	Der Lehrer war bei sprachlichen und schriftlichen Formulierungen präzise.	
wUQ51	The teacher explains slowly, especially when difficulty arises	Es wurde Zeit für die Begriffsbildung genutzt.	
wUQ59	The teacher focuses on the fundamental mathematical aspects	Der Lehrer beschränkte sich auf das Wesentliche.	
wUQ60	Teacher's explanations are suitable for the students	Der Lehrer erklärte adressatengerecht.	
wUQ56	Mathematical generalizations are developed	Es fanden Verallgemeinerungen statt.	<i>teaching-related quality</i>
wUQ50	Multiple representations are used during the lesson	Es wurden verschiedene Darstellungsformen verwendet.	
wUQ53	The teacher provides connections to students' everyday life	Es wurden Bezüge zur Lebenswelt der Schülerinnen und Schüler aufgezeigt.	
wUQ54	The teacher addresses the relevance of content	Es wurde thematisiert, wozu die Inhalte gut sind.	
wUQ55	Students may bring in their own experience and interests to class	Schülerinnen und Schüler konnten eigene Erfahrungen und Interessen einbringen.	

#### 4.4.2 Skala der reflexionsbezogenen Selbstwirksamkeit (aus Lohse-Bossenz et al., 2018)

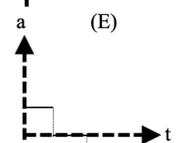
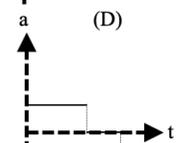
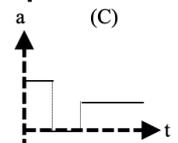
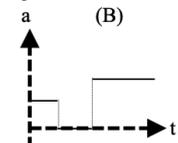
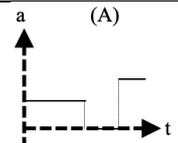
<i>Item-ID</i>	<i>Item</i>	<i>Subscale</i>
UA1	Auch wenn mehrere Kinder in eine Situation involviert waren, konnte ich mich an Handlungen einzelner Kinder erinnern.	
UA2	Obwohl die Situation schon vergangen ist, kann ich mich an die Handlungen gut erinnern.	Bes
UA7	Mir gelang es, die Unterrichtssituation genau zu beschreiben.	
UA3	Ich konnte gut einschätzen, inwiefern die Handlungen geeignet waren, die Unterrichtssituation positiv zu beeinflussen.	
UA5	Obwohl an der Unterrichtssituation mitunter mehrere Schülerinnen und Schüler beteiligt waren, gelang es mir, die Aktionen einzelner Kinder zu beurteilen.	Bew
UA10	Obwohl die Unterrichtssituation vielschichtig war, ist es mir gut gelungen, eine Bewertung der Situationen vorzunehmen.	
UA9	Mir ist es gut gelungen, Alternativen für die Gestaltung der Unterrichtssituation zu entwickeln.	
UA12	Obwohl die Unterrichtssituation komplex war, konnte ich mir gut überlegen, wie man beim nächsten Mal anders reagieren könnte.	Alt
UA8	Ausgehend von dieser Unterrichtssituation fiel es mir leicht, allgemeine Ansätze zur Gestaltung von Unterrichtssituationen abzuleiten.	
UA11	Obwohl es viele Möglichkeiten gab, seine Handlungen in der Unterrichtssituation zu ändern, gelang es mir gut, mich auf ein spezifisches Veränderungsziel festzulegen.	Kons
UA4	Es fiel mir leicht, Erklärungen für Handlungen einzelner Schülerinnen und Schüler zu formulieren.	
UA6	Trotz der Komplexität der Unterrichtssituation, konnte ich einzelne Gründe für die Handlungen identifizieren.	Beg

A4.3 Fachwissen – Skala (aus dem Mechanics Baseline Test)

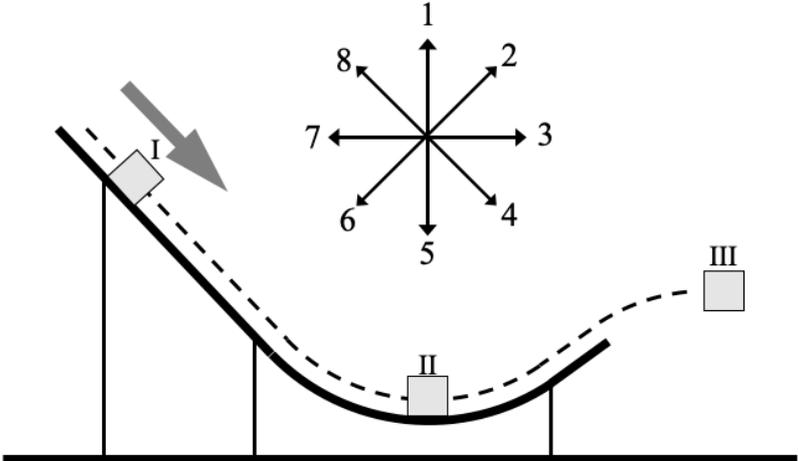
Item-ID	Item	Antworten
Strob1	<p>Beziehen Sie sich auf folgende Abbildung, wenn Sie die ersten beiden Fragen beantworten. Das Diagramm stellt eine Stroboskop-Aufnahme eines Körpers dar, welcher sich entlang einer horizontalen Ebene bewegt. Die Positionen, wie sie im Diagramm eingezeichnet sind, sind durch jeweils gleichgroße Zeitabschnitte voneinander getrennt. Die erste Aufnahme ist zu dem Zeitpunkt gemacht, an dem der Körper mit seiner Bewegung beginnt. Die letzte Aufnahme zeigt den Körper, als er gerade wieder zur Ruhe gekommen ist.</p> <div style="text-align: center;">  </div> <p>1) Welcher der folgenden Graphen stellt am besten die Geschwindigkeit des Körpers als Funktion der Zeit dar?</p>	<div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 10px;"> <p>V (A)</p>  </div> <div style="margin-bottom: 10px;"> <p>V (B)</p>  </div> <div style="margin-bottom: 10px;"> <p>V (C)</p>  </div> <div style="margin-bottom: 10px;"> <p>V (D)</p>  </div> <div style="margin-bottom: 10px;"> <p>V (E)</p>  </div> <p>„Weiß ich nicht“</p> </div>

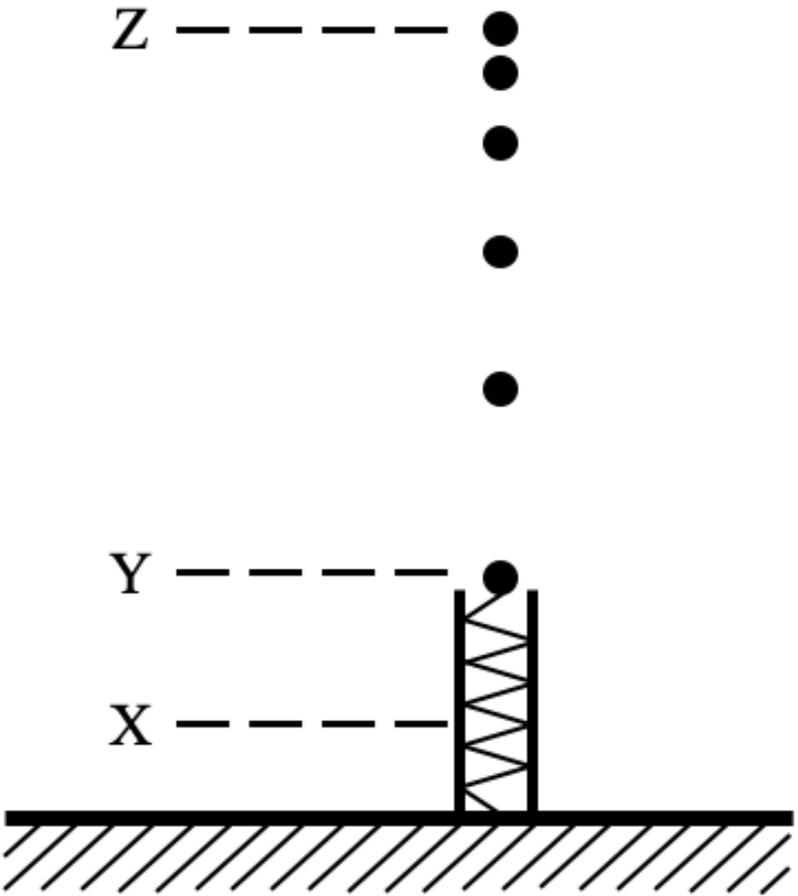
**Strob2**

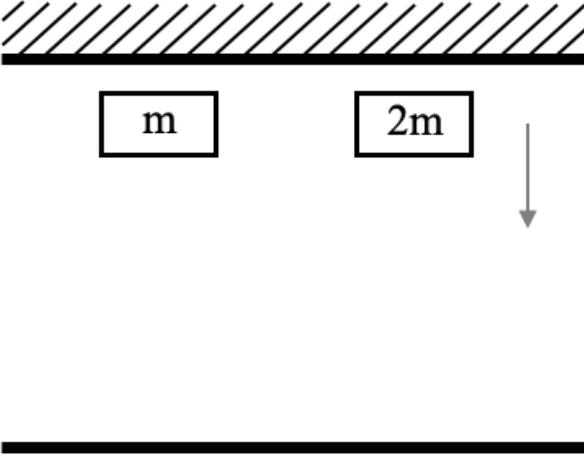
2) Welcher der folgenden Graphen stellt am besten die Beschleunigung des Körpers als Funktion der Zeit dar?



„Weiß ich nicht“

<p><b>Klotz1</b></p>	<p>Beziehen Sie sich auf die untere Grafik, wenn Sie die nächsten drei Fragen beantworten. Die Grafik stellt einen Klotz dar, der auf einer reibungsfreien Rampe gleitet. Die acht nummerierten Pfeile repräsentieren die Richtungen, auf die in den Antworten zu den Fragen Bezug genommen wird.</p>  <p>3) Durch welchen Pfeil wird die Richtung der Beschleunigung des Klotzes in Position I am besten repräsentiert?</p>	<p>„1“  „2“  „4“  „5“  „Keiner der Pfeile die Beschleunigung ist Null“  „Weiß ich nicht“</p>
<p><b>Klotz2</b></p>	<p>4) Durch welchen Pfeil wird die Richtung der Beschleunigung des Klotzes in Position II am besten repräsentiert?</p>	<p>„1“  „3“  „5“  „7“  „Keiner der Pfeile die Beschleunigung ist Null“  „Weiß ich nicht“</p>
<p><b>Klotz3</b></p>	<p>5) Durch welchen Pfeil wird die Richtung der Beschleunigung des Klotzes in Position III (nach Verlassen der Rampe) am besten repräsentiert?</p>	<p>„2“  „3“  „5“  „6“  „Keiner der Pfeile die Beschleunigung ist Null“  „Weiß ich nicht“</p>

<p><b>Senkwurf</b></p>	<p>Beziehen Sie sich auf die untere Grafik, wenn Sie die nächste Frage beantworten.</p>  <p>Die Zeichnung stellt eine Stroboskopaufnahme eines Balls dar, der von einer Feder senkrecht nach oben geschossen wird. Die Feder, mit dem Ball obenauf, wurde ursprünglich bis zum Punkt X zusammengedrückt und dann losgelassen. Der Ball verließ die Feder im Punkt Y und erreicht in Z seinen höchsten Punkt.</p> <p>6) Es wird vorausgesetzt, dass der Luftwiderstand vernachlässigt werden kann.</p>	<p>„Die Beschleunigung des Balls war kurz vor Erreichen des Punktes Y (immer noch im Kontakt mit der Feder) am größten“</p> <p>„Die Beschleunigung des Balles nahm von Punkt Y zum Punkt Z hin ab“</p> <p>„Die Beschleunigung des Balles im Punkt Z war 0“</p> <p>„Alle obigen Antworten sind richtig“</p> <p>„Die Beschleunigung des Balles auf seiner Bahn von Punkt Y zu Punkt Z ist in jedem Punkt gleich groß“</p> <p>„Weiß ich nicht“</p>
------------------------	--	---

<p><b>FF1</b></p>	<p>Beziehen Sie sich auf die untere Grafik, wenn Sie die nächsten drei Fragen beantworten.</p>  <p style="text-align: center;"><b>Boden</b></p> <p>Zwei Gegenstände sind zunächst an einer Decke aufgehängt und werden dann losgeschnitten. Beide Gegenstände fallen zu Boden (Bewegungsrichtung siehe Pfeil). Der eine Gegenstand hat dabei die doppelte Masse wie der andere Gegenstand. Der Luftwiderstand wird vernachlässigt.</p> <p>7) Was gilt für die Fallzeit?</p>	<p>„Der leichtere Gegenstand trifft zuerst auf dem Boden auf“          „Der schwerere Gegenstand trifft zuerst auf dem Boden auf“          „Beide Gegenstände treffen zu gleicher Zeit auf dem Boden auf“          „Ohne weitere Informationen zum Volumen der Gegenstände kann diese Frage nicht beantwortet werden“          „Weiß ich nicht“</p>
<p><b>FF2</b></p>	<p>8) Für den leichteren Gegenstand gilt Folgendes:</p>	<p>„In gleichen Zeitintervallen wird jeweils der gleiche Weg zurückgelegt“          „Die Geschwindigkeit ist zeitunabhängig“          „In gleichen Wegintervallen wird jeweils die gleiche Zeit benötigt“          „Die Beschleunigung ist zeitunabhängig“          „Weiß ich nicht“</p>

#### A4.4 Wert – Skala (Eigenentwicklung)

<i>Item-ID</i>	<i>Item</i>
Wert1	Das Nachdenken über eine Unterrichtssituation, wie eben geschehen, halte ich für interessant.
Wert2	Ich denke, dass es sinnvoll ist, so wie eben geschehen über Unterrichtssituationen nachzudenken.
Wert3	Ich halte es für wichtig, so wie eben geschehen über Unterrichtssituationen nachzudenken.
Wert4	Es ist aus meiner Sicht nützlich, so wie eben geschehen über Unterrichtssituationen nachzudenken.

## **A5 Tabellenverzeichnis**

- Tab. 1 Skalen, Itemanzahl, Beispielitem und Antworten
- Tab. 2 Korrelationen der Qualitätsdimensionen
- Tab. 3 Korrelationsmatrix zwischen Qualitätsdimensionen und Elementen
- Tab. 4 Korrelationen zwischen LOS und Länge zu den Qualitätsdimensionen
- Tab. 5 Qualitätsdimensionen nach Extremgruppen (bei  $N = 134$  ist pro Extremgruppe  $n = 27$ )
- Tab. 6 Faktorenanalyse nach Promax-Verfahren (oblimin)
- Tab. 7 Aufklärung der Varianz nach Variablen
- Tab. 8 Korrelationen zwischen Faktoren und zugeordneten Variablen ( $p = .001$ )
- Tab. 9 Ausreißertexte und zugehörige Faktoren bzw. Variablen
- Tab. 10 Korrelationen zwischen der Distanz zum höchsten Qualitätskorrelat und Beschreibungen\_%, LOS, Länge und Qualität
- Tab. 11 Gegenüberstellung der Qualitätsbewertung zur modellierten Distanz
- Tab. 12 Deskriptive Statistik der erhobenen Skalen (inklusive der Abweichung des Mean vom Statistischen Mittelwert einer Skala bzw. zur Skalenmitte bei Subskalen)
- Tab. 13 Lösungen einzelner Fachwissensitems
- Tab. 14 Korrelationen reflexionsbezogener Amplifiers & Filters mit ausgewählten Qualitätskorrelaten
- Tab. 15 Subskalen der rSWK nach Lohse-Bossenz et al. (2019) bezogen auf Elemente
- Tab. 16 Korrelationen zwischen absoluten bzw. relative Elementhäufigkeiten und den Amplifiers & Filters
- Tab. 17 Faktorenbestimmung nach ortho und obli in Korr. zu Amplifiers & Filters
- Tab. 18 Subskalen der wahrgenommenen Unterrichtsqualität
- Tab. 19 Subskalen der reflexionsbezogenen Selbstwirksamkeit

## **A6** *Abbildungsverzeichnis*

- Abb. 1 Modellierung der für den Lehrerberuf spezifizierten Wissenskomponenten des allgemeinen Modells professioneller Handlungskompetenz (Baumert & Kunter, 2006)
- Abb. 2 RCM of PCK for teaching science ohne Wissensfacetten (Carlson et al., 2019)
- Abb. 3 Reflexivität als Kompetenz (von Aufschnaiter et al., 2019, S. 152)
- Abb. 4 Einbettung des Modells für Reflexivität als Kompetenz (nach von Aufschnaiter et al., 2019) in das Refined Consensus Modells of Pedagogical Content Knowledge (Carlson et al., 2019) unter Berücksichtigung der Operationalisierung von Reflexion nach Nowak et al. (2019)
- Abb. 5 Visualisierung der Studienstruktur in Grundlagen, Erhebung und Analyse
- Abb. 6 Schema der satzweisen Klassifizierung einer schriftlichen Reflexion unter Verwendung eines vortrainierten ML-Modells
- Abb. 7 Schema zur Berechnung des Level of Structure (LOS)
- Abb. 8 Relativer Anteil der Diskurselemente über alle Texte
- Abb. 9 Verteilung der LOS über alle 134 schriftlichen Reflexionen und die drei Qualitätsgruppen in Gegenüberstellung der LOS-Verteilung aus Mientus et al. (im Druck)
- Abb. 10 Mittlere Unterschiede der Textlänge der Extremgruppen
- Abb. 11 Exemplarische Visualisierung der Textstruktur
- Abb. 12 Korrelationspanel der Textlänge und des prozentualen Anteils der Reflexionselemente
- Abb. 13 Korrelationspanel der LOS und der absoluten Anzahl der formulierten Reflexionselemente
- Abb. 14 Scree Plot – Faktorenanalyse
- Abb. 15 Cook-Distanzen innerhalb des Trainingsdatensatzes mit Kennzeichnung der Ausreißertexte
- Abb. 16 Korrelationsmatrix der drei Faktoren unter Einbezug der Ausreißer
- Abb. 17 Lineare (blau) versus quadratische (rot) Regression im Quantität-Qualität-Streudiagramm
- Abb. 18 Qualität nach Beschreibungen\_% mit linearer (blau) und quadratischer (rot) Regression
- Abb. 19 Qualität nach LOS mit linearer (blau) und quadratischer (rot) Regression
- Abb. 20 Qualität nach Textlänge mit linearer (blau) und quadratischer (rot) Regression

- Abb. 21 Korrelationsplott zwischen Qualität und modellierter Distanz zum höchsten Qualitätskorrelat (inkl. Regressionsgerade)
- Abb. 22 Z-standardisierte Qualitätskorrelate Länge, Beschreibung\_% und LOS aller Fremdre reflexionen zum höchsten Qualitätskorrelat (roter Punkt), unter Einbezug aller Reflexionstexte des Testdatensatzes (blau)
- Abb. 23 Histogramm der Fachwissensskala mit Indiz für Ausreißeritem
- Abb. 24 Fachwissensitem „Klotz 3“ und zugehöriges Antwortverhalten