Ziel der Studie war es, das Konstrukt des guten Erklärens im Fach Deutsch zu konzeptualisieren und empirisch zu erforschen. In FALKE-D wurden dazu Erklärungen zu deutschtypischen Themen (aus dem Bereich Argumentieren) erstellt und unterschiedlichen Statusgruppen (Schüler/-innen, Studierenden, Lehrkräften sowie Fachdidaktiker/-innen) zur Bewertung vorgelegt. Dabei konnte nachgewiesen werden, dass die aus der Theorie abgeleiteten Aspekte (z.B. Strukturiertheit) sowie der für die Studie konzeptualisierte und operationalisierte Aspekte der Erklärtiefe entscheidend für die Bewertung einer Erklärung im Fach Deutsch sind. Diese Aspekte sind also bei der Gestaltung einer Erklärung zu berücksichtigen. Die Ergebnisse aus FALKE-D werden in den Nachfolgeprojekten FALKE2 und FALKEdigital für die Lehrkräfteausbildung nutzbar gemacht.

Universität Regensburg
UNIVERSITÄTSBIBLIOTHEK

ISBN 978-3-88246-463-4
9 783882 464634

Gut erklärt?
FALKE-D: Eine empirische Studie zu Erklärungen im Deutschunterricht aus unterschiedlichen Beobachterperspektiven
Lisa Gaier

Lisa Gaier

Gut erklärt?

FALKE-D: Eine empirische Studie zu Erklärungen im Deutschunterricht aus unterschiedlichen Beobachterperspektiven

Eine Publikation der
Universitätsbibliothek Regensburg

Gut erklärt?

FALKE-D: Eine empirische Studie zu Erklärungen im Deutschunterricht aus unterschiedlichen Beobachterperspektiven

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Philosophischen Fakultät IV (Sprach-, Literatur- und Kulturwissenschaften)
der Universität Regensburg

vorgelegt von

Lisa Gaier
aus Regensburg

bei der Fakultät
Sprach-, Literatur- und Kulturwissenschaften

Regensburg 2020

Erstgutachter (Betreuer): Prof. Dr. Anita Schilcher

Zweitgutachter: Prof. Dr. Markus Pissarek

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten sind im Internet unter http://dnb.dnb.de abrufbar.

ISBN (Print): 978-3-88246-463-4

ISBN (PDF): 978-3-88246-464-1

DOI: 10.5283/epub.51210

Link zur Online-Version: https://epub.uni-regensburg.de/51210

Satz und Layout: Lisa Gaier

Umschlaggestaltung: Peter Brünsteiner

Herstellung: Universitätsbibliothek Regensburg

Erscheinungsort: Regensburg, 2022

Druck und Bindung: Digital Print Group o. Schimek GmbH, Nürnberg

Dieses Werk ist unter der Creative Commons-Lizenz Namensnennung 4.0 International (CC BY 4.0) veröffentlicht.

Für Uta und Hans, für das Glück, verwurzelt zu sein.

Für Johanna; niemand hat mir in meinem Leben mehr beigebracht als du, oft ganz ohne zu erklären.

Y para Juan Diego, mi roble fuerte y estable. Gracias por recordarme la belleza del momento.

Inhalt

1 Einleitung 7

1.1 Hinführung zum Thema *Erklärungen im Deutschunterricht* 8

1.2 Erkenntnisinteresse und Verortung in der Lehr-/Lernforschung 9

1.3 Aufbau der Studie FALKE-D 14

2 Aspekte professionellen (Deutsch-) Unterrichts 14

2.1 Qualität von (Deutsch-) Unterricht 15

2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht 17

2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht 25

2.2 Professionelle (Deutsch-) Lehrkräfte 32

2.2.1 Paradigmen zur professionellen Lehrkraft 36

2.2.2 Professionswissen von Lehrkräften 40

2.3 Erklärkompetenz von Lehrkräften 47

2.4 (Professionelle) Wahrnehmung von Unterrichtsqualität 51

2.4.1 Beobachterperspektive: Wahrnehmung der Schüler/-innen 51

2.4.2 Beobachterperspektive: Wahrnehmung der Expertinnen und Experten 54

2.4.3 Methoden und Zweck der Erfassung von Wahrnehmung 57

2.4.4 Übereinstimmung der unterschiedlichen Perspektiven 60

3 Theoretische Sichtweisen auf *Erklären* 62

3.1 (Natur-) Wissenschaftliche Sicht auf Erklärungen 62

3.2 Erklären als Sprachhandlung 64

3.3 Lernpsychologische Sicht auf Erklärungen 68

3.4 Erklärungen im (Deutsch-) Unterricht 75

4 Erklären in FALKE(-D) 81

4.1 Das Konstrukt *Gutes Erklären* 81

4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens 82

4.2.1 Strukturiertheit 84

4.2.2 Adressatenorientierung 87

4.2.3 Sprachliche Verständlichkeit 91

4.2.4 Sprech- und Körperausdruck 93

4.2.5 Erklärtiefe in FALKE-D 94

4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen 96

5 Konstruktion der Videos als Stimuli 98

5.1 Grundsätze bei der Operationalisierung von Erklärungen in FALKE-D 99

5.1.1 Themenfindung 100

5.1.2 Allgemeine Verortung „Arbeit mit einem argumentativen Text“ 102

5.1.3 Erklärprozeduren *erklären, was/ warum/ wie* 103

5.1.4 Variation der Erklärtiefe ... 105
5.2 Was-Erklärungen ... 108
5.2.1 Variante 1: Etymologischer Hintergrund ... 108
5.2.2 Variante 2: Synonyme Ausdrücke ... 110
5.3 Warum-Erklärungen ... 112
5.3.1 Variante 1: Funktional-inhaltliche Begründung ... 115
5.3.2 Variante 2: Strukturelle Begründung ... 118
5.4 Wie-Erklärungen ... 120
5.4.1 Variante 1: Methode Lautes Denken ... 121
5.4.2 Variante 2: Methode Modelltext ... 128
6 Fragestellungen ... 131
7 Forschungsmethode ... 133
7.1 Design der Studie ... 133
7.2 Erstellen und Auswahl der Videovignetten als Stimuli ... 138
7.3 Untersuchungsinstrument: Online-Fragebogen ... 139
7.4 Statusgruppen und Erhebung ... 145
7.4.1 Expertinnen und Experten ... 145
7.4.2 Adressaten: Schüler/-innen ... 152
7.4.3 Überblick über Ablauf der Erhebung ... 154
7.5 Operationalisierung des Aspektes *Erklärtiefe* ... 156
7.5.1 Items Was-Erklärungen ... 158
7.5.2 Items Warum-Erklärungen ... 159
7.5.3 Items Wie-Erklärungen ... 161
8 Quantitative Ergebnisse ... 162
8.1 Qualitätssicherung ... 163
8.1.1 Pilotierung ... 163
8.1.2 Psychometrische Gütekriterien ... 169
8.2 Deskriptive Ergebnisse ... 173
8.2.1 Globalurteile pro Videovignette ... 174
8.2.2 Skalenebene: Bewertungen der Aspekte guten Erklärens ... 181
8.2.3 Zusammenfassung der deskriptiven Ergebnisse ... 197
8.2.4 Bedeutung der variierten Erklärtiefe für die Statusgruppen ... 198
8.3 Inferenzstatistische Ergebnisse ... 200
8.3.1 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil ... 201
8.3.2 Zusammenhänge zwischen dem Globalurteil und den Aspekten guten Erklärens ... 203
8.3.3 Zusammenfassung der inferenzstatistischen Ergebnisse ... 217
9 Qualitative Ergebnisse ... 218
9.1 Vorgehen bei der qualitativen Inhaltsanalyse ... 219
9.2 Vorgehen bei der Kategoriebildung ... 220

9.2.1 Deduktive Kategoriebildung ... 220
9.2.2 Induktive Kategoriebildung ... 224
9.3 Gütekriterien ... 229
9.4 Ergebnisse der qualitativen Analysen ... 232
9.4.1 Statusgruppenübergreifende Ergebnisse der qualitativen Analysen ... 232
9.4.2 Statusgruppenspezifische Ergebnisse der qualitativen Analysen ... 241
9.4.3 Ergebnisse der qualitativen Analysen unterschieden nach Erklärtiefe ... 249
9.4.4 Zusammenfassung der qualitativen Ergebnisse ... 254
10 Diskussion der Ergebnisse zum Konstrukt des *Guten Erklärens* ... 256
10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse ... 256
10.1.1 Erklären als Teil der Lehrerkompetenz ... 257
10.1.2 Multiperspektivität: Statusgruppenübergreifende und statusgruppenspezifische Ergebnisse ... 257
10.1.3 Erklären im Deutschunterricht ... 262
10.1.4 Bedeutung der Erklärtiefe für Erklärungen im Deutschunterricht ... 263
10.2 Verallgemeinerbarkeit der Ergebnisse aus FALKE-D ... 265
10.3 Grenzen der Studie FALKE-D ... 267
11 Fazit und Ausblick ... 268
11.1 Anregungen aus FALKE-D für weitere Forschungsarbeiten ... 269
11.2 Ausblick auf FALKE II ... 271
Tabellen- und Abbildungsverzeichnis ... 278
Literaturverzeichnis ... 281
Anhang ... 294

1 Einleitung

Vorliegende Arbeit widmet sich mit dem Erklären und der Erklärkompetenz von Lehrkräften im Kontext von Schulunterricht einem Thema, das gleichermaßen als unerlässlich erachtet sowie als überholt abgelehnt wird. Fragt man nach den Eigenschaften einer guten Lehrkraft, so herrscht landläufig Einigkeit darüber, dass die Fähigkeit, gut zu erklären, für Lehrpersonen eine der wichtigsten darstellt. Auch informelle Gespräche zum Thema guter Unterricht und gute Lehrkraft, die im Zuge dieser Arbeit sowohl mit Schülerinnen und Schülern als auch mit Studierenden des Faches Deutsch und mit Kolleginnen und Kollegen im Schuldienst und aus der Lehrerbildung geführt wurden, bestätigten das Erklären als wichtige Kompetenz, – einerseits. Andererseits brachten diese Gespräche zugleich die gegenteilige Auffassung zum Vorschein: Nicht wenige Kolleginnen und Kollegen zeigten sich über das Forschungsthema Erklären im Deutschunterricht erstaunt, zum Teil weil sie es als wenig relevant empfanden. Eine Lehrkraft sagte zum Beispiel: „Ich erkläre in meinem Unterricht eigentlich gar nicht". Ein anderer Teil der Kolleginnen und Kollegen lehnte das Erklären explizit als veraltete Methode ab: „Erklären ist für mich von gestern. Damit nimmst du den Schülern die Möglichkeit, selbst zu denken". Interessanterweise zeigt sich gerade geschilderte anekdotische Uneinigkeit auch in der Forschung.

Zum einen findet sich ein produktiver Diskurs zum Erklären in der Lehr-/Lernforschung (vgl. Punkt *2.3 Erklärkompetenz von Lehrkräften* und *3.3 Lernpsychologische Sicht auf Erklärungen* in dieser Arbeit) sowie in einzelnen naturwissenschaftlichen Fachdidaktiken, allen voran in der Physikdidaktik (hier besonders die Arbeiten von Kulgemeyer, 2017; 2018; 2019). In anderen Fächern, darunter auch in der Deutschdidaktik, gibt es in der jüngeren Forschung vereinzelte Publikationen zum Erklären (z. B. Berkemeier, 2009; Berkemeier & Harren, 2016; Ernst-Weber, 2019; Morek, 2012; Spiegel, 2007). Insgesamt ist das Thema jedoch nicht sehr präsent im deutschdidaktischen Diskurs, es findet sich zum Beispiel kaum ein Themenheft zum Erklären in den einschlägigen fachdidaktischen Zeitschriften wie *Praxis Deutschunterricht* (Westermann Verlag) und auch in den Einführungen in die Sprachdidaktik fehlen Kapitel zum Thema *Erklären* (vgl. z. B. Neuland & Peschel, 2013; Ossner, 2008; J. Wild & Wildfeuer, 2019).

Andererseits nehmen gerade in den letzten Jahren Veröffentlichungen zu Unterrichtsmaterial zum Thema *Erklären* zu. Zu nennen ist hier zum Beispiel das Themenheft *Erklären* in Deutsch 5–10 (59/2019) und das Themenheft *Erklärtexte lesen und schreiben* in Praxis Deutsch (2021) sowie zahlreiche Publikationen zum Thema Erklärvideos (s. z. B. Deutschunterricht 4/2018: *YouTube & Co. - Mit Webvideos arbeiten* oder Deutschunterricht 1/2017: *Sprachwandel* (Westermann Verlag); Praxis Deutsch 265/2017: *Deutsch per Smartphone* (Friedrich Verlag) oder Praxis Grundschule 5/2015: *Märchen* (Westermann Verlag)).

Somit sind die Auffassungen zur Relevanz des Erklärens sowohl von schulpraktischer als auch von forschungstheoretischer Seite unterschiedlich. Diese Diskrepanz macht Erklären im Deutschunterricht zu einem interessanten Forschungsgegenstand.

1.1 Hinführung zum Thema *Erklärungen im Deutschunterricht*

Gerade geschilderte heterogene Wahrnehmung des Erklärens im Deutschunterricht ist einer der Gründe, weshalb diese Arbeit das Erklären zum Forschungsgegenstand hat. Ein weiterer Grund findet sich in der Forschungsgenese vorliegender Studie.

Die Idee, Erklären zum Thema eines großen interdisziplinären Projekts zu wählen, entstand an der Universität Regensburg während der Arbeit an FALKO (**Fa**chspezifische **L**ehrer**ko**mpetenzen, Krauss, Lindl, Schilcher & Tepner, 2017), bei dem sechs Fachdidaktiken (Deutsch, Englisch, evangelische Religionspädagogik, Latein, Musikpädagogik und Physik) parallele Professionswissenstests für ihre Fächer entwarfen und diese fachspezifisch und fachübergreifend untersuchten. Als zeitgemäß und gewinnbringend erwies sich dabei insbesondere die interdisziplinäre Zusammenarbeit zwischen den Fachdidaktiken, die in dieser Größenordnung eine Ausnahme darstellt. Es entstand daher die Idee, diese Zusammenarbeit weiterzuführen und zu vertiefen (s. zur genauen Forschungsgenese Punkt *2.2.2 Professionswissen von Lehrkräften* in dieser Arbeit). Dazu bietet sich das Erklären, das im Projekt FALKO bereits als Facette des fachdidaktischen Wissens erhoben wurde, und das eine Lehrerhandlung ist, die in allen Fächern eine Rolle spielt, als inter- und transdisziplinärer Forschungsgegenstand an.

Als Nachfolgeprojekt von FALKO widmet sich daher FALKE (**Fa**chspezifische **L**ehrer**k**ompetenzen im **E**rklären) mit dem Erklären einer konkreten Lehrmethode aus der Sicht von elf Fachdidaktiken (Ästhetische Kunst und Bildende Erziehung, Biologie, Chemie, Deutsch, Englisch, evangelische Religionspädagogik, Geschichte, Grundschulpädagogik und -didaktik, Mathematik, Musikpädagogik und Physik). Da außerdem fachübergreifend die deutsche Sprachwissenschaft sowie die Sprechwissenschaft an FALKE beteiligt sind, gelang mit dem Zusammenschluss von insgesamt dreizehn Disziplinen ein multiperspektivisches Projekt von im Feld der Unterrichtsforschung bisher noch nicht dagewesener Größe. Als Nachfolgeprojekt von FALKO leistet FALKE daher zum einen eine inhaltliche Fokussierung vom Professionswissensbereich des Fachdidaktischen Wissens hin zur Wissensfacette des Erklärens, und zum anderen gerade geschilderte transdisziplinäre Ausweitung.

In den ersten Besprechungen des Gesamtprojektes FALKE zeigten sich wie oben angesprochen schnell unterschiedliche Forschungsstände zum Thema Erklären. Deswegen war es notwendig, eine gemeinsame Basis zu schaffen und das Konstrukt des guten Erklärens im Gesamtprojekt und in den einzelnen Fächern näher zu bestimmen (s. Punkt *4 Erklären in FALKE-(D)* in dieser Arbeit). Dazu

wurde ein parallel angelegtes Forschungsdesign festgelegt (s. Punkt *7.1 Design der Studie* in dieser Arbeit), das sich dem guten Erklären über die Wahrnehmung unterschiedlicher Statusgruppen annähert. Hierzu wurden Erklärungen aus jedem Fach erstellt, die diesen Gruppen zur Beurteilung vorgelegt wurden. Dabei äußerten die Teilnehmer/-innen an der Studie ihre Einschätzung zu den Erklärungen zunächst ohne Vorgaben, anschließend bewerteten sie diese nach konkreten Aspekten. Gerade diese Aspekte, die zum Großteil projektübergreifend erhoben wurden, eignen sich für den interdisziplinären Vergleich.

Vorliegende Arbeit widmet sich innerhalb des Projekts FALKE dem Erklären im Fach Deutsch. In FALKE-D (FALKE-**D**eutsch) wurden dazu mündliche Erklärprozesse von Deutschlehrkräften ins Zentrum des Interesses gestellt. Unterschiedliche Gruppen wurden zur wahrgenommenen Qualität dieser gesehenen Erklärungen befragt, um Rückschlüsse darauf zu ziehen, was eine Erklärung im Deutschunterricht zu einer guten Erklärung macht.

FALKE-D leistet somit einen Beitrag zur Verbesserung der Unterrichtsqualität im Fach Deutsch, indem Spezifika für das Erklären im Fach erarbeitet werden.

1.2 Erkenntnisinteresse und Verortung in der Lehr-/Lernforschung

Den theoretischen Hintergrund für gutes Erklären unabhängig vom Fach bildet die Lehr-/Lernforschung, die somit den Rahmen dieser Arbeit darstellt. Breite Einigkeit herrscht in diesem Fachdiskurs über das Ziel von Unterricht, nämlich Lernerfolg bei den Schülerinnen und Schülern zu erzielen, und auch die Auffassung des Lernens als eigenaktiver Prozess findet momentan breite Zustimmung: „However, they [=recent socio-constructivist approaches] emphasize important characteristics of human learning such as its active and constructive nature“ (Schnotz, 2016, S. 114). Da in FALKE-D mit dem Erklären als Lehrerhandlung eine instruktive Praxis im Fokus des Forschungsinteresses steht, scheint zunächst ein Widerspruch zu diesem vorherrschenden Paradigma des gemäßigten Konstruktivismus‘ zu bestehen.

Direkte Instruktion

Im Folgenden wird daher geklärt, was mit direkter Instruktion gemeint ist. Hattie räumt in seiner Metastudie zunächst das Missverständnis aus, direkte Instruktion mit Frontalunterricht gleichzusetzen: „Too often what the critics mean by direct instruction is didactic teacher-led talking from the front; this should *not* be confused with the very successfull „Direct Instruction“ method […]“ (Hattie, 2009, S. 205).

Die direkte Instruktion läuft nach Hattie (vgl. 2009, S. 205-206) in sieben Schritten ab, von der der Vorbereitung über die Präsentation der Inhalte bis hin zum selbstständigen Üben durch die Schüler/-innen. Während aller sieben Schritte sind Kriterien wie Strukturiertheit, Klarheit und Zielorientierung

zu beachten, um Lernerfolg bei den Schülerinnen und Schülern zu gewährleisten. Eben diese Kriterien nehmen auch bei den Erklärungen im FALKE-Projekt eine entscheidende Rolle ein (s. Punkt *4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens* in dieser Arbeit). Die Wirksamkeit der direkten Instruktion als Lehrmethode wies Hattie in seiner Metastudie mit einer mittleren Effektstärke von d=.59 nach (vgl. Hattie, 2009, S. 201).

Eine klare Definition der direkten Instruktion als Unterrichtsmethode fehlt jedoch, wie Wellenreuther (2015) schreibt: „Direkte Instruktion ist keine klar umrissene Unterrichtsmethode" (S. 360). Er nutzt direkte Instruktion in seinem Grundlagenbuch zum *Lehren und Lernen* als Oberbegriff: „für alle Unterrichtsformen […], in denen der Lehrer 1. direkt das Unterrichtsgeschehen von der ersten Aneignung komplexen Wissens bis hin zum sicheren Beherrschen von Kompetenzen lenkt und kontrolliert, und 2. in denen er die Übermittlung von Informationen, die Strukturierung von Aufgaben dazu sowie die Kontrolle der Aufgabenlösungen selbst im Sinne der Unterrichtsforschung steuert" (Wellenreuther, 2015, S. 360). Im Fokus von Instruktion steht der Lerngegenstand, die Aufgabe der Lehrkraft ist es daher, „gegenstandszentrierte[...] Lernumgebungen" (Reinmann & Mandl, 2006, S. 619) zu gestalten. In einem solchen Lehr-/Lernprozess, den Reinmann und Mandl „technologische Lehrstrategie" (2006, S. 619) nennen[1], hat die Lehrkraft folgende Aufgabe: „Wissensinhalte zu präsentieren und zu erklären sowie die Lernenden anzuleiten und ihre Lernfortschritte zu überwachen" (Reinmann & Mandl, 2006, S. 619). Dagegen ist die Rolle der/des Lernenden eher passiv und rezeptiv (vgl. ebd.), obwohl Lernen auch bei direkter Instruktion nur stattfindet, wenn die Lernenden selbst tätig werden: „Direct instruction also promotes active student involvement within learning" (Bailey, Butler, LaFramenta & Ong, 2004, S. 9).

Der Unterschied zu konstruktivistischen Auffassungen liegt vor allem in der Fokussierung des Gegenstandes bei der direkten Instruktion gegenüber der Fokussierung der Rolle der Lernenden im Konstruktivismus: „In contrast to constructivist teaching, which encourages students to participate in shaping the content and concentration of the lesson, in direct instruction the teacher explicitly teaches students skills in the desired content area […]. In addition, direct instruction methods approach content material from "part to whole"" (ebd.). In diesem Zusammenhang erscheint das Erklären, das Reinmann und Mandl (2006) wörtlich aufnehmen (s. obiges Zitat), als klar instruktive Praxis.

Auch die Erklärungen, die für vorliegende Arbeit konzipiert wurden, reihen sich in das Instruieren ein, indem sie jeweils einen deutschspezifischen Inhalt fokussieren (s. zur inhaltlichen Ausgestaltung

[1] Eine Übersicht zu unterschiedlichen Ansätzen innerhalb der technologischen Lehrstrategie findet sich bei Reinmann und Mandl, 2006, S. 619-625.

der Erklärungen die Punkte *5.2 Was-Erklärungen, 5.3 Warum-Erklärungen* sowie *5.4 Wie-Erklärungen* in dieser Arbeit). Da aber Erklärprozesse untersucht werden, werden die Erklärungen von FALKE-D nicht nur vom Gegenstand her gedacht, sondern auch von der/dem Empfänger/-in her. Analog zum Ziel des Lernerfolgs bei Schülerinnen und Schülern im Unterricht ist das Ziel der Erklärungen von FALKE-D das Verstehen bei den Adressatinnen und Adressaten. Aus diesem Grund werden im Folgenden konstruktivistische Auffassungen der direkten Instruktion gegenübergestellt, die wie gesagt die Lernenden ins Zentrum der Aufmerksamkeit stellen.

Konstruktivismus

Im Konstruktivismus wird Lernen als eigenaktiver Prozess angesehen. Ähnlich wie die direkte Instruktion ist auch der Begriff Konstruktivismus nicht klar definiert, da er eine Vielzahl an konstruktivistischen Auffassungen vereint (s. z. B. Übersicht bei Reinmann & Mandl, 2006, S. 627–637). Gemein ist diesen Auffassungen die Betonung der Gefahr, über direkte Instruktion nur träges Wissen zu vermitteln und die Schüler/-innen durch ihre passive Rolle zu demotivieren (vgl. Reinmann & Mandl, 2006, S. 625), selbst wenn anerkannt wird, dass durch direkte Instruktion Lernprozesse ausgelöst werden (s. obiges Zitat von Bailey und anderen). Die Folge ist das Ablehnen der Methoden direkter Instruktion, denn im Konstruktivismus ist zunächst das Selbstentdecken der Inhalte durch die Schüler/-innen das Gebot der Stunde. Ziel ist es letztlich, die Schüler/-innen zu selbstgesteuertem Lernen zu befähigen (vgl. Klauer & Leutner, 2012, S. 16–17).

Damit einher geht die Auffassung konstruktivistischer Ansätze, dass es nicht möglich ist, Wissen und Können transmissiv zu übermitteln: „Wissen ist weder ein äußerer Gegenstand, der sich gleichsam vom Lehrenden zum Lernenden „transportieren" lässt, noch eine getreue internale Abbildung desselben" (Reinmann & Mandl, 2006, S. 626; s. dazu z. B. auch die Gegenüberstellung bei Reusser, 2006, S. 151–152). Kennzeichnend für die konstruktivistische Auffassung des Lernens „ist die Annahme der Subjektabhängigkeit des Erkennens und des Wissens als Konstruktion des menschlichen Geistes. Erkenntnis und Wissen beruhen auf keiner Abbildung einer «gegebenen» ontologischen Wirklichkeit, sondern auf der verknüpfenden und deutenden Aktivität bedeutungsbildender Subjekte" (Reusser, 2006, S. 152). Im Fokus der Aufmerksamkeit stehen also die individuellen Konstruktionsprozesse der Lernenden und deren aktive Rolle im Lernprozess. Die Lernumgebungen, die die Lehrkraft gestaltet, müssen diesen Ansprüchen genügen und entsprechend situiert sein, um die Lernenden zu einer eigenen Auseinandersetzung mit dem Gegenstand anzuregen (vgl. Reinmann & Mandl, 2006, S. 626–627).

Diese eigenständige Auseinandersetzung mit den Inhalten ist auch das Ziel, das die Erklärungen vorliegender Arbeit auslösen sollen. Insofern untersucht FALKE-D zwar mit dem Erklären eine

instruktive Praxis, die Rolle des Lernenden wird keineswegs als passiv aufgefasst, sondern gemäß dem Konstruktivismus als aktiv.

Probleme ergeben sich auch in diesen konstruktivistischen Auffassungen des Lernens: Reinmann und Mandl (vgl. 2006, S. 634-636) nennen dabei zum Beispiel die unklare empirische Datenlage bezüglich der Wirksamkeit der Lernprozesse, eine mögliche Überforderung gerade leistungsschwacher Schüler/-innen sowie ein schlechtes Kosten-Nutzungsverhältnis, im Sinne der Relation von didaktischem und zeitlichem Aufwand und dem Lernerfolg der Schüler/-innen.

Gemäßigter Konstruktivismus

Weitgehende Einigkeit besteht daher in der aktuellen Lehr/-Lernforschung und in den Fachdidaktiken inzwischen darüber, dass weder die eine noch die andere Auffassung in „Reinform" sinnvoll für das Lernen sind. Möller hält zum Beispiel das sich ausschließende Gegenüberstellen von Konstruktion und Instruktion für einen Fehlschluss: „Missverständnisse provozierte auch die Gegenüberstellung von sogenannten „instruktiven" und „konstruktiven" Lernumgebungen" (K. Möller, 2012, S. 38). Diese Gegenüberstellung „suggeriert, dass aktive Konstruktionsprozesse der Lernenden nur in einer „konstruktiven" Lernumgebung vorkommen, während eine aktive Lehrerrolle zwangsläufig mit Passivität der Lernenden verbunden ist" (K. Möller, 2012, S. 38).

Einen Ausweg aus der nicht zielführenden Abgrenzung der beiden Haltungen stellt der sogenannte gemäßigte Konstruktivismus dar, der eine „Relativierung der Gegenüberstellung von konstruktiven und instruktiven Unterrichtsformen" (K. Möller, 2012, S. 39) anstrebt und in dem sich vorliegende Arbeit verortet. Somit löst sich auch der oben beschriebene Widerspruch zwischen dem instruktiven Element des Erklärens und der konstruktivistischen Auffassung von Lernen, in der sich das FALKE-Projekt verortet.

Reinmann und Mandl (2006) sprechen in diesem Zusammenhang von wissensbasiertem Konstruktivismus und definieren in Anlehnung an Resnick und Hall (1998) wie folgt: „Im wissensbasierten Konstruktivismus wird Lernen als eine persönliche Konstruktion von Bedeutungen interpretiert, die nur dann gelingt, wenn eine ausreichende Wissensbasis zur Verfügung steht. Zum Erwerb dieser Wissensbasis kann auf instruktionale Anleitung und Unterstützung nicht verzichtet werden" (Reinmann & Mandl, 2006, S. 638).

Deutlich wird in dieser Definition die Verbindung von konstruktivistischer Auffassung des Lernens mit instruktiven Elementen, die sich in der Lehr-/Lernforschung durchsetzt:

> Konstruktion und Instruktion lassen sich nicht nach einem Alles-oder-nichts-Prinzip realisieren. Lernen erfordert zum einen immer Motivation, Interesse und Eigenaktivität seitens der Lernenden, und der Unterricht hat die Aufgabe, ihre Konstruktionsleistungen anzuregen und zu ermöglichen. Lernen erfordert zum anderen aber auch Orientierung, Anleitung und Hilfe. Ziel muss es folglich sein, eine Balance zwischen expliziter Instruktion durch den Lehrenden und konstruktiver Aktivität des Lernenden zu finden […] (Reinmann & Mandl, 2006, S. 639).

Als Beispiel kann sowohl der Ansatz des *Cognitive Apprenticeship* dienen, ein Vorgehen, bei dem die unterschiedlichen Schritte mit instruktiven (z. B. *Modelling*) und konstruktiven (z. B. *Reflection*) Elementen gestaltet sind (vgl. Reinmann & Mandl, 2006, S. 631–633 bzw. Ausführungen unter Punkt *5.4.1 Variante 1: Methode Lautes Denken* in dieser Arbeit), oder auch der Vorschlag von Reinmann und Mandl (2006), über die Problemorientierung als Leitprinzip des Unterrichts konstruktivistische Lernumgebungen mit instruktiven Elementen zu ergänzen (s. ebd., S. 639-642). Diese Verbindung von Konstruktion und Instruktion liegt inzwischen zahlreichen Veröffentlichungen zugrunde. Es fällt auf, dass in vielen Publikationen dazu die kognitive Aktivierung der Lernenden (s. Punkte *2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht* und *2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht* in dieser Arbeit) in den Vordergrund rückt, für die die Notwendigkeit instruktionaler Unterstützung betont wird (vgl. z. B. Klauer & Leutner, 2012; Mayer, 2004; K. Möller, 2012). Möller (2012) kommt zu folgendem Schluss:

> Die Aufgabe der Lehrenden scheint es daher zu sein, in Abhängigkeit von Intelligenz und Vorerfahrungen […] diese [i.e. die Lernenden] durch eine angepasste instruktionale Unterstützung in ihren kognitiven Konstruktionen zu unterstützen. […] Das so beschriebene Verhältnis zwischen Konstruktion und Instruktion lässt sich durch die Kurzformel „Konstruktion DURCH Instruktion" beschreiben, womit betont wird, dass auf der Seite der Lernenden Konstruktionsprozesse stattfinden müssen, auf der Seite der Lehrenden es dagegen um eine passende Konstruktionsprozesse fördernde Instruktion geht (K. Möller, 2012, S. 43).

Der gemäßigte Konstruktivismus bietet somit auch für vorliegende Arbeit die optimale theoretische Grundlage: Lernen ist in FALKE-D ganz in diesem Sinne ein aktiver, konstruktiver und selbstgesteuerter Prozess. Damit er gelingen kann, sind instruktive Elemente notwendig, zum Beispiel Lehrererklärungen, die allerdings den Anspruch haben, kognitive Aktivität auszulösen (vgl. Hattie, 2009 zur *direct instruction*, S. 205-207). Sie ermöglichen dadurch ein vertiefteres Verstehen und eine bessere Merkfähigkeit als rein instruktive Erklärungen.

Fazit für FALKE-D

In dieser Arbeit wird somit zum einen dargelegt, was gutes Erklären im Deutschunterricht ist und wie es beschrieben werden kann. Dazu werden Grundsätze des guten Erklärens modelliert und ein Konstrukt entwickelt (s. Punkt *4 Erklären im FALKE(-D)* in dieser Arbeit). Die Operationalisierung dieser Erklärungen erfolgt einerseits über fachunabhängige Kriterien (s. Punkte *4.2.1-4.2.4* in dieser Arbeit), andererseits über für das Fach Deutsch entwickelte Merkmale (s. Punkte *4.2.5 Erklärtiefe in FALKE-D* und *4.2.6 Fachliche Relevanz* in dieser Arbeit). Es liegt somit erstmals ein theoretisches fachunabhängiges Modell des guten Erklärens vor, das um deutschspezifische Aspekte ergänzt wird und somit für Erklärungen aus dem Fach Deutsch tragfähig ist.

Im zweiten Teil der Arbeit wird diese Tragfähigkeit empirisch überprüft, anhand eines Designs, das wiederum fachunabhängig entwickelt wurde, aber um deutschspezifische Aspekte ergänzt wurde. Dazu wurden wie angedeutet unterschiedliche Gruppen, die mit Erklärungen im Deutschunterricht theoretisch oder praktisch „zu tun haben", zu ihrer Wahrnehmung der gesehenen Erklärqualität befragt. Im Mittelpunkt des Erkenntnisinteresses dieser Arbeit steht die Frage, wie Erklärungen im Deutschunterricht gestaltet sein müssen, damit sie als qualitätvoll für den Lernprozess von deutschlernenden Schülerinnen und Schülern wahrgenommen werden.

1.3 Aufbau der Studie FALKE-D

Im Folgenden wird der Aufbau der Studie genauer vorgestellt.

In FALKE-D werden Aspekte guten Erklärens für das Fach Deutsch nicht nur theoretisch begründet, sondern auch empirisch untersucht. Eingebettet ist die Studie in das oben skizzierte Gesamtprojekt FALKE mit seinen elf Fachdidaktiken sowie der deutschen Sprachwissenschaft und der Sprechwissenschaft. Pro Fach wurden dazu sechs kurze Erklärsituationen videografiert, die in einen für alle beteiligten Didaktiken im Wesentlichen identischen online-basierten Fragebogen implementiert wurden (vgl. Punkt *7.3 Untersuchungsinstrument: Online-Fragebogen* in dieser Arbeit). Die Qualität der gezeigten Erklärungen wurde anschließend von vier Statusgruppen (Schüler/-innen, Lehramtsstudierende, Lehrkräfte und Didaktiker/-innen[2]) beurteilt. Alle Studienteilnehmer/-innen gaben zunächst eine Globalbeurteilung (als Gesamtnote der didaktischen Qualität) ab. In einem zweiten Durchlauf beurteilten sie dieselben Videos[3] noch einmal mit Bezug auf Kriterien guten Erklärens wie beispielsweise *Strukturiertheit* oder *Adressatenorientierung* (vgl. Punkt *4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens* in dieser Arbeit).

Das Ziel des Gesamtprojekts FALKE ist es, Gemeinsamkeiten und Unterschiede guten Erklärens in den verschiedenen Unterrichtsfächern empirisch zu identifizieren. Das Ziel von FALKE-D ist es, diese Merkmale für das Fach Deutsch zu spezifizieren.

2 Aspekte professionellen (Deutsch-) Unterrichts

Die Lehr-/Lernforschung stellt wie dargelegt den Rahmen dieser Arbeit dar. Das Ziel des Unterrichts, nämlich der Lernerfolg der Schüler-/innen, steht im Erkenntnisinteresse der Unterrichtsforschung, dazu werden die Ursachen für qualitätvollen Unterricht empirisch erforscht. Zur Frage, wie dieses

[2] Nach intensiven Diskussionen wurde sich im Gesamtprojekt FALKE auf den Begriff *Didaktiker/-innen* für diese Statusgruppe geeinigt. Gemeint sind für FALKE-D immer Fachdidaktiker/-innen für das Fach Deutsch.

[3] Die Begriffe Video, Vignette sowie Videovignette werden in Bezug auf die für FALKE-D erstellten Stimuli in dieser Arbeit synonym verwendet.

Ziel erreicht werden kann, haben sich in der jüngeren Forschung drei Dimensionen, nämlich *die konstruktive Unterstützung, die effektive Klassenführung* und *das Potential zur kognitiven Aktivierung* etabliert, die maßgeblich zur Qualität von Unterricht beitragen (s. Punkt *2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht* in dieser Arbeit).

Nach Duffy (2014) ist das Erklären ein zentraler Bestandteil von Unterricht. Auch in FALKE-D wird es als Mikroteaching-Element, das heißt, als Teil des Unterrichtsganzen, aufgefasst, das – sofern es gelingt – zur Unterrichtsqualität beiträgt. Wie Erklären genau funktioniert, wird in Punkt *3 Theoretische Sichtweisen auf Erklären* erläutert. Zunächst werden allerdings die gerade genannten Dimensionen von Unterrichtsqualität dargelegt, und gezeigt, inwieweit sie auf das Mikroteaching-Element Erklären übertragen werden können.

2.1 Qualität von (Deutsch-) Unterricht

Stellt die Lehr-/Lernforschung die Frage nach Kriterien für Unterrichtsqualität über die Fächer hinweg, so ist die Frage danach, was guten Deutschunterricht ausmacht, seit jeher zentraler Gegenstand der deutschdidaktischen Forschung.

Kompetenzorientierung im Deutschunterricht

Dabei lassen sich grundsätzlich zwei Herangehensweisen in der Deutschdidaktik unterscheiden: Traditionell stehen die Inhalte des Faches Deutsch und die Frage, wie diese modelliert werden, im Zentrum des Interesses der Fachdidaktik, also auf der einen Seite „die Lösung von Kanonproblemen" (Abraham & Rothgangel, 2017, S. 15), auf der anderen Seite „die Aufgabe […] Unterrichtskonzepte zu entwickeln, zu erproben und zu evaluieren" (Abraham & Rothgangel, 2017, S. 15).

Mit der Kompetenzorientierung, die im Zuge der Auseinandersetzung mit den großen internationalen Schulleistungstests wie PISA (vgl. Reiss, Weis, Klieme & Köller, 2019) und IGLU (vgl. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich, 2017) längst in die Bildungsstandards von Deutschland sowie in die Bildungs- und Lehrpläne der Bundesländer Einzug gehalten hat, nimmt die Deutschdidaktik vermehrt die fachlichen Lehr-/Lernprozesse in den Blick, so dass sich „innerhalb der Fachdidaktiken eine fachbezogene Lehr- und Lernforschung" (Abraham & Rothgangel, 2017, S. 15) herausgebildet hat. Die Schreibdidaktik widmet sich zum Beispiel nicht mehr den Merkmalen einzelner Textsorten, sondern untersucht Schreibprozesse und die dazu benötigten Text- und Schreibprozeduren. Diese Fokussierung geht nicht damit einher, die Bedeutung der genuinen Inhalte des Faches Deutsch als weniger wichtig zu erachten. Es rücken vielmehr die zu vermittelnden Kompetenzen in den Fokus, die aber selbstverständlich nur anhand eines fachlichen, das heißt in diesem Fall deutschspezifischen Gegenstandes vermittelt und erlernt werden können. Bezugswissenschaften der Deutschdidaktik

müssen daher immer die Fachwissenschaften, also Sprachwissenschaft, Literaturwissenschaft, Medienwissenschaft und so weiter, und die Bildungswissenschaften, also Psychologie, Pädagogik, Kognitionswissenschaften und so weiter, sein.

Es ist somit Aufgabe der Deutschdidaktik, die verschiedenen Kompetenzen, die der Deutschunterricht in den unterschiedlichen Lernbereichen zu vermitteln hat (d. h. im Wesentlichen Lesen, Schreiben, literarische Kompetenz, Sprechen und Zuhören sowie Sprachreflexion), zu definieren und gleichzeitig die bestmöglichen Vermittlungswege zu erforschen, um guten Deutschunterricht zu gewährleisten (vgl. Abraham & Rothgangel, 2017; Schilcher & Krauss, 2019). In dieser Auseinandersetzung mit der Qualität von Deutschunterricht ist die Suche nach geeigneten Kriterien und Messmethoden essenziel und dringend notwendig, denn es herrscht Uneinigkeit im Fachdiskurs, ob diese a) überhaupt nötig sind und b) darüber, welche Kriterien und Messmethoden geeignet sind. Zum einen ist man sich im Fachdiskurs uneins, ob sich die Deutschdidaktik überhaupt verstärkt den Kompetenzen und deren empirischer Erforschung widmen soll oder ob es nicht die zentrale Aufgabe der Deutschdidaktik bleiben muss, sich hermeneutisch mit den Inhalten des Faches zu beschäftigen (vgl. die Auseinandersetzung im Nachgang des letzten Symposions Deutschdidaktik im Mitgliederbrief des SDD 57, S. 20). Zum anderen herrscht im deutschdidaktischen Diskurs Dissens darüber, was eine gute Lehrkraft wissen und können muss (vgl. die Darstellung unterschiedlicher Testungen von Professionswissen sowie die Konsequenzen daraus bei Bremerich-Vos, 2019). Die Frage danach, was die Qualität von Deutschunterricht ausmacht, ist also ebenso virulent, wie die Frage, wie qualitätvolles Erklären im Deutschunterricht auszusehen hat.

Guter, effektiver und qualitätvoller Deutschunterricht

Für eine zunächst theoretische Annäherung an den Themenkomplex „Unterrichtsqualität" ist in diesem Zusammenhang die Unterscheidung zwischen „gutem" und „effektivem" Unterricht gewinnbringend. Diese hat in der angelsächsischen Forschung eine lange Tradition (vgl. Berliner & Rosenshine, 1987 sowie Berliner, 2005) und wird von Holzberger und Kunter (vgl. 2016, S. 40) im deutschsprachigen Diskurs aufgegriffen. Dementsprechend folgt guter Unterricht bestimmten normativen Vorstellungen zur Organisation und Umsetzung von Lernprozessen und ist geprägt von pädagogischen Werten: Ein/-e gute/-r Lehrer/-in erfüllt die Erwartungen, die in der Gesellschaft an sie oder ihn gerichtet werden, um allen Schülerinnen und Schülern gerecht zu werden.

„Effektiv" ist der Unterricht wie auch die Erklärung als Mikroteaching-Element nur dann, wenn alle Schülerinnen und Schüler die Lernziele, die in diesem Unterricht angestrebt werden, tatsächlich erreichen, wenn also Lernerfolg nachgewiesen werden kann. Dabei ist es keineswegs so, dass guter Unterricht automatisch effektiv ist. Das bedeutet, dass es Lehrkräfte gibt, die sich den Normen und

pädagogischen Wertvorstellungen entsprechend verhalten, in deren Unterricht die Schüler/-innen aber wenig oder nichts lernen. Berliner (1987, S. 94) weist in diesem Zusammenhang sogar darauf hin, dass die Frage nach Lernerfolg bei der Diskussion um guten oder schlechten Unterricht (im Original „good (and poor) teaching" S. 94) gar keine Rolle spielt, eben da diese Diskussion normative Werte und Standards des „good practice" (Berliner, 1987, S. 94) abhandelt und es zum Beispiel darum geht, wie sich eine gute Lehrkraft im Unterricht verhalten soll oder wie sie die Lehrer-Schüler-Beziehung zu gestalten hat. Andersherum gilt das Gleiche: Effektiver Unterricht ist nicht automatisch guter Unterricht. Länder mit sehr autoritären Schulsystemen wie Japan, Südkorea oder China erzielen zum Beispiel sehr gute PISA-Ergebnisse, was bedeutet, dass die Schülerinnen und Schüler dort die angestrebten Lernziele erreichen und der Unterricht hohe Lernerfolge erzielt (vgl. z. B. Ergebnisse der letzten PISA-Studie aus dem Jahr 2018 bei Reiss et al., 2019). Dennoch wird man zögern, derartige Systeme, in denen die Persönlichkeit und Individualität von Schülerinnen und Schülern eine untergeordnete Rolle spielen, als gut zu bezeichnen. Auch die Tatsache, dass direkte Instruktion eine sehr wirksame Lehrmethode ist, führt eher zu Diskussionen denn zu dem Urteil, dass ein derart gestalteter Unterricht gut sei (s. Kunter & Ewald, 2016, S. 12 bzw. Ausführungen unter Punkt *1.2 Erkenntnisinteresse und Verortung* in dieser Arbeit).
Das Ziel von Forschung und Praxis muss daher qualitätvoller Unterricht sein, das heißt ein Unterricht, der sowohl gut als auch effektiv ist (s. Berliner, 2005, S. 207), und der somit gegenstandsbezogene und lernprozessbezogene Aspekte gleichermaßen berücksichtigt.

2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht

Die Modelle der Lehr-/Lernforschung, die in der Auseinandersetzung mit qualitätvollem Unterricht im Fach Deutsch herangezogen werden, werden im Folgenden genauer dargestellt.

Angebot-Nutzungs-Modell

Breit rezipiert wird in der Deutschdidaktik das Angebot-Nutzungs-Modell von Helmke (2014, S. 69–74), das leicht modifiziert auch Kunter und Trautwein (2018, S. 15–19) als Rahmenmodell von Unterricht dient. Das Angebot-Nutzungs-Modell geht ursprünglich auf Fend (1980) zurück und betont die Rolle der/des Lernenden, die/der, um etwas zu lernen, ein Lernangebot annehmen (d. h. nutzen) muss, denn diesem Modell liegt die konstruktivistische Sicht von Lernen als aktivem, selbstgesteuertem Prozess zugrunde. Für den Schulunterricht bedeutet das, dass es die Rolle der Lehrkraft ist, möglichst optimale Lerngelegenheiten anzubieten. Dieses Angebot steht im Zentrum des Interesses der Unterrichtsforschung, denn es kann von ganz unterschiedlicher Qualität sein. Das Lernangebot hängt von den Fähigkeiten und professionellen Kompetenzen der Lehrkraft ab und beeinflusst maßgeblich das Lernen der Schüler/-innen. Ob diese allerdings das Angebot tatsächlich

nutzen, obliegt aber wie erläutert jeder Schülerin beziehungsweise jedem Schüler selbst. Der Ertrag des Lernangebots (d. h. welche Kompetenzen Schüler/-innen tatsächlich erwerben) hängt somit auch von dem Lernpotenzial der Schülerinnen und Schüler ab. Unter Lernpotenzial sind zum einen kognitive Fähigkeiten wie Sprache und Intelligenz gefasst, aber auch Lernmotivation und Anstrengungsbereitschaft (vgl. hierzu Helmke, 2014, S. 71 sowie Klieme, Lipowsky, Rakoczy & Ratzka, 2006).

Kunter und Trautwein (2018) strukturieren obiges Modell in ihrer „Psychologie des Unterrichts" leicht um: Sie unterscheiden zum Beispiel klar motivationales von kognitivem Lernpotenzial der Schüler/-innen: Zu ersterem gehören zum Beispiel das Interesse am Unterricht, zu zweitem das Vorwissen oder die Intelligenz der Schülerin/ des Schülers. Beim Unterrichtsangebot differenzieren sie zwischen Sicht- und Tiefenstrukturen (vgl. Kunter & Trautwein, 2018, S. 17). Dabei sind die Sichtstrukturen alles, was leicht beobachtbar ist, zum Beispiel mit welcher Methode ein Unterrichtsangebot gemacht wird. Entscheidend für die Qualität des Angebotes und für den Lernprozess sind dagegen die Tiefenstrukturen dieses Angebots. Sie sind im Gegensatz zu den Sichtstrukturen nicht einfach beobachtbar, sondern nur Expertinnen und Experten zugänglich. So liegt zum Beispiel wie bereits angedeutet das kognitive Aktivierungspotenzial einer Aufgabe in der Tiefenstruktur des Unterrichts (da die Tiefenstrukturen in FALKE-D eine wichtige Rolle spielen, werden sie unter Punkt *2.1.2 Bedeutung der Sicht- und Tiefenstrukturen für die Qualität von Unterricht* in dieser Arbeit genauer erläutert).

Unterrichtsqualitätsdimensionen

Die Frage danach, wie das komplexe und vielschichtige Konstrukt „Unterrichtsqualität" konkret konzeptualisiert werden kann, ist zentral für die Lehr-/Lernforschung und für die Fachdidaktiken. Eine Übersicht über unterschiedliche Konzeptionen von Unterrichtsqualität geben Taut und Rakoczy (2016) sowie Seidel und Shavelson (2007). Allgemeine Modelle zur Unterrichtsqualität legen auf der Suche nach validen, reliablen und bezogen auf den Lernzuwachs der Schülerschaft auch prädiktiv validen Teilaspekten zwischen drei und fünf Teildimensionen zugrunde. So unterscheidet zum Beispiel das „Framework for Teaching" von Danielson (2007) die vier Dimensionen (im Original als „domain" bezeichnet) „Planning and Preparation", „The Classroom Environment", „Instruction" und „Professional Responsibilities" (Danielson, 2007, S. 3–4). Seidel und Stürmer (2014) hingegen nehmen die drei Dimensionen „goal clarity and orientation, teacher support and guidance [sowie] learning climate" (S. 743) in den Blick.

Taut und Rakoczy selbst legen ihrer Konzeptualisierung von Unterrichtsqualitätsdimensionen aufgrund in der Forschung vorliegender Ergebnisse folgende vier beziehungsweise fünf Faktoren zugrunde: *Klassenführung, kognitive Aktivierung, Schülerorientierung* (aufgespaltet in die

Unterdimensionen *organisatorische Abläufe* und *unterstützende Lehrer-Schüler-Beziehung*) sowie *Bewertungen im Unterricht* (vgl. Taut & Rakoczy, 2016, S. 47; eigene Übersetzung). Sie definieren diese Dimensionen aus der Lehr-/Lernpsychologie, betonen aber die notwendige fachspezifische beziehungsweise sogar themenspezifische Ausdifferenzierung aller Dimensionen, auch derjeniger, bei denen man die Notwendigkeit einer fachlichen Ausdifferenzierung zunächst nicht vermuten würde. So sei zwar der fachliche Inhalt einer Unterrichtsstunde unbedeutend für eine *effektive Klassenführung*, bei der es um geregelte Abläufe und Prozeduren geht; die Prozeduren selbst jedoch müssten an den Fachinhalt angepasst sein (vgl. Taut & Rakoczy, 2016, S. 47). Ähnliches gilt für die Dimension *Schülerorientierung*: Organisatorische Abläufe können fachunabhängig gestaltet werden, der Umgang mit Schülerfehlern ist allerdings fachspezifisch und entscheidend für eine unterstützende Lehrer-Schüler-Beziehung (vgl. Taut & Rakoczy, 2016, S. 48). Zur letzten Dimension der *Bewertungen im Unterricht* benötigt die Lehrkraft diagnostische Kompetenz, die wiederum nur vor dem Hintergrund des Fachs bestimmt werden kann. Mit diesen Überlegungen rufen Taut und Rakoczy (2016) die Fachdidaktiken auf den Plan, die zur Klärung der allgemeinen Unterrichtsqualität beitragen müssen, indem sie fachspezifische Lösungen für generische Qualitätsmerkmale definieren.

Anders als bei den meisten der oben genannten Dimensionen wird bei der *kognitiven Aktivierung* die Notwendigkeit, diese vom Fach her zu spezifizieren, immer wieder betont ebenso wie der diesbezügliche Forschungsbedarf (vgl. Taut & Rakoczy, 2016, S. 47)[4].

Diese Spezifizierung geschieht in der Fachdidaktik Deutsch: Arbeiten aus der Deutschdidaktik rezipieren die in der Unterrichtsforschung etablierten Modelle und wenden einzelne Dimensionen daraus auf den Deutschunterricht an (s. Punkt *2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht* in dieser Arbeit).

Insgesamt ist jedoch festzustellen, dass trotz Anwendung der Modelle deutschspezifische Ausdifferenzierungen der einzelnen Dimensionen – bis auf das Kriterium *kognitive Aktivierung* (s. unten) – noch ausstehen.

[4] Erwähnenswert ist in diesem Zusammenhang noch, dass Taut und Rakoczy eine eigene Studie vorstellen, in der sie ein Unterrichtsbeobachtungsinstrument evaluieren, das an deutschen Schulen eingesetzt wird und theoretisch in folgende Faktoren unterteilt ist: „maintaining students' attention; fostering students' understanding; ensuring application to real-world contexts; further developing students' intrinsic motivation" (Taut & Rakoczy, 2016, S. 49). Diese Dimensionen differieren zum Teil erheblich von den in der Lehr-/Lernforschung etablierten. Interessant ist nun, dass sich nach empirischer Überprüfung (über eine explorative sowie eine konfirmatorische Faktorenanalyse) ganz andere, nämlich wiederum in der Lehr-/Lernforschung bekannte Faktoren ergeben: „students at the center of instruction; cognitive activation; classroom climate; structured teaching and classroom management; evaluation and repetition" (Taut & Rakoczy, 2016, S. 52).

Etablierte Dimensionen von Unterrichtsqualität

Trotz unterschiedlicher Anzahl und zum Teil auch unterschiedlicher Spezifizierung der Dimensionen für qualitätvollen Unterricht haben sich in jüngerer Zeit drei grundlegende etabliert, die im Wesentlichen auf die von Klieme und anderen (2001) in der TiMS-Studie konzeptualisierten Basisdimensionen *Unterrichts- und Klassenführung, Schülerorientierung* und *kognitive Aktivierung* zurückgehen (vgl. Klieme, Schümer & Knoll, 2001, S. 51–52) und breite Zustimmung sowie weitere Ausdifferenzierung in der Lehr-/Lernforschung erfahren (vgl. den Überblick in Kunter & Trautwein, 2018 sowie in Praetorius, Klieme, Herbert & Pinger, 2018). Diese Dimensionen konnten überdies empirisch bereits in mehreren Studien nachgewiesen werden (vgl. z. B. Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Kunter et al., 2007).

Dabei erlaubt die *Unterrichts- und Klassenführung* einen möglichst störungsfreien Ablauf des Unterrichts, in dem die zur Verfügung stehende Zeit hauptsächlich zum Lernen genutzt wird (vgl. zum ganzen Absatz Klieme et al., 2001, S. 50-52). Die *Schülerorientierung* sehen Klieme und andere als Grundpfeiler für qualitätvollen Unterricht. Das bedeutet, dass die Lehrkraft die individuelle Bezugsnorm anlegt, hohe diagnostische Kompetenz aufweist sowie mit möglichst geringem Leistungsdruck und in angemessenem Tempo durch den Unterricht führt. Für die *kognitive Aktivierung* sind ein anspruchsvolles Üben und eine hohe Motivation der Lernenden notwendig. Der Vorteil dieses Modells ist, wie Klieme und andere (2006) selbst betonen, dass es geeignet scheine, „verschiedene theoretische Konzepte und Wirkungshypothesen der Unterrichtsforschung zu integrieren" (Klieme et al., 2006, S. 130). So umfasse es zum einen die direkte Instruktion sowie die Selbstbestimmungstheorie der Motivation (nach Deci & Ryan, 1985), zum anderen auch das Konzept der kognitiven Aktivität nach Mayer (2004). Im weiteren Forschungsdiskurs wurden diese Basisdimensionen immer weiter ausdifferenziert und inhaltlich zum Teil leicht abgeändert. So sprechen Klieme und andere bereits 2006 von *strukturierter, klarer und störungspräventiver Unterrichtsführung; unterstützendem, schülerorientiertem Klima* sowie weiterhin von *kognitiver Aktivierung*. Bei letzterer weisen auch sie darauf hin, dass diese fachspezifisch betrachtet werden müsse (vgl. Klieme et al., 2006, S. 131).

In aktuellen Arbeiten werden die drei Dimensionen qualitätvollen Unterrichts wie folgt bezeichnet: *Potenzial zur kognitiven Aktivierung, konstruktive Unterstützung* sowie *effektive Klassenführung* (vgl. z. B. Kunter & Ewald, 2016, S. 15–16; Begrich, Fauth, Kunter & Klieme, 2017, S. 24–25; Kunter & Trautwein, 2018, S. 78–103 bzw. Übersicht bei Holzberger & Kunter, 2016, S. 42). Dabei versteht man unter *Klassenführung* „alle Handlungen und Strategien, die dazu dienen, Ordnungsstrukturen im Klassenzimmer herzustellen und aufrechtzuerhalten" (Kunter & Trautwein, 2018, S. 78). Sie umfasst zum Beispiel den Umgang mit und die Prävention von Störungen, eine sinnvolle

Zeitnutzung, eine klare Strukturierung des Unterrichts und die Beteiligung einer möglichst hohen Anzahl von Schülerinnen und Schülern am Unterricht. Eine genaue Beschreibung der Dimension *Klassenführung* geben Trautwein und Kunter (2018) unter Rückgriff auf die maßgeblichen Arbeiten von Kounin (1976) (vgl. Kunter & Trautwein, 2018, S. 78–83).

Die Dimension der *konstruktiven Unterstützung* stellt ähnlich wie die oben genannte Schülerorientierung die Lernenden ins Zentrum. Bei der *konstruktiven Unterstützung* geht es allerdings nicht nur um soziales Lernen beziehungsweise um die Schüler/-innen im sozialen Kontext; die Aufmerksamkeit verschiebt sich hier vielmehr auf den Lernprozess der/des einzelnen Schülerin beziehungsweise Schülers. *Konstruktive Unterstützung* bedeutet somit vergleichbar zu obiger Definition ein positives Lernklima sowie einen respektvollen und wertschätzenden Umgang zwischen der Lehrkraft und ihren Schülerinnen und Schülern. *Konstruktive Unterstützung* bedeutet aber auch konkrete Begleitung der Lernenden im Verstehensprozess sowie Hilfestellungen bei Verständnisproblemen. Insofern fokussiert die *konstruktive Unterstützung* die Art und Weise, wie die Lehrkraft mit ihren Schülerinnen und Schülern umgeht (vgl. Kunter & Trautwein, 2018, S. 94–95). Ziel der *konstruktiven Unterstützung* ist es, die Schüler/-innen zu selbstständigem Lernen zu befähigen. Wichtige Grundsätze sind eine qualitätvolle Feedbackkultur, ein positiver Umgang mit Fehlern sowie eine angemessene Unterrichtszeit mit ausreichend Wartezeit bei Fragen an die Klasse (vgl. hierzu die genaue Definition der Dimension *konstruktive Unterstützung* bei Kunter & Trautwein, 2018, S. 94–103).

Das Potenzial zur kognitiven Aktivierung

Die letzte Dimension des *Potenzials zur kognitiven Aktivierung* fokussiert die Lernmöglichkeiten: Der Unterricht muss so gestaltet werden, dass die Auseinandersetzung mit dem Lernstoff bei den Lernenden vertiefte kognitive Prozesse auslöst. Taut und Rakoczy (2016) nennen folgende Kennzeichen, die eine kognitive Aktivierung der Schüler/-innen ermöglichen: „(a) when the teacher calls students' attention to connections between different concepts and ideas and links new content with prior knowledge, (b) when the problems and tasks are challenging, (c) when students reflect on their learning and the underlying ideas, and (d) when the teacher is interested in students' thoughts, concepts, and solutions“ (Taut & Rakoczy, 2016, S. 47). Diese vertieften kognitiven Prozesse sollen nachhaltiges Lernen ermöglichen: „In der Lehr-Lernforschung ist man sich einig, dass ein solches *higher order thinking* der zentrale Schlüssel ist, um langfristig eine gut vernetzte und transferfähige Wissensstruktur aufzubauen […]“ (Kunter & Trautwein, 2018, S. 86). Die Lehrkraft kann durch die Gestaltung des Unterrichts und die Auswahl der Aufgaben kognitive Aktivierung fördern, ob diese allerdings tatsächlich „stattfindet“, liegt – ganz im Sinne des Angebot-Nutzungs-Modells – in der Hand der Schüler/-innen. Jede/-r Einzelne muss zu dieser kognitiven

Anstrengung bereit sein und kann sie nur selbst aufbringen. Deswegen wird in jüngerer Forschung vom *Potenzial zur kognitiven Aktivierung* gesprochen. Dabei ist der Hinweis wichtig, dass kognitives Aktivierungspotenzial nicht gleichzusetzen ist mit schwieriger Aufgabe, denn Schüler/-innen können unabhängig vom Schwierigkeitsgrad einer Aufgabe kognitiv aktiv werden. Entscheidend ist hier vielmehr, dass dieser Schwierigkeitsgrad der Aufgabe an den aktuellen Lernstand sowie an das Leistungsvermögen der Schülerin/des Schülers angepasst ist (vgl. Kunter & Trautwein, 2018, S. 92–93 bzw. Kunter & Ewald, 2016, S. 16–17).

Trotz dieser theoretischen Grundsätze ist es schwierig, konkret zu definieren, was *Potenzial zur kognitiven Aktivierung* bedeutet, um es für Messungen von Unterrichtsqualität zugänglich zu machen. Praetorius, Pauli, Reusser, Rakoczy und Klieme (2014) bezeichnen es als – im Vergleich zu den anderen Dimensionen – unstabiles Merkmal von qualitätvollem Unterricht (vgl. Praetorius et al., 2014, S. 9). Die Bewertungen, die ein und dieselbe Lehrkraft in dieser Dimension bei Unterrichtsbeobachtungen erhält, können erheblich schwanken. Dies liegt zum einen an Messfehlern. Kognitive Aktivierung ist schwieriger zu erkennen als zum Beispiel ein störungsfreier Ablauf des Unterrichts, da sie nicht direkt beobachtbar ist, sondern interpretativ erschlossen werden muss. Eine weitere Hürde der Messung ist, dass kognitive Aktivierung bei der Einführung eines Themas, während des Übens oder in einer Erklärung jeweils unterschiedliche Ausprägungen hat und noch dazu fachabhängig ist. Ein Beispiel ist das Problemlösen, das im Fach Mathematik als stark kognitiv aktivierend gilt. Das lässt sich aufgrund der Art der fachlichen Inhalte nicht auf das Fach Deutsch übertragen, für das kognitive Aktivierung etwas anderes bedeutet (s. dazu auch die Punkte *2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht* sowie *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit). Auch hier muss genaue theoretische Vorarbeit geleistet werden und Rater/-innen müssen daraufhin geschult werden (vgl. Praetorius et al., 2014, S. 9). Dies ist leichter, wenn es „nur" um die kognitive Aktivierung während einer Stunde oder bei der Bearbeitung eines konkreten Materials geht. Bei genauer Operationalisierung können die Rater/-innen entsprechend ausgebildet werden und das Merkmal kann gut bewertet werden.

Wird dagegen kognitive Aktivierung als Unterrichtsqualitätsmerkmal oder im Zuge der Teacher-Effectiveness-Forschung untersucht, so dass Aussagen über den Unterricht oder eine Lehrkraft insgesamt gemacht werden sollen, ist die Operationalisierung ungleich schwieriger. Um Angaben darüber zu machen, wie gut eine Lehrkraft das Merkmal *Potenzial zur kognitiven Aktivierung* in ihrem Unterricht erfüllt, sind mehrere Beobachtungsstunden notwendig; Praetorius und andere nehmen eine Mindestanzahl von neun Stunden an (vgl. Praetorius et al., 2014, S. 8). Allerdings gehen sie auch davon aus, dass die bestehenden Messinstrumente nicht ausreichen und weitere Forschung und Ausdifferenzierung des Merkmals nötig sind (vgl. Praetorius et al., 2014, S. 9). Zu einer ähnlichen Schlussfolgerung kommen Taut und Rakoczy (2016), die in ihrer Studie keine

verallgemeinerbaren Ergebnisse für die Dimension *kognitive Aktivierung* feststellen konnten und dafür mehrere mögliche Gründe sehen: „We were unable to identify a particular source of error serving as the main explanation for this lack of generalizability. Rather, rater and indicator errors as well as unspecified error seem to be responsible" (Taut & Rakoczy, 2016, S. 56). Sie weisen ebenso auf weiteren Forschungsbedarf hin und betonen gleichzeitig wie angedeutet die notwendige fachspezifische Ausdifferenzierung der kognitiven Aktivierung (vgl. Taut & Rakoczy, 2016, S. 56).
Zu überlegen ist, ob die konkrete Ausarbeitung der kognitiven Aktivierung nicht nur fachspezifisch, sondern sogar themenabhängig erfolgen muss. Die Frage, die gestellt wird, ist dann nicht nur, wie eine deutschspezifische kognitive Aktivierung aussieht, sondern wie kognitives Aktivierungspotenzial beim Lesen, beim Schreiben und so weiter gestaltet werden kann. Man kann die Frage sogar noch spezifischer formulieren und untersuchen, wie kognitive Aktivierung beim Lesen von Sachtexten gegenüber kognitiver Aktivierung beim Lesen von Literatur ausgelöst werden kann.
Auch wenn die *kognitive Aktivierung* von Schülerinnen und Schülern inzwischen als unbestrittene Anforderung an qualitätvollen Unterricht gilt, so ist dennoch ein wichtiger Hinweis von Praetorius und anderen (2014) nicht außer Acht zu lassen. Gerade wenn sie nämlich als Dimension von Unterricht bewertet wird, muss bedacht werden, dass *kognitive Aktivierung* nicht in jeder Phase und bei jedem Thema gleich wichtig ist. So sind diesbezügliche Schwankungen im Verlauf einer oder mehrerer Unterrichtsstunden normal und im Sinne der Unterrichtsqualität als positiv zu werten. Gerade bei Untersuchungen zu qualitätvollem Unterricht muss wohl überlegt werden, wie mit diesen Schwankungen umgegangen wird, um dennoch reliable und valide Ergebnisse zu erzielen (vgl. Praetorius et al., 2018, S. 422 bzw. Praetorius et al., 2014, S. 9). Hier ist genau wie für die beiden anderen Prinzipien *Klassenführung* und *konstruktive Unterstützung* weitere Forschung nötig.

Empirische Studien zu qualitätvollem Unterricht

Konkret empirisch nachgewiesen wurden die gerade erläuterten Dimensionen in der COACTIV-Studie (vgl. Baumert & Kunter, 2011), die professionelle Kompetenz von Mathematiklehrkräften untersucht. *Effektive Klassenführung* und *kognitive Aktivierung* führen demzufolge zu besseren Leistungen im Mathematikunterricht, *konstruktive Unterstützung* zu höherer Motivation und Lernfreude (vgl. Kunter, Baumert, Blum, Klusmann & Neubrand, 2011). Sie stellen ein Grundgerüst für effektiven und qualitätvollen Unterricht dar und sind dann wirksam, wenn alle drei erfüllt werden. „Diese drei Dimensionen der Unterrichtsqualität stellen eine sparsame, aber relativ umfassende Systematisierung der im Unterricht zu beobachtenden Lehr-Lernprozesse dar und sie haben sich im deutschen Sprachraum sehr gut etabliert, […]" (Kunter & Ewald, 2016, S. 16). Auch

Praetorius und andere (2018) kommen in ihrer Übersicht zu dem Schluss, dass das Modell gerade aufgrund seiner theoretischen Fundierung eine tragfähige Basis zur Erforschung von Unterrichtsqualität darstelle, weitere Differenzierung aber dringend notwendig sei (vgl. Praetorius et al., 2018, S. 423).

Fazit für FALKE-D

Betrachtet man das Erklären, den Untersuchungsgegenstand von FALKE-D, so sind prinzipiell zwei dieser Dimensionen relevant: Erklären kann ein Akt der *konstruktiven Unterstützung* sein, wenn einer einzelnen Schülerin/ einem einzelnen Schüler differenziert etwas erklärt wird. Im Gesamtprojekt FALKE allerdings stehen Lehrererklärungen vor der gesamten Klasse im Fokus des Interesses. Diese Erklärungen müssen kognitiv aktivierend für die Schüler/-innen sein, damit möglichst alle während des Erklärvorgangs mitdenken. Deswegen greift vorliegende Studie die Dimension *kognitive Aktivierung* auf. Unter Einbezug oben geschilderter Ergebnisse werden daher in FALKE-D wie in der Einleitung kurz angerissen Erklärungen mit höherem *Potenzial zur kognitiven Aktivierung* solchen mit weniger Potenzial gegenübergestellt (s. dazu Punkte *4.2.5 Erklärtiefe in FALKE-D* sowie 5*.1.4 Variation der Erklärtiefe* und für einen Überblick über die unterschiedlichen Dimensionen Abbildung 1 in dieser Arbeit).

Arbeiten zu Unterrichtsqualität	Dimensionen
TiMS (Klieme et al., 2001)	Etablierung der drei **Basisdimensionen**: - *Unterrichts- und Klassenführung* (i.e. störungsfreier Ablauf) - *Schülerorientierung* (i.e. individuelle Bezugsnorm) und - *kognitive Aktivierung* (i.e. anspruchsvolles Üben)
COACTIV (Baumert & Kunter 2011)	**Empirisch nachgewiesene Dimensionen**: - *Effektive Klassenführung* und - *kognitive Aktivierung* ➡ führen zu besseren Leistungen im Mathematikunterricht - *konstruktive Unterstützung* ➡ führt zu höherer Motivation und Lernfreude
Aktuelle Arbeiten, z.B. Kunter & Ewald, 2016; Begrich et al., 2017; Kunter & Trautwein, 2018	Zur Zeit **etablierte Dimensionen** in Lehr-/Lernforschung: - *effektive Klassenführung* (i.e. Störungsprävention, effektive Zeitnutzung, Strukturierung) - *konstruktive Unterstützung* (i.e. qualitätvolle Feedbackkultur, ein positiver Umgang mit Fehlern, Zeitmanagement) - *Potenzial zur kognitiven Aktivierung* (i.e. Auslösen von vertieften kognitiven Prozessen)
FALKE-D (Gaier 2021)	Auf Lehrerkompetenz des **Erklärens übertragene Dimension** - *Erklärtiefe* (i.e. Auslösen vertiefter Verstehensprozesse)

Abbildung 1 Dimensionen von Unterrichtsqualität

2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht

Gemeinsam ist den drei etablierten Dimensionen von Unterrichtsqualität, dass sie zu den bereits erwähnten Tiefenstrukturen des Unterrichts gehören. Wie in obigem Punkt angedeutet wird in der allgemeinen Lehr-/Lernforschung bei der Beobachtung und Bewertung von Unterricht zwischen den Sicht- und Tiefenstrukturen unterschieden. Sichtstrukturen bilden die Rahmenbedingungen für den Unterricht. Sie sind einfach zu beobachtende und leicht zu zählende Merkmale wie Sozialformen oder Methoden. Ihre Beurteilung kann unabhängig von spezifischem Fach- oder fachdidaktischen Wissen vorgenommen werden und ist damit in der Bewertung niedrig inferent.

Bei der Erforschung von Unterrichtsqualität lässt sich etwa seit den 2000er Jahren eine verstärkte Bewusstwerdung von und Hinwendung zu Tiefenstrukturen beobachten. So werden noch im Prozess-Produkt-Paradigma (s. Punkt *2.2.1 Paradigmen zur professionellen Lehrkraft* in dieser Arbeit) vor allem Sichtstrukturen erfasst und bewertet (z. B. welche Methoden eingesetzt werden oder wie sich die Sprechanteile zwischen Lehrkraft und Schüler/-innen verteilen). Auch in der Hattie-Studie werden vor allem sehr detaillierte Oberflächenmerkmale zur Unterrichtsqualität untersucht. Eine Vielzahl der Studien, die in Hatties Berechnungen einfließen, arbeitet, wie Studien zum Prozess-

Produkt-Paradigma, mit Beobachtungslisten, auf denen vor allem notiert wird, was im Unterricht gemacht wird (vgl. Hattie, 2009).

So kommen in dieser kleinteiligen Aufzählung von Merkmalen guten Unterrichts Faktoren wie *Klarheit der Lehrperson* mit d = 0.75[5] oder *Vermittlung metakognitiver Strategien* mit d = 0.69 große Bedeutung für den Lernfortschritt für Schüler/-innen zu. Auch die *Anregung zu lautem Denken* (d = 0.64) sowie *Problemlösen* (d = 0.61) erweisen sich bei Hattie als wichtige Faktoren für den Lernzuwachs (vgl. Hattie, 2009). Auch wenn gerade die letztgenannten eher den Tiefenstrukturen (s. folgenden Absatz) zuzuordnen sind, so wird doch aufgrund der Methodik in den Studien Unterricht in seiner Oberflächenstruktur untersucht. Diese wird dann jedoch sehr kleinschrittig beschrieben, denn Hattie nennt insgesamt 138 wirksame Faktoren für qualitätvollen Unterricht.

In jüngerer Zeit rückt man von diesem kleinschrittigen und oberflächlichen Beobachten von Unterricht eher ab. Auf der einen Seite wird Unterricht nicht mehr „nur" über Beobachtung evaluiert, vielmehr fließt eine Vielzahl an anderen Informationen mit in die Bewertung der Unterrichtsqualität ein, etwa Aufgaben, die Lehrkräfte einsetzen, Hefteinträge, die gestaltet werden, oder abgehaltene Prüfungen. Dadurch wird der Unterricht stärker „in Gänze" erfasst, wie zum Beispiel in der TiMS-Studie (2001) oder in der COACTIV-Studie (2011) (s. hierzu Punkt *2.4.3 Methoden und Zweck der Erfassung von Wahrnehmung* in dieser Arbeit).

Auf der anderen Seite haben sich wie bereits angedeutet nach Kunter und Trautwein (2018), Helmke (2014) sowie Klieme und anderen (2006) vor allem die Tiefenstrukturen des Unterrichts als entscheidend dafür erwiesen, was Schüler/-innen lernen und wie qualitätvoll und nachhaltig dieses Lernen ist. Wenn man die Effektivität des Unterrichts in den Blick nimmt, ist es deswegen notwendig, diese Tiefenstrukturen von Unterricht zu analysieren, die deswegen zunehmend ins Interesse der Forschungen rücken (vgl. Kunter & Trautwein, 2018; Seidel & Shavelson, 2007). Ob, wie viel und wie vertieft Schüler/-innen während einer Gruppenarbeitsphase (d. h. Sichtstruktur) lernen, ist zum Beispiel abhängig von der *effektiven Zeitnutzung*, vom *kognitiven Aktivierungspotenzial* der zu bearbeitenden Aufgaben sowie vom *konstruktiven Unterstützungsangebot* bei Lernschwierigkeiten (alle drei Tiefenstrukturen). Die Tiefenstrukturen des Unterrichts liegen im Verborgenen und werden folglich nur von Expertinnen und Experten wahrgenommen, sind also hoch inferent und haben immer einen subjektiven Anteil: Wie hoch das *kognitive Aktivierungspotenzial* einer Aufgabe aus dem Schreibunterricht ist, kann nur eingeschätzt werden. Ob die kognitive Aktivierung tatsächlich ausgelöst wird, lässt sich ebenso wenig direkt

[5] Zur Erläuterung: Mit Cohens d wird in der Statistik die Effektstärke angeben. Dabei gilt ein Wert, der kleiner als 0.5 ist, als kleiner Effekt. Werte zwischen 0.5 und 0.8 sind mittlere Effekte und Werte darüber zählen als großer Effekt.

beobachten. Hier sind also Expertinnen und Experten gefragt, das Expertenparadigma in der Lehrerforschung (s. Punkt *2.2.1 Paradigmen zur professionellen Lehrkraft* in dieser Arbeit) nimmt daher derartige Tiefenstrukturen in den Blick.

Beim Übertrag dieser Ergebnisse aus der allgemeinen Unterrichtsforschung in die Fachdidaktiken wird meist angenommen, dass sich zwei der drei Basisdimensionen, nämlich *konstruktive Unterstützung* und *effektive Klassenführung* sinnvoll aus einer fachübergreifenden Perspektive betrachten und untersuchen lassen. Inwiefern das zielführend ist, ist mit der Forderung von Taut und Rakoczy (2016) nach einer fachspezifischen Spezifizierung aller Dimensionen (s. Absatz oben zu *Etablierte Dimensionen von Unterrichtsqualität* unter Punkt *2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht*) zu überdenken und zu überprüfen. Gerade bei der *konstruktiven Unterstützung* ist ein fachspezifischer Aspekt nicht abzustreiten, Schilcher und Krauss sprechen daher sogar von „[f]achbezogene[r] konstruktive[r] Unterstützung" (2019, S. 37), da die Lehrkraft über hohe diagnostische Kompetenz verfügen muss, um sinnvolle konstruktive Unterstützung leisten zu können.

Dagegen muss *kognitive Aktivierung* wie bereits erläutert aus der Fachperspektive bestimmt werden (vgl. Praetorius et al., 2014, S. 3), denn das *Potenzial zur kognitiven Aktivierung* ist zwangsläufig an den fachlichen Gegenstand beziehungsweise an die Art der Auseinandersetzung mit diesem fachlichen Gegenstand gebunden[6]. In diesem Sinne ist es die Aufgabe der Lehrkraft „als Vertreter eines Fachs und ‚Instrukteur' die Lernumgebung so [zu] gestalten, dass den Schülerinnen und Schülern verständnisvolles Lernen und eine aktive kognitive Auseinandersetzung mit dem Lerngegenstand ermöglicht wird" (Klieme et al., 2006, S. 132).

Deutschdidaktische Arbeiten zum Potenzial zur kognitiven Aktivierung

In der deutschdidaktischen Forschung widmen sich daher einige Arbeiten diesem Merkmal qualitätvollen Unterrichts und stellen das *Potenzial zur kognitiven Aktivierung* ins Zentrum ihres Interesses. Es wird an die spezifischen Anforderungen des Faches Deutsch angepasst und aus der fachspezifischen Perspektive untersucht. Beispiele hierfür sind die Arbeiten von Hanisch (2018), Winkler (2017), Köster (2016), Lotz (2016), Kleinbub (2016) sowie Stahns (2013) (s. als Überblick Abbildung 2 in dieser Arbeit). Betrachtet man die Studien genauer, so bestätigt sich, dass kognitive Aktivierung ein „facettenreiches Konstrukt" (Lipowsky, 2015, S. 91) ist, denn die genauen Definitionen und Ausgestaltungen unterscheiden sich in den Studien. Gemein ist allen Studien das

[6] Mit der Forderung nach einer fachspezifischen Konkretisierung schließen interessanterweise auch Praetorius, Decristan, Hess und Holzberger (2020) ihren Überblick über den aktuellen Forschungsstand zu Oberflächen- und Tiefenmerkmalen von Unterricht: „Bislang zu wenig Beachtung hat in diesem Zuge der Fachinhalt erfahren" (S. 111).

Merkmal einer eigenständigen, aktiven geistigen Auseinandersetzung mit dem Lernstoff. Der Lernstoff bestimmt dabei wesentlich die Operationalisierung der kognitiven Aktivierung.

Zunächst unabhängig vom fachlichen Gegenstand untersucht Köster (2016) Aufgaben in ihrem Buch „Aufgaben im Deutschunterricht" als „Herzstück des Unterrichts" (Köster, 2016, S. 8). Sie untersucht Lern- und Leistungsaufgaben und beschäftigt sich im ersten Teil mit Merkmalen und Funktionen von Aufgaben, bevor sie für die unterschiedlichen Lernbereiche des Deutschunterrichts konkrete Aufgaben vorschlägt. Eines dieser allgemeinen Merkmale von Aufgaben sind die Schüleraktivitäten, die die entsprechende Aufgabe auslöst. Köster unterscheidet hier nämlich kognitive von exekutiven Aktivitäten:

> Aufgaben sind dazu da, Schülertätigkeit zu bewirken. In der Regel geht es darum, sie ans Denken zu bringen und die Ergebnisse des Denkprozesses zu kommunizieren. Das kann sowohl im mündlichen Vortrag oder im Klassengespräch geschehen als auch in der schriftlichen Dokumentation. Die aufs Denken gerichteten, geistigen Tätigkeiten werden auch als kognitive Aktivitäten bezeichnet und die Art und Weise der Dokumentation bzw. der Kommunikation als exekutive Aktivitäten (Köster, 2016, S. 48).

Da die geistige Anstrengung nicht direkt beobachtbar ist, werden die kognitiven Aktivitäten über die exekutiven Tätigkeiten sichtbar gemacht und überprüft. Diese „aufs Denken gerichteten, geistigen Tätigkeiten" (s. Zitat oben) können unterschiedlich intensiv sein. Bei der Gestaltung und/oder Auswahl von Aufgaben sollen die Lehrkräfte darauf achten, dass diese in einem an die Schülergruppe angemessenen Grad kognitiv aktivieren. Diese kognitive Aktivierung kann dabei unterschiedlicher Natur sein. Köster nennt hier folgende Möglichkeiten:

> Hier hat die Lehrkraft zu entscheiden, ob sie Problemlösungsaufgaben, Selbsterklärungsaufgaben oder metakognitive Fragen einsetzt […]. Als kognitiv anregend gelten sowohl Aufgaben, die an Alltagserfahrungen anknüpfen, als auch Problemlösungsaufgaben, bei denen die gegebenen Informationen eine Lücke lassen, die durch Nachdenken und Wissensaktivierung geschlossen werden muss - so wie man ein passendes Puzzleteil findet. Kognitiv anregend sind darüber hinaus Aufgaben, die eine - begründungspflichtige - Entscheidung fordern […] (Köster, 2016, S. 116).

Diese Möglichkeiten sind recht allgemein gehalten und folgen den Grundsätzen zur kognitiven Aktivierung aus der Lehr-/Lernpsychologie. Wie eine hohe geistige Tätigkeit, also eine hohe kognitive Aktivierung der Schüler/-innen gelingen und konkret aussehen kann, ist Gegenstand folgender Studien aus der Deutschdidaktik, die jeweils eine bestimmte Kompetenz in den Fokus nehmen.

Es werden zunächst zwei neuere Studien aus der Sprachdidaktik vorgestellt. Hanisch (2018) entwickelt in ihrer Studie zum Rechtschreiberwerb einen Lehr-Lernkreislauf, der folgendermaßen aussieht: „Der ‚Lehr-Lernkreislauf' eines kognitiv aktivierenden Unterrichts ergibt sich 1) aus dem kognitiv aktivierenden Lehrangebot an die Schüler, 2) aus der daraus resultierenden gesteigerten kognitiven Aktivität der Schüler sowie 3) aus den Einfluss nehmenden individuellen Mediations- und Aktivierungsprozessen, welche die interne Aktivierung beeinflussen" (Hanisch, 2018, S. 33). Sie

unterscheidet somit das kognitiv aktivierende Angebot, das die Lehrkraft bereitstellen muss, von den kognitiven Prozessen, die dieses Angebot bei den Schülerinnen und Schülern auslösen kann, vorausgesetzt diese sind dazu bereit: Die kognitive Aktivierung, für die die Lehrkräfte verantwortlich zeichnen, wird somit von der kognitiven Aktivität, die bei Schülerinnen und Schülern ausgelöst wird, differenziert (vgl. Hanisch, 2018, S. 35–39). Deutlich erkennbar ist die Anlehnung an das Angebot-Nutzungs-Modell von Helmke (2014). Bezogen auf den eigenen Forschungsgegenstand des Orthographieunterrichts verbindet Hanisch dann kognitive Aktivierung mit Rechtschreibbewusstheit und definiert dies folgendermaßen: „Kognitive Aktivierung im Rechtschreibunterricht ist dadurch gekennzeichnet, dass die Aufgaben und deren Einbettung in den Unterricht in einer Weise gestaltet werden, dass die zur Verbalisierung fähige Rechtschreibbewusstheit hinsichtlich orthographischen Wissens auf- und ausgebaut werden kann" (Hanisch, 2018, S. 134). Im Fokus der Arbeit stehen die Aufgaben des Rechtschreibunterrichts.

Auch Stahns verweist in seiner Studie zum Grammatikunterricht auf den Zusammenhang zwischen „der Entwicklung der Sprachbewusstheit und kognitiv aktivierender Unterrichtsgestaltung" (2013, S. 97). Er definiert kognitive Aktivierung allgemein so: „Die kognitive Aktivierung kann als Merkmal eines Unterrichts angesehen werden, in dem beachtet wird, dass guter Unterricht Unterricht ist, in dem die Schüler in Abhängigkeit von ihren Möglichkeiten zu eigenständigen Lernaktivitäten angeregt werden" (Stahns, 2013, S. 273). In seinen Videoanalysen von Unterrichtstunden berücksichtigt er eine Vielzahl an Merkmalen kognitiver Aktivierung zum fragend-entwickelnden Unterricht, da dies die Methodik war, die er in den videographierten Stunden vor allem vorgefunden hatte. Er untersucht die videographierten Gespräche im Unterricht auf der Grundlage von Transkripten und schätzt das kognitive Aktivierungspotenzial anhand von Indikatoren ein. Dazu betrachtet er folgende Kategorien: „(1) Länge und Häufigkeit der Schüler- und Lehrerbeiträge, (2) Art der Schüler- und Lehrerbeiträge, (3) Engführung der Schüler (Schüler als Stichwortgeber), (4) Lernbegleitung im Klassengespräch und in Schülerarbeitsphasen, (5) Umgang mit Fehlern" (Stahns, 2013, S. 142). Um der kognitiven Aktivierung auf die Spur zu kommen, vergleicht er zum Beispiel die Sprechanteile, die Gesprächsformen sowie die Sprechzeit von Schülerinnen und Schülern mit denen der Lehrkraft (vgl. Stahns, 2013, S. 197–203). Die „Art" des Sprechbeitrages wird kategorisiert in Antwort, Frage, Instruktion/Aufgabenstellung, Lesen und Andere (vgl. Stahns, 2013, S. 212), bevor die Fragen wiederum spezifiziert werden, und zwar in für die kognitive Aktivierung relevante qualitative Fragetypen: „(1) Andere (2) Inhaltliche Frage (3) Organisatorische Frage (4) Rückfrage zur Aufgabenstellung (5) Verständnisfrage (6) Keine Frage" (Stahns, 2013, S. 212). Stahns kommt in seiner Untersuchung zum Schluss, dass die videographierten Unterrichtsstunden im Ganzen eher eng geführt und wenig kognitiv aktivierend waren. Er führt dies weniger auf mangelndes Bewusstsein der Lehrkräfte für kognitiv aktivierenden Unterricht zurück, denn auf

Bemühungen, Schüleraktivität mit dem curricularen Anspruch der Lehrpläne zu verbinden (vgl. Stahns, 2013, S. 274–275). Er regt dazu an, in weiterer Forschung zu untersuchen, ob und wie sich *kognitive Aktivierung* tatsächlich auf die Lernleistungen der Schüler/-innen auswirkt (vgl. Stahns, 2013, S. 273–274).

In zwei der Studien aus der Literaturdidaktik stehen erneut die Aufgaben im Interessensfokus. Winkler und Steinmetz (2016) plädieren dafür, Methoden und Setzungen aus der empirischen Sozialforschung nach kritischer Prüfung in die Fachdidaktik zu übernehmen (vgl. Winkler & Steinmetz, 2016, S. 52–53). Sie ziehen in ihrer Studie KoALa das Modell von Kunter und Trautwein (2018) heran und untersuchen das Potenzial zur kognitiven Aktivierung, das über Aufgaben aus dem Literaturunterricht operationalisiert wird (s. auch Winkler (2017)). Dabei unterscheiden sie die Aufgabenstellung von der Aufgabenimplementation. Während erste fachspezifisch ausgestaltet werden muss, folgt die Aufgabenimplementation neben der literaturdidaktischen Spezifik auch allgemeindidaktischen beziehungsweise lehr-/lernpsychologischen Anforderungen. In ihrer Videostudie wird das Potenzial zur kognitiven Aktivierung sowohl in der Aufgabenstellung untersucht, zum Beispiel über den Aufgabenfokus, der für sie am „aktivierendsten" ist, wenn der Fokus sowohl auf den Text als auch auf den Leser gesetzt wird (vgl. Winkler, 2017, S. 84–85). Gleichzeitig wird die Aufmerksamkeit aber auch auf die Aufgabenimplementation gelenkt, wenn zum Beispiel die Verarbeitungstiefe kodiert wird, etwa über die Frage, ob von den Schülerinnen und Schülern Begründungen eingefordert werden (vgl. Winkler, 2017, S. 87). Winkler weist darauf hin, dass das weitere Forschungsinteresse der Wirksamkeit der Aufgaben gilt, und somit der Frage, ob die kognitiv aktivierenden Aufgaben höheren Lernerfolg mit sich bringen. Das Projekt dient somit der Verknüpfung von gutem mit effektivem Literaturunterricht (vgl. Winkler, 2017, S. 93–94; s. hierzu auch Ausführungen zu gutem, effektivem und qualitätvollem Unterricht unter Punkt *2.1 Qualität von (Deutsch-) Unterricht* in dieser Arbeit).

Lotz (2016) weist in ihrer Studie zum Leseunterricht in der Grundschule zunächst auf die Heterogenität des Konstruktes kognitive Aktivierung hin: „Die einzelnen Studien konzeptualisieren die kognitive Aktivierung im Detail aber sehr unterschiedlich, sodass noch von keinem einheitlichen Verständnis ausgegangen werden kann" (Lotz, 2016, S. 43). Einigkeit bestehe allerdings darüber, dass „ein kognitiv aktivierender Unterricht die Schülerinnen und Schüler herausfordert und zu höherwertigen Denkprozessen anregt" (Lotz, 2016, S. 43). Auf der Grundlage bisheriger Forschungsansätze entwickelt sie ein eigenes Modell für kognitiv aktivierende Leseübungen, das sie in die Phasen Aufgabenstellung, Aufgabenbearbeitung und Aufgabenreflexion aufteilt (vgl. Lotz, 2016, S. 76). In der Phase der Aufgabenbearbeitung nimmt Lotz folgende Merkmale von kognitiver Aktivierung an: aktivierende Fragen, die die Lehrkraft stellt; Hilfeleistungen, die zum Weiterdenken

anregen; konstruktives Feedback sowie die Bereitstellung und den kontinuierlichen Ausbau von Lesestrategien (vgl. Lotz, 2016, S. 77 bzw. S. 78-146).

Schließlich untersucht Kleinbub (2016) in ihrer Studie zum Leseunterricht in vierten Klassen (N=862 Schüler/-innen) die Dimensionen *Kompetenzorientierung, Aktivierung, Adaptivität, Konkretisierung, Motivierung* und *Strukturierung*. Sie stellt fest, dass in den videographierten Stunden vor allem das Merkmal der *Motivierung* erfüllt war, die *Kompetenzorientierung* dagegen nur in geringem Maße gegeben war (vgl. Kleinbub, 2016, S. 32). Bezogen auf die Leseleistung kann sie nur beim Merkmal *Aktivierung* (das hier als „aktive Teilhabe der Schülerinnen und Schüler" (Kleinbub, 2016, S. 19) verstanden wird) eine Bedeutung nachweisen (vgl. ebd., S. 33). Somit kann ein Zusammenhang zwischen der im Unterricht erfolgten Aktivierung der Schüler/-innen und der von ihnen erzielten Leseleistung angenommen werden. Für die anderen untersuchten Merkmale konnte kein vergleichbarer Zusammenhang nachgewiesen werden. Die Ergebnisse bedürfen allerdings der weiteren Überprüfung, da die Stichprobe zu klein für eine Verallgemeinerung war. Kleinbub folgert aus diesen wenigen nachweisbaren Zusammenhängen, dass das, „[w]as guter Unterricht ist und wie er zum Erreichen der Unterrichtsziele beträgt, […] sich nur schwerlich exakt und erschöpfend beantworten" (Kleinbub, 2016, S. 34) lässt.

Recht deutlich zeigt sich in den unterschiedlichen Konzeptualisierungen von kognitiver Aktivierung in den vorgestellten Studien die Notwendigkeit, diese an den Gegenstand anzupassen. Allen Konzeptualisierungen gemein ist wie gesagt der Grundsatz, Schülerinnen und Schülern eine vertiefte geistige Auseinandersetzung mit dem Unterrichtsgegenstand zu ermöglichen.

Mit der Konzeptualisierung von *Erklärtiefe* als Aspekt von Erklärqualität sowie der Implementation von einfachen versus vertieften Erklärungen fokussiert auch vorliegende Arbeit den Aspekt *Potenzial zur kognitiven Aktivierung* und übertragt somit diese Dimension der Unterrichtsqualität auf das Mikroteaching-Element des Erklärens.

Im Unterschied zu vielen der gerade beschriebenen Studien stehen bei FALKE-D aber keine Aufgaben im Fokus, sondern Lehrererklärungen, die sich an eine gesamte Klasse richten. Bei FALKE-D wird davon ausgegangen, dass Erklärungen, die durch das Aufzeigen verborgener Zusammenhänge *Erklärtiefe* aufweisen, über höheres Potenzial zur kognitiven Aktivierung verfügen als Erklärungen, die einen Sachverhalt eher oberflächlich erläutern (s. Punkt *4.2.5 Erklärtiefe in FALKE-D* sowie Punkt *5.1.4 Variation der Erklärtiefe* in dieser Arbeit).

Deutschdidaktische Studie	Fachspezifische Dimension der kognitiven Aktivierung
Köster (2016): Aufgaben im Deutschunterricht	Kognitive Aktivitäten als aufs Denken gerichtete geistige Tätigkeiten vs. exekutive Aktivitäten
Hanisch (2018): Rechtschreibunterricht	Kreislauf aus kognitiv aktivierendem Lehrangebot an die Schüler, kognitiver Aktivität der Schüler sowie individuellen Mediations- und Aktivierungsprozessen zum Auf- und Ausbau von Rechtschreibbewusstheit
Stahns (2013): Grammatikunterricht	Kognitive Aktivierung als Anregen zu eigenständigen Lernaktivitäten
Winkler und Steinmetz (2016) bzw. Winkler (2017) Literaturunterricht	Kognitive Aktivierung in Aufgabenstellung und Aufgabenimplementation
Lotz (2016): Lesen	Kognitive Aktivierung als höherwertige Denkprozesse
Kleinbub (2016): Lesen	Aktivierung als aktive Teilhabe der Schüler/-innen
Gaier (2021): Erklären als Lehrerkompetenz	Potenzial zur kognitiven Aktivierung als *Erklärtiefe*, d.h. Aufzeigen verborgener Zusammenhänge

Abbildung 2 Dimension der kognitiven Aktivierung in deutschdidaktischen Studien

2.2 Professionelle (Deutsch-) Lehrkräfte

In der Diskussion darüber, was qualitätvollen (Deutsch-) Unterricht ausmacht, steht neben den gerade beschriebenen Dimensionen die Lehrkraft im Zentrum des Interesses: Was muss eine kompetente Lehrkraft wissen und können? Für qualitätvollen Unterricht sieht bereits Hattie (2009) die zentrale Bedeutung der Lehrkraft, die er auch empirisch nachweist, denn in seiner Metaanalyse zu wirksamem Unterricht steht sie im Mittelpunkt: „The major message is simple – what teachers *do* matters. However, this has become a cliché that masks the fact that the greatest source of variance in our system relates to teachers – they can vary in major ways. The codicil is that what ‚some' teachers do matters – especially those who teach in a most deliberate and visible manner" (Hattie, 2009, S. 22).

Es gilt demnach dieses komplexe Konstrukt der Kompetenz des guten Unterrichtens auch für den Deutschunterricht möglichst greifbar zu machen. Vorliegende Arbeit definiert Kompetenz in Anlehnung an die Vorgängerstudien COACTIV und FALKO nach Weinert: „Dabei versteht man unter Kompetenzen die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen

Situationen erfolgreich und verantwortungsvoll anwenden zu können“ (Weinert, 2014, S. 27–28). Allen Kompetenzen ist dabei gemein, dass sie sich auf komplexe Handlungen beziehen, im Gegensatz zur Intelligenz von den jeweiligen Domänen abhängen und sie außerdem durch *deliberate practice* erlernt werden können (vgl. Weinert, 2001; Blömeke, Gustaffson & Shavelson, 2015).

Inwiefern das tatsächliche Umsetzen dieser Fähigkeiten in aktives Tun (i. e. die Performanz) Teil der Kompetenz ist, wird unterschiedlich aufgefasst. Blömeke et al. (2015) zeigen in diesem Zusammenhang folgende Dichotomie auf: Kompetenz wird einerseits als gezeigtes Verhalten aufgefasst. Das bedeutet, dass kognitive Fähigkeiten, affektiv-motivationale Aspekte sowie die Performanz miteinander verbunden sind (i. e. *behavioral perspective*). Andererseits wird Kompetenz als die Summe der kognitiven und motivationalen Ressourcen aufgefasst. In diesem Sinne bilden alle entsprechenden Teilkompetenzen zusammen die Kompetenz. Diese Teilkompetenzen können Verhalten vorhersagen (i. e. *dispositional perspective*) (vgl. Blömeke et al., 2015, S. 6).

Blömeke und andere (2015) plädieren dafür, diese beiden Perspektiven zu verbinden, um der Prozesshaftigkeit von Kompetenz gerecht zu werden und Kompetenz nicht auf Wissen oder Verhalten zu reduzieren. Daher fassen sie Kompetenz als Kontinuum auf, das von den notwendigen kognitiven und affektiv-motivationalen Dispositionen, die die situationsspezifischen Fähigkeiten *Wahrnehmung, Interpretation* sowie *Entscheidungsfindung* lenken, zur tatsächlichen Performanz führt (vgl. Blömeke et al., 2015, S. 7 bzw. die Erläuterungen zum PID-Modell in dieser Arbeit unter Punkt *2.4.2 Beobachterperspektive: Wahrnehmung der Expertinnen und Experten*). Die Messung von Kompetenz stellt dabei eine Herausforderung dar, da immer nur latente Merkmale (über verschiedene theoretische Tests) oder kurze Verhaltensweisen (bei Performanztests) bewertet werden, mit deren Hilfe dann Rückschlüsse auf die Gesamtkompetenz gezogen werden müssen. Dabei ist es nicht sinnvoll, sich auf die Messung von Wissen zu beschränken. Betrachtet werden muss vielmehr konkretes Handeln, das Rückschlüsse auf zugrunde liegende Kompetenzen erlaubt (vgl. Blömeke et al., 2015, S. 8).

Das bisher umfangreichste Modell in diesem Zusammenhang ist nun erläuterte Kaskadenmodell von Krauss und anderen (vgl. Krauss et al., 2020), dem gerade angesprochene Auffassung von Kompetenz als Kontinuum zugrunde liegt. Das Kaskadenmodell greift bestehende empirische Befunde und theoretische Modelle zu Lehrerkompetenz und Unterrichtsqualität auf und fasst sie in ein Modell zusammen. Wie in Abbildung 3 zu erkennen integriert das Kaskadenmodell sowohl das Lehrerkompetenzmodell als auch das kausale Modell zu Unterrichtsqualität, die COACTIV zugrunde liegen (vgl. Baumert & Kunter, 2011), indem nämlich Säule 1 die Lehrkräfte, Säule 3 den Unterricht und Säule 5 die Schüler/-innen fokussiert. Gleichzeitig sieht man in Säule 1 die unten erläuterten

Paradigmen der Lehrerforschung widergespiegelt (grau hinterlegt das Expertenparadigma, darunter hingegen das Persönlichkeitsparadigma, s. Ausführungen unter Punkt *2.2.1 Paradigmen zur professionellen Lehrkraft* in dieser Arbeit). Auch das gerade erwähnte Modell der Kompetenz als Kontinuum von Blömeke und anderen (2015) spiegelt sich im Kaskadenmodell (Säulen 1 bis 3). Krauss und andere (2020) validieren empirisch vier der fünf Säulen (nämlich 1, 2, 3 und 5) des Kaskadenmodells mit den Daten der COACTIV-Studie (vgl. Krauss et al., 2020). Es gelingt mit diesem Modell somit erstmals eine Vielzahl an bisheriger theoretischer und empirischer Forschung zur Lehrerprofessionalität und zur Unterrichtsqualität (im Original wird das Modell als „cascade modell of teaching and learning" Krauss et al., 2020, S. 316 bezeichnet) zu verbinden, so dass es möglich ist, „to put a stronger focus on the common ground that already exists among various projects and constructs […] rather than on their differences" (Krauss et al., 2020, S. 325).

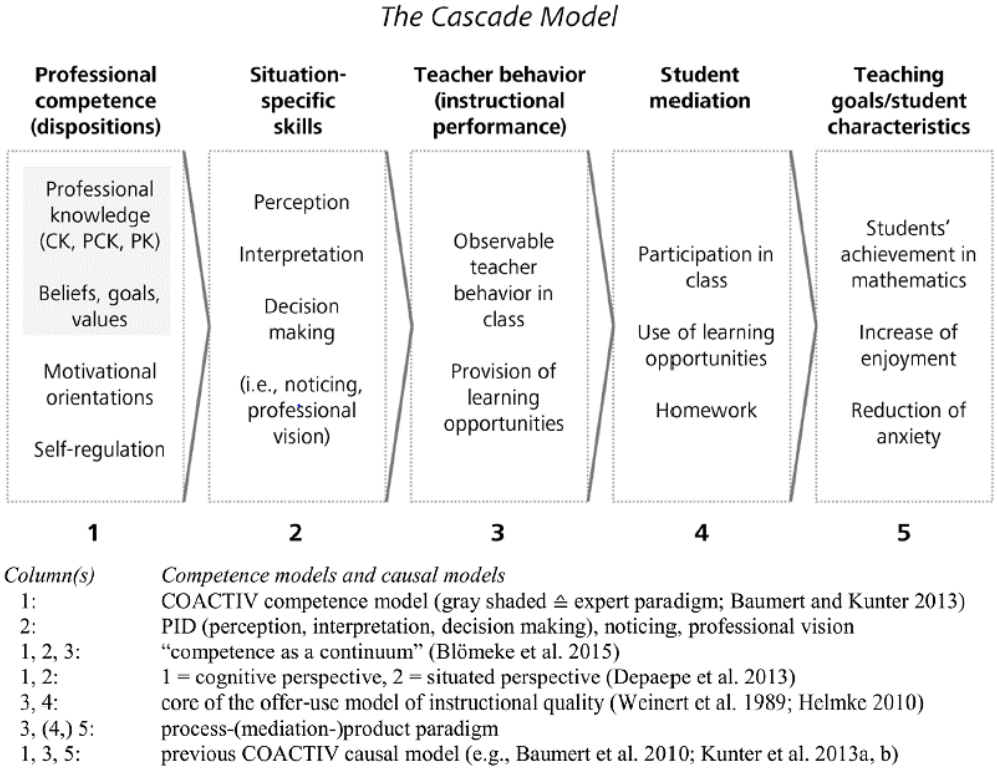


Abbildung 3: Das Kaskadenmodell
aus: Krauss et al., 2020, S. 316

Vorliegende Arbeit knüpft an das Modell der professionellen Kompetenz von Lehrkräften der COACTIV-Studie an, das leicht abgewandelt auch der FALKO-Studie zugrunde liegt (s. Abb. 4 zur *Professionellen Kompetenz*). Die professionelle Kompetenz wird dort in vier Kompetenzaspekte aufgeteilt, von denen das professionelle Wissen[7] die notwendige, aber nicht hinreichende Grundlage bildet. Um dieses Wissen auch anwenden zu können, es also in konkrete Handlungen

[7] Die Begriffe professionelles Wissen und Professionswissen werden in dieser Arbeit synonym verwendet.

zu überführen, bedarf es motivationaler und selbstregulativer Aspekte. Außerdem beeinflussen die eigenen Überzeugungen, Werthaltungen und Ziele bezogen auf den Beruf die Anwendung des professionellen Wissens (vgl. dazu auch Findeisen, 2017, S. 91–95). Diese vier Aspekte werden daraufhin weiter in einzelne Bereiche oder Facetten untergliedert. Wie für FALKO gilt auch bei FALKE das Interesse dem Professionswissen, das in die Facetten Fachwissen, fachdidaktisches Wissen, pädagogisch-psychologisches Wissen, Organisationswissen und Beratungswissen aufgeteilt wird. Die für FALKE-D entscheidenden Facetten werden unter Punkt *2.3 Erklärkompetenz von Lehrkräften* in dieser Arbeit weiter ausgeführt.

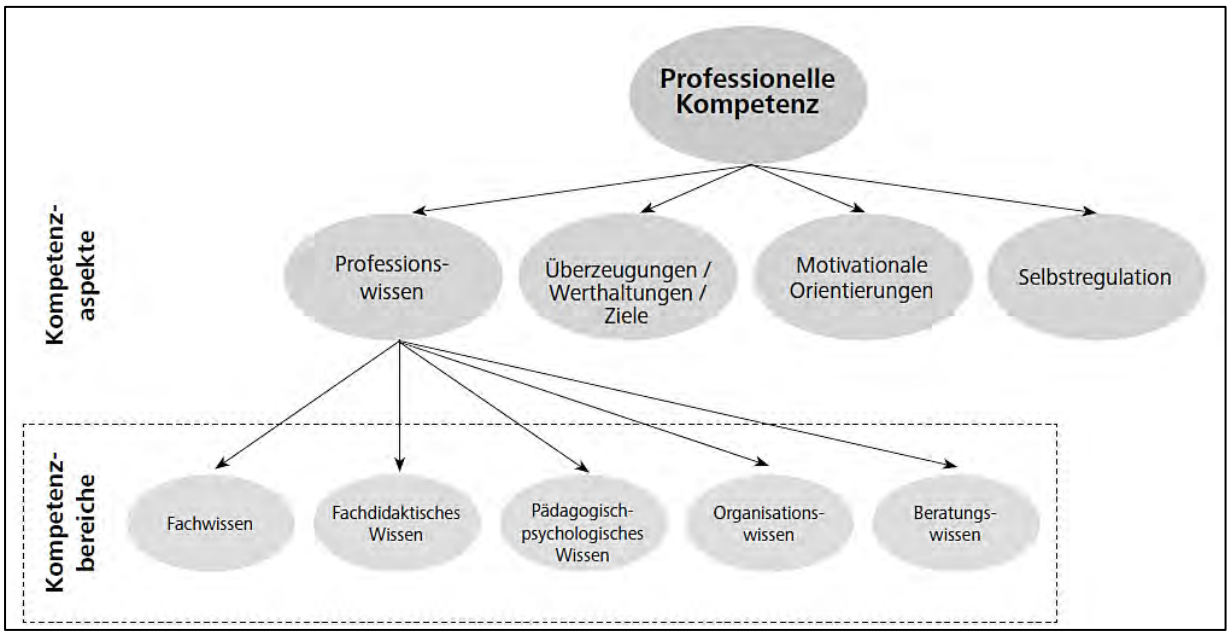


Abbildung 4: Professionelle Kompetenz
leicht abgeändert aus Krauss, Lindl, Schilcher & Tepner, 2017, S. 19

Die Bedeutung der professionellen Lehrkraft wird auch unter dem Blickwinkel der oben ausgeführten Dimensionen für qualitätvollen Unterricht deutlich, denn dieser steht und fällt mit der professionellen Lehrkraft: Sie muss für einen störungsfreien Ablauf des Unterrichts sorgen, wirksame Unterstützung bei Verstehensproblemen anbieten sowie kognitiv aktivierende Aufgaben entwickeln und vieles mehr. Als eine gute beziehungsweise professionelle oder erfolgreiche Lehrkraft wird in diesem Zusammenhang diejenige bezeichnet, die Lernzuwachs bei ihren Schülerinnen und Schülern erreicht. Mit Baumert und Kunter (2006, S. 470) gesprochen ist es die „Kernaufgabe von Lehrkräften […], Unterricht zu erteilen und verständnisvolles Lernen von Schülerinnen und Schülern systematisch anzubahnen und zu unterstützen."

Es ist in diesem Zusammenhang jedoch darauf hinzuweisen, dass Lernerfolg nicht das einzige „Qualitätskriterium" für die gute Lehrkraft ist[8], auch wenn sich vorliegende Studie besagtem Hauptaugenmerk auf den Lernzuwachs bei den Lernenden anschließt. Motivationale Aspekte werden in FALKE zwar mitbedacht (z. B. bei dem Aspekt der *Persönlichkeitswirkung* s. Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* in dieser Arbeit), der Fokus liegt jedoch nicht darauf.

Nach dieser Darstellung unterschiedlicher Modelle von professioneller Kompetenz wird im Folgenden der Blick noch genauer auf die Lehrkraft gelegt. Ähnlich zu den oben beschriebenen Dimensionen von Unterrichtsqualität weist die allgemeine Lehrerforschung auch eine längere Tradition bezüglich dessen auf, was genau eine erfolgreiche Lehrkraft können muss (vgl. z. B. Berliner, 1987; Shulman, 1986). Da in FALKE und somit in FALKE-D Lehrererklärungen im Fokus der Untersuchung stehen, ist zu überlegen, welche Kompetenzen aus der Lehrerforschung allgemein auf die Lehrerkompetenz des Erklärens übertragen werden können.

Krauss und Bruckmaier (2014) unterscheiden in Anlehnung an Bromme (1992) folgende Paradigmen in der Erforschung der guten Lehrkraft.

2.2.1 Paradigmen zur professionellen Lehrkraft

Persönlichkeitsparadigma

Bis in die 60er Jahre des vorherigen Jahrhunderts bestimmte das so genannte Persönlichkeitsparadigma, das das Vorhandensein bestimmter Persönlichkeitsmerkmale als Voraussetzung für eine gute Lehrkraft ansieht, die Forschung. Dieser Ansatz geht davon aus, dass gewisse Charaktereigenschaften (z. B. Geduld) oder Unterrichtsstile (z. B. freundlich) eine Lehrkraft dazu befähigen, ihre Schüler/-innen zu Lernerfolg zu führen. Zwar konnten einige (banale) Zusammenhänge zwischen Persönlichkeitsmerkmalen und dem Lernfortschritt der Schüler/-innen nachgewiesen werden (vgl. Besser & Krauss, 2009, S. 71–72). Ähnliches gilt für die „big five"-Skalen aus der Psychologie: Es zeigen sich hier Hinweise (wenn auch schwache), dass sich niedrige Werte bei „Extraversion" und „Gewissenhaftigkeit" und hohe Werte bei „Neurotizismus" ungünstig auf den Berufserfolg auswirken (vgl. Eckert & Sieland, 2016, S. 147). Dennoch gilt dieser Ansatz sowohl theoretisch als auch empirisch längst als überholt (vgl. Bromme, Rheinberg, Minsel, Winteler & Weidenmann, 2006, S. 299–301).

[8] Zu denken ist hier an pädagogische Fähigkeiten wie Zugewandtheit, aber auch an Geduld und erzieherische Vorbildfunktion (vgl. hierzu Bromme, Rheinberg, Minsel, Winteler und Weidenmann, 2006, S. 296-299) sowie an das Schaffen notwendiger motivationaler Voraussetzungen, um am beruflichen und gesellschaftlichen Leben teilzunehmen (vgl. Baumert und Kunter, 2006, S. 472).

Es ist sicherlich kein Spezifikum des Lehrberufs, dass ein freundlicher und sympathischer Mensch erfolgreicher in seinem Tun ist als ein depressiver unsympathischer. Allerdings ist festzuhalten, dass im außeruniversitären Bereich Aussprüche wie „Sie ist die geborene Lehrerin" durchaus üblich sind und als gültig angenommen werden. Im Gegensatz zur Wissenschaft herrscht landläufig die Meinung, dass es bestimmte Persönlichkeitsmerkmale gibt, die die gute Lehrerin oder den guten Lehrer ausmachen.
Auch in der Beratung über einen möglichen Lehrberuf scheinen einige Ergebnisse aus der Forschung zu Lehrerpersönlichkeiten weiterhin von Interesse zu sein (vgl. Bromme et al., 2006, S. 322).

Prozess-(Mediations-)Produkt-Paradigma

In der Lehr-/Lernforschung wendete sich der Fokus in den 1960er Jahren weg von der Persönlichkeit der Lehrkraft hin zu ihrem (konkreten und beobachtbaren) Verhalten. Laut diesem Prozess-Produkt-Paradigma ist eine erfolgreiche Lehrkraft die, die bestimmte, besonders wirksame Lerntechniken anwendet, so dass die Schüler/-innen aufgrund dieses Unterrichthandelns einen hohen Lernerfolg erfahren. Untersucht werden hierbei die so genannten Prozesse, zum Beispiel die Anzahl und Art der Fragestellungen, der Umgang mit Schülerfehlern oder die Art der Rückmeldung auf Schüleräußerungen. Dazu wird der Unterricht entweder beobachtet oder videographiert und die Ergebnisse dieser Untersuchungen werden mit dem Lernzuwachs der Schüler/-innen, den Produkten, in Verbindung gesetzt (vgl. Bromme et al., 2006, S. 301; Besser & Krauss, 2009, S. 72). Forschungen unter diesem Paradigma konnten belastbare Ergebnisse zu wirkungsvollen Faktoren für erfolgreiche Lehrkräfte herausarbeiten, zum Beispiel eine angemessene Unterrichtsgeschwindigkeit oder die Adaptivität des Unterrichtsstoffes an die individuellen Voraussetzungen der Schüler/-innen (vgl. Besser & Krauss, 2009, S. 72) sowie eine klare und strukturierte Präsentation von Lernmaterial (vgl. Klieme et al., 2006, S. 127–128). Vergleicht man diese Faktoren mit den kurz angerissenen Sicht- und Tiefenstrukturen von Unterricht, so fällt auf, dass im Prozess-Produkt-Paradigma vor allem Oberflächenmerkmale beschrieben werden, die leicht beobachtbar sind (s. Ausführungen unter *2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht* in dieser Arbeit).
In den frühen 1980er Jahren wurde dieser Ansatz zum Prozess-Mediations-Produkt-Paradigma weiterentwickelt, denn es stellte sich heraus, dass sich nicht jedes Lehrerverhalten bei jeder Schülerin beziehungsweise bei jedem Schüler gleich positiv auf den Lernerfolg auswirkt (vgl. Bromme et al., 2006, S. 304). Deswegen wurden fortan in Anlehnung an die kognitive Wende neben gerade genannten wirksamen Lehrtechniken der Lehrkraft auch individuelle Verarbeitungsprozesse der Schüler/-innen, die so genannten Mediationen, als Merkmale des Lernprozesses berücksichtigt. Die lerntheoretischen Ansätze des Konstruktivismus stimmen mit

diesem Paradigma überein: Die beste Lehrtechnik ist unwirksam, wenn die Lernerin oder der Lerner nicht lernen möchte oder kann.

Im Gegensatz zu oben beschriebenem Persönlichkeitsparadigma sind die Kompetenzen, die eine gute Lehrkraft ausmachen, in diesem Paradigma grundsätzlich erlernbar.

Herausforderungen für Studien unter diesem Paradigma ergaben sich folgende: Zunächst hängen die einzelnen wirksamen Verhaltensweisen untereinander zusammen. So ist es durchaus möglich und realistisch, dass Lehrkräfte insgesamt sehr hohe Lernerfolge bei ihren Schülerinnen und Schülern erzielen, obwohl sie bei bestimmten Merkmalen deutlich unter Durchschnitt abschneiden, bei anderen dafür deutlich überdurchschnittlich (oder umgekehrt: Eine Lehrkraft erzielt wenig Lernerfolg, weist aber bei einzelnen Merkmalen überdurchschnittlich gute Leistungen auf). Es lässt sich nicht „die" wirkungsvollste Verhaltensweise finden, da jedes Lehrerverhalten ein Konglomerat aus einzelnen und unterschiedlichen Faktoren ist. Ebenso ist es möglich, dass ein und dieselbe Lehrkraft je nach Thema oder Klasse unterschiedlich gut agiert und so unterschiedlichen Erfolg bei bestimmten Themen beziehungsweise in bestimmten Klassen hat. Auch die Art und Weise, wie wirksame Lehrtechniken angewandt werden (z. B. aufgrund von unterschiedlichen Charaktereigenschaften der Lehrkräfte), bestimmt den Erfolg maßgeblich. Hier ist also mit Ergebnissen zu rechnen, die sich nicht automatisch verallgemeinern lassen (vgl. Bromme et al., 2006, S. 303).

Eine wechselseitige Wirkung besteht auch zwischen dem Verhalten der Lehrkraft und dem der Klasse: Sowohl sehr erfolgreiche als auch weniger erfolgreiche Lehrkräfte handeln zum Beispiel in aufgrund ihres Verhaltens sehr schwierigen Klassen anders als in motivierten und einfach zu führenden Klassen. Beide Gruppen von Lehrkräften erzielen in ersteren Klassen weniger Lernfolge als bei einfach zu handhabenden, interessierten Schülerinnen und Schülern (vgl. Bromme et al., 2006, S. 304).

In der Forschung kristallisierte sich nach und nach außerdem die Bedeutung des Faches und des konkreten fachlichen Inhaltes heraus, weil viele Kriterien des Prozess-Produkt-Paradigmas nur fachspezifisch erfasst werden können. Ob der Unterricht zum Beispiel Klarheit aufweist oder ob der Unterrichtsgegenstand fachlich präzise und korrekt dargestellt wird, lässt sich nur aus der Fachperspektive heraus bestimmen (vgl. Seidel & Shavelson, 2007, 459 bzw. 483; Bromme et al., 2006, S. 302–303). So nimmt es nicht wunder, dass der Bedeutung des Faches immer stärkere Aufmerksamkeit zuteilwird, was sich auch im nun erläuterten Expertenparadigma zeigt.

Expertenparadigma

Im Zuge der Expertiseforschung aus der Kognitionspsychologie verschiebt sich der Fokus etwa seit Mitte der 1980er Jahre noch einmal; erneut rückt die Lehrperson in den Mittelpunkt des

Forschungsinteresses, denn eine erfolgreiche Lehrkraft handelt aufgrund ihrer Expertise professionell. Expertise zeigt sich im Lehrberuf zum Beispiel „beim Reflektieren pädagogischer Praxis [oder] bei kriteriengestützter Awareness" (Kuntze, 2016, S. 528). Das Experten-Paradigma besagt, dass eine Lehrkraft über professionalisierte Unterrichtsexpertise, das heißt über Wissen und Fähigkeiten im Bereich des Unterrichtens (und nicht etwa über bestimmte Persönlichkeitsmerkmale wie beim Persönlichkeitsparadigma) verfügt, die sie zum erfolgreichen Ausüben ihrer Profession benötigt (vgl. Besser & Krauss, 2009, S. 73–77 bzw. Krauss & Bruckmaier, 2014, S. 242–243). Professionelles Wissen (egal für welches Berufsfeld) charakterisieren Baumert und Kunter (2006) folgendermaßen:

- Professionelles Wissen ist domänenspezifisch und ausbildungs- bzw. trainingsabhängig.
- Expertenwissen ist sehr gut vernetzt und hierarchisch organisiert.
- In professionellen Domänen ist Expertenwissen um Schlüsselkonzepte und eine begrenzte Zahl von Ereignisschemata arrangiert, an die Einzelfälle, episodische Einheiten oder Sequenzen von Episoden (Skripts) angedockt sind.
- Professionelles Expertenwissen integriert Kontexte und erlaubt variantenreicheres „opportunistisches Verhalten".
- Basisprozeduren sind automatisiert, aber gleichwohl flexibel an die spezifischen Bedingungen des Einzelfalles und des Kontextes adaptierbar (vgl. Hatano & Inagaki, 1986) (Baumert & Kunter, 2006, S. 483).

Die professionelle Lehrkraft verfügt über derartiges Wissen über das Unterrichten und ist somit dafür eine Expertin/ein Experte, das heißt ein Fachmann oder eine Fachfrau, im Unterschied zu allen anderen Menschen, die diesen Beruf nicht ausüben und deshalb bezogen auf den Lehrerberuf Laien sind. Hervorzuheben ist, wie aus der obigen Definition ersichtlich, dass – im Gegensatz zum Persönlichkeitsparadigma – das zur Berufsausübung benötigte Wissen und die benötigten Fähigkeiten erlernbar sind und von Novizinnen und Novizen im Laufe ihrer Ausbildung erworben werden (vgl. Bromme et al., 2006, S. 305). Es handelt sich bei diesem Wissen vor allem um prozedurales Wissen und bei den Fähigkeiten (d. h. Können) nicht um direkt beobachtbares Verhalten (wie im Prozess-Produkt-Paradigma), sondern um erworbene Persönlichkeitsattribute (vgl. Besser & Krauss, 2009, S. 73–74).[9] Somit beschreibt das Expertenparadigma im Gegensatz zum Prozess-Produkt-Paradigma Tiefenstrukturen, die in jüngerer Forschung verstärkt im Fokus stehen (s. Erläuterungen unter *2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht* in dieser Arbeit).

[9] Abzugrenzen ist der Expertenbegriff in diesem Zusammenhang von der Bedeutung der Expertin/des Experten als einer Person, die dauerhaft herausragende Leistungen auf einem Gebiet erbringt. Dies entspricht der angelsächsischen Konzeptualisierung des *expert teacher*, bei dessen Erforschung es darum geht, herauszufinden, was Lehrkräfte, die ihren Beruf besonders erfolgreich ausüben, von Lehrkräften unterscheidet, die das weniger erfolgreich tun. Dem Verständnis nach Bromme (1992) folgend sind Expertinnen/Experten wie oben dargelegt alle Personen, die über das notwendige professionelle Wissen verfügen, um den Lehrerberuf auszuführen (vgl. dazu die ausführliche Darstellung bei Besser und Krauss (2009)).

Die Darstellung der Paradigmen erfolgte in ihrem zeitlichen Ablauf; tatsächlich handelt es sich aber nicht um drei unabhängige Ansätze, vielmehr geht ein Paradigma jeweils aus dem vorherigen hervor (vgl. zu diesem Absatz Krauss & Bruckmaier, 2014, S. 241–243).

2.2.2 Professionswissen von Lehrkräften

Es lässt sich somit noch einmal festhalten, dass professionelle Lehrkräfte über professionelles Wissen über den Lehrberuf verfügen (s. auch Abb. 4). Bei der Frage, wie genau dieses Wissen beschaffen ist und welche Teilgebiete es umfasst, wird weiterhin auf die Taxonomie von Shulman (1986, 1987) zurückgegriffen, dessen Arbeiten zum Professionswissen maßgeblich sind.

Shulmans (1986; 1987) Taxonomie der professionellen Wissensdomänen von Lehrkräften

Diese Arbeiten bilden sowohl in der COACTIV-Studie wie auch in der FALKO- und der FALKE-Studie die Grundlage für die dortige Konzeptualisierung des Professionswissens (s. Abb. 4). Shulman unterscheidet folgende Kategorien des Professionswissens für Lehrkräfte: Fachwissen, allgemeines pädagogisches Wissen, Wissen über die gültigen Curricula, fachdidaktisches Wissen, Wissen über die Lernenden und ihre Eigenschaften, Wissen über gültigen Bildungskontext sowie Wissen über Bildungsziele, -werte und deren philosophischen und historischen Hintergrund (vgl. Shulman, 1987, S. 7–8; eigene Übersetzung).

Mit dem professionellen Wissen als Grundlage entwickelt Shulman einen „Handlungsplan" (im Original: *A Model of Pedagogical Reasoning and Action* (1987, S. 15)), den Lehrkräfte in der Vorbereitung, während und nach dem Unterricht verfolgen. Dieser besteht aus einzelnen Schritten, zum Beispiel dem Aufbereiten des Unterrichtsstoffes, dem tatsächlichen Instruieren im Unterricht und dem Evaluieren während beziehungsweise nach dem Unterricht (vgl. Shulman, 1987, S. 15–16).

Fachwissen

Bereits in seinen frühen Arbeiten betont Shulman die Bedeutung des Fachwissens als Grundlage und Voraussetzung für gutes Unterrichten (vgl. Shulman, 1987, S. 9) sowie die Bedeutung des Inhaltes einer Unterrichtsstunde, wenn es um die Reflexion von konkretem Unterricht geht (vgl. Shulman, 1987, S. 20). Er kritisiert, dass sowohl Forschung zum Lehrberuf als auch Feedback an Lehrkräfte auf konkretes Unterrichtshandeln oft losgelöst von fachspezifischen Aspekten rein basierend auf allgemeinpädagogischen Beobachtungen stattfinde und die Bedeutung des Fachwissens so vernachlässigt werde (vgl. Shulman, 1987, S. 11; S. 20). Shulman hingegen stellt die Bedeutung von vertieftem, vernetzten Wissen über die Unterrichtsgegenstände für die Qualität des Unterrichts heraus, das idealerweise verschiedene Zugänge zum gleichen Inhalt bereitstellt.

Eine Zuwendung zum Fachwissen zeigt sich auch in der Lehr-/Lernforschung jüngeren Datums. Sie hat weitreichende Konsequenzen für die Konzeptualisierung der Anforderungen an den Lehrerberuf:

> Mit der Betonung, dass erfolgreiches Unterrichten vor allem aufgrund einer gut vernetzten und umfangreichen domänenspezifischen Wissensbasis ermöglicht wird und dass diese Wissensbasis im Rahmen der strukturierten Lehreraus- und Weiterbildung vermittelbar ist, kann der Lehrerberuf als Profession verstanden werden, also als ein Berufsfeld, das sich unter anderem durch eine intensive spezialisierte Ausbildung, eine Tätigkeit in autonomen und nichtroutine-basierten Situationen, eine gemeinsame theoretische Wissensbasis und spezialisierte Fertigkeiten sowie systematische Qualitätssicherung und Wissenserweiterung innerhalb des Feldes auszeichnet […] (Baumert & Kunter, 2011, S. 30).

Das Experten-Paradigma, das in der Lehrerprofessionsforschung heutzutage breite Zustimmung findet, fügt sich gut in diese Konzeptualisierung von professionellem Wissen von Lehrkräften ein, denn dort ist die Rolle des fachspezifischen Wissens nicht von der Hand zu weisen. Ein Teil des professionellen Wissens über das eine Lehrkraft, die Expertin/Experte ist, verfügt, muss Wissen aus dem jeweiligen Fach („Entwicklung des fachbezogenen Wissens“ (Besser & Krauss, 2009, S. 76)) sein: „Weil die Verknüpfung und Abstimmung zwischen den curricular-fachbezogenen und den allgemeineren didaktischen und psychologischen Aspekten des Lehr-Lern-Prozesses eine zentrale Anforderung des Lehrerberufs ist, bezieht die psychologische Forschung zum professionellen Wissen auch das fachbezogene Lehrerwissen mit ein“ (Bromme et al., 2006, S. 332).

Fachdidaktisches Wissen

Neben dem Fachwissen ist das fachdidaktische Wissen (bei Shulman (1986) *Pedagogical Content Knowledge*) entscheidend und charakteristisch für die Lehrkraft als Expertin/Experte, denn dieses unterscheidet Lehrkräfte von Expertinnen und Experten aus dem Fach und macht sie zu Expertinnen und Experten für das Unterrichten eines Faches (vgl. Shulman, 1987, S. 8). Das fachdidaktische Wissen ist abhängig vom Fach, das heißt je nach Fachinhalten wird unterschiedliches fachdidaktisches Wissen benötigt, um erfolgreich unterrichten zu können (vgl. Sevian & Gonsalves, 2008, S. 1464).

Auch die anderen oben aufgezählten Wissenskomponenten nach Shulman finden Eingang in die Lehrerforschung und Modelle zum Professionswissen, zum Beispiel im Modell bei COACTIV (vgl. Baumert & Kunter, 2011, S. 32 bzw. Abb. 4 und Ausführungen unten). Vor allem etabliert haben sich in der aktuellen Forschung die drei Facetten Fachwissen, allgemeines pädagogisches Wissen und fachdidaktisches Wissen, die auch bei FALKO angenommen werden (vgl. Krauss, Lindl, Schilcher, Fricke et al., 2017, S. 19 bzw. Ausführungen unten).

Aufgrund der stärkeren Betonung des Fachwissens in der Lehr-/Lernforschung ist es nicht verwunderlich, dass zunehmend die Fachdidaktiken Anforderungen und Kompetenzen für eine

professionelle Lehrkraft im Fach ausdifferenzieren, oft in Zusammenarbeit mit den Bildungs- und Erziehungswissenschaften, die die Notwendigkeit der fachspezifischen Ausrichtung selbst betonen (vgl. z. B. Baumert & Kunter, 2006, S. 492). Dabei entstehen zum Beispiel Wissenstests, die das professionelle Wissen von Lehrkräften prüfen. Auch die Deutschdidaktik greift auf die Ergebnisse und Methoden aus der Bildungswissenschaft zurück und entwickelt unterschiedliche Testformate, die das Professionswissen von Deutschlehrkräften bestimmen (vgl. Bremerich-Vos, Dämmer, Willenberg & Schwippert, 2011; Krauss, Lindl, Schilcher, Fricke et al., 2017; Pissarek & Schilcher, 2017; Winkler, 2017).

Wenn diese Tests allerdings von bildungspolitischen Überlegungen bestimmt sind, ist Vorsicht geboten, wie Berliner (2005, S. 206) deutlich macht: Die Bestimmung von Qualität bedeute immer ein Werteurteil, weswegen es wichtig sei, sich darüber im Klaren zu sein, von wem und zu welchem Zweck die Qualität von Unterricht bestimmt werden soll. In diesem Zusammenhang weist er zum einen auf kulturelle Unterschiede bei der Bestimmung von Professionswissen hin. So würden zum Beispiel pädagogische Fähigkeiten des *classroom management* in sehr autoritären Schulsystemen wie Indien gar nicht genannt, da dort kaum Disziplinprobleme auftreten, während in westlichen Konzeptualisierungen von Professionswissen für Lehrkräfte (z. B. Shulman, 1986) die pädagogischen Fähigkeiten immer eine eigene Facette darstellten. Zum anderen kritisiert Berliner (2005), dass Professionswissenstests für angehende Lehrkräfte und deren Ergebnisse in den USA stark politisch motiviert seien: Je mehr Bedarf an Lehrkräften ein Bundesstaat habe, umso mehr Bewerber/-innen bestünden den Test (vgl. Berliner, 2005, S. 209). Diese Kritik trifft auf Professionswissenstests, die an deutschen Universitäten zur Forschung entstehen nicht zu. Zum einen werden wie bei FALKO nicht nur angehende Lehrkräfte, sondern auch Lehrkräfte mit Berufserfahrung getestet, die bereits eine Stelle besitzen. Zum anderen dienen die Tests wissenschaftlichen Fragestellungen; sie werden gänzlich unabhängig von politischen Interessen erstellt und spielen daher bei der Vergabe von Stellen keine Rolle.

Professionswissen in der COACTIV-Studie

Wissenstests, die in den Erziehungswissenschaften beziehungsweise Fachdidaktiken entwickelt werden, haben dagegen zum Ziel, Unterricht besser (d. h. qualitätvoller) zu machen. Einen Meilenstein bildet hier die bereits erwähnte COACTIV-Studie, die professionelles Wissen von Mathematiklehrkräften testet (vgl. Kunter et al., 2011). Ausgewählt wurden Mathematiklehrkräfte, deren Klassen am PISA-Test 2003 (Fokus: Mathematik) teilnahmen, um so ein vollständigeres Bild des Unterrichtsgeschehens zu erlangen. Dazu erhielten die Lehrkräfte über ein Schuljahr verteilt Fragebögen zu motivationalen Aspekten sowie zu ihrer Berufserfahrung und subjektivem Berufserleben sowie Tests, in denen professionelles Mathematiklehrerwissen überprüft wurde. Das

Professionswissen wird in Anlehnung an Shulman (1986) in die Facetten Fachwissen und fachdidaktisches Wissen (s. auch Erläuterungen weiter oben) unterteilt. Ergänzend reichten die Lehrkräfte Arbeitsblätter und Leistungsnachweise ein, die sie im Laufe des Schuljahres in der betroffenen Klasse verwendet hatten. Durch diese sehr umfassende Erhebung sowie die vorliegenden PISA-Ergebnisse und weitere Leistungserhebungen der Schüler/-innen zu Ende zweier Schuljahre war es möglich, den Zusammenhang zwischen Lehrerwissen und -handeln und dem Lernzuwachs der Schüler/-innen zu ermitteln. Auf diese Weise konnte eine breite Basis an Daten über das Unterrichtsgeschehen erhoben werden, die Rückschlüsse auf Tiefenstrukturen im Unterricht erlaubt.

Das für vorliegende Studie interessanteste Ergebnis der COACTIV-Studie ist, dass sich das mathematikdidaktische Wissen als prädiktiv valide für den Lernzuwachs der Schüler/-innen im Mathematikunterricht erweist. Für das Fachwissen kann keine vergleichbare Vorhersagekraft nachgewiesen werden. Allerdings ist davon auszugehen, dass das Fachwissen die Grundlage für das fachdidaktische Wissen bildet, worauf auch die engen Korrelationen der beiden Facetten in COACTIV hindeuten. Zur ausführlichen Konzeptualisierung und Operationalisierung des Professionswissens von Mathematiklehrkräften sowie zu den entsprechenden Ergebnissen siehe Kunter et al. (2011) sowie Krauss et al. (2008).

Neben der Mathematikdidaktik untersuchen auch immer mehr andere Fachdidaktiken, darunter auch die Deutschdidaktik, die Frage, was eine professionelle Lehrkraft in ihrem Fach wissen muss.

Professionswissen in der Studie TEDS-LT

Zu erwähnen ist für die Deutschdidaktik in diesem Zusammenhang die TEDS-LT-Studie (vgl. Blömeke, 2011; Bremerich-Vos et al., 2011; Bremerich-Vos & Dämmer, 2013), die das professionelle Wissen von Deutschlehramtsstudierenden quantitativ untersucht. Das professionelle Wissen wird bei TEDS-LT in die linguistische, die literaturwissenschaftliche sowie die deutschdidaktische Dimension unterteilt, wobei letztere wiederum zwischen sprach- und literaturdidaktischen Kompetenzbereichen unterscheidet. Die knapp 90 Testfragen dazu orientieren sich vor allem an den Curricula der Deutschlehramtsstudiengänge an deutschen Universitäten sowie an deutschdidaktischen Standardwerken. Dass die Konzeptualisierung des professionellen Wissens von Deutschlehrkräften die deutschdidaktische Forschung vor große Herausforderungen stellt, zeigen die Ergebnisse der TEDS-LT-Studie: Bei einigen Studierenden nimmt das Fachwissen, so wie es in der Studie konzeptualisiert und operationalisiert wurde, im Studienverlauf zum Beispiel ab statt zu. Wieser (2015) nimmt die Studie zum Anlass, Professionswissenstests für Deutschlehrkräfte generell in Frage zu stellen, da innerhalb der Deutschdidaktik keine Einigkeit über die Inhalte des professionellen Wissens bestünde. Für sie ist somit „die Basis für eine kompetenzorientierte

Forschung, wie sie beispielsweise im Rahmen von TEDS-LT durchgeführt wird, letztlich nur eingeschränkt gegeben“ (Wieser, 2015, S. 25). Diesem Einwand sind zwei Argumente entgegenzuhalten: Divergentes Wissen und divergente Lösungen sind typisch und kennzeichnend für das Fach Deutsch; was für den Unterricht in der Schule gilt, gilt ebenso für die Erforschung dieses Unterrichts. Somit ist gerade ein kontroverser Diskurs über die möglichen Inhalte eines Professionswissenstests vielversprechend und zugleich Abbild der Eigenständigkeit des deutschdidaktischen Forschungsdiskurses. Zum anderen wäre die letztendliche Konsequenz aus Wiesers Argumentation das Ablehnen von Professionswissenstests für Deutschlehrkräfte aufgrund fehlender inhaltlicher Einigung im Fachdiskurs. Erneut gilt aber auch hier, dass es sicherlich gewinnbringender für die Fachgemeinschaft ist, Professionswissenstests zu entwickeln und diese zur Diskussion zu stellen, – auch wenn und gerade weil keine (schnelle) inhaltliche Einigung zu erwarten ist.

Die Studie TEDS-LT ist bislang die größte empirische Studie zum Professionswissen von angehenden Deutschlehrkräften und bietet eine wichtige Grundlage im Fach, um den Diskurs voranzubringen.

Professionswissen in der Studie FALKO-D

Eine weitere quantitative Studie, an der die Deutschdidaktik beteiligt war, ist die bereits erwähnte FALKO-Studie. In der Vorgängerstudie von FALKE entwickelten wie bereits angedeutet sechs Fachdidaktiken, darunter auch Deutsch, parallel aufgebaute Professionswissenstests für Lehrkräfte und konnten somit neben den Erkenntnissen für das eigene Fach auch fachübergreifende Vergleiche anstellen (vgl. Krauss, Lindl, Schilcher, Fricke et al., 2017). Die Konzeptualisierung der Professionswissenstests lehnt sich an die oben erläuterte COACTIV-Studie an. Untersucht werden dabei abermals Fachwissen (FW) sowie fachdidaktisches Wissen (FDW) in Anlehnung an Shulman (1986) (s. auch Abb. 4 bzw. Abb. 5). Dessen Taxonomie erweist sich für das FALKO-Projekt als besonders gewinnbringend, da in ihr das Fachwissen und das fachdidaktische Wissen jeweils aus der Perspektive des Faches bestimmt werden, sie aufgrund der gemeinsamen Kompetenzfacetten aber dennoch fachübergreifende Vergleiche ermöglicht. Die Forschergruppe untersucht zum Beispiel, welche fachdidaktischen Kompetenzen jeweils besonders wichtig für ein Fach sind (vgl. Lindl & Krauss, 2017, S. 411–422). Mit der Fokussierung auf das fachspezifische Wissen von Lehrkräften verortet sich die FALKO-Studie im oben beschriebenen Experten-Paradigma, das wie dargelegt von einer zentralen Rolle des Lehrerwissens bei der Steuerung erfolgreicher Lehr-/Lernprozesse ausgeht.

Für das Fach Deutsch entwickelten Pissarek und Schilcher (2017) einen Test zum Messen des Professionswissens von Deutschlehrkräften, der Lehrkräften mit unterschiedlich langer Berufserfahrung, Lehramtsanwärterinnen und -anwärtern sowie Lehramtsstudierenden vorgelegt

wurde. Sie erhielten 11 Fragen zum Fachwissen und 17 zum fachdidaktischen Wissen. Die Bearbeitungsdauer eines Testheftes belief sich auf 90 Minuten, von denen etwa 30 auf den fachwissenschaftlichen und 60 Minuten auf den fachdidaktischen Teil entfielen.

Das Fachwissen wird in der Gesamtstudie in Anlehnung an die Ergebnisse aus der COACTIV-Studie als vertieftes Sekundarschulwissen konzeptualisiert. Dies wird in der Studie in Analogie zum mathematischen Fachwissen bei COACTIV (vgl. Krauss et al., 2011, S. 142) als Wissen der sogenannten Ebene 3 bezeichnet. Darunter, auf Ebene 2, siedelt sich die Beherrschung des Schulstoffes durch eine/-n durchschnittliche/-n bis gute/-n Schüler/-in der jeweiligen Jahrgangsstufe an; darüber, auf Ebene 4, liegt das universitäre Wissen (vgl. Pissarek & Schilcher, 2017, S. 80–81). Eine Aufgabe zum Fachwissen ist in der Studie zum Beispiel das Zerlegen eines komplexen Satzes in die einzelnen Satzglieder (vgl. Pissarek & Schilcher, 2017, S. 83).

Das fachdidaktische Wissen in FALKO-D wird analog zu allen anderen beteiligten Fächern in drei Dimensionen aufgeteilt und zwar projektübergreifend in Anlehnung an Shulmans Ausdifferenzierung des fachdidaktischen Wissens (vgl. Shulman, 1986, S. 9–10) in die Facetten „Erklären und Repräsentieren“ und „Umgang mit Schülerkognitionen“ sowie deutschspezifisch in die Facette „Einschätzung des Potenzials von Texten“ (vgl. Pissarek & Schilcher, 2017, S. 77–79). Unter der ersten Facette „Erklären und Repräsentieren“ finden sich Fragen zur konkreten Strukturierung, Darstellung und Vernetzung von Unterrichtsgegenständen. Hier wird zum Beispiel nach Strategien gefragt, die einem Schüler bei seiner mangelhaften Kommasetzung zwischen Teilsätzen helfen können (vgl. Pissarek & Schilcher, 2017, S. 82).

Die Facette „Umgang mit Schülerkognitionen“ versammelt Aufgaben zur Beurteilung und zur Lösung von typischen Schülerproblemen bei der Vermittlung eines spezifischen Gegenstandes. Ein Beispiel hierfür ist ein Schülertext, für den die Studienteilnehmenden zwei Hauptprobleme identifizieren sollen, die sie für die unmittelbare Weiterarbeit auswählen würden (vgl. Pissarek & Schilcher, 2017, S. 82). In der deutschspezifischen Facette der „Einschätzung des Potenzials von Texten“ wird von den Studienteilnehmerinnen und -teilnehmern verlangt, das Lernpotenzial konkreter Texte (sowohl von literarischen als auch von Sachtexten) einzuschätzen. Eine Aufgabe aus diesem Bereich ist die Auswahl eines Tafelbildes zur geeigneten Darstellung des Lernpotenziales eines gegebenen Gedichts (vgl. Pissarek & Schilcher, 2017, S. 83).

Ziel von FALKO war es, prozedurales Wissen zu erheben, das zumindest theoretisch im Unterricht umgesetzt werden könnte. Deswegen war der überwiegende Teil der Fragestellungen im offenen Format gestellt, „um erheben zu können, wie viele und welche Handlungsoptionen Lehrkräften für prototypische Situationen zu Verfügung stehen“ (Pissarek & Schilcher, 2015, S. 329), auch wenn dies die Auswertungsobjektivität zwangsläufig einschränkt.

In FALKO-D können, wie in allen anderen beteiligten Fächern auch, das Fachwissen und das fachdidaktische Wissen mit seinen drei Facetten als unterschiedliche Konstrukte nachgewiesen werden, die miteinander korrelieren, aber statistisch dennoch klar zu trennen sind (vgl. Pissarek & Schilcher, 2017, S. 92). Das Fachwissen zeigt sich als notwendige, aber nicht hinreichende Bedingung für das fachdidaktische Wissen. Interessanterweise korreliert gerade die Facette „Einschätzung des Potenzials von Texten" am wenigsten mit dem Fachwissen (die COACTIV-Studie weist dazu ein analoges Ergebnis auf), obwohl man gerade bei dieser Facette einen hohen fachwissenschaftlichen Anteil annimmt (vgl. Pissarek & Schilcher, 2017, S. 92). Insgesamt ist die Korrelation der beiden Facetten im Fach Deutsch im Vergleich zu den anderen Fächern mit r= 0,51** relativ niedrig (vgl. Pissarek & Schilcher, 2017, S. 92). Pissarek und Schilcher geben dafür als möglichen Grund eine zunehmende Orientierung der Fachdidaktik an der Lehr-/Lernpsychologie und damit eine eventuelle Entkoppelung von den Fachwissenschaften an. Gerade wenn es zum Beispiel um die Vermittlung von Strategien zum Lesen oder Schreiben geht, greift die Didaktik verstärkt auf Erkenntnisse aus der Lehr-/Lernpsychologie und nicht so sehr aus den Fachwissenschaften zurück (vgl. Pissarek & Schilcher, 2017, S. 92).

Bezogen auf die unterschiedlichen erhobenen Statusgruppen zeigen die Ergebnisse der FALKO-Studie erwartungsgemäß bessere Leistungen im Fachwissen und im fachdidaktischen Wissen bei den Lehrkräften im Vergleich zu den Studierenden, wobei der Unterschied im Fachwissen erheblich größer als im fachdidaktischen Wissen ausfällt (zu den genaueren Ergebnissen siehe Pissarek und Schilcher, 2017, S. 89-99).

Lehrerkompetenz wird bei FALKO durch das richtige Lösen von Aufgaben nachgewiesen. Dabei wird in FALKO gerade nicht postuliert, dass die Items sämtliche Inhalte des Faches Deutsch oder „die" deutschdidaktische Kompetenz abbilden. Vielmehr wird davon ausgegangen, dass Lehrkräfte, die die in FALKO gestellten Aufgaben gut lösen können, auch andere Aufgaben auf einem ähnlichen Niveau lösen können. FALKO fasst Lehrerkompetenz somit als latentes Konstrukt auf, das sich der direkten Beobachtbarkeit entzieht, sich aber in der Lösung der Variablen niederschlägt. Insgesamt ist festzuhalten, dass FALKO-D einen wichtigen und validen Baustein für die Messung der fachwissenschaftlichen und fachdidaktischen Kompetenzen von Deutschlehrkräften darstellt, der als Ausgangspunkt für weitere Forschung dient.

Aktueller Forschungsstand

Der deutschdidaktische Forschungsdiskurs zu qualitätvollem Unterricht und zur Rolle der professionellen Lehrkraft ist gerade in den letzten zehn Jahren deutlich lebendiger geworden, wie man an der steigenden Anzahl der Publikationen sieht: In ihrem Sammelband zur empirischen Lehrerforschung in der Deutschdidaktik untergliedern Bräuer und Wieser (2015) die dort

vorgestellten Studien in folgende drei Oberpunkte: 1) die Untersuchung von Vorstellungen und Orientierungen von Lehrenden, 2) die Rekonstruktion und Analyse des Lehrerhandelns in Unterrichtssituationen und 3) die Erhebung von fachlich-didaktischen Kompetenzen (vgl. Bräuer & Wieser, 2015, S. 9). Bei der Durchsicht des Bandes fällt auf, dass vorwiegend (bzw. unter den Punkten 1 und 2 sogar ausschließlich) kleinere, qualitative Studien genannt werden, die normalerweise auf der Grundlage von Interviews oder Fallskizzen arbeiten (siehe dort z. B. Lindow, 2015; Pflugmacher, 2015; Reinert, 2015; Scherf, 2015). Quantitative Arbeiten stellen in der Deutschdidaktik nach wie vor die Ausnahme dar.

Neben der dargestellten Anzahl an Studien und Veröffentlichungen zeigt sich die Lebendigkeit der Lehrerforschung in der aktuellen Deutschdidaktik auch an dem Thema „Professionalisierung für den Deutschunterricht" der letzten zentralen Tagung des Symposions Deutschdidaktik 2018. Diese widmete sich zum ersten Mal konkret den professionellen Kompetenzen der Lehrkräfte sowie den Lernprozessen im Deutschunterricht.

Fazit für FALKE-D

FALKE, als Nachfolgerprojekt von FALKO, greift dessen Ergebnisse auf und fokussiert einen eigenen Schwerpunkt in der Auseinandersetzung mit professioneller Kompetenz von Lehrkräften. Während FALKO Wissenstests zum Fachwissen und zum fachdidaktischen Wissen (aufgeteilt in die drei oben genannten Teilfacetten) in den sechs beteiligten Fachdidaktiken sowie einen fachunabhängigen Test zum pädagogischen Wissen entwickelte und überprüfte, fokussiert FALKE in allen elf beteiligten Fachdidaktiken eine der Teilfacetten des fachdidaktischen Wissens, nämlich das Wissen über Erklären und Repräsentieren von Inhalten. Innerhalb der Genese von COACTIV über FALKO zu FALKE zeigt sich somit wie in der Einleitung angedeutet eine immer stärker ausgeweitete inter- und transdisziplinäre Zusammenarbeit sowie eine stärkere und detaillierte Fokussierung auf Teilaspekte professioneller Kompetenz.

Das Projekt FALKE-D leistet mit der Untersuchung der Lehrerkompetenz „Erklären" einen Baustein im weiten Feld der Professionsforschung von Deutschlehrkräften.

2.3 Erklärkompetenz von Lehrkräften

Die Grundannahme von FALKE-D ist, dass eine gute Lehrkraft qualitätvoll erklären können muss und gleichzeitig auch wissen muss, wann eine Erklärung sinnvoll und nötig ist, und wann vielmehr andere Methoden im Unterricht zielführend sind. Es wird selbstverständlich nicht postuliert, dass Erklärkompetenz die wichtigste Kompetenz von Lehrkräften ist, sondern vielmehr eine von vielen. Eine gute Lehrkraft, die qualitätvollen Unterricht erteilt, denkt nicht in Entweder–Oder–Kategorien. Qualitätvoller Unterricht kann nur eine bewusste und adäquate Mischung von Instruktion durch die

Lehrkraft und Konstruktion durch die Schülerinnen und Schüler sein, wie in der Einleitung bereits dargelegt:

> In summary, the constructivist teaching and direct instruction approaches provide a pedagogical framework for content instruction. The review of this literature shows that effective teachers use a combination of these approaches, tailoring their teaching to their perceptions of individual student learning styles. In addition, effective teachers use language that is appropriate for and effective in conveying information and ideas to students (Bailey et al., 2004, S. 11).

Dabei ist Erklären zu können eine Fähigkeit von Lehrkräften, die in den unterschiedlichsten Fächern für unterschiedliche Zwecke eine Rolle spielt: „[…] providing explanations is a fundamental teaching practice that cuts across different disciplines: regardless of subject matter, teachers typically provide explanations at several points, including introducing new content, responding to students' queries, supporting struggling students, and helping students understand their errors and misconceptions […]" (Charalambous, Hill & Ball, 2011, S. 442).

Die Erklärkompetenz ist nach Shulmans Taxonomie (1986, 1987) dem komplexen fachdidaktischen Wissen zuzuordnen, das sich aus verschiedenen Teilkomponenten zusammensetzt. Explizit erwähnt er das Wissen um Schüler(fehl)vorstellungen (vgl. Shulman, 1986, S. 9–10) sowie das Wissen um unterschiedliche Darstellungsmöglichkeiten und Erklärungsweisen für wichtige Fachinhalte:

> Within the category of pedagogical content knowledge I include, for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations - in a word, the ways of representing and formulating the subject that make it comprehensible to others. Since there are no single most powerful forms of representation, the teacher must have at hand a veritable armamentarium of alternative forms of representation, some of which derive from research whereas others originate in the wisdom of practice (Shulman, 1986, S. 9).

FALKE-D schließt sich dieser Taxonomie Shulmans und der entsprechenden Modellierung in COACTIV (Kunter et al., 2011) sowie in FALKO (Krauss, Lindl, Schilcher, Fricke et al., 2017) an und ordnet Erklären als Teil der professionellen Handlungskompetenz dem fachdidaktischen Wissen von Lehrpersonen zu. Die Modellierung dieses fachdidaktischen Wissens in FALKE-D wird aus der Vorgängerstudie FALKO (vgl. Abb. 5) übernommen: Dort wird das fachdidaktische Wissen in die oben genannten drei Facetten *Wissen über typische Schülerschwierigkeiten und -fehler, Wissen über das Potenzial von Lernmaterialien* sowie eben *Wissen über Erklären und Repräsentieren von Inhalten* unterteilt. Letztere wird als Erklärkompetenz von Lehrkräften in FALKE fokussiert und untersucht.

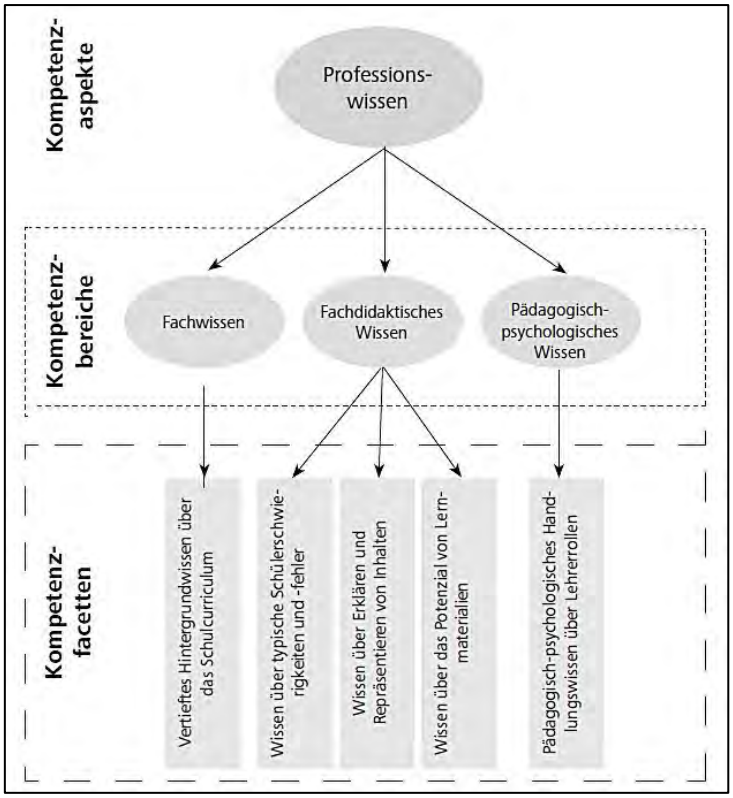


Abbildung 5: Professionswissen
leicht abgeändert aus Krauss, Lindl, Schilcher & Tepner, 2017, S. 19

Nichtsdestotrotz reicht fachdidaktisches Wissen nicht aus, um eine gute Erklärung geben zu können beziehungsweise, um über hohe Erklärkompetenz zu verfügen. Selbstverständlich ist auch fachliches Wissen dazu nötig, das die Grundlage für eine Erklärung darstellt. Es ist trivial zu erkennen, dass man nichts erklären kann, was man nicht weiß, selbst bei hohem fachdidaktischem Wissen. Fachliche Korrektheit als Kriterium für eine gute Lehrerklärung aufzunehmen, ist für diese Studie deswegen nicht reizvoll. Es ist selbstverständlich, dass eine fachlich korrekte Erklärung gelungener ist als eine fachlich falsche (vgl. Schilcher, Krauss, Rincke & Hilbert, 2017 sowie Findeisen, 2017).

Eine Lehrkraft mit hoher Erklärkompetenz verfügt aber nicht nur über vertieftes Fachwissen sowie über hohes fachdidaktisches Wissen. Auch eine dritte Facette der Taxonomie nach Shulman (1986, 1987), nämlich das pädagogisch-psychologische Wissen, muss bei einer/einem Erklärenden ausgeprägt sein. Selbst eine nach besten fachdidaktischen Prinzipien gestaltete und fachlich korrekt gehaltene Erklärung führt nicht zwangsläufig zu Verstehen bei den Adressatinnen und Adressaten. Vielmehr muss deren Vorwissen, deren kognitive Leistungsfähigkeit, deren Interessiertheit sowie Motivation bedacht werden (vgl. z. B. Sevian & Gonsalves, 2008, S. 1444 sowie Schilcher et al., 2017, S. 443) und die Gestaltung der Erklärung auch daraufhin ausgerichtet sein. Somit sind alle drei Facetten des Professionswissens notwendig, um eine gute Erklärung halten zu können (vgl.

Abb. 5 zum Professionswissen). Dies spiegelt sich auch in der in Punkt *4.1 Das Konstrukt Gutes Erklären* dieser Arbeit erläuterten Arbeitsdefinition von Erklären im FALKE-Gesamtprojekt wider.

Das Wissen über den Erklärgegenstand sowie das Wissen über gute Erklärungen muss schließlich im Unterricht umgesetzt werden können. Nur Lehrkräfte, denen dies gelingt, besitzen hohe Erklärkompetenz.[10] Wie bei jeder Kompetenz geht auch bei der Erklärkompetenz der Weg von Dispositionen über Fähigkeiten zur tatsächlichen Performanz (vgl. Blömeke et al., 2015). Ein hohes ausgeprägtes Wissen in allen dargelegten Facetten garantiert dennoch nicht eine tatsächlich und konkret gehaltene und gelungene Erklärung. Sowohl der Erwerb von Erklärkompetenz als auch wirksame Lehrmethoden, um Studierende zu guten Erklärenden auszubilden, sind weiterhin ein Desiderat der Forschung (vgl. Charalambous et al., 2011, S. 442 bzw. Erläuterungen zur *Erklärkompetenz in der Lehrerbildung* unter Punkt *11.2 Ausblick auf FALKE II* in dieser Arbeit). Die Erklärfähigkeit einer Lehrperson zeigt sich in der Performanz und kann nur im Feld, das heißt im konkreten Unterricht, betrachtet und untersucht werden.

Es liegen in der Deutschdidaktik wenige Studien zur Erklärfähigkeit von Lehrkräften oder Schülerinnen und Schülern vor. Spreckels (2009) vergleicht die Erklärungen von zwei Praktikantinnen mit denen von erfahrenen Lehrkräften. Wenig überraschend stellt sie fest, dass erfahrene Lehrkräfte besser erklären, was sie mit höherer Sachkenntnis, Prozessorientierung statt starrer Lernzielorientierung in Bezug auf den Ablauf des Unterrichts sowie mit besserer Adressatenadaptivität auf Seiten der erfahrenen Lehrkräfte erklärt.

Ernst-Weber (2019) nimmt in ihren vergleichenden Gesprächsanalysen die mündlichen Erklärfähigkeiten von Schülerinnen und Schülern der Sekundarstufe I in den Blick. Ihre Analysen ergeben, dass Schüler/-innen ganz unterschiedliche Fähigkeiten im Bereich des Erklärens aufweisen. Aufgrund der Bedeutung von Erklärfähigkeit im Alltag als „omnipräsente Gesprächspraktik" (S. 318) fordert Ernst-Weber das Erklären analog zum Argumentieren oder Erzählen in die Curricula aufzunehmen (vgl. S. 319).

Auch Berkemeier (2009) betont die Wichtigkeit der Sprechhandlung Erklären im Alltag und daher die Notwendigkeit, diese Fähigkeit auch bei Schülerinnen und Schülern zu fördern. Dabei nennt sie eine ausreichend vertiefte Sachkenntnis sowie die Fähigkeit zur Adressatenadaptivität als Voraussetzungen für eine gelungene Erklärung.

Morek, Heller und Quasthoff (2017) betonen in ihrer Arbeit die Kontextabhängigkeit von Erklärungen, die es notwendig mache, dass Schülerinnen und Schüler die im Unterricht spezifischen Anforderungen erlernen, um dort zu gelungenen Erklärungen befähigt zu werden. Schüler/-innen

[10] Die Begriffe Erklärkompetenz und Erklärfähigkeit werden in dieser Arbeit synonym verwendet.

müssten dazu zunächst erkennen, dass eine Erklärung verlangt ist, und diese dann sprachlich an das im Unterricht und vom Gegenstand geforderte Niveau anpassen.

Fazit für FALKE-D

Im Gegensatz zu den genannten Arbeiten aus der Deutschdidaktik untersucht FALKE-D die wahrgenommene Qualität von Lehrererklärungen empirisch über eine Fragebogenstudie. Somit ist der Untersuchungsgegenstand beobachtbares Unterrichtsverhalten, genauer gesagt das Bereitstellen einer Erklärung als Lerngelegenheit (Säule 3 im Kaskadenmodell, s. Punkt *2.2 Professionelle (Deutsch-) Lehrkräfte* in dieser Arbeit). Über diese Bewertungen werden Rückschlüsse über gutes Erklären (nicht über die erklärende Person) gezogen, die in weiteren Studien überprüft und zur Aus- und Weiterbildung von Erklärkompetenz herangezogen werden sollen (s. Punkt *11.1 Anregungen aus FALKE-D für weitere Forschungsarbeiten* und Punkt *11.2 Ausblick auf FLAKE II*).

2.4 (Professionelle) Wahrnehmung von Unterrichtsqualität

Nach der Darlegung der Anforderungen an qualitätvollen Unterricht und professionelle Lehrkräfte stellt sich die Frage nach der (psychometrischen) Messung von Unterrichtsqualität; das heißt konkret für FALKE-D, wie und welche Aussagen über Erklärqualität erhoben und analysiert werden. Erklärqualität kann grundsätzlich nur als Wahrnehmung dieser Qualität erhoben werden (vgl. Clausen, 2002), da die Bewertung von Unterrichtsqualität (und somit auch von Erklärqualität) zwangläufig einer Beobachterin/ eines Beobachters bedarf. In Studien im Unterrichtskontext haben sich dazu drei unterschiedliche Perspektiven etabliert: Schüler/-innen, Lehrkräfte sowie externe Beobachter/-innen, die entweder geschult werden oder die sich aufgrund ihrer Profession theoretisch mit Unterrichtsqualität beschäftigen und so zum Beispiel für die didaktische Perspektive als Expertinnen und Experten gelten. Im Folgenden werden Spezifika in der Wahrnehmung dieser Gruppen dargelegt.

2.4.1 Beobachterperspektive: Wahrnehmung der Schüler/-innen

Seit den 2000er Jahren wurden unterschiedliche Fragebogenskalen für die Erfassung der Wahrnehmung von Unterrichtsqualität auf Seiten der Schüler/-innen erstellt, wobei in den entsprechenden Studien normalerweise die selbst am Unterricht beteiligten Schüler/-innen einbezogen wurden, denn „Schüler(innen) [können] aufgrund ihrer vielfältigen schulischen Erfahrungen als Experten für Unterricht angesehen werden“ (Begrich et al., 2017, S. 25).

So erheben bereits die oben erläuterte COACTIV-Studie (vgl. Kunter et al., 2011) sowie die DESI-Studie (vgl. Klieme et al., 2008) die wahrgenommene Unterrichtsqualität in den Fächern Mathematik (COACTIV) sowie Deutsch und Englisch (DESI) aus der Schülerperspektive.

In diesem Zusammenhang wird auch untersucht, inwiefern die Aussagen aus dieser Perspektive verlässlich und gültig sind. Gruehn (2000) entwickelte zum Beispiel einen Fragebogen, mit dem die Schülerwahrnehmung des eigenen Unterrichts sowie der eigenen Lehrkraft fachspezifisch (nämlich in den Fächern Mathematik, Biologie und Physik) erhoben wird. Dessen Validität im Hinblick auf die Lernzuwächse der Schüler/-innen wurde anschließend empirisch überprüft, so dass Zusammenhänge zwischen der wahrgenommenen Unterrichtsqualität der Schüler/-innen und ihrer Leistungszuwächse hergestellt werden können (vgl. Gruehn, 2000, S. 205). Obwohl die Ergebnisse der Schülerurteile zum Teil zeigen, dass Schüler/-innen die abgefragten Faktoren von Unterrichtsqualität anders wahrnehmen als theoretisch angenommen (vgl. Gruehn, 2000, S. 141–145), verdeutlichen sie insgesamt, dass „Schüler durchaus als kompetenter [sic!] Beurteiler von Unterricht gelten können und ihrer Bewertung differenzierte Wahrnehmungen der unterschiedlichen Komponenten des Unterrichts und Lehrerverhaltens zugrunde liegen“ (Gruehn, 2000, S. 204).

Auch die Studie von W. Wagner, Göllner, Helmke, Trautwein und Lüdtke (2013) kommt zu dem Ergebnis, dass Schüler/-innen der Sekundarstufe Unterrichtsqualität wahrnehmen und bewerten können: Sie überprüfen dazu die Wahrnehmung von fünf Dimensionen von Unterrichtsqualität (nämlich *Struktur, Klassenführung, Verstehbarkeit, Motivation* sowie *Schüleraktivität*) in den Fächern Englisch und Deutsch. Die Wahmehmung der Unterrichtsqualität der Schüler/-innen erweist sich dabei als multidimensional (vgl. W. Wagner et al., 2013, S. 6). Außerdem zeigt sich, dass die Erhebung der Schülerwahrnehmung zu nur einem Messzeitpunkt ausreichend ist (vgl. W. Wagner et al., 2016).

Zu ähnlichen Ergebnissen kommen Scherer, Nilsen und Jansen (2016), die die Schülerdaten zum Mathematikunterricht aus USA, Kanada und Australien aus der PISA-Studie 2012 analysieren: „Taken together, we identified positive relations between students’ perceptions of instructional quality, their achievement, self-concept, and motivation in mathematics across countries. These findings were, by and large, robust against the different modeling approaches“ (Scherer et al., 2016, S. 12). Für Mathematik lassen sich also Schülerwahrnehmungen nutzen, um Aussagen über Lernerfolg, Selbstkonzept und Motivation der Schüler/-innen zu treffen.

Gindele und Voss (2017) befragten in ihrer Studie die Schüler/-innen zu ihren Einschätzungen bezüglich der Erklärfähigkeit und der emotionalen Erschöpfung ihrer Lehrkraft. Sie können signifikante Zusammenhänge mit dem pädagogisch-psychologischen Wissen der Lehrkräfte nachweisen: „Lehrkräfte mit höherem Wissen wurden von den Schüler/-inne/n als kompetenter beim Erklären und als weniger erschöpft wahrgenommen“ (Gindele & Voss, 2017, S. 255). Sie weisen in diesem Zusammenhang im Anschluss an Kunter und Baumert (2006) auch auf die Notwendigkeit hin, Schüler/-innen zur Erklärkompetenz ihrer Lehrkräfte zu befragen: „Bezüglich der

Erklärfertigkeiten der Lehrkräfte sollte beispielsweise insbesondere die Schülersicht entscheidend sein, da von außen vermutlich schwer zu erfassen ist, wie verständlich die Erklärungen der Lehrkraft für die unterrichteten Schüler/innen tatsächlich sind" (Gindele & Voss, 2017, S. 268). Es war daher naheliegend in FALKE-D auch Schüler/-innen zu ihrer Wahrnehmung von Erklärqualität zu befragen (s. Punkt *7.4.2 Adressaten: Schüler/-innen* in dieser Arbeit).

Schließlich zeigt auch die breit angelegte Studie von Wisniewski, Zierer, Dresel und Daumiller (2020, vorab online), in der über 15000 Schüler/-innen (der Jahrgangsstufen 5 bis 12) aus knapp 700 Klassen an Realschulen, Gymnasien sowie Berufsschulen in ganz Deutschland über ein Online-Feedback-Portal zu eigenem Unterricht befragt wurden, dass Schüler/-innen in der Lage sind, Unterricht reliabel und valide zu beurteilen. Wisniewski und andere (2020) teilen dabei Unterrichtsqualität in die bekannten drei Basisdimensionen *Classroom management, Student support* und *Cognitive activation* ein, verwenden jedoch den Fragebogen *teaCH*, der sieben Unterkategorien annimmt, die sich den gerade genannten drei Basisdimensionen zuordnen lassen (vgl. Wisniewski et al., 2020, vorab online). Interessant an den Ergebnissen ist besonders, dass sie sich bereits mit einer kleinen Stichprobe von der Größe einer Schulklasse als reliabel erweisen sowie dass die Ergebnisse der Schüler/-innen mit denen der Lehrkräfte gut übereinstimmen (vgl. ebd.). Bemerkenswert ist weiterhin, dass sich das angenommene Modell als stabil und verallgemeinerbar über Klassenstufen, Fächer sowie Schularten hinweg erweist: „We demonstrated the generalizability of the 7-factor structure not only across different subjects, but also different school types and grade levels, indicating a high robustness against contextual distortions […]" (Wisniewski et al., 2020, vorab online). Dieses Ergebnis widerspricht früherer Forschung, die zum Teil eine hohe Kontextabhängigkeit der Bewertungen von Schülerinnen und Schülern annimmt (z. B. W. Wagner et al., 2013).

In vorliegenden Studien werden wie angedeutet Schüler/-innen (normalerweise am Ende eines Schuljahres) ausnahmslos zu eigenem Unterricht und zu Lehrkräften befragt, die sie selbst im Unterricht haben. Dieses Vorgehen hat zum einen praktische Gründe, da diese Befragungen leicht zu organisieren sind. Außerdem bringt es mit sich, dass die Schüler/-innen bei der Bewertung auf eine lange Erfahrung zurückgreifen können und nicht auf Momentaufnahmen reagieren, wie es externe Beobachter/-innen tun. Man geht gerade deshalb davon aus, dass Schülerurteile zuverlässig und genau sind. In FALKE-D dagegen werden „fremde" Erklärungen zur Bewertung eingesetzt (s. zu den Gründen dieser Entscheidung Punkt *7.4.2 Adressaten: Schüler/-innen* in dieser Arbeit).

Fazit für FALKE-D

Zusammenfassend zeigen die Ergebnisse der berichteten Studien, dass es lohnt, Schüler/-innen zur Unterrichtsqualität zu befragen. Diese wurden daher als in das Gesamtprojekt FALKE und somit

auch in FALKE-D aufgenommen. Da sie die eigentlichen Adressatinnen und Adressaten der Erklärungen sind, ist ihre Perspektive besonders interessant.
Weitere Forschung zur Befragung von Schülerinnen und Schülern zur Unterrichtsqualität ist aber, gerade im Hinblick auf die Dimensionen, zu denen sie befragt werden, sowie auf den Vergleich der unterschiedlichen Perspektiven auf Unterricht, notwendig.
Dies postuliert Clausen bereits 2002. Nach ihm müssen unterschiedliche Unterrichtsmerkmale aus verschiedenen Perspektiven betrachtet werden, weil für die einzelnen Gruppen jeweils Unterschiedliches bedeutsam ist. Über diese verschiedenen Perspektiven ist es möglich, ein umfassendes Bild von Unterrichtsqualität zu erhalten.
Eingesetzt werden daher zur Beurteilung von Unterrichtsqualität neben den Schülerinnen und Schülern auch Lehrkräfte sowie externe Beobachtende als Expertinnen und Experten. Nach einem kurzen Exkurs zur Beobachterperspektive der Studierenden (als vierte Befragtengruppe bei FALKE-D), wird im Folgenden die Wahrnehmung von Expertinnen und Experten erläutert.

2.4.2 Beobachterperspektive: Wahrnehmung der Expertinnen und Experten

Studierende

Studierende können aufgrund ihres Ausbildungstandes und ihrer mangelnden Berufserfahrung (vgl. Bromme, 1992, S. 46) noch nicht als Expertinnen/Experten gelten. Sie befinden sich gerade im Erwerb eines Expertenstatus. Bei Lehramtsstudierenden ist die Erwerbsperspektive besonders interessant, da die Studierenden normalerweise vor nicht allzu langer Zeit noch Schülerinnen und Schüler waren und Unterricht aus der Adressatenperspektive wahrgenommen haben. Ihre Beobachterperspektive siedelt sich deswegen „zwischen" denen der Schüler/-innen und denen der Expertinnen/Experten (d. h. Lehrkräfte und externe Beobachter/-innen) an. In vorliegender Studie werden sie dennoch zunächst zu den Experten-Statusgruppen gezählt. Zur genauen Begründung dieser Entscheidung sowie der Befragung von Studierenden und zu den Ergebnissen dieser Gruppe siehe in dieser Arbeit die Punkte *7.4.1.3 Studierende* und *8.2.1.2, 8.2.2.2, 8.3.2.2* sowie *9.4.2* zu den Ergebnissen der Befragung der Statusgruppe Studierende.

Expertinnen und Experten

An dieser Stelle geht es um die Besonderheiten, die die Wahrnehmung von Expertinnen und Experten gegenüber Laien aufweist, denn beim Vergleich von Expertinnen und Experten mit Laien unterscheidet sich nicht „nur" das Denken und das Handeln, sondern bereits deren Wahrnehmung (vgl. Kuntze, 2016, S. 528).
In diesem Zusammenhang sind zwei Konzepte genauer zu erläutern, die postulieren, dass Expertenwahrnehmung maßgeblich von der Wahrnehmung von Laien differiert und dass erstere

ein Charakteristikum von Professionalität darstellt. Es handelt sich um die professionelle Unterrichtswahrnehmung (*professional vision*)[11] und um das Konzept *noticing.* Dabei sind aus der jüngeren Forschung vor allem die Studien zur professionellen Unterrichtswahrnehmung von Seidel und Kollegen (vgl. Seidel et al., 2010; Seidel & Stürmer, 2014 bzw. die Übersicht bei Seidel & Thiel, 2017) sowie die Studien von Sherin und anderen zum Konzept *noticing* (vgl. Sherin, Jacobs & Philipp, 2011b) zu erwähnen. Die Zielsetzung dieser Studien ist die Einteilung der Studienteilnehmenden in unterschiedliche Kompetenzstufen bezüglich ihrer Profession. Sie lassen sich also oben erläuterter Konzeptualisierung der Lehrkräfte als *expert teachers* (im Vergleich zu Novizen oder zu weniger fähigen Lehrkräften) zuordnen.

Professionelle Unterrichtswahrnehmung

Seidel und andere (2010) gliedern die professionelle Wahrnehmung in Anlehnung an Berliner (1987), Sherin und van Es (2009) und van Es (2009) in die Aspekte *Beschreiben, Erklären* und *Vorhersagen* und erfassen diese empirisch als Gesamtfähigkeit (vgl. Seidel et al., 2010, S. 297–300). Konkret untersuchen sie, wie (angehende) Lehrkräfte komplexe Unterrichtssituationen vor dem Hintergrund ihres theoretischen Wissens um Lehren und Lernen beobachten und interpretieren. Diese professionelle Wahrnehmung ist für Seidel und andere (2010) die Voraussetzung für die Reflexion des eigenen Unterrichts. Dabei zeigen sich qualitative Unterschiede in der Wahrnehmung zwischen Expertinnen/Experten und Novizinnen/Novizen und zwar auf allen drei Ebenen: Expertinnen und Experten *beschreiben* eine Lehr-/Lernsituation differenzierter, *erklären* diese Situation anhand ihres Wissens präzise, wobei sie wissenschaftliche Theorien und Befunde aus ihnen bekannten Studien heranziehen; und schließlich *sagen* Expertinnen und Experten treffsicherer und genauer Auswirkungen des Gesehenen auf den weiteren Lernprozess der Schüler/-innen *vorher*. Über die Auswertung der Ergebnisse aus Studien zum Beispiel mittels des Tools Observers (vgl. Seidel et al., 2010; Seidel & Stürmer, 2014) werden die Studienteilnehmer/-innen in gute und weniger gute Expertinnen und Experten eingeteilt, – je nachdem was sie *beschreiben, erklären* und *vorhersagen* und was nicht. In einer Nachfolgestudie erwies sich Observer überdies auch für Lehramtsstudierende (an unterschiedlichen Universitätsstandorten, mit unterschiedlichen Schularten und unterschiedlichen Fächern) als geeignetes Instrument, um die professionelle Wahrnehmung (differenziert in die Aspekte *Beschreiben, Erklären, Vorhersagen*) zu untersuchen und entsprechend der Ergebnisse gezielte Förderung anzustoßen (vgl. Jahn, Seidel, Stürmer & Prenzel, 2014).

[11] Die Begriffe professionelle Unterrichtswahrnehmung, professionelle Wahrnehmung und *professional vision* werden in dieser Studie synonym verwendet.

Noticing

Ähnlich verhält es sich mit dem Konzept des *(teacher) noticing* nach Sherin, Jacobs und Philipp (2011a), das sie explizit als eigenes Konstrukt auffassen und vom *Pedagogical Content Knowledge* nach Shulman (1987) unterscheiden. Leitfragen für ihre Forschung sind: „Where do teachers look, what do they see, and what sense do they make of what they see?" (Sherin et al., 2011b, S. 3). *Noticing* wird demnach zum einen als die Fähigkeit aufgefasst, Bedeutsames für die Lehr-/Lerninteraktion zu bemerken, aber genauso „Nicht-Bedeutsames" nicht zu bemerken oder nicht darauf zu reagieren (vgl. Sherin et al., 2011b, S. 5). Da Unterricht eine äußerst komplexe Situation mit einer Fülle an Informationen darstellt, zeichnet eben diese Fähigkeit des „Filterns" Expertinnen und Experten (erneut im Sinne von *expert teacher*) aus.

Zum anderen gehört zum *teacher noticing* auch die Fähigkeit, das Wahrgenommene zu interpretieren (im Original: „make sense" (Sherin et al., 2011b, S. 5)), denn natürlich nehmen professionelle Lehrkräfte das im Unterrichtsgeschehen Bedeutsame nicht einfach „nur" wahr, sondern interpretieren das Wahrgenommene auch vor dem Hintergrund ihres professionellen Wissens, um schließlich angemessen zu reagieren. Dabei wird der letztgenannte Aspekt des *making sense* in der Forschung unterschiedlich aufgefasst: In einigen Studien umfasst *making sense* nur das Interpretieren des zuvor Wahrgenommenen, in anderen dagegen gehört auch das Begründen der weiteren Entscheidungen zum *making sense* (vgl. Sherin et al., 2011b, S. 8). Letztere Auffassung kommt wiederum der oben bereits erläuterten Konzeptualisierung von Kompetenz als Kontinuum von Blömeke und anderen (2015) nahe, die als spezifische Fähigkeiten zwischen Dispositionen und Performanz *Wahrnehmen, Interpretieren* und *Entscheidungen treffen* annehmen, die als Kompetenzfacetten (im Original *Perception, Interpretation* und *Decision making*) in das PID-Modell zu professioneller Wahrnehmung eingegangen sind.

Wichtig ist hervorzuheben, dass *noticing* nicht immer bewusst stattfindet, sondern gerade oft unbewusst (vgl. Miller, 2011, S. 52; Erickson, 2011, S. 21–23), wodurch kognitive Kapazitäten bei der Lehrperson frei sind und für bewusst ablaufende Prozesse des Unterrichtens genutzt werden können.

Auch Studien zum *noticing* zielen darauf ab, die teilnehmenden Personen in Expertinnen und Experten unterschiedlichen Grades einzuteilen. Überprüft wird *noticing* dabei nicht nur über Videos, vielmehr gibt es auch Studien, die mit Unterrichtsskripts und schriftlichen Tests arbeiten, sowie Arbeiten im Feld (vgl. die unterschiedlichen Studien in Sherin et al., 2011a).

Kaskadenmodell

Zu erwähnen ist an dieser Stelle außerdem noch einmal das Kaskadenmodell (Krauss et al., 2020), das sowohl das PID-Modell als auch *noticing* als situationsspezifische Fähigkeiten zwischen die Dispositionen, die Lehrkräfte besitzen müssen, und das konkrete Verhalten, das Lehrkräfte zeigen müssen, integriert (s. Säule 2, Abb. 3). Die Verallgemeinerbarkeit der Schlussfolgerungen aus den Studien ist dennoch weiterhin zu überprüfen, da sogar das Wahrnehmen, das *noticing,* laut Erickson (2011) noch nicht hinreichend erforscht ist: „[…] we also need to learn more about the relativity of teacher noticing, its varieties in differing circumstances of pedagogical use and belief" (Erickson, 2011, S. 33).

Trotz dieser Einschränkungen und der damit gebotenen Vorsicht kann man davon ausgehen, dass es professionelles Wahrnehmen bei Lehrkräften gibt. Miller (2011) zitiert zum Beispiel eine Studie, in der nachgewiesen wird, dass professionelle Lehrkräfte in hoch komplexen Situationen signifikant mehr wahrnehmen als Novizinnen und Novizen. Er führt in diesem Zusammenhang das Konzept der *situation awareness* (2011, S. 51–54) ein, die Situationsbewusstheit also, die trainiert und erlernt werden muss und Expertinnen/Experten eben von Novizinnen/Novizen unterscheidet. Auch die Studien von Seidel und anderen zur professionellen Unterrichtwahrnehmung (siehe oben bzw. vgl. Seidel et al., 2010; Seidel & Stürmer, 2014) weisen in die gleiche Richtung.

Fazit für FALKE-D

Es ist noch einmal zu betonen, dass FALKE-D im Gegensatz zu obigen Studien nicht die Kompetenz der Studienteilnehmer/-innen testet. Vielmehr werden in FALKE-D Expertinnen und Experten mit unterschiedlichem Expertisegrad befragt, um deren professionelle Wahrnehmung zu nutzen, um das fachspezifische Erklären im Deutschunterricht besser zu verstehen. Dabei werden neben den Lehrkräften auch Didaktiker/-innen befragt, bei denen davon ausgegangen wird, dass sie mit ihrer Metaperspektive auf Unterricht eine eigene Expertise für das Unterrichten aufweisen, so dass ihre Wahrnehmung ebenso als professionell im Sinne der gerade erläuterten Modelle angesehen wird (s. hierzu ausführlich Punkt *7.4.1 Expertinnen und Experten* in dieser Arbeit).

2.4.3 Methoden und Zweck der Erfassung von Wahrnehmung

Um die Unterrichtsqualität zu erheben, gilt es neben der Beobachterperspektive auch die Methode sowie den Zweck der Erhebung zu reflektieren.

Erfassen von Unterrichtsqualität im schulischen Rahmen

Zum einen wird Unterrichtsqualität zur Qualitätssicherung auf Schulebene erhoben. Diese Erhebungen können oben genannten politischen Zwängen unterliegen. Sie werden meist von externen Evaluierenden durchgeführt.

Zum anderen kann die Bestimmung der Unterrichtsqualität der einzelnen Lehrkraft dazu dienen, den eigenen Unterricht zu verbessern. Dabei können Videoaufzeichnungen des eigenen Unterrichts zur Selbstreflexion eingesetzt werden. Sie haben für Lehrkräfte den Vorteil, dass sie das eigene (flüchtige) Verhalten im Unterricht aufzeichnen und ihnen einen unverfälschten Blick auf das eigene Tun ermöglichen. Zur Verbesserung des eigenen Lehrens sind Videoaufnahmen eine große Chance. Besonders effektiv ist es, wenn der eigene Unterricht zusammen mit anderen (entweder mit Kolleginnen und Kollegen oder mit externen Beurteilerinnen und Beurteilern) reflektiert wird. Dazu wird sowohl Videographie als auch Live-Beobachtung des Unterrichts eingesetzt. Feedback von außen und eigene Reflexionsfähigkeit sind nötig, da das Verändern eigener Handlungspraxis nicht automatisch ausgelöst werden kann, sondern aktiver Bewusstmachung bedarf (vgl. Quittenbaum, 2016, S. 47–48).[12]

Erfassen von Unterrichtsqualität im universitären Rahmen

Neben Untersuchungen im Kontext Schule widmet sich ein ganzes Forschungsfeld an den Universitäten und Bildungsinstituten den Bestimmungs- und Wirkfaktoren von qualitätvollem Unterricht. Ein Ziel ist es dabei (wie erläutert), die Rolle der Lehrkraft zu bestimmen und herauszufinden, was erfolgreiche Lehrkräfte anders machen als weniger erfolgreiche. Die Ergebnisse dieser Forschung werden in die Praxis übertragen, zum Beispiel indem teilnehmenden Lehrkräften Feedback gegeben wird, indem Fortbildungen abgehalten werden und indem die Forschungsergebnisse in die Ausbildung der ersten Lehrerbildungsphase einfließen. In diesem Zusammenhang werden im Allgemeinen am Unterricht beteiligte Schüler/-innen und Lehrkräfte nach dem Unterricht über Interviews oder Fragebögen zu ihren Einschätzungen und Wahrnehmungen befragt. Für externe Beobachter/-innen etabliert sich auch hier neben der Live-Beobachtung die Arbeit mit videographierten Unterrichtsausschnitten, die zum Beispiel mit Kodiermanual analysiert werden und den Vorteil haben, wiederholt angesehen werden zu können. Abgesehen vom Unterrichtsgeschehen selbst werden aus genannten Gründen immer mehr auch

[12] Der Vollständigkeit halber sei erwähnt, dass Lehrkräfte zur Verbesserung des eigenen Unterrichts verstärkt Schülerfeedback zum Beispiel in Form von Apps wie FeedbackSchule nutzen (vgl. Wisniewski, Zierer und Hattie, 2017). Dieser Aspekt spielt in vorliegender Studie jedoch keine Rolle.

das im Unterricht verwendete Material (also z. B. Schulbücher, Arbeitsblätter, Hefteinträge usw.) sowie Leistungsnachweise als zusätzliche Informationsquelle herangezogen (dies war wie gesagt z. B. bei der TIMS-Studie (2001) und der COACTIV-Studie (2011) der Fall). Dies erlaubt ein umfassenderes Bild und Einblick in die Tiefenstrukturen (z. B. in das Potenzial zur kognitiven Aktivierung von Aufgaben) des Unterrichts.

Arbeiten die Studien nicht nur mit externen Beurteilenden, sondern beziehen sie Schüler/-innen und Lehrkräfte in die Bewertung von Unterrichtsqualität mit ein, so stellt sich trotz des Vorteils, eigenen Unterricht zu bewerten, die Frage, ob nicht fremder Unterricht besser geeignet ist, um valide Ergebnisse über die Unterrichtsqualität zu erhalten, denn die Bewertung von eigenem Unterricht hat Fehlerquellen: Bei Schülerinnen und Schülern besteht die Gefahr, dass die Sympathie beziehungsweise die Ablehnung gegenüber der zu bewertenden Lehrkraft oder dem Fach (d. h. das Interesse und die Motivation am Fach genauso wie die eigene Leistung im Fach) die Beurteilung beeinflussen (vgl. z. B. Gruehn, 2000, S. 212 bzw. Punkt *7.4.2 Adressaten: Schüler/-innen* in dieser Arbeit).

Bei den Lehrkräften hingegen kann es zu „selbstwertdienlichen Verzerrungen" (Fauth et al., 2014, S. 128) kommen, wodurch sie den eigenen Unterricht positiver bewerten, als er eigentlich ist. Ebenso besteht bei der Bewertung eigenen Unterrichts durch Lehrkräfte die Gefahr, sozial erwünscht zu antworten, was die Ergebnisse ebenfalls verzerrt (vgl. Begrich et al., 2017, S. 26).

Bei externen Evaluierenden geht man dagegen davon aus, dass derartige Verzerrungen viel weniger stattfinden (vgl. Maag Merki & Werner, 2014, S. 755–757), sie werden deswegen gerne herangezogen. Abgesehen davon, dass Beobachter/-innen von außen den gesehenen Unterricht emotional unbeteiligt beurteilen können, sind sie durch Schulung und/oder Expertise Expertinnen und Experten für Unterrichtsbeurteilung und daher in der Lage, oben beschriebene Tiefenstrukturen differenziert wahrzunehmen. Gerade bei der Arbeit mit Videographien gelangen geschulte Bewerter/-innen zu sehr genauen Urteilen, weil sie relevante Ausschnitte immer wieder ansehen können (vgl. Begrich et al., 2017, S. 26). Die Bewertung von Unterricht durch externe, geschulte Beobachter/-innen gilt daher gemeinhin als verlässlichste Möglichkeit, Unterrichtsqualität zu bewerten und daraus wissenschaftlich fundierte Erkenntnisse zu gewinnen. Dennoch sind auch hier Nachteile zu nennen: Zum einen muss die Reliabilität noch besser überprüft werden, denn es gibt Studien, die zeigen, dass selbst bei externen Bewerterinnen und Bewertern die Gütekriterien nur mittelmäßig erfüllt werden (s. hierzu z. B. die Studie von Praetorius, Lenske & Helmke, 2012, die bei externen Raterinnen und Ratern einen Bias von 12-40% feststellt). Zum anderen sind diese Verfahren oft wenig praktikabel, da sie sehr kosten- und zeitaufwändig sind.

Oftmals werden daher verschiedene Perspektiven auf denselben Unterricht erfasst, um ein umfassenderes Bild von Unterrichtsqualität zu erlangen. Damit einher geht natürlich die Frage nach

dem Grad der Übereinstimmung bei den unterschiedlichen Wahrnehmungen sowie die Frage nach der Validität der Aussagen.

2.4.4 Übereinstimmung der unterschiedlichen Perspektiven

Gerade wenn es um die Wahrnehmung der Qualität von authentischem Unterricht geht, ist es reizvoll, die Perspektiven der beteiligten Akteure zu vergleichen (vgl. COACTIV (2011) sowie DESI (2006)) und zu überprüfen, ob und inwieweit sie in ihren Urteilen übereinstimmen. Dies ist auch ein zentrales Erkenntnisinteresse der FALKE-D-Studie, bei der die Perspektiven der Fachdidaktiker/-innen, der Lehrkräfte, der Studierenden und der Schülerinnen und Schüler miteinander verglichen werden.

Betrachtet man bestehende Studien zur Übereinstimmung der verschiedenen Perspektiven genauer, ergibt sich ein gemischtes Bild, das es notwendig macht, die Ergebnisse detailliert zu betrachten. In manchen Befragungen kommen die verschiedenen Gruppen zu ähnlichen Urteilen über Unterrichtsqualität (vgl. Gruehn, 2000, S. 94–95). Größtenteils weisen sie aber wenig Übereinstimmung auf, wenn es darum geht, die Unterrichtsqualität insgesamt zu bewerten. So kann etwa Erickson (2011) in einer Langzeitstudie mit Grundschullehrkräften und Lehramtsstudierenden zeigen, dass diese beim Betrachten und Bewerten von Unterrichtsvideos ganz Unterschiedliches bemerken. Er leitet das auf die jeweils unterschiedlichen Erfahrungen der Teilnehmer/-innen zurück, die dazu führen, dass Lehrkräfte das Gesehene jeweils in einen anderen Rahmen (Erickson spricht von „story frame" (Erickson, 2011, S. 28)) einordnen. Diese „frames" und die sich daraus ergebenden Urteile hängen nach Erickson (vgl. 2011, S. 28–29) vor allem vom pädagogischen Engagement der Lehrkräfte ab und führen in der Studie zu teils massiv anderen Einschätzungen der gezeigten Unterrichtsvideos selbst innerhalb der Statusgruppe Lehrkräfte (vgl. Erickson, 2011, S. 19–22). Damit stellt sich die Frage nach der Gültigkeit der Urteile der verschiedenen Statusgruppen und nach dem (unterschiedlichen?) Verständnis des abgefragten Konstruktes beziehungsweise der abgefragten Konstrukte (vgl. Kunter & Baumert, 2006, S. 243; Clausen, 2002, S. 139).

Im Gegensatz zu den niedrigen bis moderaten Übereinstimmungen zwischen den drei (oder zwischen zwei der drei, unabhängig davon, welche Kombination man wählt) Perspektiven (d. h. Schüler/-innen, Lehrkräfte oder externe Beobachter/-innen) auf Unterrichtsqualität als Gesamtkonstrukt zeigen sich für einzelne untersuchte Facetten des Gesamtkonstruktes zum Teil höhere Übereinstimmungen: Clausen findet zum Beispiel für das Merkmal *Disziplin* eine hohe Übereinstimmung zwischen Schülerinnen und Schülern sowie externen Beobachterinnen und Beobachtern (vgl. 2002, S. 139). Kunter und Baumert weisen dagegen hohe Übereinstimmung zwischen Schüler- und Lehrerbeurteilungen beim Merkmal *Klassenführung* nach (vgl. 2006, S. 231

bzw. S. 240). Beim *unterstützenden Klima* stimmen wiederum zum Beispiel vor allem Schüler/-innen und Lehrkräfte in ihren Urteilen überein, externe Beobachter/-innen dagegen weniger (vgl. Clausen, 2002, S. 129; Fauth et al., 2014, S. 129).

Dahingegen kann Clausen gar keine Übereinstimmungen bei Merkmalen wie *anspruchsvolles Üben* (als Subfacette von *kognitiver Aktivierung*) oder *Monitoring* finden (vgl. Clausen, 2002, S. 128). Auch Kunter und Baumert stellen bei den Merkmalen *kognitive Aktivierung* und *Unterrichtstempo* niedrige beziehungsweise keine Übereinstimmung zwischen Schüler- und Lehrerurteilen fest (vgl. 2006, S. 231), so dass insbesondere zur Facette der *kognitiven Aktivierung* die Ergebnisse der Forschungen heterogen sind (vgl. Clausen, 2002, S. 161–164; Fauth et al., 2014, S. 129; Kunter & Baumert, 2006, S. 240).

Die höhere Übereinstimmung bei Merkmalen wie *Klassenführung* lässt sich damit erklären, dass dies ein für alle Gruppen relativ leicht zu beobachtendes Merkmal der Sichtstruktur ist und diese Beobachtbarkeit zu höheren Übereinstimmungswerten führt; dieses Ergebnis wird durch andere beobachtbare Merkmale gestützt, die ähnliche Werte erhalten (vgl. Clausen, 2002, S. 161; Kunter & Baumert, 2006, S. 241; Fauth et al., 2014, S. 129).

Eine Erklärung für die Übereinstimmung zwischen den Urteilen von Lehrkräften und Schülerinnen und Schülern, die den eigenen Unterricht bewerten, ist die gemeinsam verbrachte Zeit: Sie greifen auf Erfahrungen vieler gemeinsamer Unterrichtstunden zurück, die in ihre Bewertungen einfließen. Damit haben sie eine breitere Basis als externe Bobachter/-innen, die normalerweise nur kurze Ausschnitte aus dem Unterricht passiv (also ohne die Möglichkeit der Interaktion mit den beteiligten Schülerinnen/Schülern und Lehrkräften) bewerten.

Festhalten lässt sich dennoch, dass Unterrichtsqualität in den einzelnen Gruppen uneinheitlich wahrgenommen wird (vgl. Fauth et al., 2014, S. 134).

Auch im Hinblick auf die Validität der Aussagen der unterschiedlichen Gruppen ist daher ein genauer Blick notwendig. Es gilt heute gemeinhin als erwiesen, dass die drei Perspektiven differentielle Validität bezüglich unterschiedlicher Aspekte des Unterrichts besitzen: Geht es um motivationale und affektive Lerneffekte des Unterrichts so erweisen sich vor allem Schülerurteile als valide. Lehrkräfte geben besonders bei der Untersuchung von Lernzielen gültige Auskünfte und externe Beobachter/-innen erkennen und referieren in ihren Bewertungen umgesetzte didaktische Prinzipien (vgl. Clausen, 2002; Kunter & Baumert, 2006). Weitere Forschung ist hier nötig. Auch die Urteile der vier Perspektiven, die in FALKE-D erhoben werden, müssen in Folgestudien auf ihre Validität überprüft werden. Klar ist aber, dass durch den multiperspektivischen Ansatz bei FALKE-D eine inhaltliche Validierung stattfindet, die bei der Befragung nur einer Statusgruppe nicht gegeben wäre.

3 Theoretische Sichtweisen auf *Erklären*

Gegenstand der FALKE-D-Studie ist das Erklären aus deutschdidaktischer Sicht. Das Erklären und Erklärungen spielen allerdings in unterschiedlichen fachlichen Diskursen eine Rolle. Lange Tradition haben sie in der Erkenntnis- und Wissenschaftstheorie (s. dazu Punkt *3.1 (Natur-) Wissenschaftliche Sicht auf Erklärungen* in dieser Arbeit) sowie in der Sprachwissenschaft (s. Punkt *3.2 Erklären als Sprachhandlung* in dieser Arbeit). Die Lehr-/Lernforschung beschäftigt sich aus einer überfachlichen Perspektive mit dem Erklären und wendet sich ihm gerade in den letzten Jahren verstärkt zu (s. Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* in dieser Arbeit) und in der didaktischen Forschung (s. Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit) lassen sich je nach Fach ganz unterschiedliche Wissensstände zum Erklären ausmachen. Im Folgenden wird daher ein kurzer Überblick über diese Perspektiven gegeben.

3.1 (Natur-) Wissenschaftliche Sicht auf Erklärungen

Die (natur-) wissenschaftliche Sichtweise auf Erklärungen wird der Vollständigkeit halber kurz dargelegt. Ihren Ursprung haben diese Erklärungen im Modell von Hempel und Oppenheim (1948) zur Kausalerklärung. Es handelt sich dabei um eine deduktiv-nomologische Erklärung, die folgende formale Struktur aufweist: Der Erklärgegenstand (Explanandum) wird von allgemeingültigen Gesetzmäßigkeiten und Ausgangsbedingungen (Explanans) logisch abgeleitet (vgl. Hempel & Oppenheim, 1948, S. 136). Die Beschaffenheit der Gesetzmäßigkeiten und Bedingungen beschreiben Hempel und Oppenheim (1948) so: Gesetze müssen wahr und von universeller Natur sein, das heißt, sie müssen allgemein gültig sein und dürfen sich nicht nur auf eine bestimmte Situation oder eine bestimmte Anzahl von Fällen beziehen. Hempel und Oppenheim sprechen hier von „condition of non-limited scope" (Hempel & Oppenheim, 1948, S. 155). Zur Verdeutlichung dieser Bedingung stellen sie folgende gesetzartige Aussagen (im Original: „lawlike sentence" (Hempel & Oppenheim, 1948, S. 153)) einander gegenüber. Zunächst nennen sie den Satz: „Alle Äpfel im Korb *k* sind zum Zeitpunkt *t* rot". Diese Aussage ist zwar wahr, aber kein allgemeingültiges Gesetz, da sie auf eine bestimmte Situation zutrifft, aber nicht verallgemeinerbar ist. Die Aussage ist nur für eine bestimmte Anzahl an Äpfeln wahr, nämlich die, die sich zu dem konkreten Zeitpunkt im Korb befinden. Dagegen ist die Aussage „Die Eier von Rotkehlchen sind grün-blau" gesetzartig, weil sie sich eben nicht nur auf eine bestimmte Anzahl von Eiern beschränkt[13] (vgl. Hempel & Oppenheim, 1948, S. 153-154).

[13] Der Vollständigkeit halber sei erwähnt, dass sich natürlich auch diese Aussage letzten Endes auf eine begrenzte Anzahl an Eiern bezieht, nämlich alle Rotkehlcheneier der Vergangenheit, Gegenwart und Zukunft. Da dazu aber viel spezielleres, nämlich empirisches, Wissen notwendig ist, gilt sie dennoch, gerade im Vergleich mit der Aussage zu den Äpfeln, als gesetzähnlich (vgl. Hempel und Oppenheim, 1948, S. 153-154).

Damit eine Erklärung richtig ist, muss das Explanans gültig sein, das heißt, es müssen fachlich richtige und anerkannte Gesetze und Bedingungen angewandt werden. Die fachliche Korrektheit ist das ausschlaggebende Kriterium für die Qualität einer wissenschaftlichen Erklärung.

Zur Verdeutlichung wird als Beispiel die deduktiv-nomologische Erklärung folgenden Erklärgegenstandes von Hempel (1977) zitiert: Die Wasseroberfläche in einem Gefäß ändert sich nicht, wenn ein darin befindliches Stück Eis schmilzt. Das Explanans setzt sich aus den folgenden Ausgangsbedingungen und Gesetzen zusammen: Die Ausgangsbedingungen, die erfüllt werden, sind: Das Wasser hat Zimmertemperatur. Das Eis im Wasser schmilzt. Der Wasserspiegel ändert sich dabei nicht. Ein zugrunde liegendes Gesetz ist das sogenannte Archimedische Prinzip, das besagt, dass ein Körper, der auf einer Flüssigkeit schwimmt, genauso viel Flüssigkeitsvolumen im Wasser verdrängt, wie er wiegt. Ein weiteres Gesetz, das hier greift, ist das Gesetz des Schmelzens von Eis bei einer Temperatur über 0 Grad Celsius. Die Erklärung wird demnach logisch hergeleitet: „[…] der Eisbrocken [wiegt] genau so viel wie das Wasser, das von seinem untergetauchten Teil verdrängt wird. Da das Schmelzen das Gewicht nicht verändert, wird aus dem Eis eine Wassermenge mit dem gleichen Gewicht – und deshalb mit dem gleichen Volumen – wie das Wasser, das durch den untergetauchten Teil ursprünglich verdrängt wurde; deshalb bleibt der Wasserspiegel unverändert“ (Hempel, 1977, S. 18). Es ist unschwer zu erkennen, dass sich diese Erklärung nicht nur auf einen konkreten Fall bezieht, sondern allgemein gültig ist, wenn die Ausgangsbedingungen erfüllt sind.

Dennoch beschränkt die rein kausal-logische Herleitung von Erklärungen natürlich deren mögliche Inhalte. Die streng logische Schlussfolgerung eignet sich vor allem für naturwissenschaftliche Phänomene und erläutert Ursache-Wirkungs-Zusammenhänge. Hempel selbst bemerkt dazu: „Die besten Beispiele von Erklärungen, die mit dem D-N-[i.e. deduktiv-nomologischen] Modell übereinstimmen, beruhen auf deterministischen physikalischen Theorien“ (Hempel, 1977, S. 24).

Allerdings gründen weder sämtliche naturwissenschaftlichen noch geisteswissenschaftlichen Erkenntnisse auf allgemeingültigen Gesetzen und Bedingungen, und können daher nicht in dieser Weise abgeleitet werden. Außerdem erweisen sich einige der Annahmen, die Hempel und Oppenheim setzen, als nicht haltbar (s. zur Kritik an diesem Schema Neumeister, 2011, S. 11–14). Nichtsdestotrotz hat sich die Unterscheidung in Explanandum, als den Erklärgegenstand, und Explanans, als alles, was das Explanandum erläutert, in der Forschung zum Erklären etabliert; sie wird nicht mehr nur auf naturwissenschaftliche Kausalerklärungen angewendet (vgl. z. B. Erath, 2017, S. 9-10).

Wie oben bereits angedeutet werden wissenschaftliche Erklärungen in dieser Arbeit nur kurz angerissen. Dies hat vor allem den Grund, dass sich das Ziel, das sie verfolgen, grundsätzlich von dem Ziel der Erklärungen, die in dieser Arbeit untersucht werden, unterscheidet: Wissenschaftliche Erklärungen stellen Zusammenhänge dar und erläutern diese (vgl. Findeisen, 2017, S. 26; Leinhardt,

2001, S. 339), generieren also neues Wissen beziehungsweise machen Zusammenhänge nachvollziehbar. Über die logische Herleitung sollen Zusammenhänge verstanden werden, die vorher noch nicht verstanden wurden. Somit sind wissenschaftliche Erklärungen normalerweise von einer konkreten Kommunikationssituation und damit auch von einer/-m konkreten Adressatin/Adressaten losgelöst und stehen vielmehr für sich. Leinhardt (2001) betont, dass wissenschaftliche Erklärungen (im Original „disciplinary explanations" (Leinhardt, 2001, S. 339)) nicht an Ort und Zeit gebunden sind, da normalerweise keine face-to-face-Kommunikation vorliegt, wenn sie gegeben werden. Vielmehr richten sie sich an unbekannte Adressatinnen und Adressaten. Leinhardt (2001) weist zudem noch auf die Fachspezifität der wissenschaftlichen Erklärungen hin: Was in einer Disziplin als Erklärung gilt, zählt in einer anderen zum Beispiel zum Beschreiben. Dennoch weisen sie eine – eben je nach Disziplin unterschiedliche – feste Form und Sprache auf, die im entsprechenden Diskurs anerkannt und bekannt sind. Wissenschaftliche Erklärungen müssen diesen Ansprüchen genügen, Schilcher und andere (2017) sprechen hier von „Gültigkeit" (S. 442), um im Fachdiskurs anerkannt zu sein. In diesem Kontext wird deswegen auch von *Erklärung* gesprochen, die dazu dient, der Fachkommunität neues Wissen bereitzustellen. Dahingegen stellen didaktische Erklärungen vielmehr den Prozess des *Erklärens* in den Mittelpunkt: Sie haben das Vermitteln von Wissen und somit das Verstehen einer/-s bestimmten Adressatin/Adressaten zum Ziel (vgl. Findeisen, 2017, S. 26; Schilcher et al., 2017, S. 442 bzw. die Erläuterungen unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit).

Dennoch dient die naturwissenschaftliche Perspektive von Erklärungen der sprachwissenschaftlichen Konzeption als Grundlage. Dies betrifft nicht nur die oben erwähnte Unterscheidung in Explanandum und Explanans, wie im folgenden Punkt dargelegt wird.

3.2 Erklären als Sprachhandlung

Eine andere Perspektive auf Erklären als die Naturwissenschaft nimmt die Sprachwissenschaft ein. Die handlungstheoretische Forschung der Sprachwissenschaft untersucht den Sprachhandlungstyp „Erklären" selbst. Hier zeigt sich zunächst, wie etabliert das oben beschriebene Schema nach Hempel und Oppenheim (1948) und die Unterscheidung in Explanandum und Explanans sind. Die Definition von *Explikation* im *Lexikon der Sprachwissenschaft* (Bußmann, 2008) greift darauf zurück:

> Explikation: Bezeichnung der Textlinguistik für ein Vertextungsmuster […], das der Erklärung von Sachverhalten oder Objekten dient. Das zu Erklärende (*explanandum*) wird dabei auf einen anderen, erklärenden Sachverhaltskomplex (*explanans*) zurückgeführt, dessen Erklärungskraft auf einer Tatsache (Anfangsbedingung) und einer verknüpfenden Gesetzmäßigkeit beruht […]. Die explikative Sequenzierung ist charakteristisch für wissensvermittelnde Texte (Bußmann, 2008, S. 183).

Es ist überraschend, dass Explikation in einem Eintrag eines sprachwissenschaftlichen Lexikons auf die Rückführung auf Bedingungen und Gesetze begrenzt ist, da diese Rückführung nicht für jeden Gegenstand geeignet ist (s. Punkt *3.1 (Natur-) Wissenschaftliche Sicht auf Erklärungen* in dieser Arbeit). Obige Definition zeigt jedoch, dass sich auch die Sprachwissenschaft lange Zeit an dem wissenschaftstheoretischen deduktiv-nomologischen Modell von Erklärungen orientiert hat (vgl. Morek et al., 2017, S. 15). Frühe Forschung nimmt dabei zunächst schriftliche Texte in den Blick, bei denen in Anlehnung an die Sprechakttheorie von explikativen Vertextungsmustern (s. obige Definition) gesprochen wird. Dabei stellt zum Beispiel Bayer (1981) sehr genaue Regeln für den Sprechakt „Erklären" auf, die sich ganz klar an dem deduktiv-nomologischen Modell orientieren (vgl. S. 39-41). Im gleichen Artikel weist Bayer allerdings bereits darauf hin, dass die Kriterien der Präzision oder der kausal-logischen Herleitung des Explanans, die für eine wissenschaftliche Erklärung gelten, für die Sprachwirklichkeit von Alltagserklärungen nicht in gleichem Maße gültig sind (vgl. Bayer, 1981, S. 35–36). Auch Jahr (2000) trennt wissenschaftliche Erklärungstexte von Alltagserklärungen ab: Letztere unterscheiden sich zum Beispiel durch weniger komplexe Inhalte und eine weniger komplexe Struktur von wissenschaftlichen Erklärungstexten (vgl. S. 385-390). Gemein ist wissenschaftlichen Erklärungen und Alltagserklärungen der Zweck der Wissensvermittlung beziehungsweise das Ziel, ein Verstehensdefizit zu beseitigen (vgl. Deppermann, 2018, S. 109 bzw. S. 111).

Da mündliche Erklärungen Gegenstand der FALKE-Studie sind, wird hier zur genaueren Erläuterung von schriftlichen Erklärungstexten auf Brinker, Cölfen und Pappert (2018), Jahr (2000) beziehungsweise Lang (1974) verwiesen. Viele der Eigenschaften von Erklärungstexten lassen sich allerdings auch auf mündliche Erklärungen übertragen, zum Beispiel die Orientierung an der Adressatin/am Adressaten (vgl. Jahr, 2000, S. 395).

Wie soeben dargelegt grenzt die Sprachwissenschaft schon früh Gestaltung und Zweck wissenschaftlicher Erklärungen von anderen Erklärungen ab. Gewinnbringend ist hier die Unterscheidung, die Morek et al. (2017) vornehmen: Sie trennen zunächst sprecherzentrierte beziehungsweise interaktionsorientierte Perspektiven auf das Erklären von Erwerbsstudien (Morek et al., 2017, S. 12–13). Erwerbsstudien beschäftigen sich dabei, wie der Name schon sagt, mit dem Erwerb von Erklärkompetenz. Sie spielen in dieser Arbeit keine Rolle.

Die sprecherzentrierte Perspektive konzeptualisiert das Erklären als monologischen Sprechakt und konzentriert sich folglich auf das Produkt, nämlich die Erklärung (vgl. Morek et al., 2017, S. 14). Die oben beschriebene naturwissenschaftliche Sichtweise auf Erklärungen lässt sich in diese sprecherzentrierte Perspektive einordnen.

Dahingegen beschäftigen sich die interaktionsorientierten Perspektiven mit dem Prozess des Erklärens in konkreten Äußerungen. Je nach Schwerpunkt stehen dabei einmal die Teilnehmer/-innen, einmal die Sprachhandlung selbst im Zentrum der Untersuchungen (vgl. Morek et al., 2017, S. 15–16). Bei den teilnehmerorientierten Studien geht es vor allem darum, zu untersuchen, welche Normen und Kriterien die Teilnehmer/-innen an eine Erklärung anlegen (vgl. Morek et al., 2017, S. 16). FALKE-D reiht sich hier in dem Sinne ein, dass der Wahrnehmung von Erklärungen durch Personengruppen, die unterschiedliche Rollen im Erklärprozess haben, auf den Grund gegangen wird. Kriterien, die diese Personengruppen bei der Bewertung von Erklärungen anlegen und als wesentlich erachten, sollen in FALKE-D sichtbar gemacht werden (vgl. Punkt *2.4 (Professionelle) Wahrnehmung von Unterrichtsqualität* sowie Punkt *8 Quantitative Ergebnisse* und Punkt *9 Qualitative Ergebnisse* in dieser Arbeit).

Auch der funktional-pragmatische Ansatz der Sprachwissenschaft untersucht mündliche Äußerungen in ihren konkreten Verwendungskontexten. Er kann nach Morek, Heller und Quasthoff (2017) ebenso zur oben beschriebenen interaktionsorientierten Perspektive gezählt werden, allerdings mit Schwerpunkt auf die Sprachhandlung, die den Sprechhandlungstyp „Erklären" genauer in den Blick nimmt. Dieser Ansatz greift auf ältere Forschung zurück: So stellt Settekorn (1991) bereits Prinzipien für mündliche Erklärhandlungen auf, zum Beispiel das Prinzip der Dialogizität (das die Interaktivität einer Erklärung betont) oder der Ausdrucksvariabilität (das besagt, dass der/die Erklärende im Falle von Nichtverstehen der Adressatin/des Adressaten auf alternative Ausdrucksweisen zurückgreifen kann) und verweist explizit darauf, dass für Erklärungen auch nonverbale Mittel (also Gestik, Mimik usw.) zur Verfügung stehen (vgl. Settekorn, 1991, S. 240–242).

Einigkeit herrscht in der Sprachwissenschaft darüber, dass Erklären ein komplexer Sprechakt (vgl. Klein, 1987) ist: Zum einen ist die notwendige Struktur des Sprechaktes Erklären komplex: „Zum ERKLÄREN bedarf es neben einer Proposition für das, was ERKLÄRT werden soll (Explanandum), mindestens einer weiteren Proposition für den Sachverhalt, der die Erklärung liefert (Explanans)" (Klein, 2009, S. 28). Das bedeutet, dass Erklären propositional mindestens zweistellig ist. Klein (2009) stellt das Erklären zur Verdeutlichung dem einstelligen Sprechakt Versprechen gegenüber. Dieser ist mit einem Satz wie *Ich verspreche Ihnen, dass ich die Klausur morgen zurückgebe* vollständig abgeschlossen. Dahingegen ist es nicht ausreichend für den Sprechakt Erklären zu sagen: *Ich erkläre Ihnen, wie man eine Klausur aufbaut*, denn die eigentliche Erklärung, das

Explanans, fehlt und muss mit mindestens einer weiteren Proposition erläutert werden, etwa *Nach einer kurzen Einleitung strukturieren Sie den Hauptteil, so wie es die Fragestellung vorgibt, und schließen dann mit einem Fazit* (vgl. zu diesem Abschnitt Klein, 2009, S. 27-28).

Komplex ist der Sprechakt Erklären außerdem, weil es ein illokutives Verb ist, dessen perlokutive Wirkung nicht automatisch erfolgt. Während der illokutionäre Akt in der Erklärung selbst besteht (d. h. jemand erklärt jemandem etwas), so ist die beabsichtigte Wirkung (d. h. das Verstehen der Adressatin oder des Adressanten) nicht automatisch erfüllt. Hier vergleicht Klein (2009) das Erklären mit dem Überzeugen: Wenn jemand überzeugt wurde, ist es nicht möglich, dass sie/er nicht glaubt, wovon sie/er überzeugt wurde. Dahingegen ist es sehr wohl möglich, dass jemandem etwas erklärt wurde, sie/er es aber nicht versteht. Zwischen dem Erklären und dem Verstehen ist also zu trennen (vgl. Klein, 2009, S. 27).

Komplex ist als Drittes der Gegenstand, der eine Erklärung erfordert. Um einfache Inhalte zu vermitteln, genügen wissensvermittelnde Sprechakte wie Mitteilen, Informieren (vgl. Hohenstein, 2009, S. 37–41) oder auch Beschreiben (vgl. Becker-Mrotzek, 2004, S. 176–180). Beim Erklären wird grundsätzlich etwas dargestellt, dessen Zusammenhänge nicht offenkundig sichtbar sind, und das eben deswegen einer Erklärung bedarf: „Der [durch eine Erklärung deutlich gemachte] Zusammenhang kann dabei von sehr unterschiedlicher Qualität sein und so verschiedene Qualitäten wie soziale Beziehungen oder technische Funktionsweisen erfassen. Ein zentrales gemeinsames Merkmal der erklärten Aspekte ist ihre Nicht-Wahrnehmbarkeit an der Oberfläche“ (Becker-Mrotzek, 2004, S. 180).

Eine mögliche Kategorisierung unterschiedlicher Arten des Erklärens (je nach erklärtem Gegenstand) ist die Einteilung in Erklärprozeduren. Klein (1987, 2009) unterscheidet die Erklärprozeduren nach dem sich anschließenden Fragewort *was, wie* und *warum*, denn *Erklären* hat für ihn keine einheitliche Bedeutung, es besitzt vielmehr ein „Verwendungs*spektrum*“ (Klein, 2009, S. 25). Mit dieser Kategorisierung gelingt es Klein, die unterschiedlichen Erklärarten auf syntaktischer, pragmatischer und semantischer Ebene zu unterscheiden und zu definieren. Das bedeutet, dass aus pragmatischer Sicht alle drei Arten eine Antwort auf eine Frage darstellen, die sich syntaktisch definieren lässt durch das entsprechende Fragewort, *was, warum* oder *wie*. Diese Fragewörter bestimmen jeweils die syntaktische Struktur der Frage, gleichzeitig machen sie auch den semantischen Unterschied der Erklärung aus. Eine Was-Erklärung bezieht sich auf die Eigenschaften oder konstitutiven Elemente des Erklärgegenstandes, eine Warum-Erklärung erfordert das Nennen von Gründen oder Ursachen, während eine Wie-Erklärung den Ablauf eines Prozesses oder einer Handlung deutlich machen soll (vgl. Klein, 2009). Neumeister erwähnt in ihrer Studie in Anlehnung an Fiehler (2007, unveröffentlicht) die Erweiterung dieser Kategorisierung durch andere

Fragewörter, zum Beispiel ‚erklären, wozu' oder ‚erklären, wodurch' und abstrahiert dies als ‚erklären, *x*' (vgl. Neumeister, 2011, S. 34).

Jüngere Forschung widmet sich verstärkt der Frage, was den Sprechakt des Erklärens genau ausmacht und was ihn von anderen Sprechakten unterscheidet. Dabei zählt das Erklären ebenso wie das Begründen oder das Beschreiben zu den explanativen Sprechakten (vgl. Hohenstein, 2009, S. 37). Ihnen ist gemein, dass sie eine Wissensdifferenz ausgleichen sollen (vgl. Hohenstein, 2009, S. 39), was im Alltagsverständnis und auch in einigen linguistischen Arbeiten dazu geführt hat, die Sprechakte nicht zu differenzieren. Dennoch unterscheiden sie sich klar voneinander, denn ihnen wird jeweils eine eigene illokutive Kraft zugewiesen. Der Zweck der Sprechakte sowie die Art des zu vermittelnden Wissens sind unterschiedlich (vgl. Morek, 2013b, S. 72–73): Beim Beschreiben steht im Vordergrund, dass die Adressatin/der Adressat eine (räumlich-zeitliche) Vorstellung des Gegenstandes gewinnt, während beim Begründen (über die Vermittlung eines Grundes oder einer Information) die Basis für gemeinsames Handeln von Sprecher/-in und Adressat/-in hergestellt wird. Das Erklären hat demgegenüber zum Ziel, dass der Adressatin/dem Adressaten Wissen vermittelt wird, das einen tieferen Einblick in den Gesamtzusammenhang des Erklärgegenstandes gibt (vgl. Hohenstein, 2009, S. 39–41). Die Wahl des Sprechaktes muss also sowohl zum Gegenstand als auch zum Ziel passen, das die/der Erklärende verfolgt. Als Konsequenz dieser Unterscheidung gehört es für Hohenstein (2009) zu hoher Erklärkompetenz, in einer Kommunikationssituation den jeweils passenden Sprechakt wählen zu können (vgl. Hohenstein, 2009, S. 38).

Mit der naturwissenschaftlichen und der sprachwissenschaftlichen Perspektive wurde ein Überblick über zwei verschiedene theoretische Herangehensweisen an Erklären und Erklärungen gegeben. Selbstredend sind für einen tieferen Einblick weitere Differenzierungen notwendig, an die auf dieser Stelle verzichtet wird, da sie nicht im Forschungsinteresse von FALKE-D liegen. Verwiesen wird auf den Überblick zum Erklären im unterrichtlichen Kontext bei Findeisen (2017), die Übersicht zu diskursiven Praktiken des Erklärens im Mathematikunterricht bei Erath (2017) und die Darstellung der unterschiedlichen Perspektiven auf Erklären bei Neumeister (2011).

Im Folgenden wird hingegen eine Sichtweise auf Erklären in den Blick genommen, die dem Kontext der Erklärungen, die in FALKE-D untersucht werden, relativ nahekommt: die lernpsychologische Perspektive auf Erklärungen in Lehr-/Lernkontexten.

3.3 Lernpsychologische Sicht auf Erklärungen

Erklärungen in Lehr-/Lernkontexten werden in der psychologischen Lernforschung als instruktionale Erklärungen beziehungsweise *instructional explanations* bezeichnet, die im *Dorsch Lexikon der Psychologie* wie folgt definiert werden:

> Erklärungen, die in Lehr-Lern-Kontexten gegeben werden, um das Wissen oder Verstehen der Lernenden zu erweitern, werden i.E. genannt. Wenn von i.E. gesprochen wird, meint das i.d.R. nicht längere expositorische Texte (Lehrtexte), sondern kürzere Darlegungen einer Lehrkraft, eines Tutors oder eines Lernprogramms zu einer implizit oder explizit aufkommenden Frage […] (https://portal.hogrefe.com/dorsch/instruktionale-erklaerungen/, 01.05.20).

Das Ziel von Erklärungen in Lehr-/Lernkontexten ist es, eine Grundlage zu schaffen, auf der neues Wissen aufgebaut werden kann, Verständnislücken zu schließen, Fehlvorstellungen zu korrigieren beziehungsweise kognitive Ressourcen freizumachen, um einen tieferen Verstehensprozess zu ermöglichen (vgl. Wittwer & Renkl, 2008, S. 57).

Das Erreichen dieser Ziele wurde in einigen Studien zur Effektivität von instruktionalen Erklärungen überprüft; die Ergebnisse dieser Studien machen es notwendig, diese Erklärungen differenzierter zu betrachten, da in den meisten kaum Lernerfolg nachgewiesen werden konnte (vgl. z. B. Schworm & Renkl, 2006, S. 439–440; Stark, Gruber, Mandl & Hinkofer, 2001, S. 34; VanLehn, Siler, Murray, Yamauchi & Baggett, 2003, S. 244–245)[14]. Es scheint so zu sein, dass das Vorhandensein von instruktionalen Erklärungen Lernende dazu verleiten kann, selbst weniger kognitive Anstrengung zu vollbringen und deswegen auch weniger zu lernen, als wenn sie sich selbst ein Problem erarbeiten müssen (vgl. Wittwer & Renkl, 2008, S. 59–60). Sinnvoll sind instruktionale Erklärungen daher nur, wenn neue Konzepte eingeführt werden oder Gegenstände zu schwierig sind, als dass sich Lernende diese selbst erklären könnten, sowie immer dann, wenn der Lernprozess ins Stocken gerät (vgl. Wittwer & Renkl 2008, S. 50; Charalambous et al., 2011 S. 442; vgl. hierzu auch die Diskussion um das Verhältnis von Instruktion und Konstruktion in Punkt *1.2 Erkenntnisinteresse und Verortung* in dieser Arbeit). Für diese Situationen wurden Grundsätze der Gestaltung entwickelt, die eingehalten werden müssen, damit die gegebenen Erklärungen effektiv sein können (wobei unter Effektivität hier ein Lernzuwachs der Adressatin/des Adressaten zu verstehen ist). Diese Grundsätze werden nun dargelegt.

[14] Hier ist anzumerken, dass sich die Mehrheit der Studien aus der Lehr-/Lernforschung auf schriftliche instruktionale Erklärungen beziehen, die FALKE-D-Studie dagegen mündliche Erklärungen untersucht. Die Ergebnisse scheinen allerdings ähnlich (vgl. Studie zu mündlichen Erklärungen von VanLehn, Siler, Murray, Yamauchi und Baggett, 2003), obwohl sich die kontextuellen Bedingungen von schriftlichen und mündlichen Erklärungen stark unterscheiden. Siehe für FALKE-D ausführlicher Punkt *4 Erklären in FALKE-(D)* in dieser Arbeit.

Das Prinzip der Vorwissensadaptivität

Dabei ist zunächst die Vorwissensadaptivität zu nennen: Bereits 1987 weisen Nickerson, Baddeley und Freeman darauf hin, dass in jeglicher Kommunikationssituation das Gesagte an den vermuteten Wissensstand des Gegenübers angepasst wird. Sie nennen unter anderem das Beispiel des Physikers, der über ein physikalisches Phänomen mit Kolleginnen und Kollegen ganz anders diskutiert als im Gespräch mit Laien (vgl. Nickerson et al., 1987, S. 245-246). Die Vorwissensadaptivität ist also ein grundlegendes Charakteristikum der (Alltags-) Kommunikation.

Die Bedeutung der korrekten Einschätzung des (Vor-)Wissens der Adressatin/des Adressaten ist daher schon seit langem Forschungsgegenstand in der Experten-Laien-Kommunikation. Sie gilt in Gesprächen zwischen Experten und Laien als grundlegende Voraussetzung, damit die Kommunikation beziehungsweise die Erklärung gelingen kann. Dies haben zum Beispiel Bromme, Rambow und Nückles in ihren Studien zu Experten-Laienkommunikation nachgewiesen (vgl. Bromme et al., 2001, S. 317).

Zusammenfassend ist zu sagen, dass sich die Vorwissensadaptivität in Erklärsituationen als essenziell erweist. Die/der Erklärende muss das Vorwissen der/des Lernenden diagnostizieren und die Erklärung daraufhin an eben dieses anpassen, damit überhaupt ein Lernerfolg möglich ist (vgl. Helmke, 2014; Wellenreuther, 2015; Renkl & Wittwer, 2006; Wittwer & Renkl, 2008). Renkl und Wittwer (2006, S. 209) bezeichnen dieses Vorgehen deswegen als „notwendige Bedingung für Effektivität von Erklärungen". Diese Ergebnisse aus der Lehr-/Lernforschung haben dazu geführt, dass im FALKE-Gesamtprojekt *Adressatenorientierung*[15] als ein Kriterium[16] bei der Untersuchung der Wahrnehmung von Erklärungen fachübergreifend aufgenommen wurde. Zur entsprechenden Operationalisierung dieses Aspektes siehe Punkt *4.2.2 Adressatenorientierung* in dieser Arbeit.

Nutzen des Expertenwissens

Hiermit einher geht die Bedeutung des eigenen Wissens beziehungsweise die Einschätzung des eigenen Wissens: Nickerson und andere (1987) konnten in ihrer Studie zeigen, dass Gesprächsteilnehmer/-innen dazu neigen, das eigene Wissen als Grundlage anzusehen für das Wissen, das auch die/der Gesprächspartner/-in besitzt; gerade bei Themen, die dem Allgemeinwissen zugeordnet werden, scheinen die Gesprächsteilnehmer/-innen davon auszugehen, dass das eigene Wissen in etwa auch den anderen zur Verfügung steht (vgl. Nickerson et al., 1987,

[15] Adressatenorientierung umfasst mehr als die Vorwissensadaptivität, schließt diese aber mit ein. Näheres s. Punkt *4.2.2 Adressatenorientierung* in dieser Arbeit.

[16] Im Folgenden werden die Begriffe *Kriterium, Merkmal* und *Aspekt* synonym verwendet und beziehen sich auf die Qualität einer Erklärung.

S. 257–259). Nachfolgestudien bestätigen dieses Ergebnis (vgl. z. B. Fussell & Krauss, 1992). In Gesprächen zu Alltagsthemen ist diese Annahme logisch und sicherlich in den meisten Fällen auch zutreffend und somit eine verlässliche Basis für gelingende Alltagskommunikation (vgl. Nückles, 2001, S. 10).

Im Gegensatz zu den Themen des Allgemeinwissens ist es jedoch bei spezielleren (Fach-) Themen keineswegs so, dass Gesprächspartner/-innen, die sich als Expertin/Experte ansehen, das eigene Wissen als Grundlage bei dem anderen voraussetzen. Vielmehr sind sie sich ihres Status als Expertin/Experte bewusst und unterscheiden, ob die/der Gesprächspartner/-in zur gleichen Expertengruppe zählt oder eben nicht (vgl. Nickerson et al., 1987, S. 258). Der Grund für die mögliche fehlende Vorwissensadaptivität von Erklärungen von Expertinnen und Experten liegt also nicht in einem fehlenden Bewusstsein darüber, dass sie als Expertin/Experte spezielles Wissen besitzen, über das Laien nicht verfügen.

Ob Expertinnen und Experten trotz dieses Bewusstseins von einem zu hohen Vorwissen bei Laien in Bezug auf das Expertenthema ausgehen, haben unter anderem Bromme und andere (2001) anhand ihrer „expertise hypothesis" untersucht. Diese besagt: „persons who possess an extraordinarily high level of knowledge in a certain domain will tend to overestimate what other persons know about this domain" (Bromme et al., 2001, S. 318). In ihrer Studie wurden den Expertinnen und Experten Fragestellungen vorgelegt, bei denen sie einschätzen sollten, ob Laien diese richtig beantworten konnten. Die einzelnen Fragen waren zum Teil dem Expertenwissen („specialist knowledge" (Bromme et al., 2001, S. 318)) und zum Teil dem Allgemeinwissen entnommen, wobei eine Gruppe der Teilnehmer/-innen die Fragen ohne Kennzeichnung in gemischter Reihenfolge bekam, für die andere Gruppe jedoch angegeben war, welche Fragen dem Expertenwissen zuzuordnen waren und welche als Allgemeinwissen galten. Laien erhielten die gleichen Fragen zur Beantwortung, so dass ein direkter Vergleich der Antworten der Laienbefragung mit den Einschätzungen der Expertinnen und Experten vorgenommen werden konnte. Die Ergebnisse der Studie waren allerdings weniger eindeutig als erwartet, da das Laienwissen von Expertinnen/Experten sowohl über- als auch unterschätzt wurde. Allerdings wurde bei knapp der Hälfte der Teilnehmer/-innen gar keine Tendenz, weder in die eine noch in die andere Richtung, festgestellt; vielmehr schätzten die Fachleute bei einzelnen Items das Wissen von Laien richtig ein (vgl. zu den Ergebnissen Bromme et al., 2001, S. 325-326). Bromme und andere (2001) interpretieren diese Ergebnisse folgendermaßen: Zum einen hängt die Einschätzung der Expertinnen/Experten davon ab, ob und wieviel Kontakt sie mit Laien haben. Je mehr sie mit Laien kommunizieren, umso besser können sie deren Vorwissen beurteilen. Zum anderen verbesserte sich die Einschätzung der Expertinnen/Experten in der Studie, wenn bestimmte Items explizit als Expertenwissen ausgewiesen wurden. Sie konnten dann besser beurteilen, wieviel Wissen Laien in

Bezug auf diese Items haben würden. Die bewusste Zuordnung von Inhalten zu Expertenwissen hilft also offenbar bei der richtigen Einschätzung des Vorwissens von Laien (vgl. Bromme et al., 2001, S. 326–328 sowie die Charakterisierung von professionellem Wissen nach Baumert & Kunter, 2006, die unter Punkt *2.2.1 Paradigmen zur professionellen Lehrkraft* in dieser Arbeit zitiert wurde). Nückles (2001) hebt in seiner Studie zur Experten-Laienkommunikation über Computerprobleme die Notwendigkeit der Perspektivenübernahme durch die Expertinnen/Experten hervor. Diese müssen sich ihrer Rolle im Gespräch bewusst sein und müssen in der Lage sein, sich in das Wissen des Laien hineinzuversetzen und diesem in adäquater, nämlich hörergerechter Weise zu begegnen (vgl. Nückles, 2001, S. 16).

Ein weiterer Erklärungsansatz für mögliche fehlende Vorwissensadaptivität liegt in der unterschiedlichen Struktur und Vernetzung von Expertenwissen gegenüber Laienwissen. Expertenwissen unterscheidet sich nämlich nicht nur quantitativ von Laienwissen, sondern auch in seiner Qualität. Es ist vernetzter und kohärenter, die Begrifflichkeiten und Konzepte sind abstrakter und vertiefter verarbeitet (vgl. Bromme, 1992, S. 49–51). Mangelnde Vorwissensadaptivität in Erklärungen kann also ebenso der Unfähigkeit geschuldet sein, das eigene Wissen auf ein Laien zugängliches Niveau zu vereinfachen.

Damit ergibt sich ein weiterer Aspekt, der auf den ersten Blick kontraintuitiv scheint: Zu hohes Fachwissen auf Seiten der/des Erklärenden kann die Fähigkeit zur Vorwissensadaptivität hemmen, so dass Lehrende mit sehr hohem eigenen Vorwissen es weniger gut schaffen, an das Vorwissen ihrer Lerner/-innen anzuknüpfen als Lehrende, die über weniger Vorwissen verfügen (vgl. Renkl & Wittwer, 2006, S. 207-209). Diese Tatsache mag zunächst erstaunen. Eine mögliche Erklärung für das „Hindernis", zu hohes Fachwissen zu besitzen, ist der Umstand, dass die Studien, die zu diesem Ergebnis kamen, im Bereich der Experten-Laien-Kommunikation durchgeführt wurden (vgl. z. B. Bromme et al., 2001), in einem Kontext also, in dem die Erklärenden keine fachdidaktische Ausbildung erhalten haben, was bei Lehrkräften dagegen selbstverständlich der Fall ist. Auch Renkl und Wittwer (2006) weisen auf diese Tatsache hin. Interessant ist in diesem Zusammenhang allerdings auch, dass Studien zur Unterrichtsqualität hohes Fachwissen, im Gegensatz zum fachdidaktischen Wissen, oft nicht als prädiktiv validen Faktor ausweisen konnten (vgl. z. B. Kunter et al., 2011; K. Möller, Kleickmann & Lange, 2013; zum Teil auch Kirschner et al., 2017 bzw. Ausführungen oben unter Punkt *2.2.2 Professionswissen von Lehrkräften*).

Das Prinzip der Funktionen-/Prinzipienfokussierung

Das nächste wichtige Charakteristikum von guten Erklärungen ist nach Wittwer und Renkl (2008) die Funktionen-/Prinzipienfokussierung, welche ein vertieftes Verstehen überhaupt erst ermöglicht (vgl. Wittwer & Renkl, 2008, S. 53–55; Renkl & Wittwer, 2006, S. 214-215). Erklärungen sollen so

gestaltet sein, dass sie je nach Inhalt die zugrundeliegenden Prinzipien (z. B. bei naturwissenschaftlichen Gesetzen) beziehungsweise die Funktionen (z. B. von einzelnen Handlungsschritten) mit erläutern. Diese Prinzipien beziehungsweise Funktionen sind nicht sichtbar, so dass die Erklärungen Einblicke in verborgene Zusammenhänge bieten. Das Erkennen zugrundeliegender Mechanismen versetzt die Lernenden in die Lage, die Inhalte auf neue Kontexte zu übertragen. Diese sind dann bei einer neuen Aufgabe anwendbar. Renkl und Wittwer (2006; S. 214) sprechen in diesem Zusammenhang auch von „konzeptuellen Erklärungen".

Die Forderung nach Funktionen-/Prinzipienfokussierung wurde in fachdidaktischen Studien bereits aufgenommen: So konstatieren Charalambous und andere (2011) in ihrer mathematikdidaktischen Studie zum Erlernen des Erklärens durch Lehramtsanwärter/-innen: „In mathematics, good instructional explanations are also more than simple descriptions of the steps involved in a procedure; rather, they give meaning to these steps" (Charalambous et al., 2011, S. 443). Schopf, Zwischenbrugger und Baier (2015) führen im Rahmen von Erklärungen aus der Wirtschaftswissenschaft an, dass in jeder Erklärung die Fragen was, wie, warum und wozu geklärt werden müssen. Gerade die Frage nach dem „Wozu" entspricht der Funktionenfokussierung: „Mit der Beantwortung der Frage „Wozu braucht man es?" soll die Bedeutung des Inhaltes geklärt werden, d.h. welcher (aktuelle) Stellenwert ihm in der Wirtschaft/Gesellschaft zukommt und welchen Sinn/Zweck beziehungsweise welche Anwendungsbereiche er hat" (Schopf et al., 2015, S. 26). Es fällt bei der Durchsicht der Studien auf, dass keine Beispiele aus geisteswissenschaftlichen Fächern gegeben werden. Auch Wittwer und Renkl (2008) nennen naturwissenschaftliche beziehungsweise technische Inhalte in ihren Erläuterungen. Inwieweit dieser Punkt auf deutschdidaktische Erklärungen übertragen werden kann, wird daher unter Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit ausgeführt und begründet.

Notwendigkeit der minimalistischen Gestaltung einer Erklärung

Ein weiterer Grundsatz, der zur Gestaltung von Erklärungen genannt wird, ist deren minimalistische Gestaltung. In Studien zum Umgang mit dem Computer wurde zum Beispiel nachgewiesen, dass Erklärungen effektiver genutzt werden, je präziser und knapper sie formuliert sind, deswegen sollte zunächst immer eine möglichst kurze Erklärung gegeben werden und ausführlichere nur dann bereit gestellt werden, falls ein/-e Lernende/-r das Bedürfnis hat, einen Sachverhalt genauer zu ergründen. Die Studien konnten nachweisen, dass Erklärungen, die sehr lang sind (z. B. bei Hilfestellungen im Umgang mit Computerproblemen), kaum genutzt werden (vgl. Renkl & Wittwer, 2006, S. 212-213 bzw. S. 215). Der Kontext der Studien, nämlich Experten-Laien-Kommunikation, ist hier erneut mit zu bedenken: Es handelte sich um schriftliche Erklärungen, die die Studienteilnehmenden als Hilfe suchten, wenn sie auf ein Problem stießen. Sie hatten also ein konkretes Ziel vor Augen, nämlich

das Überwinden ihres aktuellen Problems. Der Grundsatz der minimalistischen Gestaltung ist sicher zu überdenken, wenn das Ziel einer Erklärung vertieftes Verstehen bei Adressatinnen und Adressaten ist, – so wie es im Schulunterricht oft der Fall ist.

Entscheidend für die Effektivität von instruktionalen Erklärungen ist außerdem das, was unmittelbar danach passiert: So müssen Lernende nicht nur während des Aufnehmens der Erklärung kognitiv aktiv bleiben, sondern das Erklärte im besten Fall direkt nach der Erklärung anwenden. Auf diese Art und Weise bleibt die kognitive Aktivität auf hohem Niveau und gleichzeitig gelingt die Verankerung des neuen Wissens im Langzeitgedächtnis durch das unmittelbare Weiterverarbeiten besser (vgl. Renkl & Wittwer, 2006; Wittwer & Renkl, 2008, S. 215-216 bzw. S. 56-58; Webb & Mastergeorge, 2003, S. 417–422).

Die sprachliche Gestaltung der Erklärung

Ein weiterer Aspekt ist die Sprache, die in Erklärungen verwendet wird. Überraschenderweise wird die sprachliche Gestaltung in der Lehr-/Lernforschung nur selten hervorgehoben. Diese ist aber insofern von Bedeutung, als der Zusammenhang von fachlichem Lernen und sprachlicher Kompetenz der Schülerinnen und Schüler in zahlreichen Studien nachgewiesen wurde (vgl. z. B. Grießhaber, 2010; Heinze, Herwartz-Emden & Reiss, 2007; Erath, 2017; siehe auch Übersicht in Schmölzer-Eibinger, 2013). Wo die sprachliche Gestaltung im Zusammenhang mit Erklärungen erwähnt wird, wird relativ knapp die Notwendigkeit beschrieben, die Sprache einfach zu halten. Diese Einfachheit bezieht sich sowohl auf die Wort- als auch auf die Satz- und die Textebene. Hier wird dafür plädiert, so verständlich und klar wie möglich zu formulieren (vgl. Schopf et al., 2015, S. 45).

Diese Forderung steht allerdings im Spannungsverhältnis zu dem konzeptionell schriftsprachlichen Register, in dem Erklärungen üblicherweise formuliert sind, denn das Erklären gehört, wie auch das Beschreiben oder das Vergleichen, laut Bailey und anderen (2004) zu Sprachhandlungen, die in Lehr-/Lernkontexten zu erwarten sind und die typischerweise in Bildungssprache formuliert werden (vgl. Bailey et al., 2004, S. 5 bzw. S. 30). Bildungssprache wiederum ist laut Gogolin und Lange (2011) ein eigenes Register der Sprache, das sich gerade nicht durch Einfachheit auszeichnet:

> Mit Bildungssprache ist also ein bestimmter Ausschnitt sprachlicher Kompetenz bezeichnet. Gemeint ist ein formelles Sprachregister, d. h. eine Art und Weise Sprache zu verwenden, die bestimmte formale Anforderungen beachtet. Sehr grob charakterisiert, kann man sagen, dass Bildungssprache auch dann, wenn sie im Mündlichen vorkommt, an den Regeln des Schriftsprachgebrauchs orientiert ist. Besonderes Gewicht besitzt das Register im Bildungskontext: Es wird bei Lernaufgaben, in Lehrwerken und anderem Unterrichtsmaterial verwendet; es wird in Prüfungen und vielen Unterrichtsgesprächen eingesetzt (Gogolin & Lange, 2011, S. 111).

Zur Bildungssprache gehört Fachvokabular, das disziplinspezifisch ist, genauso wie verallgemeinerbares Vokabular, das disziplinübergreifend verwendet wird; Bailey und andere (2004, S. 6) sprechen hier von „nonspecialized language that cuts across content areas".

Alle von Bailey und anderen (2004) untersuchten mündlichen Erklärungen wurden im Register der Bildungssprache gegeben. Damit ist im Umkehrschluss das Beherrschen der Bildungssprache durch die Adressatinnen und Adressaten Voraussetzung, um die Erklärung überhaupt verstehen zu können (vgl. Bailey et al., 2004, S. 11). Die Merkmale der Schriftlichkeit, die das Register der Bildungssprache kennzeichnen (vgl. Feilke, 2013, S. 119), stehen somit im Widerspruch zu der geforderten sprachlichen Einfachheit von Erklärungen.

Auch wenn die untersuchten Beispiele aus naturwissenschaftlichen Disziplinen stammen, so ist das Register der Bildungssprache auch bei Erklärungen in geisteswissenschaftlichen Fächern angemessen. Dies stellt Lehrkräfte und Schüler/-innen vor sprachliche Herausforderungen: Denn für gelingenden Unterricht bedeutet das zum einen, dass die Lehrkräfte Bildungssprache vermitteln müssen (vgl. Schmölzer-Eibinger, 2013, S. 25). Zum anderen müssen die Lernenden Bildungssprache verstehen und sie auch selbst anwenden. „Bildungssprache ist […] ein Werkzeug, das als Potential für bestimmte Handlungstypen zur Verfügung steht" (Feilke, 2013, S. 119).

Konkret auf das Erklären bezogen heißt das, dass es neben themenspezifischem Fachvokabular spezifische, fachübergreifende Formulierungen und Satzstrukturen gibt, die unabhängig von der Disziplin und vom Thema der Erklärung typisch für das Erklären sind. Genaueres über die sprachlichen Eigenheiten von Erklärungen ist bei Thim-Mabrey und Lindl (2021, in Vorb.) sowie Thim-Mabrey (in Vorb.) nachzulesen.

3.4 Erklärungen im (Deutsch-) Unterricht

Instruktionale Erklärungen im Sinne der Lehr-/Lernforschung stellen eine Methode jeden Unterrichts dar. Erklärungen im Schulunterricht haben ebenso das Ziel, dass die Adressatinnen und Adressaten den Inhalt der Erklärung verstehen und eventuell das neue Wissen auch selbst anwenden können. Im Gegensatz zu wissenschaftlichen Erklärungen wird kein neues Wissen generiert, sondern Wissen vermittelt. Die Gegenstände sind aber denen wissenschaftlicher Erklärungen ähnlicher als denen von Alltagserklärungen. Laut Leinhardt (2001, S. 339) schlagen didaktische Erklärungen deshalb eine Brücke von Alltags- zu wissenschaftlichen Erklärungen.

Die oben beschriebenen Gelingensbedingungen können zunächst für Erklärungen im Schulunterricht angenommen werden. Die Gestaltungserklärungen und Kriterien für gutes Erklären, die in der FALKE-D-Studie angenommen wurden, werden unter Punkt *4 Erklären in FALKE-(D)* in dieser Arbeit genauer ausgeschärft und erläutert.

Trotz dieser Gemeinsamkeiten mit instruktionalen Erklärungen machen die kontextuellen und institutionellen Bedingungen der Kommunikationsformen es notwendig, Erklärungen im Schulunterricht gesondert zu betrachten. Schulische Erklärungen unterscheiden sich sowohl von wissenschaftlichen als auch von Alltagserklärungen, denn es handelt sich dabei um „institutionell spezialisierte kommunikative Gattungen, die auf den primären Alltagsgattungen aufbauen und sie in je spezifischer Weise ausgestalten. Zu nennen sind hier […] die Gattungen der Unterrichtskommunikation […]" (Deppermann, 2018, S. 112).
Da in der FALKE-Studie mündliche Erklärungen untersucht werden, beziehen sich die folgenden Ausführungen nur auf mündliche unterrichtliche Erklärsituationen.

Sprachliche Besonderheiten von Erklärungen in der Schule

Die institutionellen Bedingungen betreffen als erstes die sprachlichen Handlungsmuster des Unterrichts, denn von Schülerinnen und Schülern wird das Beherrschen einer Vielzahl von unterschiedlichen Handlungsmustern verlangt. Diese kennen sie aus dem Alltag, zum Beispiel das Argumentieren, das Erzählen oder eben auch das Erklären. Die spezifischen Anforderungen, die im Kontext Schule mit diesen Mustern einhergehen, können die Schülerinnen und Schüler allerdings nicht mit Hilfe ihres Alltagswissens erfüllen, sondern sie müssen sie in der Institution erst erlernen (vgl. Ehlich, 2009, S. 10-11; Vogt, 2009, S. 203-204). Bei Erklärungen ist das vor allem die oben beschriebene Bildungssprache, die durch ihre konzeptionelle Schriftlichkeit Herausforderungen auf allen Ebenen des Sprachsystems aufweist. Schüler/-innen müssen sie verstehen und auch anwenden (s. dazu die Ausführungen in dieser Arbeit unter Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* beziehungsweise zur sprachlichen Komplexität des Sprechaktes Erklärung die Ausführungen unter Punkt *3.2 Erklären als Sprachhandlung*). Während die institutionellen Bedingungen von Schülerinnen und Schülern im Laufe der ersten Schuljahre erlernt und an den weiterführenden Schulen normalerweise von allen Beteiligten akzeptiert werden, erstreckt sich der Lernprozess der spezifischen sprachlichen Anforderungen über die gesamte Schulzeit.

Zeitpunkt des Erklärens im Unterricht

Die Einordnung von Erklärungen in den Gesamtverlauf einer Unterrichtsstunde ist ebenfalls von den institutionellen Rahmenbedingungen geprägt. So sind Erklärungen oft kommunikative Handlungen, die sich relativ klar innerhalb einer längeren Kommunikation isolieren lassen. In Alltagsgesprächen stellen sie häufig Nebenkommunikationen dar, mit Hilfe derer eine Wissenslücke gefüllt wird, bevor zum eigentlichen Gesprächsthema zurückgegangen wird (vgl. Settekorn, 1991, S. 255). Settekorn (1991) unterscheidet bei Erklärungen so genannte lokale Sequenzen von globalen Sequenzen, wobei die lokalen wie oben dargestellt die eigentliche Kommunikation aufgrund von

Verständigungsschwierigkeiten unterbrechen, um sie im Folgenden weiter zu ermöglichen. Globale Sequenzen stellen hingegen Erklärungen dar, die selbst das eigentliche Thema einer Kommunikation sind. Typische Beispiele hierfür sind Wegerklärungen oder Spieleanleitungen (vgl. Settekorn, 1991, S. 255-256).

Diese Unterscheidung kann auf Erklärungen im Unterricht übertragen werden (vgl. Findeisen, 2017; A. Wagner & Wörn, 2011): Gerade Erklärungen, die sich während des Stundenverlaufes aufgrund von Rückfragen, Missverstehen oder Nichtverstehen von Schülerinnen und Schülern ergeben, stellen solche Nebenkommunikationen dar. Die Lehrkraft, sofern sie den Erkläranlass aufgreift, weicht vom geplanten Skript ab, um ein Wissensdefizit zu beseitigen und so den Lernprozess weiter zu ermöglichen. Diese Adhoc-Erklärungen können im Unterricht jederzeit notwendig sein.

Daneben treten geplante und vorbereitete Erklärungen im Unterrichtsverlauf ebenso an verschiedenen Stellen auf und haben je nachdem unterschiedliche Funktionen. Einführungen zu neuen Themen zu geben beziehungsweise gleiche Voraussetzung für das Erarbeiten eines neuen Themenbereichs zu schaffen, ist ein Zweck von Erklärungen, der in der Lehr-/Lernforschung immer wieder genannt wird (vgl. Wellenreuther, 2015, S. 177–178). Diese Erklärungen finden sich vor allem zu Beginn einer Unterrichtseinheit.

Daneben gibt es in der Schule aber auch Erklärprozesse, die im Zentrum des Unterrichtsgeschehens stehen, weil die Erklärgegenstände gleichzeitig (Haupt-)Thema des Unterrichts sind. Dies ist zum Beispiel in den Stunden der Fall, die Neumeister (2011) in ihrer Dissertation zur Frage „(Wie) Wird im Deutschunterricht erklärt?“ (Neumeister 2011) untersucht hat[17]. Die Arbeit beschäftigt sich vor allem mit der Frage, wie diese Erklärungen organisiert und strukturiert sind.

Vogt (2009) geht in seiner Forschung noch einen Schritt weiter, indem er unterrichtliches Erklären explizit nicht als „die Realisierung einer einzelnen sprachlichen Handlung [versteht], sondern [als] Prozess“ (Vogt, 2009, S. 203). Er bezeichnet den „Erklärprozess […] im Vergleich zur direkten Interaktion zwischen zwei Personen [als] zerdehnt […]“ (Vogt, 2009, S. 209). Dieser Erklärprozess sei von den Lehrkräften vorbereitet, unterliege allerdings in der Umsetzung immer wieder Störungen, zum Beispiel wenn Schülerinnen oder Schüler etwas nicht verstehen und nachfragen oder den Erklärprozess unterbrechen (vgl. Vogt, 2009, S. 209). Eine so „weite“ Definition ist allerdings mit Vorsicht zu betrachten: Letzten Endes wird das Erklären hier mit dem Unterrichten im Sinne von Wissensvermitteln gleichgesetzt. Vogt spricht auch von einer Unterrichtsstunde als „organisierte[m] Erklärprozess,“ (Vogt, 2009, S. 204), dem er die Großzahl an unterrichtlicher Kommunikation, mit wenigen Ausnahmen wie etwa Aufgaben, die die Kreativität fördern sollen,

[17] Es handelt sich um Stunden der Sekundarstufe 1, in denen grammatikalische Kategorien, Sprichwörter/Redensarten bzw. Metaphern erklärt wurden. Die während der Stunden vorkommenden Erklärungen stellten keine Nebenkommunikation, sondern vielmehr den Kern des Unterrichtsgesprächs dar.

zuschreibt (vgl. Vogt, 2009, S. 204; Becker-Mrotzek & Vogt, 2009, S. 144–151). Dies ist eine Auffassung, die der vorliegenden Arbeit nicht zugrunde liegt, da Erklären ein spezifisches sprachliches Muster darstellt, das sich klar von anderen sprachlichen Mustern im Unterrichtsgeschehen unterscheiden lässt (s. hierzu auch die Definition von Erklären im Gesamtprojekt FALKE unter Punkt *4.1 Das Konstrukt Gutes Erklären* in dieser Arbeit).

Aufbau einer schulischen Erklärung

Dieses sprachliche Muster weist im Schulunterricht einen bestimmten Aufbau auf. Schulische Erklärungen werden in die Sequenzen Erkläranlass, Erklärinitiierung, Erklärprozess sowie Erklärcoda unterteilt (vgl. A. Wagner & Wörn, 2011, S. 23–26[18]). Ein Erkläranlass besteht immer dann, wenn ein Wissensdefizit vorhanden ist. Im Unterricht bedeutet das, dass die Lehrkraft Erklärbedarf bei ihren Schülerinnen und Schülern voraussetzt[19]. Wird der Erkläranlass aufgegriffen, so kommt es zur Initiierung der Erklärung. Dies geschieht von Schülerseite in Form einer Nachfrage oder auch nur einer Mimik (z. B. Stirnrunzeln). Von Lehrerseite wird zum Beispiel durch Aufzeigen eines Widerspruchs oder durch eine Frage eine Erklärinitiierung vorgenommen (vgl. auch die Ausführungen weiter unten). Durch die Initiierung wird für alle Beteiligten deutlich, dass Erklärbedarf besteht. Sie leitet den eigentlichen Erklärprozess ein (zu möglichen Abläufen s. u.), der schließlich mit einer Erklärcoda beendet wird (vgl. Findeisen, 2017; A. Wagner & Wörn, 2011, S. 23-26)[20].

[18] Neumeister (2011, S. 108-111) spricht in ihrer Untersuchung von Initiierung, Kern und Beendigung; die Sequenzierung deckt sich aber im Wesentlichen mit der von A. Wagner & Wörn (2011), bis auf das Zusammenfallen von Erkläranlass und Erklärinitiierung bei Neumeister.

[19] Natürlich kommt es im Unterricht auch vor, dass Erklärbedarf auf Seiten der Schülerinnen und Schüler besteht, das Wissensdefizit jedoch von der Lehrkraft nicht wahrgenommen oder nicht aufgegriffen wird. In diesem Fall liegt zwar ein Erkläranlass vor, es kommt aber nicht zu einer Erklärinitiierung (vgl. Findeisen, 2017, S. 20).

[20] Vogt (2009) nimmt eine andere Unterteilung vor, die sich aus seiner Auffassung des Unterrichts als zerdehntem Erklärprozess ergibt: In den von ihm untersuchten Unterrichtsstunden macht er die Phasen Aufmerksamkeitsausrichtung, Phänomenisolierung, abstrahierende Fokussierung sowie Exemplifizierung (vgl. Vogt, 2009, S. 205 bzw. Becker-Mrotzek und Vogt, 2009, S. 144) aus, die die Makrostruktur des Erklärprozesses darstellen. Eine ähnliche Einteilung nimmt auch Kiel (1999) in seinem Grundmodell des Erklärens vor, das die Phasen Erkenntnis durch die Sinne, Erkenntnis durch den Verstand mittels Begriffsbildung, Erkenntnis der Vernunft durch Herstellen von Begriffsrelationen und Legimitation der Vernunfterkenntnis (vgl. Kiel, 1999, S. 402-403) unterscheidet. Für Kiel (1999) gilt diese Phasierung unabhängig von den unterschiedlichen Erklärarten, die er definiert, als „Übertragung von Wissen" (vgl. Kiel, 1999, S. 100-159); „Entwicklung von Wissen" (vgl. Kiel, 1999, S. 160-229) und „Aushandlung von Wissen" (vgl. Kiel, 1999, S. 230-266). Von den Erklärungen, die für diese Arbeit konzipiert wurden, lassen sich die Was-Erklärungen am ehesten der Übertragung von Wissen und die Warum- und Wie-Erklärungen eher dem Entwickeln von Wissen zuordnen (s. dazu auch Punkt *5.1.3 Erklärprozeduren erklären, was/ wie/ warum* in dieser Arbeit). Dennoch wird Erklären in der vorliegenden Arbeit enger gefasst als bei Vogt (2009) und Kiel (1999), weswegen deren Sequenzierung hier nicht sinnvoll erscheint.

Betrachtet man diese Phasen der Erklärung genauer, so besteht ein wichtiger Unterschied zu Alltagserklärungen in der Initiierung der unterrichtlichen Erklärung: Bei Alltagserklärungen drückt normalerweise die/der künftige Adressat/-in der Erklärung selbst ihr/sein Bedürfnis nach einer Erklärung aus, indem sie/er eine Frage stellt oder ihr/sein Nichtverstehen auf eine andere Weise kundtut. Die/der Erklärungsgebende greift dieses Bedürfnis als Erkläranlass auf und *erklärt*. Beide Aktanten akzeptieren jeweils die Rolle des anderen.

In der Schule freilich wird dieser Erklärbedarf insbesondere bei vorbereiteten Erklärungen von der Lehrkraft künstlich erzeugt, sei es eben durch das Stellen einer provokanten Frage oder das simple Ankündigen einer Erklärung. Das liegt ebenso wie die veränderte Rollenverteilung an den institutionellen Bedingungen: Die Lehrkräfte haben ein Curriculum zu erfüllen, in dem die zu vermittelnden Themen und der damit verbundene Erklärbedarf festgelegt ist. Den Schülerinnen und Schülern ist dieses Curriculum selten bekannt, ebenso wenig ist ihnen zu Beginn der Stunde ihr eigenes Wissensdefizit bewusst, was das Thema der Unterrichtseinheit anbelangt, und schließlich empfinden sicherlich nicht alle Schülerinnen und Schüler in gleichem Maße das Bedürfnis, dieses Wissensdefizit aufzulösen (vgl. Neumeister, 2011, S. 358-359; Neumeister & Vogt, 2015, S. 569). Anders sieht es natürlich bei Adhoc-Erklärungen (also bei spontanen Erklärungen) aus; diese können durchaus von der beziehungsweise dem künftigen Adressatin/Adressaten ausgelöst werden: Klassischerweise geschieht das in Form einer Schülerfrage, so dass die/der Adressat/-in die Initiierung vornimmt (s. dazu auch Ausführungen weiter oben).

Zudem ist in der Schule nicht mit einem störungsfreien Ablauf der Erklärung zu rechnen, – anders als bei Alltagserklärungen, bei denen die/der Adressat/-in selbst das Bedürfnis hat, eine vorhandene Wissenslücke zu schließen und deswegen nach einer Erklärung fragt. Aufgrund eben genannter institutioneller Bedingungen sowie aufgrund der Tatsache, dass sich eine Erklärung im Unterrichtsplenum immer an eine Adressatengruppe richtet, bei der ein Teil mit dem Inhalt über-, ein anderer Teil unterfordert sein wird, ist von Unterbrechungen des Erklärprozesses auszugehen (vgl. Becker-Mrotzek & Vogt, 2009, S. 143).

Erklärende im Unterricht

Neben dem Ablauf ist auch die Person, die erklärt, in unterrichtlichen mündlichen Erklärsituationen zu beachten. Die institutionellen Rahmenbedingungen bringen es mit sich, dass das asymmetrische Kommunikationsverhältnis, das bei Erklärungen sowieso besteht, im Schulunterricht noch stärker markiert ist, da die Lehrkraft sich aufgrund ihrer Rolle in der Institution Schule in einer „Machtposition“ befindet. Sie vergibt nämlich die Rederechte und entzieht sie auch wieder (vgl. Michalak, Lemke & Goeke, 2015, S. 19; Ehlich, 2009, S. 8-9).

Die oder der Erklärende kann eine Schülerin oder ein Schüler sein. In diesem Falle richtet sich die Erklärung normalerweise an die Klasse beziehungsweise eine Mitschülerin oder einen Mitschüler, aber auch die Lehrkraft ist Adressat/-in der Erklärung, wenn auch ein „künstlicher", da bei ihr meist kein Erklärbedarf besteht. Schülererklärungen sind im Unterricht einerseits zu finden, um die Kompetenz des Erklärens bei Schülern zu fördern (vgl. z. B. Berkemeier, 2009; Harren, 2009; Schmölzer-Eibinger & Fanta, 2014), andererseits können Schülererklärungen zur Wiederholung des Lernstoffes dienen. Schülerinnen und Schüler können von Erklärungen Gleichaltriger durchaus profitieren (vgl. Alber & Neumeister, 2009, S. 154), so dass am Ende einer Lerneinheit gerne ein/-e Schüler/-in gebeten wird, den Stoff noch einmal zusammenfassend zu erklären (vgl. Schmölzer-Eibinger & Fanta, 2014, S. 158-159)[21]. Schülererklärungen dienen aber auch dem Nachvollziehen und Überprüfen des Verstehensprozesses bei der Schülerin/dem Schüler, die/der die Erklärung gibt. Die eigene Erklärung gibt Aufschluss darüber, ob und wie die/der Schüler/-in den Lernstoff verstanden hat, also auch darüber, ob noch weiterer Erklärbedarf besteht (vgl. z. B. Alber & Neumeister, 2009; Appel, 2009). Schülererklärungen sind ein gewinnbringendes Thema für die Untersuchung von Lehr-/Lerninteraktionen. Aufgrund des Designs vorliegender Studie werden sie hier jedoch außer Acht gelassen.

Bei den anderen Arten von Erklärungen sind die Erklärenden entweder Schüler/-innen gemeinsam mit der Lehrkraft bei Erklärungen mit verteilten Rollen (dies geschieht vor allem bei der Auffassung von Unterricht als zerdehntem Erklärprozess s. o.) oder die Lehrkraft allein im Lehrervortrag.

Die in verschiedenen Untersuchungen analysierten Erklärungen mit verteilten Rollen entsprachen im Wesentlichen dem fragend-entwickelnden Unterrichtsgespräch nach dem IRE-Schema (vgl. z. B. Koole, 2009; Neumeister & Vogt, 2015), das Ehlich und Rehbein bereits 1986 als Aufgaben-Stellen/Aufgaben-Lösen-Muster beschrieben haben (vgl. Ehlich & Rehbein, 1986). Bei Erklärungen nach diesem Muster ist die Lehrkraft darum bemüht, möglichst viele Schülerinnen und Schüler am Erklärprozess zu beteiligen und übernimmt deswegen die Gesprächssteuerung, wohingegen sich

[21] An dieser Stelle muss angemerkt werden, dass die Befunde der Lehr-/Lernforschung zu den Lerneffekten durch das Geben von Erklärungen nicht einheitlich sind. In seiner Studie verglich Renkl (1997) z. B. Lernende, die nach einer Textrezeption eine Erklärung geben mussten, mit Lernenden, die dieser Erklärung zuhörten. Entgegen vorheriger Annahmen schnitten die Lernenden, die die Erklärung gaben, in einem anschließenden Wissenstest nicht besser ab als diejenigen Lernenden, die der Erklärung folgten. Als mögliche Gründe hierfür wird einerseits ein erhöhtes Stresslevel bei Lernenden angegeben, die wissen, dass sie nach der Textrezeption eine Erklärung halten müssen, andererseits scheinen die Lernenden mit den Anforderungen, die Erklärungen stellen, überfordert. Dabei handelt es sich nicht unbedingt um inhaltliche Anforderungen, sondern vielmehr um erklärspezifische wie Vorwissensadaptivität, Finden von relevanten Beispielen etc. Außerdem kann nicht davon ausgegangen werden, dass Lernschwierigkeiten automatisch dazu führen, dass die Lernenden diese durch alternative Lösungswege selbst überwinden. Diese Lernschwierigkeiten können daher noch bestehen, während die Lernenden den Sachverhalt erklären, und so zu fehlerhaften Erklärungen führen (vgl. Renkl 1997, S. 137-138). Es ist also weder ein Automatismus, dass Schüler/-innen vom Geben einer Erklärung profitieren, noch dass Schüler/-innen als Adressaten von einer Erklärung Gleichaltriger profitieren können. Weitere Forschung hierzu ist notwendig.

die Schülerinnen und Schüler in der reagierenden Position befinden (vgl. Neumeister & Vogt, 2015). Erklärungen dieser Art werden aufgrund des Studiendesigns in dieser Untersuchung ebenso außer Acht gelassen.

In Studien wurde immer wieder nachgewiesen, dass es hauptsächlich Lehrkräfte sind, die im Unterricht mündlich erklären (vgl. Neumeister, 2011; Spreckels, 2009). Auch wenn diese Erklärungen als eher monologische Gesprächshandlungen gelten, so sind die Adressatinnen und Adressaten dennoch beteiligt: Deren Kommunikationsanteile können allein aus Verstehens- oder Zuhörsignalen (sogenanntes Backchanneling wie Kopfnicken, Stirnrunzeln etc.) bestehen, aber auch aus Zwischenfragen oder Kommentaren.

Adressatinnen und Adressaten von Erklärungen im Unterricht

Die Adressatin/der Adressat der Erklärungen ist im schulischen Kontext normalerweise die Schülerschaft, bei der ein Wissensdefizit beseitigt werden soll. Situationen, in denen ein/-e Schüler/-in über Wissen verfügt, das die Lehrkraft nicht besitzt, so dass diese einen echten Erklärbedarf gegenüber dieser/-m Schüler/-in zum Ausdruck bringt, kommen zwar auch vor, sind aber eher die Ausnahme und nicht Gegenstand dieser Betrachtung. Oben bereits angesprochene Besonderheit, dass im Unterricht oft die gesamte Klasse und damit eine heterogene Adressatengruppe vorliegt, stellt Lehrkräfte bei der Gestaltung von Erklärungen vor besondere Herausforderungen.

4 Erklären in FALKE(-D)

Erklären in den Fokus einer interdisziplinären Studie zu stellen, bietet sich wie bereits angedeutet aufgrund der oben aufgezählten Charakteristika dieser Sprachhandlung an. Für das FALKE-Projekt eignet sich das Erklären besonders, da es eine „fächer- wie modalitätsübergreifende sprachlich-kommunikative Schlüsselkompetenz" (Morek, 2013a, S. 71) darstellt. Da im Schulalltag wie gerade erläutert vor allem Lehrkräfte mündlich etwas erklären, wird Erklären in dieser Studie als Lehrerhandlung konzipiert. Dass der Erklärprozess selbst oft in Interaktion mit einer/einem oder mehreren Adressatin/-nen oder Adressaten abläuft, wurde unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit ebenfalls bereits ausgeführt.

4.1 Das Konstrukt *Gutes Erklären*

Im Sinne der Lehr-/Lernforschung wird Erklären bei FALKE-D als instruktionales Erklären aufgefasst, das die Wissens- und Verstehenserweiterung der Adressatin/des Adressaten zum Ziel hat. Darüber hinaus können Erklärungen auch dazu dienen, Fähigkeiten beziehungsweise Fertigkeiten (im Sinne des tatsächlichen Ausübens einer Fähigkeit) zu vermitteln (vgl. z. B. Kiel, 1999; Macaro, 2013). Je

nachdem, was die/der Erklärende beabsichtigt, wird die Erklärung unterschiedlich gestaltet beziehungsweise eine andere Erklärprozedur (s. dazu Punkt *5.1.3 Erklärprozeduren erklären, was/ warum/ wie* in dieser Arbeit) nötig (vgl. z. B. Morek, 2012; Prediger & Erath, 2014). Erklärungen haben also unterschiedliche Ziele, wobei die Vermittlung eines Erklärgegenstandes an Adressatinnen und Adressaten mit weniger Wissen aber grundlegend ist.

Insgesamt wird Erklären, so wie es in dieser Studie sowie im gesamten Projekt FALKE untersucht wird, folgendermaßen definiert:

> Unterrichtliches Erklären einer Lehrkraft ist ein komplexer, vorbereiteter oder sich situativ ergebender sowie interaktiver Kommunikationsprozess, der auf eine Fähig- bzw. Fertigkeitsvermittlung sowie die Initiierung eines Verstehensprozesses bei den Adressatinnen und Adressaten abzielt (vgl. Kiel, 1999; Leinhardt, 2001; Prediger & Erath, 2014; A. Wagner & Wörn, 2011).

Wesentliche Gedanken dieser Arbeitsdefinition finden sich in Arbeiten zum Erklären aus der Deutschdidaktik. Die soeben erwähnte Komplexität des Erklärgegenstandes ist auch typisch für Erklärungen im Deutschunterricht. So erläutern Neumeister und Vogt (2015) in ihrem Kapitel „Erklären im Unterricht", dass es beim Erklären im Deutschunterricht darum geht, Zusammenhänge darzustellen, was bei der Vermittlung der Gegenstände durch andere Sprechhandlungen, wie zum Beispiel dem Informieren, nicht nötig ist (vgl. Neumeister & Vogt, 2015, S. 563). Außerdem schreiben sie, dass die Lehrkraft ihre „in der wissenschaftlichen Ausbildung erworbenen Wissensbestände situationsadäquat und adressatengerecht einbringen können" (Neumeister & Vogt, 2015, S. 562) muss. Für sie sind Fachkompetenz („erworbene Wissensbestände") sowie Reflexivität (im Sinne von situationsadäquatem und adressatengerechtem Einbringen, s. obiges Zitat) Voraussetzungen für eine gute Erklärung (vgl. Neumeister & Vogt, 2015, S. 562). Beide Aspekte werden in der FALKE-D-Studie berücksichtigt, zum einen mit der Entscheidung, keine fachlich falschen Entscheidungen zu präsentieren (s. Punkt *2.3 Erklärkompetenz von Lehrkräften* in dieser Arbeit), zum anderen mit der expliziten Nennung des Verstehensprozesses bei den Adressatinnen und Adressaten (s. obige Definition sowie Genaueres dazu zu dem unter folgendem Punkt erläuterten Aspekt *Adressatenorientierung*).

4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens

Die gemeinsame Arbeitsdefinition von Erklären diente als Grundlage, um im FALKE-Gesamtprojekt disziplinübergreifende Kriterien für gutes Erklären zu erarbeiten und diese dann jeweils um fachspezifische zu ergänzen. Wie bereits angedeutet zeigte sich in der Diskussion dieser Kriterien schnell, dass zwischen den einzelnen Disziplinen ein sehr heterogener Forschungsstand herrscht: Während zum Beispiel in der Physik- und Mathematikdidaktik Erklären ein sehr präsentes Thema ist (vgl. z. B. Kulgemeyer, 2017; Erath, 2017; Koole, 2009) und auch zahlreiche aktuelle Studien zum Erklärenkönnen (z. B. Kulgemeyer & Tomczyszyn, 2015; Prediger & Erath, 2014; Wagner & Wörn,

2011; Schmidt-Thieme, 2009) vorliegen, gibt es in anderen Fächern wie Englisch, Geschichte oder Biologie eher vereinzelte Studien und Arbeiten zum Erklären (vgl. z. B. Macaro, 2013; Wenzel, 2009; Feigenspan & Michalak, 2018). In der Musik- oder Kunstpädagogik hingegen existieren gar keine Vorarbeiten zum Erklären im Fach[22].

Zu dieser Diversität kommt dazu, dass das Konstrukt „Erklären" beziehungsweise die Kriterien für gutes Erklären in den einzelnen Fächern mitunter recht verschieden sind (vgl. gerade genannte Arbeiten). Trotz dieser Unterschiede finden sich in der Lehr-/Lernforschung, aber auch in Studien zum Erklären aus den verschiedenen Fächern gemeinsame Ansätze. Als Voraussetzungen für das Gelingen von Erklärungen werden zum Beispiel fachliche Richtigkeit (vgl. Findeisen, 2017; Leinhardt, 2001), angemessene Visualisierungen (vgl. Lehner, 2018; Findeisen, 2017; Schopf et al., 2015) und passende Beispiele (vgl. Lehner, 2018; Kulgemeyer & Tomczyszyn, 2015; Wellenreuther, 2015; Bailey et al., 2004) beschrieben. Ebenso werden das Anknüpfen an das Vorwissen der Adressatinnen und Adressaten (vgl. Findeisen, 2017; Schopf et al., 2015; A. Wagner & Wörn, 2011; Charalambous et al., 2011), verständnissichernde Wiederholungen (vgl. Bailey et al., 2004) sowie ein strukturierter Aufbau (vgl. Lehner, 2018, Charalambous et al., 2011; Wenzel, 2009) genannt (vgl. auch die Übersicht bei Findeisen, 2017, S. 62–66).

In einem aufwendigen disziplinübergreifenden Einigungsprozess wurden in FALKE Kriterien für eine gelungene Erklärung gewählt, die identisch in allen Teilstudien abgefragt wurden: Zum einen sind das *Strukturiertheit* und *Adressatenorientierung*, die in den einzelnen Fächern ausreichend Überschneidungen aufweisen, um als disziplinunabhängige Kriterien aufgenommen werden zu können. Zum anderen wurden die *Sprachliche Verständlichkeit* sowie der *Sprech- und Körperausdruck* als Kriterien festgelegt. Die Bedeutung von Sprache (vgl. Becker-Mrotzek, Schramm, Thürmann & Vollmer, 2013; Gogolin & Lange, 2011) und sprech-sprachlicher Gestaltung in der Unterrichtskommunikation (vgl. Pabst-Weinschenk, 2016) wird in der Lehr-/Lernforschung seit einiger Zeit immer stärker betont; gerade die sprech-sprachliche Gestaltung spielt allerdings in der Forschung zum Erklären selten eine Rolle (eine Ausnahme bietet der Leitfaden zum Erklären von Scheffel, 2019; s. hierzu auch Punkt *3 Theoretische Sichtweisen auf Erklären* in dieser Arbeit). Deswegen ist es gewinnbringend, diesen Aspekt ebenfalls fachübergreifend zu untersuchen.

Die Bedeutung anderer Kriterien, wie zum Beispiel Visualisierungen oder passende Beispiele, erwies sich als stark von der jeweiligen Disziplin beziehungsweise vom Thema der Erklärung abhängig. Sie wurden deshalb nicht als disziplinunabhängige Kriterien in FALKE konzeptualisiert und operationalisiert.

[22] Interessanterweise wird aber in Penzel (2015) die Methode des *Iconic Concept Mapping* aus der Kunstpädagogik zum anschaulichen Erklären für andere Fächer vorgestellt.

Im Folgenden werden die einzelnen Aspekte, die im Gesamtprojekt FALKE untersucht wurden, kurz erläutert. Die deutschspezifischen Aspekte, die zusätzlich nur in der Studie FALKE-D erhoben und untersucht wurden, werden unter Punkt *4.2.5 Erklärtiefe in FALKE-D* und Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* (in dieser Arbeit) genauer ausgeführt.

An dieser Stelle sei bereits kurz auf die empirische Untersuchung der Aspekte hingewiesen: Zur möglichst breiten Erfassung der Aspekte wurden zu jedem mehrere Items entwickelt, aus denen für die Berechnungen eine Skala erstellt wurde (vgl. Punkt *8.2.2 Skalenebene: Bewertungen der Aspekte guten Erklärens* in dieser Arbeit). Die Skalen sind der Übersichtlichkeit halber identisch benannt wie die Aspekte, also *Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit, Sprech-/Körperausdruck* und *Erklärtiefe* sowie *Persönlichkeitswirkung* und *Fachliche Relevanz*. Die Ergebnisse der empirischen Untersuchung der Skalen werden ebenfalls unter gerade genanntem Punkt *8.2.2 Skalenebene: Bewertungen der Aspekte guten Erklärens* in dieser Arbeit dargestellt.

4.2.1 Strukturiertheit

Eine klare Struktur hat sich längst als Merkmal guten Unterrichts etabliert (vgl. Helmke, 2014, S. 197–199). Sie ist ebenso ein Kriterium für eine gute Erklärung (vgl. Kiel, 1999; Leisen, 2007), denn sie ermöglicht es den Zuhörerinnen und Zuhörern, der Erklärung bewusst folgen zu können, da sich die Struktur sowohl am Inhalt der Erklärung als auch am Lernprozess (vgl. Schopf et al., 2015, S. 44) orientieren muss.

Im FALKE-Gesamtprojekt wurde auf der Grundlage von bisheriger Forschung sowie in Absprache mit allen beteiligten Fächern ein Modell zur *Strukturiertheit* erstellt, dessen zentrale Unteraspekte *Gegenstandsbestimmung, Fokussierung, Aufbau* und *Transparenz* sind und das in Abbildung 6 sowie in Anhang 1 abgebildet ist.

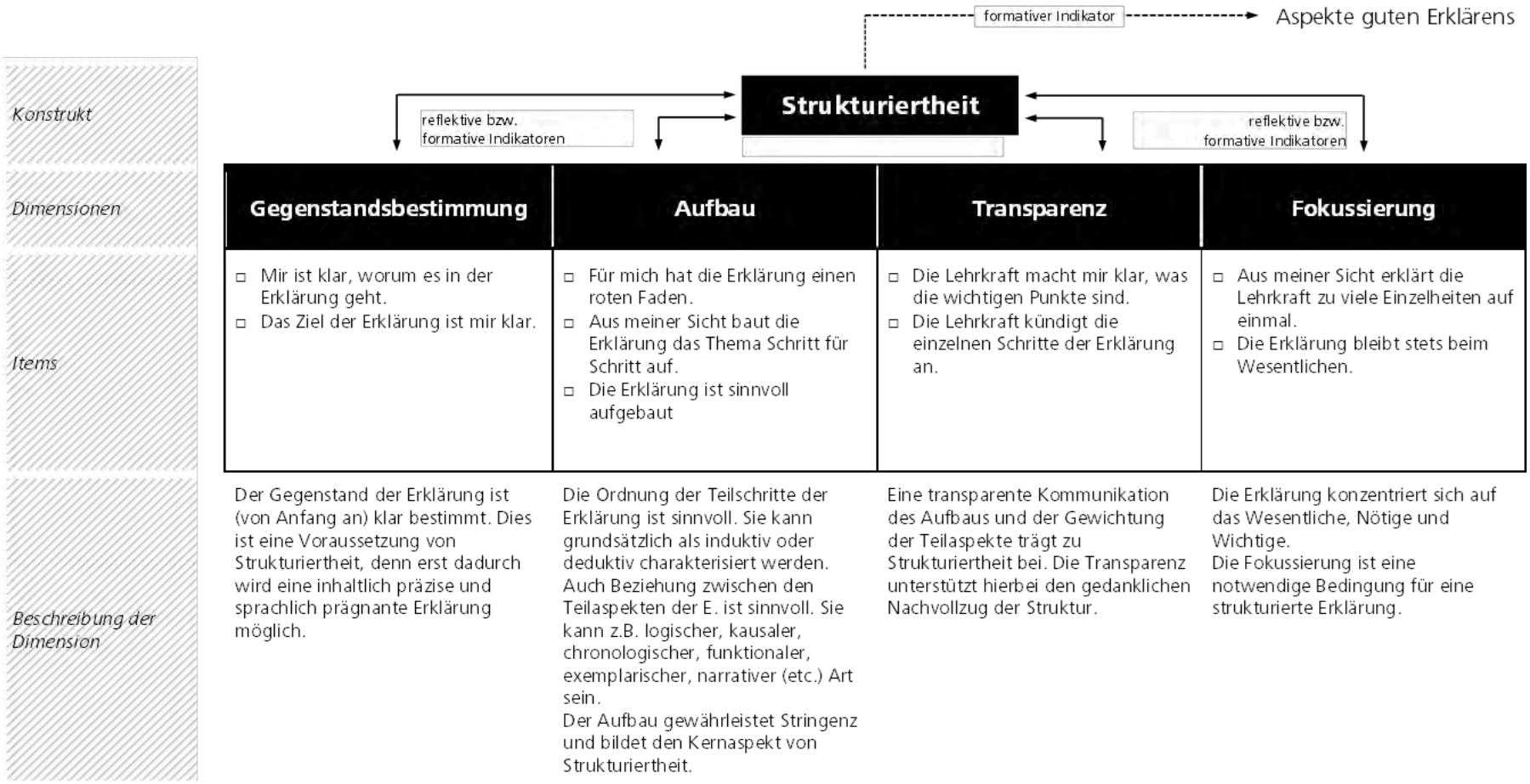


Abbildung 6: Modell zur Strukturiertheit bei FALKE

Unter *Gegenstandsbestimmung* wird dabei verstanden, dass der Gegenstand der Erklärung von Anfang an klar bestimmt ist. Die *Gegenstandsbestimmung* wird als Voraussetzung für *Strukturiertheit* aufgefasst, denn sie ermöglicht erst eine inhaltlich präzise und sprachlich prägnante Erklärung (vgl. Kiel, 1999, S. 59–61; Wellenreuther, 2015, S. 178). Dies wurde mit dem Item „Es ist klar, was genau die Lehrkraft erklären wollte" operationalisiert.

Fokussierung besagt, dass sich die Erklärung auf das Wesentliche, Nötige und Wichtige konzentriert. Die *Fokussierung* ist eine notwendige Bedingung für eine strukturierte Erklärung (vgl. Helmke, 2014, S. 193–195; Wragg, 1993, S. 135), denn Abweichungen vom Thema verhindern eine klare Struktur. Sie wurde mit dem Item: „Die Lehrkraft hat zu viele Einzelheiten erklärt" erhoben, dessen Werte für die Berechnungen umkodiert wurden.

Mit *Aufbau* ist der Unteraspekt gemeint, der spontan in den Sinn kommt, wenn über die Struktur einer Erklärung gesprochen wird. Der Aufbau ist dann klar und stringent, wenn die Reihenfolge der Teilschritte der Erklärung sinnvoll ist. Auch die Darstellung der Beziehung zwischen den Teilschritten der Erklärung muss dem Gegenstand angemessen sein. Sie kann logischer, kausaler, chronologischer, funktionaler, exemplarischer oder narrativer Art sein. Der *Aufbau* bildet den Kernaspekt von *Strukturiertheit* (vgl. Kiel, 1999, S. 134; Leisen, 2007, S. 461; A. Wagner & Wörn, 2011, S. 27). Er wurde mit den Items „Die Erklärung hatte einen roten Faden" sowie „Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut" operationalisiert.

Der letzte Teilbereich der *Strukturiertheit*, die *Transparenz*, bezieht sich zum einen darauf, den Aufbau der Erklärung ausdrücklich klarzumachen, zum anderen auf die explizite Gewichtung der Teilaspekte. Durch dieses Sichtbarmachen des Aufbaus kann die Struktur gedanklich besser nachvollzogen werden (vgl. Wragg, 1993, S. 112; Allhoff & Allhoff, 2006, S. 64). Da für Letzteres ein sprachliches Explizitmachen unabdingbar ist, wurde dieses Kriterium unter dem Merkmal *Sprachliche Verständlichkeit* aufgenommen und dort erhoben (s. Punkt *4.2.3 Sprachliche Verständlichkeit* in dieser Arbeit).

Zur Validierung der einzelnen Aspekte guten Erklärens wurde zu jedem zusätzlich ein Globalurteil erhoben. Für die *Strukturiertheit* lautet es „Die Erklärung war gut strukturiert".

Für die genaue Ausarbeitung des Gesamtmodells der *Strukturiertheit*, aus dem die Konzeptualisierung sowie die Operationalisierung im Gesamtprojekt FALKE hervorgehen, sei erneut auf Anhang 1 verwiesen.

Sämtliche Items dieses Aspekts sind in Anhang 3.1 nachzulesen.

4.2.2 Adressatenorientierung

Das Anpassen einer Äußerung an die Adressatin/den Adressaten ist zunächst ein Grundsatz in jeglicher Kommunikation: „Wenn wir einen Gesprächsbeitrag produzieren, müssen wir unweigerlich Annahmen darüber machen, welches Wissen über die Referenten und Sachverhalte, auf die sich der zu produzierende Turn bezieht, unsere Gesprächspartner bereits haben" (Deppermann, 2018, S. 112). Für eine Erklärung im schulischen Unterricht gilt das in besonderem Maße.

Die *Adressatenorientierung* ist, genau wie die *Strukturiertheit*, ein anerkanntes Prinzip guten Unterrichts allgemein, das sich auf die Mikrodimension des Erklärens übertragen lässt. Für eine gute Erklärung im Unterricht ist sie unabdingbar und gleichzeitig der zentrale Unterschied zu wissenschaftlichen Erklärungen: Während bei Erklärungen im Lehr-/Lernkontext die Vermittlung eines Gegenstandes im Vordergrund steht, ist das zentrale Ziel von wissenschaftlichen Erklärungen, den Gegenstand selbst zu erklären. Wissenschaftliche Erklärungen stehen deshalb für sich und wenden sich nicht explizit an einen Adressaten oder eine Adressatin (s. Ausführungen unter Punkt *3.1 (Natur-) Wissenschaftliche Sicht auf Erklärungen* sowie unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit). In Lehr-/Lernkontexten und gerade im schulischen Unterricht stellt sich durch das Ziel der Vermittlung des Gegenstandes an die Adressatinnen und Adressaten die Bedeutung der *Adressatenorientierung* dagegen als essenziell heraus.

In der Forschung zu Unterrichtsqualität beziehungsweise zur Qualität von Erklärungen wird dieses Prinzip verschieden bezeichnet, so sprechen zum Beispiel Helmke (2014, S. 248–262) und Wellenreuther (2015, S. 180) von „Passung" beziehungsweise Wellenreuther später auch von „Adaptivität" (2015, S. 358), Findeisen (2017, S. 57–62) von „Lernendenzentrierung" oder Appel (2009, S. 36) von „Adressatenbezug". Gemeint ist grundsätzlich die Anpassung der Erklärung an die Adressatin/ den Adressaten, und zwar in sämtlichen, im Folgenden ausgeführten Punkten[23]. Im FALKE-Gesamtprojekt wurde der Begriff *Adressatenorientierung* gewählt, da dieser Begriff in einem Großteil der beteiligten Didaktiken bekannt ist und von allen Fächern mitgetragen werden kann.

Die Erklärung in Lehr-/Lernkontexten muss wie gesagt immer die Zielgruppe im Blick haben (vgl. Findeisen, 2017, S. 17-18; Helmke, 2014, S. 236), denn es gibt nicht „die" gute Erklärung per se. Das bedeutet, dass die *Adressatenorientierung* nicht nur ein Kriterium für eine gute Erklärung darstellt, sondern sogar Voraussetzung dafür ist, dass die Erklärung überhaupt einen Lernprozess auslösen kann (vgl. Schopf & Zwischenbrugger, 2015, S. 22-23; Renkl & Wittwer, 2006, S. 209-212).

[23] Die oben bereits erläuterte Vorwissensadaptivität (die auch als „Vorwissensangepasstheit" bezeichnet wird, z. B. bei Renkl und Wittwer, 2006, S. 209–212) ist der Kernbereich der Adressatenorientierung, nämlich die Anpassung der Erklärung an das Vorwissen der Adressatin/des Adressaten.

Analog zur *Strukturiertheit* wurde im FALKE-Gesamtprojekt auch für die *Adressatenorientierung* ein fächerübergreifendes Modell auf der Basis von Literaturrecherche und eigenen Überlegungen entwickelt, das in Abbildung 7 sowie in Anhang 2 einzusehen ist. Da die *Adressatenorientierung* grundsätzlich alle Teilaspekte einer Erklärung betrifft und deswegen ein sehr komplexes Merkmal ist, ist eine Anordnung der Aspekte nach den Oberpunkten *Inhalt* der Erklärung (im Modell als *Sache* bezeichnet), *Adressat/-in* der Erklärung beziehungsweise *Performanz* des Erklärens (im Modell als *adressatenbezogenes Erklären* bezeichnet) sinnvoll. Der Inhalt muss zum Beispiel an das Vorwissen der Lernenden angepasst sein (vgl. z. B. Kunter & Trautwein, 2018, S. 93-94) und didaktisch so reduziert sein (vgl. z. B. Kunter & Trautwein, 2018, S. 149–151), dass die Lernenden ihn kognitiv verarbeiten und behalten können. Gleichzeitig muss die Erklärung sprachlich, sprecherisch und medial (vgl. z. B. Schopf et al., 2015, S. 36–37) so umgesetzt werden, dass die Adressatinnen und Adressaten gut folgen können. Die einzelnen Aspekte werden nun genauer dargestellt.

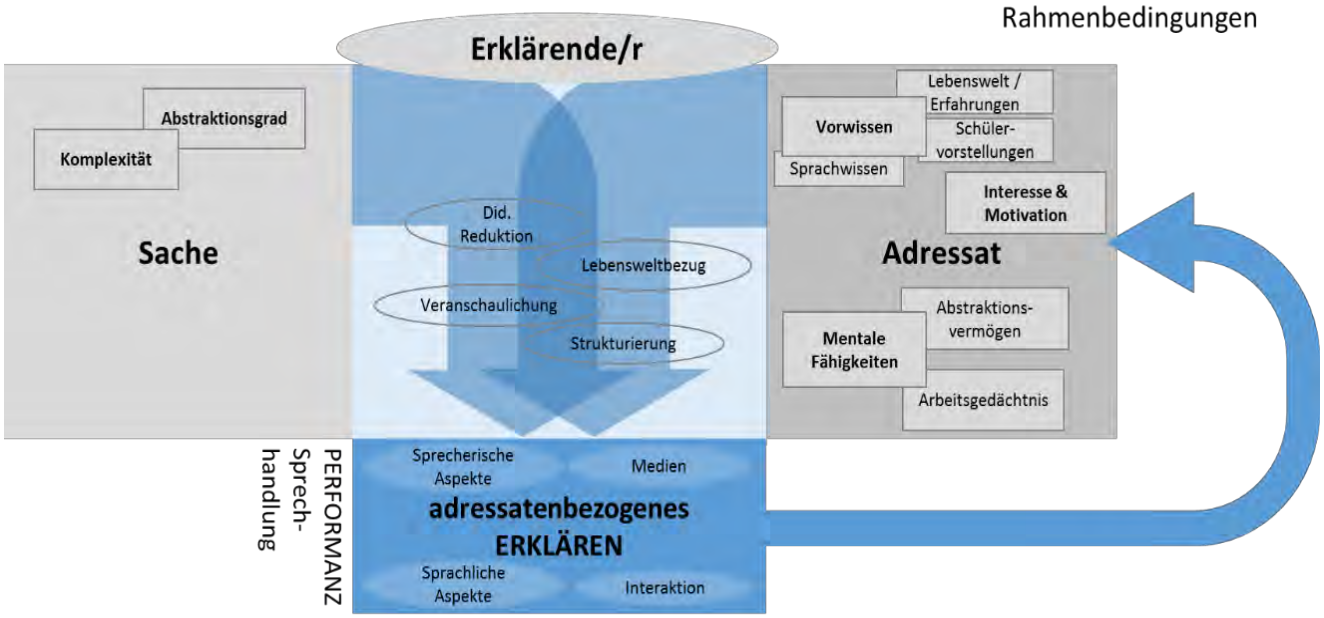


Abbildung 7: Modell zur Adressatenorientierung bei FALKE

Inhaltliche Aspekte

Das FALKE-Modell zur *Adressatenorientierung* nimmt zunächst den Inhalt der Erklärung in den Blick: Die/der Erklärende muss den Gegenstand fachlich in all seiner Komplexität und seinem Abstraktionsgrad erfasst haben (vgl. Findeisen, 2017, S. 104–107; Josefy, 2009, S. 90; Spreckels, 2009, S. 135–136), bevor sie oder er damit beginnen kann, eine Erklärung dazu vorzubereiten. Das fachliche Wissen ist, wie der Name sagt, fachspezifisch; je nach Disziplin ist es ganz

unterschiedlicher Natur und verlangt deswegen auch unterschiedliche Erklärungen (vgl. Leinhardt, 2001, S. 336).

Unter Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* (in dieser Arbeit) wurde bereits der Unterschied zwischen Experten- und Laienwissen dargelegt und ebenso dargestellt, dass zu hohes fachliches Wissen für eine Erklärung hinderlich sein kann. Dennoch ist eine solide fachliche Basis des zu erklärenden Inhaltes notwendig und das fachliche Wissen somit unabdingbar für eine gute Erklärung (s. dazu auch Punkt *2.2.2 Professionswissen von Lehrkräften* in dieser Arbeit). Die Herausforderung auf den Gegenstand bezogen bedeutet, dass dieser in der Erklärung fachlich richtig dargestellt, aber in seiner Komplexität so reduziert werden muss, dass er dem Abstraktionsvermögen der Adressatinnen und Adressaten Rechnung trägt. Fachliches Wissen allein reicht nicht aus, um eine gute Erklärung geben zu können.

Adressat/-in der Erklärung

Neben dem eigentlichen Erklärgegenstand muss die/der Erklärende die Adressatinnen und Adressaten im Blick haben, denn die Aufbereitung muss auf ihre Eigenschaften abgestimmt sein. Das ist eine besondere Herausforderung, da diese Eigenschaften im Gegensatz zu den Eigenschaften der Sache veränderlich sind und in jeder einzelnen Erklärung immer wieder neu angepasst werden müssen; eine an die Adressatin/den Adressaten angepasste Erklärung ist deswegen immer zeit- und situationsgebunden.

Zu den Eigenschaften der Adressatin/des Adressaten zählen einmal deren/dessen mentale Fähigkeiten, das heißt im Wesentlichen die kognitive Leistungsfähigkeit des Arbeitsgedächtnisses sowie das Abstraktionsvermögen. Diese sind altersabhängig, dabei aber individuell verschieden. Das konkret auf den Erklärgegenstand bezogene Vorwissen zählt ebenso zu diesen Aspekten: Die Bedeutung des Vorwissens wurde in der Lehr-/Lernforschung immer wieder nachgewiesen. Lernende verarbeiten neues Wissen leichter und behalten es auch länger, wenn sie es mit vorhandenen Wissensbeständen verknüpfen können (s. Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* in dieser Arbeit). Das Vorwissen darf nicht überschätzt werden, denn dann entstehen zu schwierige Erklärungen für die Adressatin/ den Adressaten, so dass der neue Inhalt nicht erfasst werden kann. Das Vorwissen darf aber auch nicht unterschätzt werden, denn dann entstehen zu leichte Erklärungen, die unnötige kognitive Ressourcen in Anspruch nehmen und schnell zu Langeweile führen (vgl. Findeisen, 2017, S. 59).

Zu berücksichtigen ist in diesem Zusammenhang auch die sprachliche Kompetenz der Adressatinnen und Adressaten. Die sprachlichen Herausforderungen beim Geben einer Erklärung im Schulunterricht wurden unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* (in dieser Arbeit) bereits dargelegt. Damit die Adressatinnen und Adressaten die Erklärung korrekt nachvollziehen

können, muss die sprachliche Gestaltung ihrem Sprachniveau entsprechen, und zwar sowohl auf der Wort- als auch auf der Satz- sowie auf der Textebene. Auch die sprachlichen Fähigkeiten der Adressatinnen und Adressaten sind individuell unterschiedlich.

Neben diesen mentalen Fähigkeiten gehören zu den Eigenschaften der Adressatinnen und Adressaten subjektive individuelle Merkmale. Damit ist zum Beispiel die motivationale Einstellung zum Gegenstand gemeint; eine positive motivationale Grundstimmung sowie die damit verbundene Anstrengungsbereitschaft sind Voraussetzung für effektives Lernen (vgl. Kunter & Trautwein, 2018, S. 17). Falls die Adressatin/der Adressat den Gegenstand als wenig interessant erachtet, muss versucht werden, das Interesse daran zu wecken. Auch Vorerfahrungen mit dem Erklärgegenstand beziehungsweise der Grad an Lebensweltbezug des Gegenstandes gehören zu diesen individuellen Merkmalen der Adressatinnen und Adressaten.

All diese Merkmale und Eigenschaften des Gegenstandes sowie der Adressatin/des Adressaten werden bei der Planung der Erklärung analysiert und münden schließlich während des Haltens der Erklärung zu im Folgenden beschriebenen Gestaltungsprinzipien.

Grundsätze bei der Performanz der Erklärung für die/den Erklärende/-n

Grundsätze bei der Aufbereitung des Gegenstandes sind *didaktische Reduktion, Lebensweltbezug, Veranschaulichung* und *Strukturierung*. Damit die eigentliche Erklärung tatsächlich auf die Adressatin/den Adressaten zugeschnitten werden kann, müssen diese Grundsätze zur Anwendung kommen.

Die *Strukturiertheit* wurde als eigenes Merkmal der Erklärung abgefragt (s. Punkt *4.2.1 Strukturiertheit* in dieser Arbeit) und ist in dieser Arbeit nicht der *Adressatenorientierung* zugeordnet.

Die *didaktische Reduktion* gelingt dann, wenn der Gegenstand der Erklärung fachlich richtig, aber bezogen auf die Adressatinnen und Adressaten weder zu einfach noch zu schwierig dargestellt wird (s. obige Ausführungen zu *Adressat/-in der Erklärung*). Operationalisiert wurde dieser Aspekt bei FALKE durch die Items: „Für Schüler/-innen war die Erklärung zu schwierig" beziehungsweise für die Statusgruppe der Schüler/-innen „Die Erklärung war zu schwierig für mich" und „Die Erklärung war angemessen didaktisch reduziert" (nur bei den Expertenstatusgruppen), wobei das Item „..zu schwierig" bei den Berechnungen entsprechend umkodiert wurde. Zur *didaktischen Reduktion* gehört auch, dass das Vorwissen der Adressatinnen und Adressaten bei der Erklärung so berücksichtigt wird, dass diese den neuen Gegenstand einordnen und an bestehendes Wissen anknüpfen können. Dieser Punkt wurde mit dem Item „Schüler/-innen konnten der Erklärung mit ihrem Wissen gut folgen" beziehungsweise für die Statusgruppe Schüler/-innen „Ich konnte der Erklärung mit meinem Wissen gut folgen" untersucht.

Gleichzeitig wird den Adressatinnen und Adressaten in der Erklärung verdeutlicht, dass der Erklärgegenstand für sie bedeutsam ist, indem *Lebensweltbezug* hergestellt wird. Untersucht wurde das mit dem Item: „Die Erklärung berücksichtigte die Lebenswelt der Schüler/-innen" beziehungsweise für die Statusgruppe der Schüler/-innen „Die Erklärung hatte etwas mit meinem Leben zu tun."

Schließlich ist zur Absicherung des Aspektes auch hier ein Globalurteil nötig. Dieses lautet für die Statusgruppen der Studierenden/Lehrkräfte/Didaktiker/-innen: „Die Erklärung war adressatengerecht". Da davon auszugehen ist, dass die Statusgruppe der Schüler/-innen den didaktischen Fachbegriff „adressatengerecht" nicht kennt, wurde das Globalurteil der *Adressatenorientierung* außerdem mit dem Item: „Die Erklärung passte für Schüler/-innen dieser Jahrgangsstufe" beziehungsweise für die Schüler/-innen „Die Erklärung passte für mich" erfragt.

Weitere Aspekte der *Adressatenorientierung* wie die Veranschaulichung (vgl. Deppermann, 2018, S. 111), zum Beispiel durch verwendete Medien, aber auch die Interaktion mit den Adressatinnen und Adressaten sind zwar im Modell berücksichtigt, weil sie eine wichtige Rolle beim Prozess des Erklärens spielen. Bei der gemeinsamen Operationalisierung im Gesamtprojekt wurden sie jedoch aus den in Punkt *4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens* in dieser Arbeit genannten Überlegungen außen vor gelassen.

Nur wenn die/der Erklärende die Voraussetzungen korrekt analysiert und die Gestaltungsprinzipien unter Berücksichtigung dieser Analyse umsetzen kann, gelingt es, eine Passung zwischen der Sache und der Adressatin/dem Adressaten herzustellen. Und nur unter dieser Bedingung kann überhaupt ein Verstehensprozess in Gang gesetzt werden, der eine Veränderung des Vorwissens oder den Aufbau neuen Wissens bei der Adressatin/dem Adressaten bewirkt.

Das Modell zur *Adressatenorientierung* des FALKE-Gesamtprojektes findet sich wie gesagt in Anhang 2, die Items dazu in Anhang 3.1.

Für die Aspekte der *Sprachlichen Verständlichkeit* sowie des *Sprech- und Körperausdrucks* wurden unter Einbezug der entsprechenden Fachwissenschaften eigene Konstrukte entwickelt, die im Folgenden kurz vorgestellt werden.

4.2.3 Sprachliche Verständlichkeit

Die *Sprachliche Verständlichkeit* wurde neben der *Strukturiertheit* und der *Adressatenorientierung* als drittes fachübergreifendes Kriterium zur Bewertung von Erklärungen im FALKE-Projekt aufgenommen. Gerade aus dem interdisziplinären Blickwinkel heraus ist die sprachliche Gestaltung der Erklärungen ein gewinnbringender Untersuchungsaspekt, da Sprache in allen Fächern und bei nahezu allen Themen das zentrale Lernmedium (vgl. Schmölzer-Eibinger, 2013, S. 25) ist. In den Bildungsstandards des Faches Deutsch für die Primarstufe ist diese zentrale Bedeutung der Sprache

explizit aufgenommen: „Sprache ist in allen Fächern Medium des Lernens“ (KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004b, S. 6). In Studien aus den Naturwissenschaften zeigt sich der Zusammenhang zwischen sprachlichen Fähigkeiten und Leistungen im Fach (vgl. Heinze et al., 2007; Gebhardt, Rauch, Mang, Sälzer & Stanat, 2013).

Was die Sprache in Erklärungen betrifft, so wurde auf das Spannungsverhältnis zwischen dem für Erklärungen üblichen Register der Bildungssprache und der gleichzeitig geforderten Einfachheit der Sprache bereits unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* dieser Arbeit hingewiesen. Eine Erklärung kann nur dann gelingen, wenn deren Sprache für die Adressatin/den Adressaten verständlich ist. Im FALKE-Gesamtprojekt ist deswegen die deutsche Sprachwissenschaft vertreten. Sie beleuchtet diesen Aspekt aus dezidiert fachwissenschaftlicher Perspektive.

Oben genannte Anpassung der Erklärung an die Adressatin/den Adressaten gilt auch für die verwendete Sprache: „Annahmen über das Partnerwissen sind für die Turngestaltung entscheidend, weil von ihnen abhängt, welche Ausdrücke und Formulierungen wir benutzen können, um einen Beitrag zu produzieren, der einerseits hinreichend explizit und somit verständlich, andererseits aber auch nicht zu redundant mit Blick auf das schon bestehende Wissen des Partners ist“ (Deppermann, 2018, S. 112). Damit die Zuhörenden eine Erklärung verstehen, müssen sie zum Beispiel den in der Erklärung verwendeten Wortschatz kennen. Dies wurde mit dem Item „Bei manchen Wörtern wissen die Schüler/-innen eventuell nicht, was sie bedeuten“ beziehungsweise für die Statusgruppe der Schüler/-innen „Bei manchen Wörtern habe ich nicht gewusst, was sie bedeuten“ erhoben. In der Auswertung wurde dieses Item umkodiert.

Auf der Satzebene muss der Satzbau für das Hörverständnis angemessen einfach gebaut sein, was erhoben wurde mit „Manche Sätze hat die Lehrkraft zu lang gemacht“. Auch dieses Item wurde in der Auswertung umkodiert.

Außerdem wurde, wie bei der Erläuterung der *Strukturiertheit* bereits angeklungen, der Teilaspekt der Transparenz bei der sprachlichen Verständlichkeit erhoben. Unter Transparenz wird hier das sprachliche Explizitmachen der Gewichtung einzelner Inhalte verstanden. Operationalisiert wurde das folgendermaßen: „Manchmal hat die Lehrkraft extra gesagt, dass etwas sehr wichtig ist.“[24] Wie

[24] Der Vollständigkeit halber sei erwähnt, dass dieses Item in der Publikation zum Gesamtprojekt FALKE (Schilcher et al., 2021, in Vorb.) aufgrund der schlechten Reliabilitätswerte in vielen beteiligten Fächern aus der Skala *Sprachliche Verständlichkeit* herausgenommen wurde. In FALKE-D wurde es in der Skala belassen, da die Werte dieses Items nicht ganz so problematisch (Cronbachs Alpha von 0,02 bis 0,75, s. Übersicht in Anhang 10) sind, und da es außerdem einen spezifischen Aussagewert besitzt und den inhaltlichen Überlegungen vor statistischen Ausschlusskriterien Vorrang gegeben wird. Dadurch ergeben sich für die Skala *Sprachliche Verständlichkeit* in vorliegender Arbeit naturgemäß leicht andere deskriptive und inferenzstatistische Werte als in den Ausführungen zu FALKE-D im Gesamtprojekt (vgl. Gaier & Schilcher, 2021, in Vorb. sowie die Ergebnisse der Regressionen unter Punkt *8.3* in dieser Arbeit).

bei den vorherigen Aspekten wurde auch zur sprachlichen Verständlichkeit ein Globalurteil erhoben, und zwar: „Die Lehrkraft hat sich gut verständlich ausgedrückt."
Die einzelnen Items finden sich noch einmal aufgelistet in der Übersicht in Anhang 3.1.
Zu den einzelnen Merkmalen der *Sprachlichen Verständlichkeit* und den angewendeten Kriterien der Analyse im FALKE-Projekt sei auf die entsprechende Publikation von Thim-Mabrey & Lindl (2021, in Vorb.), für ausführlichere Analysen außerdem auf Thim-Mabrey (in Vorb.) verwiesen; in letzterer finden sich detaillierte Untersuchungen zum Beispiel weiterer satzgrammatischer Merkmale (wie Schwierigkeit oder Geläufigkeit des verwendeten Wortschatzes), aber auch Analysen aus textlinguistischer Sicht (z. B. hinsichtlich der Typen inhaltlicher Textstruktur, der Erklärfrage oder der Anzahl und der Art von Verknüpfungen zwischen Haupt- und Nebentopiks).

4.2.4 Sprech- und Körperausdruck

Erst in jüngerer Zeit nimmt die Lehr-/Lernforschung die sprecherische Gestaltung der Unterrichtskommunikation in den Blick (z. B. Burger, 2018). Da im FALKE-Projekt mündliche Lehrererklärungen im Fokus der Untersuchung stehen, muss neben den genannten Aspekten die Performanz dieser Erklärung ebenfalls als Gelingensbedingung in den Fokus gerückt werden (vgl. Schopf et al., 2015, S. 47–55). Dabei ist der *Sprech- und Körperausdruck*, genau wie die sprachliche Gestaltung, gerade auch unter dem interdisziplinären Blickwinkel interessant.
Zum *Sprech- und Körperausdruck* zählen im Wesentlichen ein angemessener sprech-sprachlicher, gestischer und mimischer Ausdruck. Dazu wurden auf unterschiedlichen Ebenen Items erhoben, und zwar zunächst zur Stimme: „Die Lehrkraft hatte eine angenehm klingende Stimme"; außerdem zur Sprechweise und Artikulation: „Die Lehrkraft hatte eine deutliche Aussprache", „Die Lehrkraft hatte eine angemessene Sprechgeschwindigkeit", „Die Lehrkraft hat die Sprechpausen so gesetzt, dass ich gut folgen konnte" sowie „Die Lehrkraft hatte eine abwechslungsreiche Sprechweise". Bezogen auf die Mimik, Gestik und Körperhaltung wurde gefragt: „Die Lehrkraft hatte einen Körperausdruck, der die Erklärung unterstützt hat" und „Die Lehrkraft hat sich den Schülerinnen und Schülern zugewendet".
Schließlich wurde auch zu diesem Merkmal ein Globalurteil erfragt: „Die Art und Weise, wie die Lehrkraft spricht, gefällt mir." Alle Items zu diesem Konstrukt finden sich noch einmal aufgelistet in Anhang 3.1.
Zur genauen Konzeptualisierung des *Sprech- und Körperausdrucks*, dessen Operationalisierung und den Ergebnissen der Analyse sei auf die entsprechenden Publikationen von Gunga, Schilcher, Kranich und Gegner (2021, in Vorb.) und Gunga (2022, in Vorb.) verwiesen. Dort finden sich zum Beispiel akustische Messungen zur Sprechgeschwindigkeit, zur Pausensetzung und zu stimmlichen

Klangeigenschaften sowie auditive Analysen zum Stimmklang, zur Artikulation, aber auch Analysen zu visuell wahrnehmbaren Merkmalen des Körperausdrucks.

4.2.5 Erklärtiefe in FALKE-D

Nachdem die Kriterien für gutes Erklären dargelegt wurden, die im FALKE-Projekt fachübergreifend untersucht wurden, wird im Folgenden der Aspekt der *Erklärtiefe* erläutert. Dieser wurde fachspezifisch nur im Fach Deutsch und damit explizit für die Studie FALKE-D konzipiert und erhoben. Dazu wird das Unterrichtsqualitätsmerkmal der vertieften Auseinandersetzung und Verarbeitung auf Erklärungen angewendet.

Die mögliche Verarbeitungstiefe spielt in Arbeiten zum Erklären immer wieder eine Rolle: So legt zum Beispiel Josefy (2009) ihrer Untersuchung zu Erklärungen im Englischunterricht die Abgrenzung von *Erklären im weiteren Sinne* von *Erklären im engeren Sinne* zugrunde. Dabei umfasst ersteres sämtliche Sprachhandlungsmuster, die dazu dienen, Wissen zu vermitteln, unter diese Definition fallen somit auch Erläuterungen oder Beschreibungen (vgl. Josefy, 2009, S. 88). *Erklären im engeren Sinne* hingegen ist eine wissensvermittelnde Sprachhandlung, die darauf ausgerichtet ist, Zusammenhänge deutlich zu machen, so dass ein vertieftes Verstehen des Gegenstandes ermöglicht wird (vgl. Josefy, 2009, S. 88). Findeisen (2017) verwendet für die gleiche Unterscheidung interessanterweise den Begriff der Erklärtiefe: Sie unterscheidet Erklären von Erläutern, indem sie von einer größeren Erklärtiefe beim Erklären spricht, weil im Zuge dessen auch Hintergründe und Zusammenhänge des Gegenstands dargelegt werden (vgl. Findeisen, 2017, S. 27). In FALKE-D wird der Begriff der Erklärtiefe anders, nämlich wie folgt dargelegt, verstanden.

Der Aspekt der *Erklärtiefe* in FALKE-D lehnt sich an unterschiedliche Konzepte aus der Lehr-/Lernpsychologie sowie aus der Deutschdidaktik an. Zum einen beruht er auf der kognitiven Aktivierung, die als Qualitätsdimension von Unterricht unter Punkt *2.1 Qualität von (Deutsch-) Unterricht* in dieser Arbeit erläutert wird. Mit kognitiver Aktivierung bezeichnen Kunter und Trautwein „den intellektuellen Anforderungsgehalt im Unterricht“ (2013, S. 86). In qualitätvollem Unterricht haben die Lernenden also die Möglichkeit, sich mental aktiv mit dem Inhalt zu beschäftigten; ob sie es tatsächlich tun, hängt natürlich von ihnen selbst ab. Durch dieses „Potenzial zu kognitiver Aktivierung“ (Kunter & Trautwein, 2018, S. 86) wird nachhaltiges, vertieftes Lernen möglich gemacht. Dieses Prinzip ist somit von zwei Seiten zu beleuchten: Zunächst ist das Potenzial zu kognitiver Aktivierung im Unterricht ein Merkmal dessen Qualität, auf der Seite der Lernenden ist die kognitive Aktivierung gleichzeitig Voraussetzung für einen eigenständigen Lernprozess: „In line with the active nature of human learning, a key to successful learning and deeper understanding is the individual’s cognitive activity. Accordingly, cognitively activating teaching is

considered as an essential requirement for knowledge acquisition and comprehension" (Schnotz, 2016, S. 114)[25].

Auch Erklärungen können kognitive Aktivierung und damit vertiefte Verstehensprozesse auslösen, wenn sie entsprechend gestaltet sind:

> The various benefits of instructional explanations for learning provide insightful counterexamples to the possible misconception that instructional explanations generated by others and knowledge-construction activities carried out by learners are mutually exclusive forms of learning and instruction. Instead, they support the view that instructional explanations are a useful vehicle for engaging in meaningful learning. If they promote active processing, they can nicely dovetail with the learners' self-explanatory activities, making them a powerful tool for learning. Moreover, the discussion about the relationship between receiving instructional explanations and actively constructing knowledge on one's own parallels—more or less—current controversies about the impact of direct instructional guidance on discovery learning [...]. The observation that instructional explanations can be combined with learners' cognitive activities in fruitful ways to enhance learning is in line with recent empirical findings which show that instructional support is indeed an important ingredient in constructivist learning environments […] (Wittwer & Renkl, 2008, S. 57–58).

In FALKE-D werden all diese Erkenntnisse genutzt, indem Erklärungen mit unterschiedlich starkem Potenzial zur kognitiven Aktivierung gestaltet wurden. Dabei spielen in den Erklärungen besonders zwei der oben erläuterten Kennzeichen von kognitiver Aktivierung nach Taut und Rakoczy (2016) eine Rolle, und zwar (a) das Verknüpfen unterschiedlicher Konzepte und Ideen sowie (b) die herausfordernde Aufgabenstellung (vgl. hierzu die Ausführungen unter Punkt *2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht* zum Potenzial zur kognitiven Aktivierung bzw. die Ausführungen zu den einzelnen Erklärungen in den Punkten *5.2 - 5.4* in dieser Arbeit).

Außerdem kann mit Erklärtiefe an die oben geschilderte Funktionen- und Prinzipienfokussierung angeknüpft werden. Wittwer und Renkl (2008) betonen selbst, dass gerade dadurch ein vertieftes Verstehen sowie der Transfer des neuen Wissens auf andere Kontexte ermöglicht wird. In einigen für FALKE-D konzipierten Erklärungen werden daher, wo es möglich ist, Funktionen betont (s. Genaueres bei den Ausführungen zu den einzelnen Inhalten der Videos der Warum- und Wie-Erklärungen, Punkte *5.3 - 5.4* in dieser Arbeit).

Schließlich reiht sich der Begriff der Erklärtiefe bei FALKE-D in die Bestimmung von *Erklären*$_2$ ein, die Neumeister (2011) in Anlehnung an Fiehlers Ausführungen zum Architerm *Erklären* geprägt hat. Dabei ist *Erklären*$_1$ das Hyperonym für eine Vielzahl an wissensvermittelnden Handlungen, zum Beispiel für das Begründen und das Erläutern, aber auch für das *Erklären*$_2$ (vgl. Neumeister, 2011). Sie definiert es folgendermaßen:

> Mit *Erklären*$_2$ wird dagegen eine besondere Form der wissensvermittelnden Handlung verstanden, die durch eine größere Komplexität und Tiefe gegenüber anderen Formen der Wissensvermittlung gekennzeichnet ist. In diesem Sinn stellt das Erklären ein Einsichtigmachen verborgener

[25] Hanisch (2018) weitet eben diese doppelte Perspektivierung in ihrer Arbeit zu oben beschriebenem Dreischritt aus.

> Zusammenhänge bzw. das Verbalisieren eines Funktionszusammenhangs dar, das durch ein Mehr gegenüber den anderen wissensvermittelnden Handlungen gekennzeichnet ist (Neumeister, 2011, S. 96).

Somit steht *Erklären*$_1$ dem alltagssprachlichen Verständnis von *Erklären* nahe, *Erklären*$_2$ jedoch weist eine spezifischere Bedeutung auf, die impliziert, dass das Ziel die Vermittlung von tieferliegenden funktionalen Zusammenhängen ist (vgl. Neumeister, 2011)[26]. Neumeister (2011) legt in ihrer deutschdidaktischen Studie diese Unterscheidung der Untersuchung von realen Erklärungen im Unterricht zugrunde. Die für FALKE-D konzipierten Erklärungen lassen sich jeweils dem Erklären$_1$ beziehungsweise dem Erklären$_2$ zuordnen.

Zusammenfassend wird in dieser Arbeit – in Analogie zu Ergebnissen aus der Lehr-/Lernforschung (vgl. z. B. Dutke & Reimer, 2000; Wittwer & Renkl, 2008; Wellenreuther, 2015 sowie Findeisen, 2017) – davon ausgegangen, dass Erklärungen mit größerer Erklärtiefe zu besserem Verständnis und höherer Merkfähigkeit führen. Denn durch das Erläutern verborgener Zusammenhänge bei Erklärungen mit Erklärtiefe werden die Zuhörenden stärker kognitiv aktiviert, so dass eine aktive geistige Auseinandersetzung mit dem Erklärgegenstand bewirkt wird und das neue Wissen besser in vorhandenes integriert werden kann; sie tragen so zu einer verbesserten Merkfähigkeit bei.

Mit welchen Items der Aspekt *Erklärtiefe* konkret operationalisiert wurde, ist unter dem gesamten Punkt *7.5 Operationalisierung des Aspektes Erklärtiefe* in dieser Arbeit nachzulesen, eine Übersicht über die einzelnen Items findet sich außerdem in Anhang 3.2. Aufgrund der unterschiedlichen Erklärgegenstände war es notwendig, die Operationalisierung der *Erklärtiefe* jeweils ganz konkret für jede Erklärung anzupassen. Die Items wurden deswegen pro Video leicht unterschiedlich formuliert, obwohl sie inhaltlich immer auf das Gleiche abzielen (s. Genaueres dazu ebenfalls unter den Punkten *5.2 - 5.4 Was-/ Warum-/ Wie-Erklärungen* in dieser Arbeit).

4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen

Neben den erläuterten Aspekten zur Qualität einer Erklärung wurden in FALKE-D zwei weitere Aspekte erhoben, die keine Kriterien von Erklärqualität darstellen, aber dennoch einen Einfluss auf deren Bewertung haben beziehungsweise der inhaltlichen Validität der zu bewertenden Erklärungen dienen. Dabei wurde die *Persönlichkeitswirkung* analog zu den unter *4.2.1 Strukturiertheit* bis *4.2.4 Sprech- und Körperausdruck* in dieser Arbeit erläuterten Aspekte

[26] Hohenstein (2006) nimmt eine noch stärkere Differenzierung vor, die kurz aufgeführt werden soll. In ihrer Studie zu mündlichen Erklärungen im wissenschaftlichen Vortrag unterteilt sie den Architerm *Erklären* in *Erklären*$_1$ als performatives Verb, zum Beispiel in der Wendung *seinen Rücktritt erklären,* in *Erklären*$_2$ als sprachliches Handlungsmuster, das zum Beispiel dem Begründen nahesteht, sowie in *Erklären*$_3$ als Illokutionsverbund der mündlichen Erklärungen im wissenschaftlichen Vortrag (vgl. Hohenstein, 2006). Diese Unterteilung ist für vorliegende Studie allerdings nicht brauchbar.

fachübergreifend im Gesamtprojekt FALKE erhoben, wohingegen die *Fachliche Relevanz* nur in FALKE-D erfragt wurde.

Fachübergreifend: Persönlichkeitswirkung

Mit *Persönlichkeitswirkung* ist die Wirkung der erklärenden Person auf die/den Beurteilende/-n gemeint. Dieser Aspekt bezieht sich wie der *Sprech- und Körperausdruck* auf die Performanz der Erklärung. Aus der Sozialwissenschaft ist seit dem Doktor-Fox-Experiment von 1973 bekannt, dass ein sicheres und interessantes Auftreten der/des Vortragenden inhaltliche Defizite überdecken kann (vgl. Naftulin, Ware & Donnelly, 1973). Ebenso weiß man aus der Teacher-Enthusiasm-Forschung, dass das konkrete Verhalten sowie die Wirkung der unterrichtenden Person in Lehr-/Lernkontexten eine entscheidende Rolle für deren Bewertung spielen (vgl. Keller, Hoy, Goetz & Frenzel, 2016, S. 760–762). Um diese Wirkung in FALKE zumindest ansatzweise zu erfassen, wurden folgende Items erhoben: „Die Lehrkraft wirkte in diesem Video auf mich begeistert" und „Die Lehrkraft wirkte in diesem Video auf mich natürlich". Abgesehen von der wahrgenommenen Begeisterung und Authentizität wurde auch die empfundene Sympathie für die erklärende Person erfragt. Der Halo-Effekt (vgl. Thorndike, 1920) ist einer der häufigsten Beurteilungsfehler. Gerade bei der Einschätzung mündlich gegebener Erklärungen bei FALKE ist davon auszugehen, dass die Sympathie beziehungsweise die Antipathie gegenüber der erklärenden Person die Bewertung der Erklärungen beeinflusst. Deswegen wurde auch die Sympathie als Kontrollvariable erhoben, indem im Gesamtprojekt FALKE die Teilnehmenden gefragt wurden: „Die Lehrkraft wirkte in diesem Video auf mich sympathisch". Alle drei Items wurden als Kovariaten erhoben, in der Auswertung der Befragungsdaten stellte sich heraus, dass die drei Items Indexcharakter für eine gemeinsame Skala aufweisen, sie werden daher als reflexives Konstrukt *Persönlichkeitswirkung* behandelt; die einzelnen Items finden sich in der Übersicht in Anhang 3.1.

Fachliche Relevanz in FALKE-D

Neben der *Persönlichkeitswirkung,* die eine Kontrollvariable im Gesamtprojekt FALKE darstellt, untersucht die Studie FALKE-D die inhaltliche Validität der Erklärungen aus dem Fach Deutsch auch aus Sicht der befragten Personen. Dieser Aspekt steht nicht für die Qualität der Erklärung, sondern soll Aufschluss darüber geben, ob es gelungen ist, realistische beziehungsweise relevante Inhalte zu implementieren. Die *Fachliche Relevanz* hat somit das Ziel, zu überprüfen, ob die befragten Personen die Inhalte und Vorgehensweisen in den Erklärungen als wichtig und typisch für den Deutschunterricht ansehen oder nicht. Daher dient er der inhaltlichen Validierung.

Dazu wurde die Statusgruppe der Expertinnen und Experten Folgendes gefragt: (1) „Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt", (2) „Erklärungen wie

die eben gesehene kommen im Deutschunterricht vor“ sowie (3) „[…][27] sollte regelmäßig im Deutschunterricht verwendet werden“. Die Statusgruppe der Schüler/-innen wurde Item (1) aufgrund mangelnder fachdidaktischer Kenntnisse nicht gefragt. Auch diese Items wurden zunächst als Kovariaten erhoben. Während Item (2) direkt nach der Vertrautheit der Studienteilnehmenden mit den umgesetzten Themen und Erklärweisen fragt, beziehen sich die Items (1) und (3) eher auf Überzeugungen, die die befragten Personen im Hinblick auf die Erklärungen haben. In der Auswertung erwiesen sich diese Items dennoch als (wenn auch schwaches) reflexives Konstrukt mit Indexcharakter, so dass die Skala *Fachliche Relevanz* gebildet wurde (eine Übersicht über die Items findet sich außerdem in Anhang 3.2). Diese dient wie dargelegt nicht der Bestimmung der wahrgenommenen Erklärqualität und fließt somit nicht in die Berechnungen dieser mit ein, sondern wird deskriptiv berichtet (s. zu den deskriptiven Ergebnissen zur *Fachlichen Relevanz* Punkt *8.2.2.2 Statusgruppenspezifische Bewertungen der Aspekte* in dieser Arbeit).

Es ist festzuhalten, dass in FALKE-D insgesamt sieben Aspekte zum Erklären erfasst wurden. Davon sind fünf Merkmale guten Erklärens, nämlich die oben erläuterten Aspekte der *Strukturiertheit*, der *Adressatenorientierung*, der *Sprachlichen Verständlichkeit*, des *Sprech- und Körperausdrucks* sowie der *Erklärtiefe*. Die beiden anderen Merkmale dienen der inhaltlichen Absicherung des Gesamtdesigns. Die Ergebnisse des Merkmals *Persönlichkeitswirkung* fließen aufgrund der dargelegten Argumente in die Berechnungen zur wahrgenommenen Erklärqualität mit ein, die Ergebnisse zum Merkmal *Fachliche Relevanz* dagegen nicht.

5 Konstruktion der Videos als Stimuli

Nachdem die unterschiedlichen Dimensionen von Unterrichtsqualität, der Diskurs zur professionellen Kompetenz von Lehrkräften und deren Messung sowie die theoretischen Grundsätze zum Erklären und zur Erklärqualität dargelegt wurden, wird nun die konkrete Anlage der FALKE-D-Studie vorgestellt. Begonnen wird mit den Videovignetten, die in FALKE als Stimuli konstruiert wurden. FALKE fasst Vignette als kurzes Video auf, in dem eine Unterrichtsszene dargestellt wird, es handelt sich also nicht um Unterrichtsmitschnitte. Die genaue Definition von Vignette, die im Gesamtprojekt zugrunde liegt, findet sich in dieser Arbeit unter Punkt *7.1 Design der Studie*.

Anders als in den meisten Studien, die mit Videovignetten arbeiten (z. B. Seidel et al., 2010), geht es in FALKE nicht darum, die Kompetenz der Studienteilnehmenden zu testen. Über die in der Befragung geäußerten Einschätzungen der Studienteilnehmer/-innen zu den Vignetten sollen

[27] […] steht für das jeweilige Thema und die jeweilige Erklärvariante. Die Formulierung dieses Items wurde an jede Vignette einzeln angepasst. Eine Übersicht dazu findet sich in Anhang 3.2.

Aussagen über den Gegenstand selbst, das gute Erklären, getroffen werden. Von allen Statusgruppen wird daher fremdes Verhalten bewertet, eine eigene Involviertheit besteht nicht (vgl. dazu die Referenten-Perspektiven-Matrix, die Göllner, Praetorius, Wagner, Lenske und Fauth (2020, S. 144) anlegen, die der Einordnung von Beobachtungsitems aus unterschiedlichen Perspektiven dient und in der die eigene „Nicht-Involviertheit" aller Teilnehmenden bei FALKE-D klar erkennbar wird).

5.1 Grundsätze bei der Operationalisierung von Erklärungen in FALKE-D

In dieser Studie stehen Erklärungen im Mittelpunkt, die die Lehrkraft in Form eines kurzen Lehrervortrags – ohne Interaktion mit den Schülerinnen und Schülern – gibt. Die Kürze hat designtechnische Gründe: Die befragten Personen müssen mehrere Erklärungen nach ihrer Qualität einschätzen, damit Rückschlüsse auf allgemeine und deutschspezifische Kriterien guten Erklärens gezogen werden können; Erklärungen mit verteilten Rollen sind zeitlich deutlich länger als monologische Lehrererklärungen. Somit macht die Tatsache, dass alle sechs unterschiedlichen Vignetten mehrmals angesehen und bewertet werden sollen, eine zeitliche Beschränkung im Sinne der Zumutbarkeit für die Studienteilnehmer/-innen notwendig. Daher wurden für FALKE-D kurze, etwa dreiminütige, monologische Erklärvignetten erstellt. Diese Kürze der Vignetten entspricht außerdem dem lehr-/lernpsychologischen Kriterium der minimalistischen Gestaltung von instruktionalen Erklärungen. Minimalistische Erklärungen ohne Aus- und Abschweifungen führen laut Studien zu höherem Lernerfolg (s. Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* in dieser Arbeit).

Zum anderen haben auch methodische Gründe zur Entscheidung für das Erklären als Lehrervortrag geführt: Erstens ist *erklären,* wie unter Punkt *3.2 Erklären als Sprachhandlung* in dieser Arbeit ausgeführt, ein illokutives Verb, dessen perlokutive Wirkung, das Verstehen, nicht von vornherein in der Bedeutung impliziert ist. Da zwischen dem Erklären und dem Verstehen also sowieso zu trennen ist, wird ein mögliches Verstehen oder Nichtverstehen der Adressatinnen und Adressaten in den Vignetten nicht abgebildet. Zweitens wird in dieser Studie das Erklären als Lehrerhandlung untersucht. Somit ist es ob der Kontextreduzierung geboten, tatsächlich „nur" Erklären der Lehrkraft ohne Interaktion mit Schülerinnen und Schülern abzubilden, damit sich die Befragten ganz auf das Erklärhandeln der Lehrkraft konzentrieren können. Und schließlich sollen drittens die Studienteilnehmenden unvoreingenommen die Qualität der gesehenen Erklärung einschätzen. Bei Erklärungen mit verteilten Rollen könnten zum Beispiel positive Schülerantworten die Bewerter/-innen dazu verleiten, die Erklärung besser zu beurteilen, als sie sie tatsächlich fanden, weil die Schülerreaktion zeigt, dass die Schülerinnen und Schüler das Erklärte offensichtlich verstanden

haben. Um dies zu verhindern, wurde auf Interaktion in den Vignetten verzichtet, wohlwissend, dass die Erklärungen dadurch weniger authentisch erscheinen.
Schließlich ist angesichts der immensen Komplexität des Unterrichtens das Zerlegen dieser Gesamtkompetenz in kleine Elemente ein vielversprechender Weg, um verlässliche Ergebnisse darüber zu erhalten, was qualitätvolles Unterrichten bedeutet. Diese Ergebnisse tragen in kleinen Schritten zur Verbesserung des Unterrichts und der Lehrerausbildung bei (vgl. Leinhardt, 2001, S. 337). Auch vor diesem Hintergrund ist es sinnvoll, Erklären „nur" als Lehrerhandlung zu untersuchen.
FALKE-D wählte als Rahmenthema für alle sechs Videovignetten die *Arbeit mit einem argumentativen Text* (s. Punkt *5.1.2 Allgemeine Verortung „Arbeit mit einem argumentativen Text"* in dieser Arbeit). Für die Konstruktion der Videos waren zwei heuristische Prinzipien leitend, nämlich zum einen die Unterscheidung verschiedener Erklärprozeduren nach Klein (1987 bzw. 2009) in *erklären, was; erklären, warum* und *erklären, wie* (s. Punkt *5.1.3 Erklärprozeduren erklären, was/ warum/ wie* in dieser Arbeit), zum anderen eine systematische Variation der Erklärtiefe (s. Punkt *4.2.5 Erklärtiefe in FALKE-D* bzw. Punkt *5.1.4 Variation der Erklärtiefe* in dieser Arbeit).

5.1.1 Themenfindung

Die erste Herausforderung bei der Konstruktion der Videovignetten als Stimuli war die Themenfindung für die einzelnen Erklärungen. Die Beschaffenheit als schwach strukturierte Domäne sowie die Mannigfaltigkeit der Themen des Deutschunterrichts gehen mit einer Vielzahl an unterschiedlichen Erkläranlässen einher. Diese Erkläranlässe erfordern jeweils unterschiedliche Erklärprozeduren: Die S-Schreibung wird auf eine andere Art erklärt als die Unterscheidung von Adverb und Adverbiale; das Deuten von Metaphern wiederum auf eine andere Weise als die Entstehungsbedingungen eines Gedichts; ein bestimmtes Textmuster zu erkennen oder zu produzieren erneut völlig anders als eine bestimmte Lesestrategie anzuwenden und so weiter. Es war von daher notwendig, sich für diese Studie auf bestimmte Erkläranlässe zu beschränken, ohne mit der Auswahl andere Erkläranlässe abzuwerten beziehungsweise eine Vollständigkeit zu suggerieren, die das gesamte Erklären im Deutschunterricht abbildet (s. hierzu auch Punkt *10.2 Verallgemeinerbarkeit der Ergebnisse aus FALKE-D* in dieser Arbeit). Der erste Grundsatz, der während der Themenfindung befolgt wurde, war die Gewährleistung einer möglichst hohen Authentizität. Es sollten Gegenstände erklärt werden, die, abgesehen davon, dass sie natürlich konform mit den Bildungsstandards sowie den Lehr- beziehungsweise Bildungsplänen der deutschen Schulen sein müssen, auch im tatsächlichen Schulunterricht als erklärbedürftig angesehen werden. Es lag daher nahe, Expertinnen und Experten zu erklärbedürftigen Themen zu

befragen. Dazu wurden Seminarlehrkräfte[28] gewählt, die aufgrund ihrer Funktion, ihrer Arbeit mit Referendarinnen und Referendaren sowie ihrer eigenen Unterrichtstätigkeit als Expertinnen beziehungsweise Experten für Erklärbedarf im Deutschunterricht anzusehen sind.

Im November 2016 wurden deshalb zwei Seminarlehrerinnen (Lehramt Gymnasium bzw. Lehramt Realschule) sowie ein Seminarlehrer (Lehramt Gymnasium) an die Universität Regensburg zu einem Experteninterview in Form eines Leitfadeninterviews (vgl. Flick, 2019) in der Gruppe eingeladen, das von zwei Didaktikerinnen des Lehrstuhls für Didaktik der deutschen Sprache und Literatur[29] geleitet wurde. Gelenkt wurde das Gespräch anhand von Leitfragen, die allerdings nicht von den Interviewenden in fester Reihenfolge gestellt wurden, sondern einzeln in gedruckter Form vorlagen[30]. Die Seminarlehrkräfte wählten zunächst selbstständig Fragen aus; nicht ausgewählte Fragen wurden im Laufe des Gesprächs von den Interviewerinnen eingebracht, so dass im Verlauf auf alle Fragen eingegangen wurde[31]. Die Seminarlehrkräfte wurden zum Beispiel konkret nach Themen und Lernbereichen im Deutschunterricht gefragt, bei denen sie selbst erklären beziehungsweise bei denen ihrer Meinung nach Berufsanfänger/-innen tendenziell zu wenig erklären. Ebenso wurden sie explizit nach Bereichen gefragt, in denen selbstentdeckendes Lernen nicht funktioniert und bei denen folglich Erklärbedarf besteht.

Die Lehrkräfte benannten neben anderen Themen vor allem die Schreibdidaktik und innerhalb dieser das Interpretieren von literarischen Texten sowie das eigenständige Argumentieren. Die Lehrkräfte waren sich einig, dass hier hoher Erklärbedarf besteht, den Berufsanfänger/-innen zum Teil noch nicht erkennen.

Es wurden auch weitere Themen genannt, zum Beispiel das Erschließen der Bedeutung von Metaphern oder das Erstellen eines Schreibplanes. Interessanterweise wurde auf die Bereiche Rechtschreibung und Grammatik gar nicht eingegangen, anders als bei Gesprächen mit Studierenden, die diese beiden Themen zum Beispiel in Seminaren immer als erstes als erklärbedürftig nannten. Auch bei den in der Einleitung bereits angesprochenen informellen Gesprächen mit Kolleginnen und Kollegen aus dem aktiven Schuldienst wurden, nachdem diese zunächst oft von sich wiesen, im Deutschunterricht überhaupt zu erklären, mehrheitlich Erklärthemen aus den Bereichen Rechtschreibung oder Grammatik aufgezählt. Ohne diese Themen des Faches abzuwerten, scheint der Themenbereich „Schreiben“ aus deutschdidaktischer Perspektive doch herausfordernder und entscheidender.

[28] Seminarlehrkräfte sind in Bayern mit der Ausbildung der Referendarinnen und Referendare während der zweiten Lehrerbildungsphase betraut.
[29] D. h. die Lehrstuhlinhaberin Prof. Dr. Anita Schilcher sowie die Verfasserin dieser Arbeit.
[30] Die vollständige Auflistung der Fragen dieses Gesprächs findet sich in Anhang 4.
[31] Das Gespräch wurde audiographiert. Die Audioaufnahme liegt vor.

Mit dem Wissen um die praktische Relevanz von Erklärungen beim Lernbereich „Schreiben", die die Seminarlehrkräfte klar zum Ausdruck brachten, wurde zur Themenfindung außerdem deutschdidaktische Fachliteratur zum Thema „Erklären" herangezogen. Interessanterweise stammen alle Arbeiten, die sich mit dem Erklären im Deutschunterricht beschäftigen, aus dem Bereich der Sprachdidaktik (vgl. z. B. Berkemeier, 2009; Berkemeier & Harren, 2016; Neumeister, 2011; Spreckels & Trojahn, 2009). Literaturdidaktische Arbeiten, die das Erklären untersuchen, wurden keine gefunden, mit Ausnahme der Dissertation von Neumeister (2011), die neben Erklärungen zu grammatikalischen Kategorien und Redensarten auch Erklärungen zur Bedeutung von Metaphern untersucht. Somit kristallisierte sich im Verlauf dieser fachdidaktischen Literaturrecherche der Teilbereich der Sprachdidaktik als grober Rahmen heraus, da so sowohl das Schreiben berücksichtigt als auch an einige genuin deutschdidaktische Arbeiten (s. die oben genannten) angeknüpft werden kann.
Im FALKE-Gesamtprojekt erstellte jede Disziplin mehrere (normalerweise sechs) videographierte Erklärungen, damit zwischen den Bewertungen Vergleiche angestellt und daraus breitere Schlüsse auf die Erklärqualität gezogen werden können. Um den befragten Personen die Bearbeitung des Fragebogens zu erleichtern und um noch einmal deutlich zu machen, wie vielfältig die Erkläranlässe im Deutschunterricht selbst bei der Arbeit mit „einem" Thema sind, wurde den einzelnen Themen in FALKE-D ein gemeinsamer Rahmen gegeben.

5.1.2 Allgemeine Verortung „Arbeit mit einem argumentativen Text"

Die Arbeit mit einem argumentativen Text bot sich daher aus den genannten Gründen als Rahmenthema an: Das Argumentieren ist in Bayern im Deutschunterricht der Mittel- und Oberstufe in allen Jahrgangsstufen und Schularten genauso relevant (vgl. Staatsinstitut für Schulqualität und Bildungsforschung, LehrplanPlus) wie in den länderübergreifenden Bildungsstandards zum Hauptschulabschluss, zum mittleren Schulabschluss sowie zur allgemeinen Hochschulreife (vgl. KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2003; 2004a, 2012). Außerdem deuten die einzelnen Erkläranlässe innerhalb dieses Rahmenthemas oben erwähnte Mannigfaltigkeit der Erklärgegenstände an. Gleichzeitig konnten die oben genannten unterschiedlichen Prozeduren des *Erklärens, was, Erklärens, warum* sowie *Erklärens, wie* (s. Punkt *5.1.3 Erklärprozeduren erklären, was/ warum/ wie* in dieser Arbeit) an Themen aus der Sprachdidaktik umgesetzt werden.
Um möglichst hohe Authentizität gerade im Hinblick auf die Statusgruppe der Schüler/-innen zu gewährleisten, wurde für das Rahmenthema ein authentischer Text gewählt. Bei der Textauswahl wurde zum einen auf Schülernähe und -interesse geachtet. Auf der anderen Seite sollte der Text aus oben genannten Gründen unterschiedliche Erkläranlässe bieten.

Daher wurde ein argumentativer Text der Kolumne *Streit* des Internetauftritts der Zeitschrift *Fluter* gewählt, die von der Bundeszentrale für politische Bildung herausgegeben wird und sich mit aktuellen politischen Themen an jugendliche Leserinnen und Leser richtet (s.www.fluter.de, 01.05.20).

Das Thema, das aus der Kolumne gewählt wurde, lautet: „Sollten Schüler und Lehrer auf Facebook befreundet sein?" Die Frage, wie eng und von welcher Art der Kontakt zwischen Schülerinnen und Schülern und ihren Lehrkräften außerhalb der Schule sein soll, hat mit den Möglichkeiten der neuen Medien eine ganz neue Tragweite gewonnen. Facebook und Messenger-Dienste sind aus dem Alltag von Schülerinnen und Schülern nicht mehr wegzudenken und greifen natürlich auch in Lehrer-Schüler-Beziehungen ein. Das Thema des Rahmentextes und dessen inhaltliche Aufbereitung erfüllt grundlegende Voraussetzungen, um Argumentieren zu schulen: Ob Lehrkräfte und Schüler/-innen auf Facebook miteinander befreundet sein sollen, ist eine Fragestellung, die Schülerinnen und Schüler unmittelbar betrifft. Bei der Zielgruppe der Erklärungen, Schüler/-innen der 9./10. Jahrgangsstufe, ist davon auszugehen, dass sie über ausreichend Sachwissen zum Thema Facebook und Freundschaften auf Facebook verfügt, um sich angemessen mit dem Gegenstand auseinandersetzen zu können. Es kann ebenso davon ausgegangen werden, dass das Thema auch unter Jugendlichen kontrovers aufgenommen wird, so dass tatsächlich Argumentationsbedarf besteht (s. zu den Bedingungen guter Argumentationsaufgaben Rezat, 2011).

Ausgewählt wurde für die Studie der „pro-Text", also die befürwortende Antwort auf die Frage; er wurde von einer Lehrerin verfasst und ist im Anhang einzusehen (s. Anhang 5 beziehungsweise den Originaltext unter http://www.fluter.de/like-den-lehrer, 01.05.2020).

Streitbar ist dieser Text nicht nur aufgrund des behandelten Themas, sondern auch aufgrund seiner argumentativen Qualität, so dass drei Erklärgegenstände am Text ausgemacht werden konnten. Auf der einen Seite ist der Stil des Textes einfach gehalten. Auf der anderen Seite bietet die verwendete Redewendung „um etwas viel Aufhebens machen", die vielen Schülerinnen und Schülern selbst in der Oberstufe unbekannt ist, Anlass zu erklären, was sie bedeutet. Die argumentative Kraft des Textes überzeugt nicht durchgehend, so dass er sowohl als Anlass dafür genutzt werden kann, zu erklären, (2) warum ein bestimmtes Argument aus dem Text nicht gelungen ist. Außerdem kann ein anderes Argument aus dem Text dazu herangezogen werden zu erklären, (3) wie man ein vorliegendes Argument widerlegt (s. zu den einzelnen Erklärungen die Punkte *5.2 - 5.4* in dieser Arbeit).

5.1.3 Erklärprozeduren *erklären, was/ warum/ wie*

Alle unter *3.2 Erklären als Sprachhandlung* in dieser Arbeit dargestellten Prozeduren spielen im Deutschunterricht eine Rolle, – je nach zu erklärendem Gegenstand. Was-Erklärungen stellen

Eigenschaften des Gegenstandes dar, darunter fallen zum Beispiel Worterklärungen. Wie-Erklärungen beziehen sich auf Prozeduren wie Lesen und Schreiben. Warum-Erklärungen erläutern Ursachen und Gründe, beispielsweise für ein Rechtschreibprinzip, eine Kommasetzung oder ein gelungenes Argument.

Die Was-Erklärung und die Wie-Erklärung vorliegender Arbeit reihen sich gut in diese Definition ein. Warum-Erklärungen im Deutschunterricht sind komplexer und schwieriger zu definieren, weswegen hier ein genauerer Blick notwendig ist.

Die Warum-Erklärung bei Klein (1987) ist deutlich abzugrenzen von der oben beschriebenen Kausalerklärung aus den Naturwissenschaften (s. Punkt *3.1 (Natur-) Wissenschaftliche Sicht auf Erklärungen* in dieser Arbeit). Aus dieser sprachwissenschaftlichen Perspektive hat *erklären, warum* nichts mit kausallogischer Herleitung zu tun. Die kausale Relation gibt vielmehr einen Grund für etwas an, wobei die Beziehung zwischen dem Sachverhalt und dem Grund ganz unterschiedlicher Natur sein kann. Sie ist semantisch unspezifisch und somit nicht von vornherein klar definiert. Klein beschreibt die kausale Beziehung wie folgt: „a ist jeweils ein lexematisch markierter, als geltend angenommener Grund für c" (Klein, 1987, S. 207). Betrachtet man diese kausalen Beziehungen genauer, so stellt man fest, dass „ein Kausal-Indikator wie *weil* sowohl ein finales Verhältnis […] als auch ein instrumentales […] als auch beliebige andere Gründe-Typen […] ausdrücken kann" (Klein, 1987, S. 207). Er gibt dazu folgendes Beispiel: „Weil ich Geld verdienen will, arbeite ich [1]. Weil ich arbeite, verdiene ich Geld [2]. Weil heute Dienstag ist, arbeite ich [3]" (Klein, 1987, S. 207–208). Dabei ist die Angabe des Grundes durch *weil* in Beispiel (1) finaler Natur, in Beispiel (2) instrumentaler und in Beispiel (3) sonstiger Natur. Kausalität ist in grammatischen Systemen ein Überbegriff über eine Vielzahl an Beziehungen. In vorliegender Arbeit stellt die Warum-Erklärung keine kausal-logische Erklärung dar, sondern gibt inhaltlich-funktionale beziehungsweise strukturelle Gründe für ein misslungenes Argument an (s. Näheres unter Punkt *5.3 Warum-Erklärungen* in dieser Arbeit).

Die drei Erklärprozeduren *erklären, was/ warum/ wie* sind in FALKE-D als Lehrererklärungen umgesetzt, die aufgrund des Designs losgelöst vom Verlauf einer Unterrichtsstunde bewertet werden. Dennoch lohnt sich ein theoretischer Blick in die Dramaturgie einer Stunde, denn die Erklärungen sind möglichst authentisch gehalten. Wie unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit dargestellt wurde, stellen Erklärungen im Verlauf einer Unterrichtsstunde entweder eine Nebenkommunikation (d. h. eine lokale Sequenz) dar oder der Gegenstand der Erklärung ist selbst Thema der Stunde (d. h. globale Sequenz).

In der vorliegenden Arbeit sind nach dieser Einteilung die Was-Erklärungen zur Bedeutung der Redewendung lokale Sequenzen, die eben das Weiterarbeiten mit dem Textinhalt ermöglichen sollen. Sie stellen also eine Nebenkommunikation dar, die notwendig ist, um das Verständnis der

Schüler/-innen zu sichern und sich so dem eigentlichen Thema, nämlich dem Argumentieren, widmen zu können.

Dahingegen sind die Warum- und Wie-Erklärungen zum Schreibprozess des Argumentierens globale Sequenzen und keine Nebenkommunikation, weil die Erklärgegenstände selbst das zu Vermittelnde sind. Ähnlich wie bei den von Settekorn (1991) beschriebenen Spieleerklärungen sind diese beiden Erklärungen „globale Erklärungssequenzen […], die an einem übergeordneten Zweck ausgerichtet sind: Sie sollen es den Beteiligten (dem beziehungsweise den Erklärenden und dem (den) Adressaten der Erklärung) ermöglichen, in eine gemeinsame Handlungspraxis einzutreten" (Settekorn, 1991, S. 255). Dabei zielen die vorliegenden Warum- und Wie-Erklärungen weniger auf eine gemeinsame Handlungspraxis von erklärender Lehrkraft und den Schülerinnen und Schülern als Adressatinnen/Adressaten ab, sie sollen die Schüler/-innen vielmehr dazu zu befähigen, Argumente zu bewerten beziehungsweise zu widerlegen, damit sie an der Argumentationskultur teilhaben können (s. dazu die entsprechenden Ausführungen unter *5.3 Warum-Erklärungen* und *5.4 Wie-Erklärungen* in dieser Arbeit).

Festzuhalten ist, dass FALKE-D anhand des Rahmenthemas „Arbeit mit einem argumentativen Text" sowohl die drei Erklärprozeduren *erklären, was/ warum* und *wie* abdeckt als auch die unterschiedlichen Stellenwerte *Nebenkommunikation* beziehungsweise *Unterrichtsthema* einer Erklärung im Deutschunterricht.

5.1.4 Variation der Erklärtiefe

Da der unter Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit erläuterte Aspekt der *Erklärtiefe* im Zentrum des Forschungsinteresses von FALKE-D steht, wurde die *Erklärtiefe* nicht nur als zusätzliches Merkmal der Erklärqualität neben den anderen Merkmalen wie *Strukturiertheit, Adressatenorientierung* und so weiter aufgefasst und deren Bedeutung im Fragebogen erfragt.

In den Vignetten wird die Erklärtiefe zudem unterschiedlich umgesetzt und dient so als Variation zwischen zwei Erklärungen zum gleichen Thema. Auf diese Weise ist es möglich, nicht nur die Bewertungen zum Aspekt *Erklärtiefe* selbst in die Analysen einfließen zu lassen, sondern bei jedem abgefragten Aspekt sowie bei den Globalurteilen zu untersuchen, ob die vertiefte und die einfache Variante unterschiedlich wahrgenommen und/oder unterschiedlich bewertet werden. Dies erlaubt deutlichere Rückschlüsse auf die Bedeutung der *Erklärtiefe* für die Qualität einer Erklärung. Die Variation der Erklärtiefe in den Vignetten wird nun erläutert.

Es entstanden bei FALKE-D folglich drei Vignettenpaare, die jeweils den gleichen Gegenstand (s. Punkt *5.1.2 Allgemeine Verortung „Arbeit mit einem argumentativen Text"* in dieser Arbeit) und die gleiche Prozedur (s. Punkt *5.1.3 Erklärprozeduren erklären, was/ warum/ wie* in dieser Arbeit) aufweisen, die sich aber in ihrer Erklärtiefe unterscheiden. Konkret bedeutet das, dass die je erste

Variante, das heißt vertiefte Erklärungen (nämlich Video A-C, s. Tab. 1 *Übersicht über Videovignetten*), vertiefte Verstehensprozesse ermöglichen soll, während die zweite Variante, das heißt einfache Erklärungen (nämlich Video D-F s. Tab. 1 *Übersicht über Videovignetten*), eher oberflächliche Zusammenhänge darstellt. Auch die zweite Variante ermöglicht Verstehen, allerdings bietet sie nicht so viele Anknüpfungspunkte wie Variante eins, so dass davon auszugehen ist, dass die kognitive Aktivierung und die Merkfähigkeit bei der zweiten Erklärung geringer ausfällt[32].

Die zentrale Unterscheidung innerhalb der Vignettenvarianten liegt daher in ihrer Tiefenstruktur (s. zur Unterscheidung zwischen Sicht- und Tiefenstrukturen Punkt *2.1.2 Bedeutung von Sicht- und Tiefenstrukturen für die Qualität von Unterricht* in dieser Arbeit). Bemerkenswert ist, dass die Sichtstrukturen und die Tiefenstrukturen von Unterricht unabhängig voneinander zu sein scheinen, das heißt, dass das, was in derselben Sichtstruktur passiert, in zwei unterschiedlichen Settings Lernprozesse von ganz anderer Qualität auslösen kann (vgl. Kunter & Trautwein, 2018). Diese Erkenntnis macht sich FALKE-D zunutze, indem sich jeweils die beiden Varianten einer Erklärung in ihrer Tiefenstruktur unterscheiden: Nach Kunter und Trautwein sind beide Varianten der Erklärung in ihrer Sichtstruktur gleich, da es sich beide Male um direkte Instruktion handelt, die als Lehrervortrag vor der gesamten Klasse konzipiert wurde (vgl. Kunter & Trautwein, 2018). Die Vignettenpaare unterscheiden sich aber in ihrer Tiefenstruktur, da eben Variante eins stärkere kognitive Aktivierung hervorrufen soll als Variante zwei (vgl. Kunter & Trautwein, 2018, S. 63). Somit reihen sich die Erklärungen von FALKE-D in die Unterscheidung von *Erklären$_1$* und *Erklären$_2$* nach Neumeister (2011) ein, die ebenfalls im Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit bereits dargelegt wurde. In ihrer qualitativen Untersuchung von 17 Deutschstunden stellt Neumeister (2011) fest, dass die Mehrzahl der beobachteten Erklärungen dem *Erklären$_1$* (also dem eher alltagssprachlichen Verständnis von *Erklären*) zuzuordnen ist. Für sie besteht kein grundsätzlicher Qualitätsunterschied zwischen *Erklären$_1$* und *Erklären$_2$*, sie weist vielmehr darauf hin, dass in weiteren Studien untersucht werden müsse, von welcher Art von Erklärung Schüler/-innen mehr profitieren (vgl. Neumeister, 2011). Nach dieser Unterscheidung werden in vorliegender Studie Erklärungen, die als vertiefte Variante konzeptualisiert werden, dem *Erklären$_2$* (d. h. dem

[32] Allerdings muss die Effektivität von Erklärungen differenziert gesehen werden, wie unter Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* in dieser Arbeit bereits angeklungen. Es kommen z. B. VanLehn et al. (2003) in ihrer Studie zum Ergebnis, dass „tiefere" Erklärungen weniger effektiv waren als oberflächliche: „Given that a tutorial explanation is finally necessary, what kind of explanation should be given? It turned out in this study that explanations that were just deep enough to allow students to solve the posttest problems were more effective than deeper explanations" (VanLehn et al. (2003, S. 245)). Der Kontext und das Ziel von Erklärungen müssen also unbedingt mit bedacht werden. In der Studie von VanLehn ging es um Erklärungen von Tutoren an Tutanden, deren Effektivität in einem Posttest abgeprüft wurde. In einem derartigen Setting reichen oberflächliche Erklärungen aus und sind auch effektiver als „tiefere". Weitere Forschung ist, besonders zu Erklärungen in anderen Settings, notwendig, wie VanLehn und andere selbst betonen (vgl. VanLehn et al., 2003, S. 246).

Erklären von verborgenen Zusammenhängen) zugerechnet. Dagegen entsprechen Erklärungen, die bei FALKE-D als einfache Variante konzeptualisiert werden, dem *Erklären*$_1$ (d. h. dem Erklären oberflächlicher Zusammenhänge). Zudem wird in vorliegender Studie im Gegensatz zu Neumeister (2011) (wie oben bereits dargelegt) vermutet, dass die Varianten des Erklärens$_2$, die also ein „Mehr an Wissensvermittlung" (Neumeister, 2011, S. 65) aufweisen, günstiger für den Lernprozess sind als die Varianten des Erklärens$_1$. Diese Vermutung stützt sich auf Postulate der Lehr-/Lernforschung (s. Erläuterungen oben unter Punkt *3.3 Lernpsychologische Sicht auf Erklärungen* in dieser Arbeit bzw. Wittwer & Renkl, 2008). In vorliegender Arbeit stellt dies für Erklärungen aus dem Deutschunterricht aber nur eine Hypothese dar. In FALKE-D wird überprüft, ob die in der Hypothese als günstiger angesehenen Varianten von den Teilnehmenden auch besser eingeschätzt werden. Dabei sind die Bewertungen zur *Erklärtiefe* (als potenzieller kognitiver Aktivierungsgrad) von besonderem Interesse. Obwohl die Bewertung des Potenzials von kognitiver Aktivierung herausfordernd ist (s. obige Erläuterungen dazu unter Punkt *2.1* zu den Dimensionen der Unterrichtsqualität), ist davon auszugehen, dass dessen Bewertung innerhalb jeweils einer konkreten Erklärung als Mikroteaching-Element bei FALKE-D den Studienteilnehmenden leichter fällt als die Einschätzung bezogen auf eine ganze Unterrichtsstunde und somit verlässliche Ergebnisse darüber erzielt werden können, wie das kognitive Aktivierungspotenzial der einzelnen Erklärungen wahrgenommen wird. Auf Grundlage der Ergebnisse müssen dann weitere Untersuchungen folgen, um herauszufinden, ob als besser wahrgenommene Erklärungen tatsächlich auch zu besserem Verstehen führen. Dazu sind Verstehenstests nötig, die nach den vertieften und nach den einfachen Erklärungen eingesetzt werden müssen (s. Punkt *11.1 Anregungen aus FALKE-D für weitere Forschungsarbeiten* in dieser Arbeit).

Folgende Übersicht über die sechs implementierten Erklärvideos sowie die zugrundeliegende Variation in der Erklärtiefe dient der Illustration der Anlage der Stimuli (s. Tab. 1).

Tabelle 1: Übersicht über Videovignetten

Erklärtiefe	**(1) Erklären, was**	**(2) Erklären, warum**	**(3) Erklären, wie**
vertiefte Erklärung, d.h. Erklärung mit mehr Erklärtiefe (angelehnt an erklären$_2$)	Video A: Redewendung wird anhand des etymologischen Hintergrundes erklärt	Video B: anhand der fehlenden funktionalen Passung wird erklärt, warum ein Argument nicht gelungen ist	Video C: mit der Methode des lauten Denkens wird erklärt, wie man ein Argument widerlegt
einfache Erklärung, d.h. Erklärung mit geringerer Erklärtiefe, (angelehnt an erklären$_1$)	Video D: die Redewendung wird anhand von synonymen Ausdrücken erklärt	Video E: anhand der fehlerhaften Struktur wird erklärt, warum ein Argument nicht gelungen ist	Video F: anhand eines Modelltextes wird erklärt, wie man ein Argument widerlegt

Tab. angelehnt an Gaier & Schilcher, 2021, in Vorb.

Nachdem die Grundsätze, die bei der Ausarbeitung der Erklärungen in FALKE-D maßgeblich waren, dargelegt wurden, werden nun die konkreten Konzeptualisierungen der einzelnen Vignetten dargelegt. Die Erläuterungen erfolgen dabei geordnet nach den unterschiedlichen Erklärprozeduren *erklären, was; erklären, warum* sowie *erklären, wie,* zu denen wie dargelegt jeweils zwei Erklärungen zum gleichen Gegenstand erstellt wurden, Variante eins als vertiefte Erklärung in Anlehnung an erklären$_2$ und Variante zwei als einfache Erklärung in Anlehnung an erklären$_1$.

5.2 Was-Erklärungen

Ausgehend vom Rahmenthema „Arbeit mit einem argumentativen Text" wurden am konkreten oben beschriebenen Ausgangstext zum Thema *Sollten Schüler und Lehrer auf Facebook befreundet sein?* unterschiedliche Erklärgegenstände bestimmt. Der erste Erkläranlass, der im Text identifiziert wurde, findet sich in Zeile 48-49: „Um die Frage nach Nähe oder Distanz von Lehrern und Schülern wird meines Erachtens zu viel Aufhebens gemacht" (s. Originaltext in Anhang 5). Die Autorin verwendet die Redewendung „um etwas zu viel Aufhebens machen" an dieser Stelle, um zu verdeutlichen, dass ihrer Meinung nach dem Thema (i. e. Facebookfreundschaften zwischen Schülerinnen/Schülern und Lehrkräften) zu viel Bedeutung beigemessen wird und es dementsprechend zu viel Aufmerksamkeit erlangt. Dass hier ein echter Erkläranlass besteht, wurde in zwei 10. Klassen einer bayerischen Realschule sowie in einer 10. Klasse und in einem Q11-Kurs an einem bayerischen Gymnasium überprüft: In beiden Jahrgangsstufen war die Redewendung nur einzelnen Schülerinnen und Schülern bekannt. Der Gegenstand erfordert nach Klein (1987, 2009) eine Was-Erklärung, da die Bestandteile der Bedeutung erläutert werden. So wurde im ersten Vignettenpaar erklärt, was die Redewendung „um etwas zu viel Aufhebens machen" bedeutet.

5.2.1 Variante 1: Etymologischer Hintergrund

Die erste Variante der Was-Erklärung bezieht den etymologischen Hintergrund mit ein. Die Redewendung „um etwas viel Aufhebens machen" stammt aus dem Mittelalter. Vor einem Zweikampf führten die Ritter die Zeremonie des Aufhebens der Waffe aus, während der sie laut mit ihren Fähigkeiten und Kampferfahrungen prahlten. Auf diese Weise versuchten sie, den Gegner einzuschüchtern. So lautet die Definition im *Etymologisches Wörterbuch*: „'reißerisch in den Vordergrund stellen' […] Ursprünglich Ausdruck der Fechtersprache für das zeremonielle *Aufheben* oder *Aufgehebe* der Waffen am Anfang des Kampfes (*Praeludium*); dann teilweise übertragen auf 'Anfang', teilweise auf 'protziges Gehabe' […]" (Kluge & Seebold, 2002, S. 72). Die Definition im *Lexikon der sprichwörtlichen Redensarten* betont etwas stärker die heutige Bedeutung: „viele Worte darum machen, damit prahlen. Die Rda. stammt aus der Sprache der (Schau-) Fechter, die

vor Beginn des Kampfes ihre Degen mit umständlichen Zeremonien und prahlerischen Worten vom Boden 'aufhoben'" (Röhrich, 2003, S. 109). Aus der ursprünglich wörtlichen Bedeutung des Aufhebens der Waffe entstand so die übertragene Bedeutung des „Aufhebensmachens" im Sinne von Übertreiben. Das Sprichwort hat heute eine negative Konnotation. Diese bahnte sich dadurch an, dass es zunehmend in verneinter Version (also *nicht viel* bzw*. kein Aufhebens um etwas machen*) verwendet wurde: „Da der Waffenbrauch der Fechter als umständlich, unnötig und überholt angesehen wurde, ist die Rda. im Laufe ihrer Entwicklung immer mehr nur noch in negativer Wndg. gebraucht worden […]" (Röhrich, 2003, S. 109).

Einige Arbeiten, die sich mit dem Erklären von Redewendungen beschäftigen, betonen, dass das Erläutern des etymologischen Hintergrundes sinnvoll ist, um Anknüpfungspunkte für die neu zu erlernende Bedeutung zu schaffen: Die etymologischen Hintergründe „stellen […] gerade bei den Redensarten eine wichtige Komponente der Bedeutung dar" (Neumeister, 2011, S. 200) . In den von Neumeister (2011) untersuchten Unterrichtsstunden zu Redensarten bezieht mit einer Ausnahme keine der Lehrkräfte diesen Hintergrund mit ein. Für Neumeister ist dieses Fehlen der Grund, die im Unterricht videographierten Erklärungen dem Erklären$_1$, also einem eher oberflächlichen Erklären, zuzuordnen. Die Erklärungen mit etymologischem Hintergrund sind – durch den Mehrwert an Bedeutungen und die offen gelegten tiefer liegenden Zusammenhänge – dem Erklären$_2$ zuzurechnen (vgl. Neumeister, 2011).

In vorliegender Arbeit wird diese Variante (im Studiendesign Variante 1) der Was-Erklärung deswegen als Erklärung mit größerer Erklärtiefe angesehen. In der Erklärung wird zunächst die ursprünglich wörtliche Bedeutung aufgedeckt, denn es wurde tatsächlich etwas *aufgehoben.* Dann wird bereits der Bogen zur heutigen, übertragenen Bedeutung geschlagen: Dieses Aufheben geschah auf eine bestimmte Art und Weise, nämlich mit unnötig viel Aufwand, indem währenddessen geprahlt und angegeben wurde:

> und (.) !WÄH!rend man dann diese waffen !AUF!gehoben hat, dann hat ma die nicht !EI:N!fach nur (.) AUFgehoben, sondern schon WÄHrenddessen hat ma angefangen, (.) zu PROTZen und zu PRAHlen, den GEGner eigentlich (.) mit WORten KLEIN zu machen und zu ZEIgen, wie TOLL man is. (s. Anhang 6.1, Transkript[33] zu Vignette A; Etymologischer Hintergrund, Z. 48-59)

[33] Die Transkripte aller Videos des FALKE-Projekts wurden für die Zwecke der FALKE-Projektgruppe in einer vereinfachten GAT-Version von einer sprachwissenschaftlichen Arbeitsgruppe unter der Leitung von Christiane Thim-Mabrey erstellt (vgl. Absatz *2.5.2 Transkription und Auswertung der gesprochenen Erklärungen* in Thim-Mabrey und Lindl, 2021, in Vorb.).

In der heutigen Bedeutung ist diese Art und Weise noch präsent, allerdings ist die wörtliche Bedeutung des Aufhebens verloren: Die Redewendung wird heute für Beliebiges verwendet, für das ein zu großer Aufwand betrieben wird:

> also, man hat den !VO:R!gang des auf!HE:!bens (--) GRÖßer gemacht, WICHTich gemacht, da ne RICHtich GRO:ße SAche draus gemacht und geNAU das is=es; also, den !VO:R!gang des AUFhebens (.) sehr (.) WICHtich (.) NEH:men, sehr wichtich MACHen und !EI!gentlich muss ma SAGen; ETwas (.) !ZU! (.) WICHtich nehmen. also wurde eine !RIE:!sige aktION draus gemacht; (s. Anhang 6.1, Transkript zu Vignette A; Etymologischer Hintergrund, Z. 60-73)

In dieser Erklärung werden den Schülerinnen und Schülern die verborgenen Zusammenhänge der wörtlichen mit der heutigen übertragenen Bedeutung klar gemacht. Es wird deshalb davon ausgegangen, dass die Zuhörenden durch das Aufdecken des etymologischen Hintergrundes zahlreiche Anknüpfungspunkte für das neue Wissen erhalten und so kognitiv aktiviert werden (vgl. Kennzeichen a) des Verknüpfens verschiedener Gesichtspunkte nach Taut & Rakoczy, 2016, S. 47). Dadurch können sie sich die Redewendung besser merken (vgl. Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit).

5.2.2 Variante 2: Synonyme Ausdrücke

Als zweite Variante der Was-Erklärung wurde das Erklären über synonyme Ausdrücke gewählt.

Im Laufe des Spracherwerbs erlernen Kinder unterschiedliche Arten von semantischen Relationen, wozu auch das Bilden von Synonymen gehört (vgl. Ulrich, 2011). Im Unterricht wird diese Begriffsfindung bewusst gemacht[34], so dass davon auszugehen ist, dass den befragten Schülerinnen und Schülern dieser Studie aufgrund ihres Alters und ihrer bereits zurückliegenden Schullaufbahn (alle Schüler/-innen befinden sich in der Sekundarstufe, s. Punkt *7.4.2 Adressaten: Schüler/-innen* in dieser Arbeit) das Finden und Geben von Synonymen bekannt ist.

Im Bereich der Begriffsarbeit ist dies eine Erklärstrategie, die schon Quasthoff & Hartmann, 1982 als „in Alltagsdiskursen anzutreffende[s] Erklärungsschema" (S. 107) beschreiben. Die Besonderheit dieser Strategie liegt darin, dass zwischen dem Explanandum und dem Explanans eine „tatsächliche Äquivalenz" (ebd.) besteht, da der zu erklärende Begriff sowie das gegebene Synonym „auf der gleichen begrifflichen Hierarchieebene [liegen … und somit] die gleiche extensionale Ausdehnung" (Quasthoff & Hartmann, 1982, S. 107) abdecken; darin unterscheidet sie sich von der Strategie des Gebens von Beispielen, durch das eine hierarchische Beziehung zum Explanandum hergestellt wird

[34] Dies geschieht bei systematischer Wortschatzarbeit, wenn die Schülerinnen und Schüler aufgefordert werden, synonyme Begriffe zu einem Ausdruck zu finden, ebenso wie in dienender Funktion, z. B. im Schreibunterricht, um den Kindern beizubringen, dass ein guter Schreibstil auch durch abwechslungsreiche Wortwahl gekennzeichnet ist.

(vgl. ebd.). Eine weitere Besonderheit dieser Erklärstrategie ist, dass sie den Gegenstand, in diesem Fall die Bedeutung des Begriffs, gerade nicht in seine Einzelheiten zerlegt, obwohl das eigentlich ein typisches Vorgehen bei Erklärungen ist. Vielmehr wird die Gesamtbedeutung durch ein Synonym wiedergegeben (vgl. Quasthoff & Hartmann, 1982, S. 110). Im Datenkorpus von Quasthoff und Hartmann (1982) von 117 Alltagserklärungen war die Strategie, einen bedeutungsgleichen Ausdruck zu nennen, die zweithäufigste (nach dem Nennen von Beispielen). Vermutlich kennen die Schülerinnen und Schüler daher diese Strategie nicht nur aus dem schulischen Kontext, sondern auch aus dem außerschulischen Alltag. Auch Alber und Neumeister (2009) fanden in ihrer Studie zum Vergleich von Begriffserklärungen zwischen Erwachsenen und Kindern das Nennen von Synonymen, interessanterweise vor allem als Strategie, die Erwachsene heranziehen (Alber & Neumeister, 2009, S. 150). Im Fremdsprachenunterricht ist diese Art von Erklärung etabliert und wird regelmäßig verwendet (vgl. z. B. Macaro, 2013).

Für diese Studie wurde diese Variante als weniger „erklärtief" im Vergleich zur Erklärung mit etymologischem Hintergrund gewählt. Obwohl das Geben von Synonymen eine häufige und anerkannte Strategie des Wortschatzerwerbs und der Begriffsarbeit ist, so ist es doch im Bereich der Redewendungen sicherlich eher ungewöhnlich. Die Schwierigkeit des Verständnisses einer Redewendung liegt ja darin, dass die Gesamtbedeutung sich nicht aus den Einzelbedeutungen der Wörter erschließen lässt (vgl. Ulrich, 2011). Da für die Erklärung dieser Redewendung andere Redewendungen herangezogen werden, ist es für das Verständnis nötig, dass die Schülerinnen und Schüler mindestens eine dieser Redewendungen kennen, denn die Bedeutungen lassen sich nicht durch Zugriff auf das mentale Lexikon erschließen[35]. In der Erklärung wurde deswegen auf Redewendungen zurückgegriffen, die geläufiger sind und den Schülerinnen und Schülern vermutlich bekannt sind:

> ich verSUCH euch das jetz mal (.) MIT (.) verSCHIEdenen (.) !AN!deren REdewendungen, von denen ich GLAUB, sie sind EINfacher zu erklären. und ZWAR gibt=s DIEses (.) eine (.) MÜcke …/ von einer MÜcke zu einem eleFANten …/ also AUS einer MÜcke …// (---) (s. Anhang 6.2, Transkript zu Vignette D; Synonyme Ausdrücke, Z. 11-17)

> O::der (-) ähm(---) viel !WIR:!bel um etwas machen. auch DES is etwas, was ihr !SCHON!, glaub ich, be!NUTZT!, (s. Anhang 6.2, Transkript zu Vignette D; Synonyme Ausdrücke, Z. 22-28)

[35] Bei Einzelwörtern ist das Erschließen unter Umständen einfacher, denn die Schülerinnen und Schüler können hier zum Beispiel auf verwandte Wörter oder eventuell bekannte Begriffe aus anderen Sprachen zurückgreifen. Dennoch gilt auch hier: Da die Bedeutung des Begriffs in Gänze mit einem synonymen Ausdruck gleichgestellt wird, sind die Anknüpfungsmöglichkeiten geringer als bei anderen Erklärstrategien, wie zum Beispiel beim Zerlegen des Begriffs in einzelne Bestandteile.

Da bei keiner der synonymen Wendungen davon ausgegangen werden kann, dass sie allen Schülerinnen und Schülern bekannt ist, wird eine Vielzahl an Ausdrücken (insgesamt fünf verschiedene, s. Transkript zu Vignette D im Anhang 6.2) gegeben. Über diese Wendungen kommt die Lehrkraft letzten Endes zur Erklärung, was die ursprüngliche Redewendung bedeutet:

> wenn ich=s !GANZ! (.) EINfach versuche zu sagen, kommt !EI!gentlich ʁaus; (-) ich NEHme eine SAche zu WICHtich, ich !MES!se ihr (.) ZU viel beDEUtung bei. und die be!TO!nung liegt auf dem !ZU!, (.) also auf dieser ver!STÄR!kung. (---) (s. Anhang 6.2, Transkript zu Vignette D; Synonyme Ausdrücke, Z. 42-48)

Das Geben von synonymen Ausdrücken der Redewendung entspricht dem Erklären$_1$. Während beim Erklären des etymologischen Hintergrunds eine Bewusstwerdung über den Weg von der wörtlichen Bedeutung zur übertragenen Bedeutung stattfindet, so dass die heutige abstrakte Bedeutung fassbar wird, ist dies beim Erklären über Synonyme nicht der Fall. Hier müssen die Adressatinnen und Adressaten vielmehr eine übertragene Bedeutung mit einer anderen gleichsetzen, was weniger herausfordernd (vgl. Kennzeichen b) nach Taut & Rakoczy, 2016, S. 47) für sie ist. Es wird daher angenommen, dass die Verarbeitungstiefe und die kognitive Aktivierung in dieser Erklärung geringer ausfallen als bei Variante eins.

5.3 Warum-Erklärungen

Innerhalb des Rahmenthemas „Arbeit mit einem argumentativen Text" lässt sich ein zweiter Erklärgegenstand, nämlich die Frage, warum ein Argument aus dem Text nicht gelungen ist, identifizieren. Damit wird das Argumentieren selbst in den Fokus gerückt und der Gegenstand der Erklärung wird selbst zum Unterrichtsthema (d. h. es handelt sich bei der Erklärung um eine globale Sequenz s. Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit). Die Kompetenz des Argumentierens definieren Becker-Mrotzek und andere wie folgt: „Unter argumentativer Kompetenz soll […] ein spezifischer Ausschnitt der sprachlichen Handlungsfähigkeit verstanden werden, der mündlich wie schriftlich realisiert werden kann. Dabei geht es im Kern um die Fähigkeit, sich in angemessener Weise verbal über strittige Sachverhalte auseinanderzusetzen" (Becker-Mrotzek, Schneider & Tetling, 2010, S. 5). Diese Kompetenz ist hochkomplex (vgl. z. B. Becker-Mrotzek et al., 2010; Ludwig & Spinner, 2000; Rezat, 2011), und erfordert sowohl rezeptive als auch produktive Fähigkeiten, denn bestehende Argumente müssen verstanden und aufgenommen oder widerlegt werden und eigene Argumente müssen gefunden und formuliert werden (vgl. Becker-Mrotzek et al., 2010, S. 5–6).

Die Warum-Erklärung in FALKE-D bezieht sich auf die rezeptive Fähigkeit, die Qualität von Argumenten zu beurteilen, um anschließend entsprechend darauf reagieren zu können. Dazu müssen Schülerinnen und Schüler in einer mündlichen Diskussion fähig sein (vgl. Ludwig & Spinner,

2000). Aber auch wenn sie mit der Herausforderung konfrontiert werden, schriftlich zu einem Argument Stellung zu nehmen, müssen sie dessen Qualität zunächst bewerten, um sich dazu adäquat in ihren eigenen Texten positionieren zu können. Dies fällt Schülerinnen und Schülern schwer, es handelt sich dabei um eine äußerst anspruchsvolle Tätigkeit: „Hinzu kommen komplexere Fähigkeiten dergestalt, dass die so entstehende Argumentation reflektiert und in ihrer Geltung (neu) bewertet wird. So kann beispielsweise dem Gesprächspartner oder Autoren das Recht abgesprochen werden, bestimmte Positionen zu vertreten, oder die Gültigkeit von Argumenten bestritten werden“ (Becker-Mrotzek et al., 2010, S. 6). Insgesamt ist es Bildungsziel aller Schularten, dass Schülerinnen und Schüler jeglicher Argumentation, egal ob sie diese hören oder lesen, mit einer kritischen Grundhaltung begegnen. Diese Haltung befähigt sie im Sinne der Mündigkeit dazu, die Qualität von Argumenten einzuschätzen.

Das Analysieren und Bewerten von gelungenen und weniger gelungenen Argumenten wird deswegen immer wieder trainiert und geübt (vgl. z. B. Becker-Mrotzek et al., 2010). Lohnend ist hier der Blick auf die Textprozeduren nach Feilke (2014). In älteren Arbeiten von Feilke wurden die Begriffe *literale Prozedur* beziehungsweise *Textroutine* verwendet, inzwischen hat sich der Begriff *Textprozedur* etabliert: „Nach der Diskussion des Begriffs beim Symposion Deutschdidaktik 2012 hat sich der Terminus "Textprozedur" als am besten geeignet zur Bezeichnung des Gemeinten herausgestellt: Er hebt mit "Prozedur" die Mittlerstellung zwischen dem Prozess- und dem Produktaspekt des Schreibens hervor und macht deutlich, dass es um eine auf Texte und deren Komponenten bezogene Kompetenz geht“ (Feilke, 2014, S. 11). Was unter Textprozedur genau zu verstehen ist, hat Feilke 2012 unter dem zum damaligen Zeitpunkt gültigen Begriff *Textroutine* definiert:

> *Textroutinen* sind textkonstituierende sprachlich konfundierte literale Prozeduren, die jeweils ein textliches Handlungsschema *(Gebrauchsschema)* und eine saliente Ausdrucksform *(Routineausdruck)* semiotisch koppeln. Sie können soziale Typen von Sprachhandlungsmotiven indizieren, haben ein genrekonstitutives Potential und sind ausdruckseitig durch rekurrent kookkurrente Ausdruckskomponenten ausgezeichnet. Sie können lexikalisch als Kollokationen, syntaktisch als grammatische Konstruktionen und textlich als Makroroutinen auftreten und in vielfacher Weise ineinander eingebettet sein. Pragmatisch funktionieren sie als Kontextualisierungshinweise, die auf der Grundlage einer pars-pro-toto-Semantik ein reiches Kontextwissen für die Textrezeption und Textproduktion indizieren können (Feilke, 2012, S. 11–12).

Es sind eben diese Textprozeduren, bei denen das Lehren von Argumentationskompetenz ansetzen muss, da diese prinzipiell erlernbar sind: „Im Unterschied zum je individuellen Schreibprozess, der vielfältigen kontingenten Einflüssen ausgesetzt ist, und ebenso im Unterschied zum je einmaligen Textprodukt sind es die literalen Prozeduren, die das lehr- und lernbare Schreibwissen bilden“ (Feilke, 2012, S. 10). Es ist also sinnvoll, die Prozedur des Argumentierens in der FALKE-D-Studie zum Gegenstand des zweiten Paares von Erklärungen zu machen und dabei zunächst von der Bewertung eines fremden Arguments auszugehen. Der Weg über die häufige und vielfältige

Textrezeption ist notwendig, denn „[n]ur was Gegenstand der Rezeption und impliziten Verarbeitung war, kann auch produktiv verfügbar werden“ (Feilke, 2014, S. 26). Mit Textrezeption ist in diesem Zusammenhang nicht „nur“ das Lesen, sondern das intensive Beschäftigen und die vertiefte Auseinandersetzung sowohl mit der sprachlichen Gestaltung als auch mit dem Inhalt des Textes gemeint. Ziel ist es, den Schülerinnen und Schülern Erfahrungen zu ermöglichen, die zu Routinewissen über Texte führen.

> Da das Prozedurenwissen implizit ist, zielt die didaktische Förderung dieses Wissens zunächst auf die Erzeugung von geteilter Aufmerksamkeit für Textqualitäten. Das Sprechen über für das jeweilige Genre einschlägige Texte und deren Gestaltung ist eine wichtige Grundlage dafür, dass geteilte Aufmerksamkeit für Textprozeduren erzeugt werden kann […]. Was ist eine gute, was eine schlechte Anleitung, worin unterscheiden sich spannende von weniger spannenden Erzählungen? Eine Textprozedurendidaktik folgt einem reading-to-write-Ansatz (Feilke, 2014, S. 27).

Konkret geht es in der Warum-Erklärung daher um die Frage, inwieweit die Textprozedur des Argumentierens bei vorliegendem Beispiel erfolgreich umgesetzt ist. Deswegen wird untersucht, inwieweit es gelingt, die aufgestellte strittige These so zu begründen, dass sie unstrittig wird (s.u.); dies bezeichnet Feilke als „textliches Handlungsschema“ (Feilke, 2012, S. 11). Neben der inhaltlichen Ebene spielt dabei auch die sprachliche Formulierung eine entscheidende Rolle, da die Makrostruktur an bestimmte Routineausdrücke gekoppelt ist: „Textroutinen sind salient aufgrund rekurrenter Kookkurrenz“ (Feilke, 2012, S. 11). Das bedeutet, dass zum Beispiel das Begründen an bestimmte Formulierungen und Strukturen gebunden ist, die immer wieder auftreten, und einer/einem kompetenten Leser/-in beziehungsweise Schreiber/-in bekannt sind. Bei der Beurteilung der Qualität eines Arguments wird dieses Wissen als Bewertungskriterium herangezogen.
Dass dabei für Schülerinnen und Schüler die Notwendigkeit besteht, sich mit der Qualität von Argumenten selbstständig auseinanderzusetzen und diese eigenständig zu analysieren, wird in dieser Arbeit nicht in Frage gestellt. Das Ziel des Schreibunterrichtes ist es gerade, diese Textprozeduren zu vermitteln und dies kann nur durch Routine geschehen, das heißt durch viele Lese- und Schreibübungen: „Entscheidend ist […], unterrichtlich die Schreib- und Leseerfahrung zu ermöglichen, die den Boden für die begriffliche Aneignung der Prozeduren erst schaffen kann“ (Feilke 2014, S. 29). Entdeckendes Lernen, das Schülerinnen und Schülern diese Erfahrungen ermöglicht, ist also notwendig. Dennoch wird die Frage nach der Qualität eines Arguments hier als Erkläranlass aufgegriffen. Zum einen weil es gerade bei komplexen Aufgabenstellungen ein bekannter und verbreiteter Schülerwunsch ist, dass die Lehrkraft „einfach mal sagt, wie es ist“. Auch in den Interviews mit den Seminarlehrkräften wurde dieser Wunsch bestätigt. Entscheidend ist aber vielmehr, dass die Erklärung in FALKE-D genutzt wird, um Schülerinnen und Schülern vorzumachen, wie man ein Argument bewertet, indem die Lehrkraft diese Bewertung laut ausspricht.

Dieser zweite Erkläranlass greift wie oben dargestellt gemäß der Einteilung nach Klein (2009) eine Warum-Erklärung auf. Somit wird in der Studie nach den Was-Erklärungen der ersten beiden Vignetten eine zweite Erklärprozedur aufgenommen. Warum-Erklärungen enthalten eine Begründung, indem sie Gründe oder Ursachen des zu erklärenden Gegenstands nennen (vgl. Klein, 2009, S. 26). Im Gegensatz zum naturwissenschaftlichen Unterricht lassen sich im Deutschunterricht wie bereits dargelegt die Gründe einer Warum-Erklärung selten kausal-logisch herleiten. Dennoch benötigt eine Vielzahl an Erklärgegenständen des Deutschunterrichts das Erläutern von Gründen oder Ursachen in einer Warum-Erklärung, so auch der zweite am Text ausgemachte Gegenstand. Die Lehrkraft sollte erklären, warum folgendes Argument aus dem gewählten Text (s. Anhang 5) nicht gelungen ist:

> […] Ein Verbot [von Freundschaften zwischen Schülern und Lehrern auf Facebook] finde ich unsinnig, da man mit seinen Schülern ohnehin in irgendeiner Form in Kontakt steht. Und wer eben gerne kumpelhaft ist, soll das doch bitte sein dürfen. Was ist so schlimm daran, wenn man ein entspanntes Verhältnis zu den Menschen hat, die einen täglich umgeben? (s. Anhang 5, Zeile 59-62)

Auch dieser Erkläranlass wird auf zwei unterschiedliche Arten erklärt, nämlich in Variante eins mit größerer Erklärtiefe als in Variante zwei.

5.3.1 Variante 1: Funktional-inhaltliche Begründung

In der ersten Variante dieses Vignettenpaares wird dargelegt, dass das Argument aufgrund inhaltlicher Schwächen seine Funktion zu überzeugen nicht erfüllt. Die Erklärung stellt die Funktion eines argumentierenden Textes in den Mittelpunkt ihrer Erläuterung und setzt somit um, was von der Schreibdidaktik seit der kognitiven Wende gefordert wird: Es geht im Schreibunterricht nicht mehr darum, starre Muster und Normen einzuhalten, um ein vorher genau definiertes Produkt zu erstellen. Vielmehr soll der Schreibprozess selbst, mit seinen Teilprozessen des Planens, Schreibens und Überarbeitens im Fokus des Unterrichts stehen. Dabei ist es wichtig, dass den Schülerinnen und Schülern während des gesamten Schreibprozesses klar ist, was die Funktion des Textes ist, den sie verfassen. Das heißt auch, dass ihnen Adressat/-in und Ziel des Textes bewusst sind und sie wissen, warum und wozu sie diesen Text schreiben. Den Unterschied zwischen der früheren Produktorientierung des Aufsatzunterrichts und den Zielen des Schreibunterrichts bringt Feilke (2010) so auf den Punkt: „Wem helfen starre Muster und globale Produktvorgaben, wenn die Qualität des Produkts empirisch nachweislich von Prozesseigenschaften abhängt, etwa vom Planungs- und Überarbeitungswissen, das den Experten auszeichnet? Deshalb spielen in der Schreibdidaktik heute Funktionsbewusstheit, Planen und Überarbeiten und ein darauf bezogener

Unterricht eine zentrale Rolle“ (Feilke, 2010, S. 8). Gerade diese Funktionsbewusstheit steht im Zentrum der Warum-Erklärung in der Variante eins.

Die Funktion eines Arguments ist es, wie gesagt, zu überzeugen. Um Überzeugungskraft zu haben, ist es notwendig, die im Argument aufgestellte These zu begründen. Die Begründung muss inhaltlich in nachvollziehbarer Weise auf die These bezogen werden, Grundler spricht hier davon, dass die Begründungen aus „als gemeinsam eingeschätzte[n] Konstrukten der Welt“ (Grundler, 2011, S. 47) kommen müssen. Das Ziel ist es, laut Becker- Mrotzek und anderen, „das Strittige durch Bezug auf Unstrittiges zu klären, d.h. den Dissens in einen Konsens zu überführen; […]“ (Becker-Mrotzek et al., 2010, S. 3). Dabei wird der Konsens nicht zwangsläufig als Kompromiss verstanden; vielmehr kann am Ende auch Konsens darüber bestehen, dass man sich eben nicht einig ist und wird (vgl. Becker-Mrotzek et al., 2010, S. 8). Die Funktion der Argumente zu überzeugen bedeutet also nicht, dass der Kommunikationspartner unbedingt tatsächlich überzeugt werden muss; vielmehr kann man davon sprechen, dass gelungene Argumente potenziell überzeugungsfähig sind. Dies ist zum Beispiel der Fall, wenn zur Begründung Ergebnisse aus Umfragen oder Studien herangezogen werden.

Das Berufen auf Topoi ist ein ebenso anerkannter Argumentationstyp (vgl. Becker-Mrotzek et al., 2010, S. 8); bei vorliegendem Argument wirkt es zunächst so, also würde die Lehrerin ein Topoi, also ein anerkanntes Denkmuster, heranziehen, um ihre Meinung zu stützen, wenn sie sagt: „Was ist so schlimm daran, wenn man ein entspanntes Verhältnis zu den Menschen hat, die einen täglich umgeben?“ (s. obiges Argument). Tatsächlich entspringt diese Begründung wohl allein der persönlichen Erfahrung der Lehrkraft. Sie bezieht sich also nur scheinbar auf Bekanntes (s. obiges Argument). Bezogen auf die vorliegende Vignettenvariante bedeutet das also, dass der Erklärer zeigt, dass die Begründung der These nicht überzeugt, weil eben nicht auf gemeinsam Anerkanntes Bezug genommen wird. Um wirklich zu überzeugen, also um die eigentliche Funktion des Textes zu erfüllen, hätte die Behauptung, ein Verbot von Freundschaften auf Facebook sei „unsinnig“ (s. obiges Argument) gestützt werden müssen, zum Beispiel eben durch das Bemühen von anerkannten Normen, durch das Berufen auf Autoritäten oder das Heranziehen von Studienergebnissen (vgl. Becker-Mrotzek et al., 2010, S. 8). Die Übertragbarkeit der Erfahrungen der Lehrerin auf andere kann nicht automatisch angenommen werden[36].

[36] Becker-Mrotzek und andere listen als einen möglichen Argumentationstyp „Berufung auf Erfahrung“ (Becker-Mrotzek et al., 2010, S. 8) auf. Sie führen dann ein Beispiel (Skifahren mit Helm verringert das Verletzungsrisiko) an, das nahelegt, dass es sich dabei wohl nicht nur um eigene Erfahrung handelt, denn sie weiten es aus auf „alle meine Freunde“, was sicherlich der Verallgemeinerbarkeit dienen soll (vgl. Becker-Mrotzek, Schneider und Tetling, 2010, S. 8). Dennoch ist von dieser Art, argumentieren zu unterrichten, abzuraten, da die Schülerinnen und Schüler so dazu verleitet werden, persönliche, subjektive Erfahrungen anzuführen, die in den meisten Fällen gerade nicht zu verallgemeinern sind und zu einer sehr oberflächlichen

Bei näherer Betrachtung wird außerdem deutlich, dass die Lehrkraft in ihrer Argumentation im Gegenteil eine allgemein anerkannte besondere Art der Beziehung, nämlich das Lehrer-Schülerverhältnis, nicht berücksichtigt. Sie spricht vielmehr recht allgemein von Beziehungen zwischen Menschen, die regelmäßig Kontakt haben, lässt aber außen vor, dass die Gesellschaft und auch das System der schulischen Institutionen gewisse Erwartungen an das besondere Verhältnis zwischen Schüler/-in und Lehrkraft stellen: So handelt es sich eben nicht um eine freundschaftliche Beziehung, sondern um eine, in der die Rollen in gewissem Maße gesellschaftlich festgelegt sind: Lehrkräfte haben die Aufgabe, Schüler/-innen zu unterrichten, das heißt auch, dass sie Schüler/-innen bewerten müssen und für sie Verantwortung übernehmen. Ein „kumpelhaftes" Verhältnis, von dem die Lehrerin schreibt, verbieten eigentlich die institutionellen Erwartungen an das Rollenverständnis von professionellen Lehrkräften. Gerade diese Erwartungen sind allgemein bekannt und anerkannt und müssten daher unbedingt berücksichtigt werden. In der Vignette geht der erklärende Lehrer vor allem auf diese Tatsache ein[37]:

> soDASS wir INSgesamt SAgen können, dass ALle diese AUSsagen kaum auf das SCHUL …/ sysTEM SCHUle (.) passen, !SEHR! allgeMEIN sind, die beSONderheit des SCHULsystems nicht beINhalten und von DAher auf gar keinen fall IRgendeiner ART und WEIse überZEUgen können. (s. Anhang 6.3, Transkript zu Vignette B; Funktional-inhaltliche Begründung, Z. 122-129)

Das Ziel des Arguments ist somit nicht erreicht: Dadurch, dass die These nach wie vor strittig ist, hat das Argument keine Überzeugungskraft für die Leserin oder den Leser. Deswegen muss es als nicht gelungen bezeichnet werden.

Durch den Fokus auf die Funktion und den Inhalt des Arguments werden in dieser Erklärvariante (analog zu der Was-Erklärung in Variante 1) verborgene, nicht auf der Textoberfläche zu findende Zusammenhänge thematisiert. Diese Variante der Erklärung knüpft somit an das Merkmal der Verknüpfung unterschiedlicher Gedankengänge nach Taut und Rakoczy (2016) und die von Renkl und Wittwer (2006) geforderte Prinzipien- und Funktionenfokussierung an (vgl. Ausführungen unter *3.3 (Lern-)Psychologische Sicht auf Erklärungen* bzw. *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit).

Weise des Argumentierens führen. Allenfalls als Heranführen an das Begründen von Thesen mag diese Art zu rechtfertigen sein, aber auch in diesem Fall ist es eine wenig überzeugende Art des Argumentierens.

[37] Bei der Konzeption dieser Vignette stand neben der inhaltlichen Auseinandersetzung die Funktion von Argumenten im Vordergrund. In der Erklärung, die der erklärende Lehrer als Orientierung bekam, wird mehrmals betont, dass es beim Bewerten eines Arguments wichtig ist, sich darüber im Klaren zu sein, was das Argument bezweckt, um so beurteilen zu können, ob dieser Zweck erfüllt wird oder nicht. Beim Erstellen der Vignette stellte sich dieser Erklärgegenstand jedoch als recht komplex und anspruchsvoll für eine dreiminütige Erklärung heraus. Das hat zur Folge, dass auf die Funktion des Arguments in der Erklärung nur ganz am Ende, aber nicht in der erforderlichen Explizitheit eingegangen wird (s. Anhang 6.3, Transkript zu Vignette B; Funktional-inhaltliche Begründung). Zu den Auswirkungen, die diese Einschränkung auf die Bewertungen dieser Vignette haben kann, s. Punkt *10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse* in dieser Arbeit.

Die Frage, ob die Funktion des Überzeugens erfüllt wird, lässt sich nicht auf einen Blick beantworten oder anhand einer Merkmalsliste abhaken. Vielmehr wird die Begründungsweise im Hinblick auf die These analysiert und deren inhaltliche Passung überprüft. Es wurde bereits dargelegt, dass dies eine komplexe Tätigkeit darstellt. Diese komplexe Tätigkeit wird von der Lehrkraft vorgeführt, so dass davon ausgegangen wird, dass die Zuhörenden, um zu folgen, mitdenken müssen. Die Erklärung weist somit Erklärtiefe auf, die die Zuhörenden kognitiv aktiviert.

5.3.2 Variante 2: Strukturelle Begründung

Die zweite Variante dieses Erklärgegenstandes, nämlich die Begründung des Nicht-Gelingens des Arguments aufgrund seiner Struktur, ist eine Herangehensweise an Argumente, die die Schreibdidaktik schon länger kritisiert. Bereits 2000 bemängeln Ludwig und Spinner die starke Orientierung am Aufbau von Argumenten beziehungsweise argumentativen Texten im Schulunterricht anstelle einer inhaltlichen Auseinandersetzung damit:

> In der Schule wird allerdings oft des Guten zu viel getan und eine Explizitheit verlangt, die unnötig ist und dann unkommunikativ wirkt. Solche realitätsfernen Anforderungen entstehen zum Beispiel dann, wenn die Struktur der Erörterung nach einem starren Schema (etwa in Anlehnung an den logischen Syllogismus) erklärt wird und der Unterschied zwischen logischer Tiefenstruktur und sprachlicher Oberfläche zu wenig bedacht wird (Ludwig & Spinner, 2000, S. 22).

Auch Winkler macht darauf aufmerksam, dass diese Art, sich mit Argumenten auseinanderzusetzen, überholt ist: „In der Argumentationsforschung hat man die rein formal-strukturelle Betrachtungsweise von Argumentationen längst hinter sich gelassen“ (Winkler, 2006, S. 159). Es sei stattdessen notwendig, eine sinnvolle inhaltliche Beziehung zwischen Aussagen und ihren Begründungen herzustellen (vgl. Winkler, 2006, S. 159–160), was bei Variante eins (s. Punkt *5.3.1 Variante 1: Funktional-inhaltliche Begründung* in dieser Arbeit) umgesetzt wurde.

Allerdings findet Winkler in ihrer Untersuchung zu Aufbereitungen des Themas Argumentieren in Schulbüchern eine starke Dominanz der formal-strukturellen Betrachtung: „Im Zentrum der Unterrichtseinheiten stehen die Bausteine, aus denen eine Argumentation besteht, nicht aber die notwendige Verbindung dieser Bausteine“ (Winkler, 2006, S. 160). Obwohl die Untersuchung seit über zehn Jahren vorliegt, kann aufgrund von Unterrichtsbeobachtungen und Gesprächen mit Lehrkräften davon ausgegangen werden, dass auch im aktuellen Unterricht diese formal-strukturelle Art der Betrachtung von Argumenten noch üblich ist. Aus diesen Gründen wurde in dieser Vignette die in Schulkreisen weit verbreitete Strukturierung eines Arguments in die Bestandteile Behauptung, Begründung und Beispiel herangezogen, um zu erklären, warum das Argument nicht gelungen ist.

Die Lehrkraft leitet die Erklärung deswegen explizit mit dem Verweis auf einen sinnvollen Aufbau eines Arguments ein und wiederholt gleich zu Beginn die drei Bestandteile:

> wir HAben den SINNvollen aufbau einer geLUNgenen argumentaTION kennengelernt und ich möchte jetz einfach EUCH mal an einem !SCHLECH!ten beispiel zeigen (.) äh (.) wie=s NICHT funktionieren kann […] ahm (.) wir wiederHOlen ganz kurz; eine geLUNgene argumentaTION beSTEHT (.) norMAlerweise in EInem argument aus einer beHAUPtung – KÜRze ich kurz AB – aus einer beGRÜNdung und dann aus einer verANschaulichung, wir nennen das oft !BEI!spiel, (s. Anhang 6.4, Transkript zu Vignette E, Strukturelle Begründung, Z. 5-24)

Das Argument wird daraufhin auf diese drei Bestandteile (i. e. Behauptung, Begründung, Veranschaulichung) hin untersucht. Selbst wenn die Lehrkraft auch in dieser Erklärung auf die inhaltliche Qualität des Arguments eingeht, so stellt sie doch immer wieder die Struktur in den Vordergrund. Die Lehrkraft betont zwar die Frage, ob sich die Begründung auf die Behauptung und die Veranschaulichung auf die Begründung bezieht, führt dabei allerdings weniger den Inhalt als vielmehr das starre Muster des Arguments und die Struktur als entscheidendes Merkmal eines Arguments an:

> JETZT stellt sich aber die FRAGE; bezieht sich die beGRÜNdung auf diese beHAUPtung? denn DAS hab ich euch ja AUCH beigeBRACHT, dass die beGRÜNdung sich immer exAKT darauf beZIEHEN muss. die LAUTet hier; DA man mit seinen SCHÜlern OHNEhin in irgeneiner form in konTAKT steht. ist DAS jetz ein GRUND dafür, dass ein verbot UNsinnig ist? (-) ich würde sagen; NEIN. es beGRÜNdet das nicht, (.) (s. Anhang 6.4, Transkript zu Vignette E, Strukturelle Begründung, Z. 55-68)

Diese Variante der Warum-Erklärung weist weniger Erklärtiefe auf als die oben geschilderte funktional-inhaltliche. In der strukturellen Begründung wird ein vorgegebenes Schema als gültig für ein Argument angenommen. Es wird in der Erklärung vor allem überprüft, ob dieses Schema formal erfüllt wird. Es geht also eher um ein Abgleichen des Arguments mit einer vorgegebenen Form, denn um eine vertiefte inhaltliche Auseinandersetzung mit der Qualität des Arguments, so dass die Erklärung zwangsläufig an der Oberfläche bleibt. Die Funktion eines Arguments zu überzeugen steht in dieser Erklärung nicht im Mittelpunkt. Die Übertragungsmöglichkeiten auf andere Argumente sind daher sicherlich gering. Aber auch die kognitive Aktivierung der Zuhörenden fällt bei dieser Variante der Warum-Erklärung niedriger aus als bei Variante eins[38].

[38] Zu den Schwierigkeiten bei der Umsetzung der Vorgaben und zu den Auswirkungen, die diese Einschränkung auf die Bewertungen dieser Vignette haben kann, s. erneut Punkt *10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse* in dieser Arbeit.

5.4 Wie-Erklärungen

Neben den Erklärprozeduren des *Erklärens, was* und des *Erklärens, warum* wurde auch die Prozedur des *Erklärens, wie* in FALKE-D aufgenommen. Instruktionale Wie-Erklärungen sind Erklärungen, die auf prozesshaftes Wissen ausgerichtet sind und sich besonders häufig durch Multimodalität auszeichnen (vgl. Klein, 2009). Eine Voraussetzung für das Geben einer solchen Erklärung ist es, dass die/der Erklärende selbst über das entsprechende prozedurale Wissen verfügt (vgl. Becker-Mrotzek & Vogt, 2009, S. 143). Die Adressatinnen und Adressaten einer solchen Erklärung sollen verstehen, wie eine Handlung oder ein Prozess abläuft[39] beziehungsweise sollen in die Lage versetzt werden, diese Handlung selbst zu vollziehen; eine Unterscheidung, die auch Settekorn (1991) vornimmt:

> Hinsichtlich der in den jeweiligen Situationen gegebenen Zwecksetzungen ist noch danach zu unterscheiden, ob bei der Erklärung von Handlungen, Handlungskomplexen und Handlungszusammenhängen eine rein kognitive oder darüber hinaus eine handlungspraktische Zielsetzung verfolgt wird, d. h. ob über eine Handlung informiert wird (z. B. wie man etwas tut) oder der Adressat der Erklärung dazu befähigt werden soll, die erklärte Handlung selbst auszuführen (Settekorn, 1991, S. 257).

Eine Erklärung kann für Letzteres nur eine Grundlage sein, denn das Wissen über den genauen Ablauf einer Handlung ist klar zu trennen von der korrekten Durchführung (vgl. Keßler, 2009; Klein, 2009).

Hohenstein (2009) definiert das anders: Für sie handelt es sich nur bei ersterem um eine Wie-Erklärung; soll die/der Adressat/-in die Handlung tatsächlich durchführen, handelt es sich für sie nicht um eine Erklärung, sondern um eine Instruktion: „Für das Instruieren ist kennzeichnender Zweck, dass es den Hörer zum Selber-Handeln, d. h. zur anschließenden selbstständigen Durchführung der thematischen Handlung befähigt" (Hohenstein, 2009, S. 39).

Für diese Studie wird der Auffassung Settekorns (1991) gefolgt, und die Wie-Erklärung somit als Basis angesehen, um Wissen in Können zu überführen.

Dass Textprozeduren einen wichtigen Lerngegenstand in der Schreibdidaktik bilden, wurde im Punkt *5.3 Warum-Erklärungen* in dieser Arbeit bereits dargelegt. Inhaltlich beschäftigen sich die Wie-Erklärungen von FALKE-D mit der Fähigkeit des Konzedierens als Teil der Textprozedur Argumentieren (vgl. Abraham, Baurmann & Feilke, 2015, S. 10); somit ist auch dieser Erklärgegenstand das eigentliche Thema des Unterrichts und es handelt sich bei der Wie-Erklärung um eine globale Sequenz. Das Konzedieren, also das Einräumen und Widerlegen der Gegenposition,

[39] Klein (2009) merkt allerdings an, dass dies normalerweise nicht als ausreichend erachtet wird, sondern dass das vorrangige Ziel von Wie-Erklärungen sei, dass der/die Adressat/-in in die Lage versetzt wird, diese Handlung dann auch selbst auszuführen (vgl. Klein, 2009, S. 29-30). Es wäre zu überprüfen, ob das im schulischen Kontext auch so gilt oder ob dort nicht in den einzelnen Disziplinen, man denke zum Beispiel an Chemie oder Biologie, auch ersteres das Ziel von Wie-Erklärungen darstellt.

stellt laut Rezat ein bedeutendes Kennzeichen einer guten Argumentation dar (vgl. Rezat, 2011, S. 50–51). Beim schriftlichen Konzedieren müssen Schülerinnen und Schüler nicht nur mit Texten umgehen, sondern auch eigenständig Texte produzieren. Sie müssen den Inhalt eines Arguments verstehen und bewerten; dazu gehört, dass sie Gültiges anerkennen, aber auch einhaken und Gesagtes anhand eigener Argumente widerlegen. Es handelt sich um hochkomplexe, intertextuelle Fähigkeiten, die, wie Abraham und andere postulieren, auch im späteren Berufsleben gefordert sind (vgl. Abraham et al., 2015, S. 4–5). Ähnlich wie bei den Warum-Erklärungen wird mit den Wie-Erklärungen eine Textprozedur als Erklärgegenstand bestimmt.

Eine weitere Parallele zu den oben beschriebenen Warum-Erklärungen (s. Punkt *5.3* in dieser Arbeit) stellt das Ziel dieser Wie-Erklärungen dar. Auch bei den Erklärungen zur Frage, wie man ein Argument widerlegt, ist das Ziel letzten Endes natürlich, dass die Schülerinnen und Schüler selbst in der Lage sein sollen, dies in ihren eigenen Texten überzeugend zu tun. Sie sollen also Können erwerben, wofür ihnen das vermittelte Wissen als Basis dienen soll. Dass dies nicht innerhalb einer dreiminütigen Erklärung und nicht ohne eigenständiges Üben geschehen kann, bedarf keiner Erläuterung. Ähnlich wie bei den Warum-Erklärungen kann auch bei diesen beiden monologischen Lehrererklärungen nur ein Schritt im Lernprozess präsentiert und untersucht werden. Die Wie-Erklärungen haben daher zum Ziel, den Adressatinnen und Adressaten Wissen darüber zu vermitteln, wie der Schreibprozess der Widerlegung abläuft (Variante 1) beziehungsweise abgelaufen ist (Variante 2). Aufbauend auf diesem *Wissen* sollen sie – nach weiteren Erklärungen und vor allem nach eigenständigem Schreiben – das schriftliche Widerlegen von Argumenten immer besser *können.*

Im Gegensatz zu obigen Erklärungen, bei denen eher eine rezeptive Fähigkeit vermittelt wurde, steht in den folgenden beiden Vignetten mit der Fragestellung *Wie widerlege ich ein Argument?* eine produktive Fähigkeit im Vordergrund. Somit ist dieser Erklärgegenstand der anspruchsvollste, der in dieser Studie untersucht wurde. Es ist daher logische Folge, dass die Wie-Erklärung im Fragebogen erst an dritter Stelle kommt: Die Erklärungen sind aus Gründen der Forschungslogik vom einfachsten Thema (Was-Erklärung zur Redewendung) zum komplexesten (Wie-Erklärung zur Textprozedur des Konzedierens) angeordnet.

5.4.1 Variante 1: Methode Lautes Denken

Die erste Variante der Erklärung, wie man ein Argument widerlegt, wird über die Methode lautes Denken umgesetzt. Diese Methode ist im deutschdidaktischen Forschungsdiskurs vor allem aus der Lesestrategievermittlung bekannt (vgl. z. B. Schoenbach, 2006, S. 94–96); sie hat sich als sehr günstig erwiesen, um Schülerinnen und Schülern das Anwenden von Strategien vor, während und nach dem Lesen zu verdeutlichen.

Ursprünglich ist das laute Denken eine Methode für den eigenen Problemlöseprozess. In der Lehr-/Lernpsychologie wird davon ausgegangen, dass durch das laute Aussprechen der Gedanken diese wiederum beeinflusst werden. So definiert Fleischer in seiner Dissertation das laute Denken wie folgt:

> Die verbale Arbeitsweise ist eine Form der Bearbeitung von Aufgaben, bei der alle Bewußtseinsinhalte, wie Gedanken, Gefühle, Vorstellungen und Wahrnehmungen, die sich während des Lösungsprozesses einstellen, in hörbare Worte gesetzt werden. Von der monologisch-verbalen Arbeitsweise […] soll dann gesprochen werden, wenn die lautsprachliche Äußerung der Bewußtseinsinhalte nicht an einen Zuhörer, sondern an sich selbst gerichtet ist. Für dieses Verfahren, das auch als Methode des lauten Denkens […] oder Methode des hörbaren Selbstgesprächs bezeichnet wird, ist wesentlich, daß die in hörbare Worte gefaßten (monologisch-verbalisierten) Bewußtseinsinhalte zeitgleich zum Denkprozeß wiedergegeben werden, um eine direkte Beeinflussung des Denkens durch das Sprechen zu ermöglichen. Es ist jedoch davon auszugehen, daß ein Verbalisierer bei der monologisch-verbalen Bearbeitung von Aufgaben nicht alle seine Bewußtseinsinhalte in Worte faßt (Fleischer, 1987, S. 10).

In Lehr-/Lernsituationen werden diese „in hörbare Worte gefaßten […] Bewußtseinsinhalte" (Fleischer, 1987, S. 10) genutzt, um Laien Problemlösestrategien zugänglich zu machen, die normalerweise verborgen ablaufen, da sie einen mehr oder weniger automatisierten Denkprozess darstellen. Fleischer erwähnt diese Möglichkeit bereits:

> Die Förderung des Denkens durch das hörbare Selbstgespräch erfährt eine Steigerung […], wenn die Rede mit sich selbst zur Rede vor anderen wird. Das soll nicht heißen, daß ein Gespräch geführt wird, sondern, daß über etwas vor anderen gesprochen wird, die Rede also kommunikativ sinnvoll wird (Fleischer, 1987, S. 10).

Für Kiel gehört diese Methode zum „Erklären als Entwickeln von Wissen" (Kiel, 1999, S. 160), das sich für das Aneignen von „[d]omänenspezifischen Handlungen in Problemlösungskontexten" (Kiel, 1999, S. 214) eignet. Für das Entwickeln von Problemlösungskompetenz erklärt er es gar zur Notwendigkeit, dass Novizinnen und Novizen begreifen, „was in den Köpfen der Experten an Denkprozessen vor sich geht" (Kiel, 1999, S. 205). Dazu muss die/der Lehrende „dafür sorgen, daß das, was normalerweise nicht sichtbar passiert, möglichst sichtbar gemacht wird" (Kiel, 1999, S. 207), wobei die Prozesse mit dieser Methode genauer gesagt vor allem hörbar gemacht werden. Dies begründet Kiel (1999) unter anderem damit, dass Expertinnen und Experten Probleme bereits ganz anders – nämlich strukturierter – wahrnehmen und Lösungswege oft automatisiert abrufen können. Dadurch fällt es ihnen aber schwer, dieses Wissen Novizinnen und Novizen, die diese Strukturen noch nicht besitzen, zugänglich zu machen (vgl. Kiel, 1999, S. 205 bzw. die Ausführungen zum „Hindernis" des zu hohen Fachwissens unter Punkt *3.3* in dieser Arbeit)[40]. Das Modellieren des eigenen Problemlösewegs kann hier ein Ausweg sein, da es zur Bewusstmachung

[40] Kiel (1999) erläutert in diesem Teilkapitel *cognitive apprenticeship* als eine Art des Erklärens. Der erste Schritt, das Modellieren, deckt sich mit der Methode, die die Lehrkraft in dieser Videovignette C von FALKE-D anwendet. Die weiteren Schritte des *cognitive apprenticeship* spielen in vorliegender Studie keine Rolle mehr.

desselben führt, – und zwar bei Expertinnen/Experten und bei Novizinnen/Novizen. Über die laut ausgesprochene Metakognition der Expertin/des Experten sollen die Novizinnen und Novizen in die Lage versetzt werden, ihren Problemlöseweg dem der Expertinnen/Experten anzupassen und ihn metakognitiv zu überwachen und letztendlich selbstständig anzuwenden (vgl. Kiel, 1999, S. 205–207).

In der Wie-Erklärung (Variante 1) wird die Methode auf die Schreibdidaktik übertragen. Obwohl sie in der deutschen Schreibdidaktik nicht den Stellenwert hat, den sie in der Lesedidaktik einnimmt, gibt es Ansätze, die darauf hinweisen, dass Lernende auch beim Schreiben von der Methode profitieren können. Hierfür lohnt sich zunächst ein Blick in die angelsächsische Forschung, in der diese Methodik auf dem Gebiet des Schreibens bereits eine längere Tradition hat. So arbeiten bereits Scardamalia und Bereiter (1983) in ihrem Schreiblernprogramm mit sogenannten *prompts*, die im Sinne des Scaffoldings ein Gerüst zur Verfügung stellen, anhand dessen die Schülerinnen und Schüler den Prozess des Schreibens in kleine, bewältigbare Schritte zerlegen. Nach und nach werden die Hilfestellungen nicht mehr gebraucht, das Gerüst also „abgebaut", weil die Schülerinnen und Schüler es nicht mehr benötigen. Der erste Schritt dieses Programms besteht im Modellieren der Phase des Schreibprozesses, die gerade gelehrt wird (z. B. Planungsphase). Modellieren wird in der angelsächsischen Forschung wie folgt definiert:

> *Modeling* involves an expert's carrying out a task so that students can observe and build a conceptual model of the processes that are required to accomplish the task. In cognitive domains, this requires the externalization of usually internal (cognitive) processes and activities - specifically, the heuristics and control processes by which experts make use of basic conceptual and procedural knowledge (Collins, Brown & Newman, 1989, S. 481).

Zu Beginn des Lernprozesses modelliert die Lehrkraft die Bearbeitung dieser *prompts,* indem sie ausspricht, was ihr während der Bearbeitung durch den Kopf geht. Scardamalia und Bereiter (1983) wenden diese Technik in ihrem CDO (i. e. Compare, Diagnose, Operate)-Modell an (vgl. Scardamalia & Bereiter, 1983, S. 73–75).[41] Collins und andere fassen die didaktischen Vorteile dieser Methode wie folgt zusammen:

> First of all, […] their methods help students build a new conception of the writing process. Students initially consider writing a linear process of knowledge telling. By explicitly modeling and scaffolding expert processes, they are providing students with a new model of writing that involves planning and revising (Collins et al., 1989, S. 468).

Auch Graham und Perin (2007) und Harris, Graham, Mason und Friedlaender (2010) wenden diese Methode an; in ihren vielbeachteten Arbeiten zu wirksamen Schreibstrategien wird das

[41] In Studien zu ihrem CDO-Programm lassen Scardamalia & Bereiter (1983) außerdem auch Schülerinnen und Schüler die Methode des lauten Denkens anwenden, damit sowohl sie selbst als auch die Lehrkraft bzw. die Forscher/-innen Einblick in den (eigenen) Denkprozess bekommen. Diese Einblicke eignen sich für die Forscher/-innen zur Überprüfung des Schreibprogramms, für die Lehrkraft zur Diagnose und für die lernenden Schülerinnen und Schüler zur Metakognition (vgl. Scardamalia & Bereiter, 1983, S. 89-91).

Schreiblernprogramm SRSD (self-regulated strategy development) vorgestellt, das in Schulen der USA weite Verbreitung gefunden hat. Die einzelnen Schritte dieses Programms sehen wie folgt aus:

> Develop Background Knowledge: Students are taught any background knowledge needed to use the strategy successfully.
> Describe It: The strategy as well as its purpose and benefits is described and discussed.
> Model It: The teacher models how to use the strategy.
> Memorize It: The student memorizes the steps of the strategy and any accompanying mnemonic.
> Support It: The teacher supports or scaffolds student mastery of the strategy.
> Independent Use: Students use the strategy with few or no supports (Graham & Perin, 2007, S. 15).

Im Verlauf dieses Programms, genauer gesagt beim Schritt des Modellierens, wenden sie die Technik „think aloud" (Harris et al., 2010, S. 15–16) an. Sie beschreiben dort die Methode des lauten Denkens als Schritt im guten Schreibunterricht. Das laute Denken wird als dritter (von insgesamt sechs Schritten) nach dem Erarbeiten von Hintergrundwissen zu sowie dem Reflektieren und Diskutieren der Bedeutung von Schreib- und Selbstregulationsstrategien angewendet (vgl. Harris et al., 2010, S. 14–16). Dabei schlagen sie vor, über die Methode lautes Denken sowohl Schreibstrategien als auch Selbstregulationsstrategien zunächst von Lehrkräften beziehungsweise starken Schreiberinnen und Schreibern vorführen zu lassen. So wird die Aufmerksamkeit der Schülerinnen und Schüler auf konkrete Teilschritte, zum Beispiel das Planen, gelenkt. Durch lautes Denken erfahren sie, was einer guten Schreiberin/einem guten Schreiber während dieser Phase durch den Kopf geht, und – durch die Selbstregulationsstrategien – wie sie oder er sich selbst überwacht. Besonders wichtig ist dabei, dass auch „gezeigt" wird, wie mit Problemen umgegangen wird, die während des Prozesses auftauchen können. Außerdem wird dazu geraten, die Schülerinnen und Schüler in den Prozess des lauten Denkens miteinzubeziehen, indem sie zum Beispiel eigene Ideen bereits während des Vormachens äußern sollen. Anschließend wird der Prozess mit den Schülerinnen und Schülern reflektiert. Das langfristige Ziel ist es, sie dazu zu befähigen, ihren eigenen Schreibprozess sinnvoll zu strukturieren, durchzuführen und zu überwachen (vgl. Harris et al., 2010, S. 15–19)[42].

Im deutschdidaktischen Diskurs gibt es zum lauten Denken in der Schreibdidaktik erste Ansätze. Es wird in diesem Zusammenhang meist von *Modellieren* gesprochen (vgl. z. B. Philipp, 2014; Sturm, 16.09.15), das Sturm als „didaktisiertes lautes Denken" (Sturm, 16.09.15, Folie 17) definiert. Sie macht so den Unterschied deutlich zwischen dem lauten Denken als Problemlösungsstrategie für das eigene Lernen (s.o.) und dem lauten Denken als didaktische Methode, die das Lernen anderer unterstützt. Das Modellieren wird von der Lehrkraft dergestalt vorbereitet, dass in den Prozess absichtlich Fehler oder Problemstellen eingearbeitet werden (vgl. Sturm, 16.09.15, Folie 17). Auf

[42] Harris et al. (2010) weisen darauf hin, dass es auch möglich ist, den Prozess des lauten Denkens zu audio- oder videographieren und die Aufnahme im Unterricht vorzuspielen (vgl. S. 16).

diese Weise wird ein (erhöhter) Lerneffekt für Schülerinnen und Schüler erzielt. Voraussetzung für ein wirkungsvolles Einsetzen des Modellierens ist es, dass die Lehrkraft das Vorwissen und den Lernstand ihrer Schüler/-innen kennt und den Prozess entsprechend gestaltet (vgl. Sturm, 16.09.15). Die Vorteile des Modellierens liegen auf der Hand, und zwar für alle Schülerinnen und Schüler: „Modellieren verschafft Einblick in Verborgenes (Beobachten = Lernen)" (Sturm, 16.09.15, Folie 19); das Vormachen unterstützt schwächere Schreiber/-innen und eröffnet stärkeren neue Varianten (vgl. Sturm, 16.09.15, Folie 19).

Auch Philipp (2014) unterscheidet zwischen Modellieren und lautem Denken. Das Modellieren hat für ihn zum Ziel, Schülerinnen und Schülern „Einblicke in die Anwendung von Schreibstrategien" (Philipp, 2014, S. 28) zu ermöglichen. So wird das, was Expertenschreibern durch den Kopf geht, explizit hörbar gemacht. Dies stößt einen Lernprozess bei den Schülerinnen und Schülern an. Die Lehrkraft geht dabei didaktisch geplant vor, indem sie – wie bei Sturm – bereits im Vorfeld Probleme identifiziert und in den Prozess einbaut beziehungsweise ihr Vorgehen begründet und reflektiert. Das laute Denken hingegen gibt laut Philipp (2014) vor allem Aufschluss über die Denkprozesse der Person. Das eigene Vorgehen aber wird nicht reflektiert oder begründet, da mit dem lauten Denken nach dieser Definition kein didaktischer Anspruch verbunden ist (vgl. Philipp, 2014, S. 29).

Nach den oben zitierten Auffassungen kommt der Begriff des Modellierens der Umsetzung in der Videovignette C bei FALKE-D deutlich näher als der Begriff des lauten Denkens: So dient das laute Denken in der Erklärung selbstverständlich nicht dazu, den eigenen Lösungsprozess der Lehrkraft zu verbessern (s.o.), sondern dem „Hörbarmachen" des Lösungsprozesses während der Widerlegung eines Arguments. Die Methode wird angewendet, um den Schülerinnen und Schülern das Vorgehen einer Expertin deutlich zu machen, analog zur Strategieanwendung in der Lesedidaktik. Die Erklärung in der Videovignette ist selbstverständlich von der Lehrkraft gedanklich vorbereitet und verfolgt natürlich didaktische Zielsetzungen; es handelt sich also mit Sturm tatsächlich um „didaktisiertes lautes Denken" (Sturm, 16.09.15, Folie 17). Dennoch wurde in vorliegender Studie der Begriff *lautes Denken* verwendet, da davon auszugehen ist, dass er – im Gegensatz zu *Modellieren* – zumindest den Expertinnen und Experten (also den Didaktikerinnen und Didaktikern, Lehrkräften und Studierenden) bekannt ist. Da die Forschungsfrage bei den Wie-Erklärungen genau auf die unterschiedlichen Methoden (lautes Denken vs. Modelltext) abzielt und der Begriff sowohl in der Erklärung (s. unten stehendes Zitat aus Transkript zu Vignette C) als auch in den Items (s. Anhang 3.2) auftaucht, war es notwendig, einen Begriff zu wählen, der den Studienteilnehmenden bekannt ist, auch wenn dieser – je nach Forschungsauffassung – nicht genau das trifft, was im Video gezeigt wird.

Ganz konkret verfolgt die Lehrkraft mit der Wie-Erklärung das Ziel, prozedurales Wissen zu vermitteln. Die Lehrkraft knüpft damit an das deklarative Wissen an, das Schülerinnen und Schüler

über das Widerlegen eines Gegenarguments haben, also zum Beispiel, dass man mit einer Einräumung beginnt und dann einen konkreten Inhalt auswählt, an dem man Widerspruch anbringt. Dieses Wissen reicht jedoch nicht aus, die Schülerinnen und Schüler dazu zu befähigen, dies auch tatsächlich überzeugend umzusetzen; deswegen macht die Lehrkraft implizite Denkprozesse über das laute Denken explizit (zur Unterscheidung zwischen prozeduralem/implizitem sowie deklarativem/explizitem Wissen vgl. Keßler, 2009).

Dazu entwickelt die Lehrkraft mit dieser Methode konkret einen Schreibplan zur Widerlegung des Arguments, den sie stichpunktartig und vorstrukturiert an der Tafel festhält. Während der dreiminütigen Sequenz in der Vignette können die Schülerinnen und Schüler also den Gedankengängen der Lehrkraft bei der Planung der Widerlegung folgen. Zu Beginn kündigt sie an, was sie nun tun wird:

> ich möcht=euch gern heut ZEIgen, wie man ein arguMENT der GEgenseite, des einen NICHT besonders überZEUGT, widerLE:gen kann. […] um (-) zu ner widerLEgung zu kommen, muss ich mir des arguMENT nochmal genau ANschauen (.) und darüber NACHdenken, (.) was ich daran GUT finde und wo ich mit meim WIderspruch anknüpfen KANN. ne? und !DAS! mach ich euch jetzt VOR, indem ich !LAUT! denke, sodass=ihr geNAU mitkriegt, WIE meine geDANken ABlaufen. (-) (s. Anhang 6.5, Transkript zu Vignette C, Methode lautes Denken, Z. 1-22)

Im Folgenden spricht die Lehrkraft ihre Gedanken zum Inhalt des vorliegenden Arguments laut aus, zum Beispiel:

> und bin (.) DA:mit eigentlich ganz zuFRIEden. hier wird geSAGT, dass (.) äh VIEle kollegen heutzutage nachmittags noch konTAKT pflegen mit den SCHÜlern über FACEbook (.) und das STIMMT auch mit meiner erFAHrung zuSAMmen (.) und !LIEGT! wahrscheinlich WIRKlich daran, dass die kommunikaTION relativ EINfach funktioNIERT über die sozialen NETZwerke.(.) da HAB ich eigentlich nix daGE:gen. des (.) find=ich, ka=man so STEHen lassen. (s. Anhang 6.5, Transkript zu Vignette C, Methode lautes Denken, Z. 25-39)

Somit wird, wie oben dargelegt, Implizites in Explizites verwandelt. Auf der anderen Seite spricht die Lehrkraft aber auch aus, was sie genau tut beziehungsweise tun wird; dabei handelt es sich um metakommunikative beziehungsweise selbstregulierende Passagen, die den didaktischen Anspruch ganz deutlich hervortreten lassen. Dies ist zum Beispiel bei folgenden Ausschnitten der Fall:

> ich LES=mir die ersten ZEIlen nochmal durch (s. Anhang 6.5, Transkript zu Vignette C, Methode lautes Denken, Z. 24)

> also noTIER=ich mir mal (-) in meinem PLA:N für die wiger …/ widerLE:gung den geDANken; verEINfachung der kommunikaTION. (--) und wei=ich des (.) oKAY finde, setz=ich daHINter n gelben HAken (.) (s. Anhang 6.5, Transkript zu Vignette C, Methode lautes Denken, Z. 40-44)

Passagen, in denen die Lehrerin inhaltlich argumentiert beziehungsweise metakommunikativ das eigene Verhalten kommentiert, wechseln sich über die gesamte Erklärung ab.

Parallel dazu werden die Überlegungen unterstützt durch die gleichzeitige Entwicklung des Schreibplans an der Tafel. Die Lehrkraft arbeitet dazu mit zwei Farben und notiert mit weißer Kreide die Stichpunkte des Originalarguments, auf die sie eingehen möchte, und mit gelber Kreide die eigenen Erwiderungen darauf. Auch dies kommentiert sie zum Teil:

> ich setz=a=mal a FRAgezeichen drüber, (.) (s. Anhang 6.5, Transkript zu Vignette C, Methode lautes Denken, Z. 78)

> und (-) erGÄNz (.) des mal HIER (.) als ÜBERschrift. (---) (s. Anhang 6.5, Transkript zu Vignette C, Methode lautes Denken, Z. 112-113)

Wie aus den Ausführungen deutlich wird, handelt es sich bei diesem Vorgehen um mehr als bloßes Vormachen. Die Widerlegung des Arguments beruht auf komplexen Zusammenhängen, eben dem Aufnehmen der Argumentation des Gegenübers und der Reaktion darauf, die aus Zugeständnissen und Widerlegung besteht. Beides verdeutlicht die Lehrkraft multimodal, indem sie kommentierend einen Schreibplan an der Tafel entwickelt. Die komplexen Denkprozesse werden so explizit und verständlich gemacht. Gerade dieses Verdeutlichen komplexer Zusammenhänge über die Methode des lauten Denkens (mit den Besonderheiten in der Verwirklichung, die oben beschrieben wurden) führt dazu, dass die in der Vignette gezeigte Methode sich gut einreiht in die oben beschriebenen Vorgehensweisen der Schreibdidaktik nach Scardamalia und Bereiter (1983), Harris und anderen (2010) sowie Sturm (2015). Die metakommunikativen Kommentare zeigen den Problemlöseprozess der Lehrkraft, so dass den Schülerinnen und Schülern über diese Erklärung als „Entwickeln von Wissen“ (Kiel, 1999, S. 160) der Prozess des Widerlegens zugänglich gemacht wird. Durch dieses Verdeutlichen komplexer und normalerweise verborgener Zusammenhänge (vgl. Merkmal a) und b) der kognitiven Aktivierung nach Taut und Rakoczy, 2016) wird diese Wie-Erklärung dem Erklären mit größerer Erklärtiefe beziehungsweise dem Erklären$_2$ zugeordnet. Das kognitive Aktivierungspotenzial wird als hoch eingeschätzt, da den Zuhörenden verdeutlicht wird, wie sich Expertinnen und Experten dem komplexen Vorgang des Widerlegens eines Arguments stellen. Es wird in der Erklärung nicht „bloß“ eine Widerlegung präsentiert, sondern die Zusammenhänge und Denkvorgänge im Zuge des Prozesses der Widerlegung werden aufgezeigt und die Funktion (vgl. Renkl & Wittwer, 2006) des Widerlegens steht im Fokus der Erklärung.

Dieser Erklärung wird als Variante zwei eine Erklärung mit Hilfe eines Modelltextes gegenübergestellt.

5.4.2 Variante 2: Methode Modelltext

Die zweite Variante der Wie-Erklärung unterscheidet sich deutlich von der Methode lautes Denken. Die Lehrkraft präsentiert hier nämlich einen vorgefertigten, selbst verfassten Text als Muster für eine Widerlegung, den sie den Schülerinnen und Schülern zeigt und den sie kommentiert. Im Gegensatz also zu Variante eins, bei der den Lernenden normalerweise verborgene Gedankengänge während der Planung einer Widerlegung zugänglich gemacht werden, wird ihnen in dieser Vignette eine fertige Widerlegung (s. Anhang 7) vorgestellt. Das Arbeiten mit konkreten Beispielen zu komplexen Aufgaben wie Schreibaufgaben fordert zum Beispiel Wellenreuther (2015). Der Vorteil liegt für ihn darin, dass Schülerinnen und Schüler über ein konkretes Beispiel abstrakte Aufgaben besser nachvollziehen können (vgl. Wellenreuther, 2015, S. 183–184). Das Ziel der zweiten Variante ist ebenso wie in der ersten, dass die Adressatinnen und Adressaten die Textprozedur des Widerlegens erlernen sollen.

Die Arbeit mit Modelltexten ist dabei in der Deutschdidaktik nicht neu, allerdings findet sie sich besonders in einem etwas anderen Zusammenhang: Gerade der aktuelle Diskurs misst dem materialgestützten Schreiben, das in der neuen Generation der Lehrpläne in Bayern auch eine wichtige Prüfungsform darstellt, einen hohen Stellenwert zu. In diesem Zusammenhang ist es nötig, mit fremden, normalerweise authentischen Texten zu arbeiten, die zum eigenen Schreiben genutzt werden sollen. Abraham und andere (2015) stellen dazu in ihrem Basisartikel aus Praxis Deutsch 251 fest: „Die didaktischen Herausforderungen und Potentiale betreffen Schreiben und Lesen in allen Schulstufen“ (Abraham et al., 2015, S. 4). Auch in den anderen deutschsprachigen Ländern hat das materialgestützte Schreiben an Bedeutung gewonnen. So fordern zum Beispiel Witschel & Saxalber (2012) im Zuge ihrer Ausführungen zu den neuen Anforderungen der österreichischen Reifeprüfung „[…] die Vielfalt der Textsorten auch im Unterrichtsangebot zu nutzen. In didaktischer Hinsicht bedeutet dies, dass den SchülerInnen […] sowohl Kenntnisse zu textsortenspezifischen Merkmalen vermittelt werden, sie diese an Mustertexten reflektieren und sich mit Hilfe von Textgestaltungsübungen in der Anwendung üben sollen“ (Witschel & Saxalber, 2012, S. 82). Die didaktischen Herausforderungen sind dazu wie gesagt vor allem in der Verbindung von Lesen und Schreiben zu sehen; es geht im Kern um die Kompetenz, fremde Texte nutzbar für die Produktion eigener Texte zu machen. Auch wenn in vorliegender Arbeit ein fremder Text als Ausgang gewählt wird, liegt der Fokus nicht auf dem materialgestützten Schreiben; in der Vignette wird vielmehr ein Text präsentiert, der als mögliche Lösung zu verstehen ist. Die didaktische Methode ist an dieser Stelle das Lernen am Modell (s. dazu auch die Ausführungen von Lehnen (2012) zu Textprozeduren).

Das Präsentieren von Modelltexten ist wie gerade ausgeführt der Verbindung von Lesen und Schreiben (d. h. mit Feilke (2014) dem reading-to-write-Ansatz) zuzuordnen; diese Methode spielt gerade beim Vermitteln von Textprozeduren eine wichtige Rolle: „Die unterrichtliche Unterstützung des Erwerbs [von Textprozeduren] wird durch eine enge Verbindung von Lesen und Schreiben begünstigt, sie profitiert von der Reflexion über Textprozeduren und von einer stärkeren Einbindung der Klassen/Kurse in die Rezeption der verfassten Texte“ (Abraham et al., 2015, S. 10). Eine Voraussetzung für das Erlernen von Textprozeduren ist Spracherfahrung (vgl. Abraham et al., 2015, S. 10). Durch rezeptive Spracherfahrung, also durch Lesen, soll die produktive Sprachverwendung gefördert werden.

Eine verstärkte Fokussierung auf Textroutinen als sprachliche Werkzeuge des Schreibens ist in der aktuellen Schreibdidaktik wie gesagt sehr präsent, so fordert auch Lehnen (2012) im Hinblick auf die Entwicklung von Schreibkompetenz von Studierenden Folgendes:

> Die Hinlenkung auf die sprachlichen Werkzeuge und die gezielte Wahrnehmung von Textroutinen, die sich insbesondere auch durch die Lektüre einschlägiger Texte und zunehmende Rezeptionserfahrung im Studienverlauf aufbauen, stellen deshalb eine wichtige Schaltstelle im Lernprozess dar (Lehnen, 2012, S. 39).

Wie oben bereits erwähnt kann die Rezeption von Texten das eigene Schreiben verbessern, Lehnen postuliert das für die Entwicklung von wissenschaftlicher Schreibkompetenz bei Studierenden:

> Eine wichtige Perspektive gerade mit Blick auf die von uns fokussierten Textroutinen stellt die gezielte Auseinandersetzung mit guten Textbeispielen dar. Die Analyse professioneller Texte macht Handlungsschemata, Muster und ihre Variation erkennbar wie auch Textroutinen identifizierbar. Texte werden zu sprachlichen Vorbildern. Aus der Analyse erwachsen Kriterien für die eigene Textproduktion (Lehnen, 2012, S. 57).[43]

Schwierig ist dabei für die Lehrenden, diese Muster klar festzulegen, ein Problem, auf das auch Dannerer (2012) in ihrem Artikel zur Rolle von Textroutinen in der Erzählentwicklung von Jugendlichen aufmerksam macht: „Grundlage dafür, dass Routinen überhaupt entstehen können, ist die Existenz von Formelhaftem, von Mustern, sowie die Fähigkeit, sie „automatisiert" einsetzen zu können“ (Dannerer, 2012, S. 101). Es fehlt laut Dannerer an klaren Kriterien und Definitionen dieser Muster und Routinen (vgl. 2012, S. 102–104).

Das Problem der Mustertexte ist in der Schule allgemein bekannt: Wünschenswert ist der Einsatz authentischer Texte, die weitaus motivierender sind, als von Lehrkräften formulierte Musterlösungen. Dabei ist jedoch eine Schwierigkeit, vom Anspruchsniveau und vom Thema her geeignete Texte zu finden. Die andere Schwierigkeit liegt in der Natur der authentischen Texte: Sie halten sich nicht an vom Deutschunterricht beziehungsweise von der Deutschdidaktik Postuliertes, sondern sind individuell gestaltet, so dass es tatsächlich oft schwierig ist, Textroutinen im Sinne von

[43] Selbstverständlich ist diese Forderung bei Lehnen (2012) nur ein kleiner Schritt im Aufbau von Schreibkompetenz. Zu den weiteren Schritten und Methoden, die sie vorschlägt, siehe Lehnen (2012).

sprachlichen Formulierungsroutinen auszumachen, die die Schülerinnen und Schüler für eigene Texte nutzen könnten. Der Einwand von Dannerer ist insofern berechtigt. Allerdings stellt sich die Sachlage je nach Textsorte und Textprozedur anders dar. Gerade im Bereich der Prozedur des Konzedierens ist es nämlich so, dass sich durchaus sprachliche Formulierungen finden lassen, die als prototypisch für diese Prozedur angesehen werden können. Rezat nimmt diese Spracherfahrung als Voraussetzung für konzessives Argumentieren an (vgl. Rezat, 2011, S. 64). Feilke beschreibt dies ausführlich in seiner Arbeit zu literalen Prozeduren:

> Eine Konjunktion wie *geschweige denn* […] oder eine Fügung wie *zwar […] aber* sind mehr als bloße Konjunktionen; es sind Textbildungsmittel, die Leerstellen für Sätze enthalten, die sie in eine textsemantische Beziehung bringen und zugleich pragmatisch gewichten. Als komplexe Konjunktionen sind sie erwachsen aus rekurrenten, wiederkehrenden Sequenzen von Texthandlungen in argumentativen Texten, in denen diese Sequenzen ein restriktives bzw. konzessives Muster realisieren. *Zwar […] aber* ist ein satzübergreifendes Werkzeug des Argumentierens. Es funktioniert auf der Ebene des komplexen Satzes ebenso wie auf der textueller Makrostrukturen und der textsemantischen Organisation ganzer Absätze. Dem Gegenüber etwas zu konzedieren, ja, den Adressaten in seiner Haltung zu einer kontroversen Frage sogar partiell zu bestätigen, indem mögliche Gegenargumente oder Einstellungen vorweggenommen werden, und gleich anschließend den so bereiteten textlichen ‚Sympathieboden' für ein eigenes Gegenargument zu nutzen, das ist ein Beispiel für eine *textbildende argumentative Prozedur* (Feilke, 2010, S. 11).

Der für diese Studie verfasste Modelltext enthält eben diese Formulierungen (s. Anhang 7 Modelltext für Vignette F) und folgt außerdem der oben genannten Struktur. Gerade die Struktur kommentiert die Lehrkraft dann auch in ihrer Erklärung, und zwar zunächst, indem sie deutlich macht, dass sie dem Argument aus dem Ausgangstext bis zu einem gewissen Punkt Gültigkeit zuspricht:

> un=WENN ich so=ne widerLEgung !MACH!, dann schau=ich IMMER drauf, dass der TEXT, den ich widerLEgen MÖCHte, DA auch (.) öfters mal drin VORkommt, damit ich den !WÜR!dige (.) und (.) dem (.) verTREter der GEgenseite ZEIge; ich hab des schon verSTANden, was du sagst, ich NEHM=des auch !AUF!, UND (-) MANches davon find=ich auch GANZ überZEUgend. (s. Anhang 6.6, Transkript zu Vignette F, Methode Modelltext, Z. 13-25)

Im weiteren Verlauf des Textes wird dann - ganz wie bei Feilke (2010) beschrieben - das eigentliche Gegenargument angebracht; auch diesen Schritt kommentiert die Lehrkraft:

> JA, (.) aber DA muss ich jetz ANsetzen mit meiner EIgentlichen widerLE:gung, (-) weil ich jetzt glaube, dass ge!NAU! !DAS! – diese !STAR!ke beZIEhungsarBEIT – EI:gentlich nicht NÖT…/ nicht MÖ:Glich is, WENN des nur über FACEbook läuft und OH:ne perSÖNlichen konTAKT. ALso (.) muss ich jetz hier beGRÜNden, waRUM ich glaub, des is nicht MÖ:Glich, (.) so wie die auTOrin eigentlich beHAUPtet hat. und HIER sind jetzt die KERNpunkte !MEI!ner begründung drin, (.) (s. Anhang 6.6, Transkript zu Vignette F, Methode Modelltext, Z. 45-59)

Durch das mündliche Kommentieren der einzelnen Textabschnitte wird den Schülerinnen und Schülern ebenfalls Einblick in die Planungs- und Schreibprozesse einer Expertin gegeben. Ziel dieser Erklärung ist nicht imitatives Schreiben, sondern ebenso wie bei Variante eins letztendlich prozedurales Wissen; hier nur eben über das Präsentieren eines möglichen Endproduktes, an dem durch metakommunikative Kommentare erklärt wird, aus welchen Schritten der Prozess bestand, bis das Endprodukt in der präsentierten Form vorlag. Analog zur Variante eins arbeitet die Lehrkraft auch hier multimodal, indem sie einzelne Bestandteile während ihrer Erläuterungen farblich markiert.

Im Vergleich zu Variante eins weist diese Wie-Erklärung weniger Erklärtiefe auf: Die Zuhörenden werden nicht am Entstehensprozess der Widerlegung beteiligt, sie bekommen stattdessen die fertige Widerlegung präsentiert. Trotz dieser metakognitiven Kommentare können die Zuhörenden den Prozess nicht „live" mitverfolgen, so dass sie den eigentlichen Widerlegungsprozess nur anhand des fertigen Produktes erahnen können (d. h. diese Variante ist nach Taut und Rakoczy (2016) weniger herausfordernd als oben erläuterte Variante eins). Die verborgenen Zusammenhänge und Denkprozesse werden dadurch weit weniger deutlich und die Funktion (vgl. Renkl & Wittwer, 2006) des Widerlegens steht im Gegensatz zu obiger Variante nicht im Fokus. Die Wie-Erklärung in der Variante zwei ist deswegen vielmehr dem Erklären$_1$ zuzuordnen. Es wird davon ausgegangen, dass der Lern- und vor allem der Merkeffekt geringer ist als in oben geschilderter Variante eins.

6 Fragestellungen

Nach der Darlegung der theoretischen Einordnung sowie der Konstruktion der Videovignetten als Stimuli für vorliegende Studie wird im Folgenden das Forschungsinteresse vorliegender Arbeit geklärt. FALKE-D widmet sich empirisch der Fragestellung, was gutes Erklären im Deutschunterricht ist. Die oben geschilderten theoretischen Grundlagen zum fachübergreifenden sowie zum deutschspezifischen Erklären und der dargelegte Forschungsbedarf zu gutem Erklären im Deutschunterricht führen zu den nun erläuterten Forschungsfragen.

Zunächst wird für die Gesamtstichprobe untersucht, wie die gesehenen Erklärungen global bewertet werden (Forschungsfrage 1a) und wie diese die einzelnen Aspekte der Erklärungen bewertet (Forschungsfrage 1b).

Zudem können über die mehrperspektivische Anlage der Studie Aussagen darüber gemacht werden, wie die einzelnen Statusgruppen (Schüler/-innen, Studierende, Lehrkräfte und Didaktiker/-innen) die Erklärungen global bewerten (Forschungsfrage 2a) und wie diese Statusgruppen die abgefragten Kriterien beurteilen, wobei insbesondere interessiert, welche Gemeinsamkeiten und Unterschiede sich im Antwortverhalten der Statusgruppen zeigen und inwieweit innerhalb der

einzelnen Statusgruppen Einigkeit über die wahrgenommene Qualität der Erklärungen herrscht (Forschungsfrage 2b).

Dabei ist außerdem von Belang, inwiefern die Variation der Erklärtiefe für die einzelnen Statusgruppen ausschlaggebend für die unterschiedliche Bewertung der beiden Varianten der gleichen Erklärung ist (Forschungsfrage 2c).

Außerdem wird die wahrgenommene Qualität der oben beschriebenen Kriterien *Strukturiertheit, Adressatenorientierung, Sprech- und Körperausdruck, Sprachliche Verständlichkeit* sowie *Erklärtiefe* (und als Kontrollvariable *Persönlichkeitswirkung*) in Bezug zur Einschätzung der Gesamtqualität der Erklärung gesetzt. Hier interessieren zunächst wiederum die Zusammenhänge in der Gesamtstichprobe, wobei besonders der Einfluss des deutschspezifisch konzipierten Aspektes der *Erklärtiefe* von Interesse ist (Forschungsfrage 3). Dazu kommt erneut die Analyse des statusgruppenspezifischen Antwortverhaltens, das heißt, Gemeinsamkeiten und Unterschiede im Antwortverhalten der Statusgruppen sowie die Frage, inwieweit innerhalb der einzelnen Statusgruppen Einigkeit über die Bedeutung einzelner Kriterien für die Erklärqualität besteht; insbesondere interessiert ein weiteres Mal die Frage, welche Rolle der fachspezifische Aspekt der *Erklärtiefe* bei der Aufklärung der Gesamtvarianz spielt (Forschungsfrage 4) (vgl. Gaier & Schilcher, 2021, in Vorb.).

Da die Erklärtiefe auch zur Variation in den Konzeptionen der Vignettenpaare unterschiedlich (nämlich einfach vs. vertieft) umgesetzt wurde, wird sie noch einmal gesondert untersucht, indem für die einfachen und für die vertieften Vignetten getrennt dargelegt wird, welche Kriterien für die Bewertung der Erklärqualität ausschlaggebend sind (Forschungsfrage 5).

Schließlich werden diejenigen Kriterien untersucht, die die Studienteilnehmer/-innen spontan und ohne die Konzeptualisierung und Operationalisierung bei FALKE-D zu kennen, als Begründungen für das Vergeben einer Note auf die Erklärqualität der einzelnen Vignetten nennen; dabei wird erneut zunächst das Antwortverhalten der Gesamtstichprobe (Forschungsfrage 6a) analysiert, und anschließend genauer das der einzelnen Statusgruppen dargelegt (Forschungsfrage 6b), um schließlich wiederum das Antwortverhalten für die einfachen Varianten demjenigen für die vertieften Varianten gegenüberzustellen (Forschungsfrage 6c).

Mit Beantwortung der Forschungsfragen ergibt sich ein Gesamtbild, das sich aus den geschlossenen und den offenen Antworten zusammensetzt, und das maßgelbliche Aspekte des Konstruktes *Gutes Erklären* erläutert (vgl. Punkt *10 Diskussion der Ergebnisse zum Konstrukt des Guten Erklärens* in dieser Arbeit).

7 Forschungsmethode

Bevor die Ergebnisse der Erhebung erläutert werden, wird die Forschungsmethode der Studie FALKE-D dargestellt. Dazu wird im Folgenden das Design, das Erstellen der Stimuli und des Fragebogens, die einzelnen Statusgruppen, der Ablauf der Erhebung und die Operationalisierung der *Erklärtiefe* erläutert.

7.1 Design der Studie

Wie bereits angedeutet wurde zur Beantwortung der Forschungsfragen im FALKE-Projekt ein gemeinsames Design entwickelt. Jede beteiligte Fachdidaktik erstellte sechs Videovignetten[44] mit Lehrererklärungen aus dem eigenen Fach. Diese Vignetten wurden als Stimuli in einen Online-Fragebogen implementiert, dessen Anlage in allen beteiligten Fächern identisch war: Zu zwei Messzeitpunkten wurde die Erklärqualität von den Statusgruppen Schüler/-innen, Lehrkräfte, Studierende sowie Didaktiker/-innen eingeschätzt. In Anlehnung an Kuntze (vgl. 2016, S. 530) kann das Design von FALKE folgendermaßen zusammengefasst werden: Das Konstrukt, das untersucht wird, ist das fachspezifische Erklären, in FALKE-D Erklären im Deutschunterricht. Dazu wurden vertiefte Erklärungen und einfachere Erklärungen konzipiert und als Stimuli für die Wahrnehmung von Erklärqualität eingesetzt. Als Expertinnen und Experten (mit unterschiedlichem Expertisegrad) wurden die drei Statusgruppen Didaktiker/-innen, Lehrkräfte sowie Studierende befragt, wobei FALKE wie bereits mehrfach betont die Qualität der Erklärungen untersucht und nicht den Expertenstatus der befragten Personen. Außerdem wurden Schüler/-innen als „Expertinnen und Experten" für Unterricht aus der Perspektive der Adressatinnen/Adressaten befragt. Als Situierungsformat wurden Videovignetten gewählt, die nach Art der *staged videos* extra für den Online-Fragebogen von FALKE-D erstellt wurden. Als Frageformat wurden mit Ausnahme der Begründung für das Globalurteil geschlossene Frageformate gewählt, bei denen die Teilnehmer/-innen ihre (Nicht-) Zustimmung über Ankreuzen ausdrücken konnten. Im Folgenden werden die einzelnen Entscheidungen genauer begründet.

Videovignetten

Da die Erklärungen in ihrer Performanz bewertet werden sollten, also das Erklärhandeln der Lehrkraft, zu dem neben dem Inhalt der Erklärung auch die verwendete Sprache sowie der Sprech- und Körperausdruck der erklärenden Person gehört, wurden Videos als Stimuli gewählt. In anderen Formen der Stimuli, etwa Textvignetten oder Comics, fehlen die Aspekte der Performanz (vgl. Seidel

[44] Aus designtechnischen Gründen wurden im Fach Musikpädagogik sieben Vignetten erstellt (vgl. Frei, Hofmann und Puffer, 2021, in Vorb.).

& Thiel, 2017, S. 10), weswegen sie für FALKE nicht in Frage kommen, auch wenn sie mit deutlich weniger Aufwand erstellt werden können und in Studien zum Erheben von Lehrerkompetenz eingesetzt werden und reliable Ergebnisse erzielen (vgl. z. B. Friesen, Kuntze & Vogel, 2018).

Als Vignette werden kurze, exemplarische Unterrichtsausschnitte bezeichnet, die vor allem zur Diagnose von professioneller Lehrerkompetenz eingesetzt werden: „Kurze Unterrichtssituationen, sogenannte *Vignetten*, werden als besonders geeignet betrachtet, um berufsbezogene Kompetenzen von Lehrkräften zu erfassen, da sie durch den Bezug zu konkreten Situationen eine unterrichtsnahe Erhebung fachdidaktischer Kompetenzaspekte ermöglichen" (Friesen et al., 2018, S. 157). Die Erhebung dient in FALKE nicht dazu, Kompetenzen von Lehrkräften zu erfassen, vielmehr wird das konkrete Erklären (also die Lehrerhandlung) in seiner Qualität eingeschätzt und bewertet. Aussagen sollen also nicht über die Fähigkeit der erklärenden Person getroffen werden, sondern über das Erklären, allerdings wie erläutert in seiner Performanz als Erklärprozess. In diesem Sinne wird nach Kuntze (2016) das Material analysiert, aber eben nicht, um im Nachhinein Aussagen über die Teilnehmer/-innen an der Studie zu machen, sondern über das abgebildete Konstrukt des Erklärens. Dazu wurden bewusst keine authentischen Erklärungen aus „echtem" Unterricht gewählt, um nicht auf die Bereitschaft von Lehrkräften angewiesen zu sein, das eigene Erklärverhalten aus verschiedenen Perspektiven bewertet zu bekommen. Obwohl es reizvoll ist, authentisches Erklären im Unterricht zu untersuchen, läge bei der Entscheidung für authentische Erklärungen im FALKE-Projekt zwangsläufig eine Positiv-Auswahl von Lehrkräften vor, die das Unterrichtsgeschehen doch nur zum Teil abbilden (vgl. Wisniewski et al., 2020). Abgesehen davon sollten die Erklärungen und die verwendeten Methoden in den Vignetten deutlich sichtbar sein, so dass die Erklärungen zwar selbstständig von Lehrkräften gehalten wurden, aber eben mit den unter Punkt *5 Konstruktion der Videos als Stimuli* und Punkt *7.2 Erstellen und Auswahl der Videovignetten als Stimuli* (in dieser Arbeit) beschriebenen Vorgaben.

Die für FALKE-D erstellten Vignetten sind also nicht normativ, sondern dienen dazu, ein Rahmenmodell für gutes Erklären im Fach Deutsch zu entwickeln. Eine logische Schlussfolgerung dieses Anspruches ist es, Vignetten zu erstellen, in denen eine den Studienteilnehmenden fremde Person erklärt. Die Entscheidung für „unbekannte" Erklärende fügt sich in den advokatorischen Ansatz ein, der die „Notwendigkeit einer gewissen Distanz zum eigenen Handeln [betont], um Analyseprozesse zu ermöglichen […]. Unterschiede in der Interpretation und Bewertung des beobachteten Lehrerhandelns zu bestimmten Kriterien werden dabei als Indikatoren für spezifische Kompetenzaspekte betrachtet" (Friesen et al., 2018, S. 158, s. dazu außerdem die Gründe für das Bewerten von fremden Erklärungen, die unter Punkt *2.4.3 Methoden und Zweck der Erfassung von Wahrnehmung* in dieser Arbeit bereits erläutert wurden). FALKE-D verfolgt analog zu diesem Ansatz

das Ziel, über die Bewertungen der unterschiedlichen Statusgruppen Indikatoren für gutes Erklären herauszuarbeiten, die im Wesentlichen auf der Expertise der befragten Personen beruhen.

Mit einer Länge von circa drei Minuten pro Vignette wählt FALKE-D einen (zeitlichen) Zwischenweg. Zum einen werden vergleichsweise kurze Erklärvignetten erstellt, da längere Videosequenzen, die zum Beispiel zum Messen der professionellen Kompetenz einer Lehrkraft notwendig sind, um verlässliche Ergebnisse zu erzielen (vgl. etwa Praetorius et al., 2014), in FALKE-D nicht notwendig sind: In FALKE wird zwar Unterrichtsqualität bewertet, innerhalb des komplexen, multiperspektivischen Unterrichtsgeschehens konzentrieren sich die teilnehmenden Personen aber auf die Erklärung an sich. Diese wird deswegen mit so viel Kontext wie nötig und gleichzeitig mit so wenig Kontext wie möglich dargestellt (vgl. Friesen et al., 2018, S. 158 bzw. Kuntze, 2016, S. 529). Zum anderen ist die Dauer von drei Minuten länger als die etwa 30sekündigen Thin-Slices, die sich zum Beispiel nach Begrich und anderen (2017) „für eher globale Qualitätsbeurteilung" (S. 44) des Unterrichtsgeschehens oder die Beurteilung einzelner Dimensionen von Unterricht als geeignet erweisen. In FALKE-D wird mit dem Erfragen eines Globalurteils (in Form einer Schulnote) für jede Erklärung zunächst ein solches allgemeines Qualitätsurteil erhoben. Mit der detaillierteren Befragung zu einzelnen Aspekten wird zum zweiten Messpunkt allerdings auch eine genauere Analyse der Erklärungen von den Studienteilnehmerinnen und -teilnehmern verlangt. Dazu muss die Erklärung inhaltlich abgeschlossen abgebildet werden, was in 30 Sekunden nicht möglich ist, so dass die Vignetten länger als die typischen Thin-Slices ausfallen (vgl. dazu z. B. Begrich et al., 2017, S. 44). Abgesehen von diesen methodischen Gründen spricht auch das Nebengütekriterium der Zumutbarkeit für die zeitliche Begrenzung auf etwa drei Minuten pro Vignette: Die Bearbeitungszeit der sechs Vignetten, die zwei Mal angesehen werden mussten, musste für die befragten Personen annehmbar gehalten werden.

In einem fächerübergreifenden Austausch einigte sich das Projekt FALKE auf eigens für das Projekt erstellte Vignetten, was im Folgenden noch etwas genauer begründet wird. Wie angesprochen wurde nicht auf Unterrichtsmitschnitte zurückgegriffen, um das Erklären fokussiert abzubilden. Durch den Fokus auf das Erklären kann mit verlässlichen Ergebnissen gerechnet werden: Die Konzentration auf die Erklärhandlung (vgl. Erläuterungen oben) wurde in den eigens dafür hergestellten Vignetten so umgesetzt, dass die befragten Personen einerseits mit so wenig zusätzlichen Informationen wie möglich (z. B. Schüleräußerungen) konfrontiert wurden, und anderseits aber durch das gewählte Format einen möglichst authentischen Eindruck der Erklärung erhalten sollten. FALKE wählt somit bewusst so genannte *staged videos*, keine authentischen Unterrichtsmitschnitte, und macht sich damit die oben beschriebene Unterrichtsnähe von Vignetten zunutze. Durch das Entwerfen und Drehen der Erklärungen konnten diese im Hinblick auf die

Erklärtiefe (s. Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit) bewusst gestaltet werden, so dass diese klarer hervortritt als bei „echten" Unterrichtsmitschnitten (vgl. Seidel & Thiel, 2017, S. 14).
Die Erklärungen wurden aus dem gleichen Grund monologisch durch die Lehrkraft gehalten, deswegen wurden alle sechs Erklärungen in den Vignetten von der Lehrkraft selbst initiiert und auch im Verlauf des Erklärprozesses wurde wie bereits angedeutet auf Interaktion mit Schülerinnen und Schülern bewusst verzichtet: In ihren Bewertungen, insbesondere beim Globalurteil, sollten die Studienteilnehmer/-innen unbeeinflusst von jeglicher Reaktion von den Adressatinnen und Adressaten in der Vignette ihr Urteil abgeben. Wäre Interaktion mit Schülerinnen und Schülern abgebildet, so könnte dies zu genannten Verzerrungen in den Urteilen führen, zum Beispiel, weil das in der Vignette abgebildete Verständnis der Adressatinnen und Adressaten die Teilnehmer/-innen zu einer guten Bewertung verleitet, obwohl sie die Erklärung nicht für gelungen halten, oder umgekehrt (s. zu Nutzen von bzw. Behinderung durch Kontextinformationen auch Friesen et al., 2018, S. 161).
Die Vignetten bilden das Konstrukt „Erklären" relativ abstrakt ab (eben durch die Entscheidung, monologisch erklären zu lassen), gleichzeitig betten sie es aber in eine konkrete Unterrichtssituation ein, so dass die Befragten idealerweise die Relevanz der Erklärungen nachvollziehen konnten (vgl. Kuntze, 2016, S. 529–530).
So wird trotz der oben erläuterten Einschränkungen davon ausgegangen, dass die Merkmale der Authentizität (d. h. Immersion, Resonanz und Motivation, vgl. Friesen et al., 2018) mit der Wahl dieses Befragungsformats erfüllt werden. Gerade in den offenen Antworten der FALKE-D-Befragung wird zu überprüfen sein, ob die Statusgruppe der Lehrkräfte die beabsichtigte Immersion (d. h. „wie stark sich eine Lehrkraft in die Unterrichtssituation hineinversetzt fühlt bzw. in sie eintauchen kann" Friesen et al., 2018, S. 159) sowie die beabsichtigte Resonanz (d. h. „die Verbindung der vorgelegten Unterrichtssituation mit eigenen Unterrichtserfahrungen" Friesen et al., 2018, S. 159) wahrgenommen und ob sie die Befragung als motivierend empfunden hat. Es wird davon ausgegangen, dass Immersion und Resonanz in abgeschwächter Form bei Didaktikerinnen und Didaktikern (aufgrund ihrer Unterrichtstätigkeit an der Universität sowie ihres Metablicks auf Unterricht) sowie bei Studierenden (aufgrund ihrer Erfahrungen im Schulunterricht bzw. in ihrer Rolle als angehende Unterrichtende) beziehungsweise aus der Adressatenperspektive ebenfalls bei den Schülerinnen und Schülern hervorgerufen wird.

Übersicht über Design

Die gerade dargelegten Grundlagen führten zur Entscheidung für das nun konkret erläuterte Design. Abgesehen von den Videovignetten sowie den eigenen fachspezifischen Items und einigen

kleineren Besonderheiten war das Design in allen elf beteiligten Fachdidaktiken identisch aufgebaut. Einen Überblick bietet Abbildung 8.

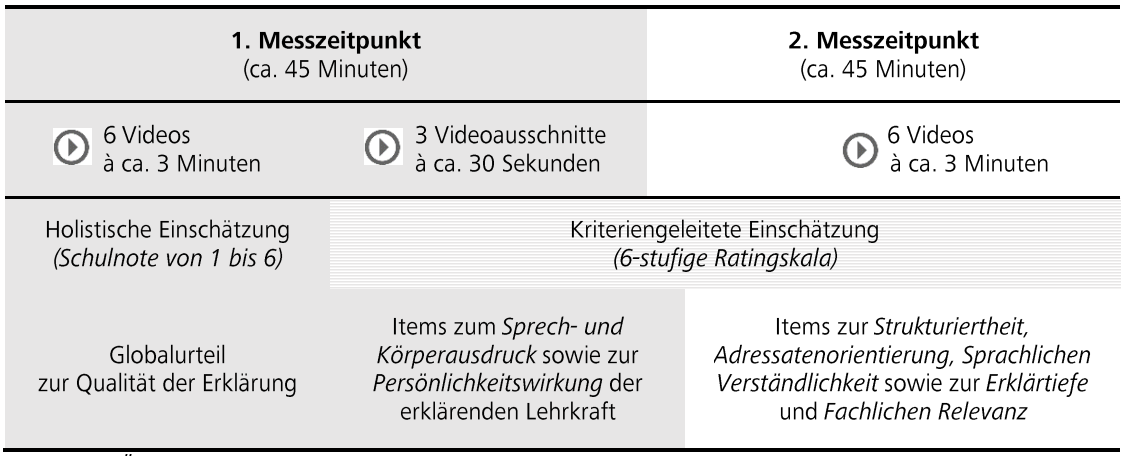


Abbildung 8: Übersicht über Design der Studie
Zu beiden Messzeitpunkten werden dieselben Videos von den Statusgruppen Schüler/-innen, Studierende, Lehrkräfte und Didaktiker/-innen eingeschätzt; Abb. angelehnt an Gaier & Schilcher, 2021, in Vorb.

Die für FALKE-D erstellten sechs Videovignetten standen alle unter dem Rahmenthema „Arbeit mit einem argumentativen Text" (vgl. Punkt *5.1.2 Allgemeine Verortung „Arbeit mit einem argumentativen Text"* in dieser Arbeit). Die Studienteilnehmer/-innen bewerteten alle sechs Vignetten im Online-Fragebogen an zwei Messzeitpunkten, was aufgrund der Befragungslänge von jeweils etwa 45 Minuten aus Gründen der Zumutbarkeit entschieden wurde. Zum ersten Zeitpunkt erteilten sie den Erklärungen ein holistisches Globalurteil in Form einer Schulnote mit der Möglichkeit, diese mit einer positiven oder negativen Tendenz zu versehen. Die Statusgruppen der Expertinnen und Experten hatten zudem die Möglichkeit, ihr Urteil offen zu begründen. Anschließend daran bewerteten die Befragten drei Ausschnitte von etwa 30 Sekunden Länge aus den gesehenen Vignetten (und zwar von jeder/jedem Sprecher/-in einen Ausschnitt) in Bezug auf den wahrgenommenen *Sprech- und Körperausdruck* sowie auf die *Persönlichkeitswirkung* der erklärenden Person. Dies erfolgte analog in allen beteiligten Fächern des Gesamtprojektes FALKE.
Der zweite Messzeitpunkt fand an einem anderen Tag statt als der erste. Zwischen beiden Messzeitpunkten lagen höchstens drei Wochen. Zum zweiten Zeitpunkt sahen die Studienteilnehmer/-innen die Videos erneut. Sie bewerteten jede Vignette kriteriengeleitet nach den oben erläuterten Aspekten des Gesamtprojekts FALKE *Sprachliche Verständlichkeit, Strukturiertheit* und *Adressatenorientierung* sowie in der Studie FALKE-D *Erklärtiefe* und zur Kontrolle der inhaltlichen Validität die *Fachliche Relevanz* (s. Punkt *4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens* in dieser Arbeit).

7.2 Erstellen und Auswahl der Videovignetten als Stimuli

Unter Punkt *5 Konstruktion der Videos als Stimuli* in dieser Arbeit wurden die theoretischen Überlegungen, die den für FALKE-D erstellten Erklärungen zugrunde liegen, bereits dargestellt. Im Folgenden wird das konkrete Erstellen der Vignetten geschildert. Als Erklärer/-in konnten ein Lehrer und zwei Lehrerinnen gewonnen werden, die alle drei an einem bayerischen Gymnasium unterrichten. Die Auswahl von drei unterschiedlichen Sprecherinnen und Sprechern dient dazu, Abhängigkeiten zwischen den Vignetten zu vermeiden (so wurden bewusst Lehrkräfte unterschiedlichen Geschlechts und unterschiedlichen Alters gewählt, vgl. Friesen et al., 2018, S. 164) und Hinweise auf den Einfluss von Persönlichkeitsmerkmalen auf die Bewertung der Erklärqualität zu bekommen, die mit drei Items als Kontrollvariable erfragt wurden (vgl. Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* in dieser Arbeit).

Die Lehrkräfte stellten sich für den Videodreh unentgeltlich und während ihrer Freizeit zur Verfügung. Jede Lehrkraft erklärte einen Gegenstand in den zwei oben beschriebenen unterschiedlichen Varianten der Erklärtiefe. Die Tatsache, dass die gleiche Lehrkraft die beiden Varianten einer Erklärung übernahm, verhinderte, dass etwaige Unterschiede in den Beurteilungen auf die *Persönlichkeitswirkung* der Lehrkraft zurückzuführen sind. Die Aufteilung der Lehrkräfte erfolgte somit nach Erklärprozedur (wobei die beiden Lehrerinnen *erklären, was* bzw. *erklären, wie* durchführten, der Lehrer *erklären, warum*).

Zur Vorbereitung auf den Dreh erhielten die Lehrkräfte zu ihrem Gegenstand jeweils zwei videographierte Erklärungen als Muster. Diese veranschaulichten zum einen die Vorgaben des monologischen Erklärens sowie die Dauer von etwa drei Minuten. Zum anderen dienten sie der Verdeutlichung der Variation innerhalb der Vignettenpaare. Zusätzlich erhielten die Sprecher/-innen die theoretischen Grundlagen für die Variation der Erklärtiefe bezogen auf ihren Gegenstand in schriftlicher Form (vgl. hierzu auch Gaier & Schilcher, 2021, in Vorb.). Das erkennbare Umsetzen der unterschiedlichen Erklärtiefe war das Hauptanliegen beim Erstellen der Vignetten.

In FALKE-D wurde bewusst auf wörtliche Skripts als Vorlage für die Erklärungen verzichtet, denn zur Erhöhung der Authentizität sollten die Lehrkräfte ihren eigenen Erklär- und Sprechstil umsetzen. Ebenfalls aus Gründen der Authentizität wurden sämtliche Vignetten der Lehrkräfte an Schulen in einem Klassenzimmer gedreht. Während des Drehs waren Schüler/-innen als Schauspieler/-innen anwesend, die auf den Vignetten auch zu sehen sind. Auch dies erhöht die ökologische Validität, außerdem fiel es den Lehrkräften leichter, die Erklärung vor Publikum und nicht in einem leeren Raum zu halten. Auch in der Wahl der Hilfsmittel und Veranschaulichungen waren die Lehrkräfte frei. Zum Einsatz kamen während der unterschiedlichen Erklärungen die Tafel und/oder der Overheadprojektor.

Gefilmt wurde in etwa in der Halbnahe, während des Drehs wurde die Einstellung der Kamera nicht verändert. Auf diese Weise ist durchgehend die Lehrkraft sowie deren Gestik erfasst. Abgesehen davon sind die entsprechende Tafelanschrift beziehungsweise die Projektion der Overheadfolie zu sehen. Textstellen, die für die Erklärung entscheidend waren, wurden den Studienteilnehmerinnen und -teilnehmern außerdem vor dem Ansehen der Vignette in Textform eingeblendet. Da auch der Sprechausdruck der Lehrkräfte bewertet werden sollte, bekamen die Lehrkräfte vor dem Dreh ein Ansteckmikrofon (der Marke Sennheiser), um eine hinreichende auditive Qualität zu gewährleisten. Das Mikrofon ist in den Vignetten kaum zu sehen und stört die wahrgenommene Authentizität nicht.

Jede Erklärung wurde mehrmals aufgenommen und mit den Lehrkräften angesehen und besprochen. Nur Vignetten, mit denen die Sprecher/-innen einverstanden waren, kamen in die nähere Auswahl für den Fragebogen. Diese Aufnahmen wurden anschließend mit Deutschdidaktikerinnen und -didaktikern der Universität Regensburg begutachtet und diskutiert. Als jeweils gelungenste Variante wurde diejenige gewählt, die die Vorgaben und implementierten Konzepte (siehe oben) am besten umsetzt. Diese wurde dann in den Fragebogen implementiert, dessen Entstehung im Folgenden erläutert wird (vgl. zu diesem Punkt Gaier & Schilcher, 2021, in Vorb.).

7.3 Untersuchungsinstrument: Online-Fragebogen

Die Wahl des Designs eines Online-Fragebogens hat methodisch den Vorteil, dass eine hohe Durchführungs- und Ausführungsobjektivität gewährleistet ist.

Da in FALKE-D die Qualität von Erklärungen als Lehrerhandlung untersucht wird, ist die Gestaltung eines Fragebogens das probate Mittel dazu: Analog zur Unterrichtsqualität kann auch die Qualität von Erklärungen nur indirekt erfasst werden, über die Wahrnehmung von Beteiligten oder externen Beobachtenden (vgl. Clausen, 2002, S. 43). Die Auswahl der befragten Statusgruppen wurde unter Punkt *2.4 (Professionelle) Wahrnehmung von Unterrichtsqualität* in dieser Arbeit bereits begründet.

Um die Wahrnehmung der Studienteilnehmer/-innen zu erfragen, eignen sich Interview- oder Fragebogenstudien. Mit der Entscheidung für letztere hat FALKE ein klassisches quantitatives Design gewählt, dessen Ergebnisse, gerade durch den Vergleich der verschiedenen Statusgruppen, mit der gebotenen Vorsicht aufgrund der Größe der befragten Stichprobe verallgemeinerbar sind. Durch das Fragebogen-Design und die hauptsächlich geschlossenen Items sind die Ergebnisse exakt quantifizierbar und ermöglichen es, statistische Zusammenhänge zu ermitteln. Das ist besonders interessant im Hinblick auf das Gesamtprojekt FALKE, bei dem fachübergreifende Analysen zur *Sprachlichen Gestaltung,* zum *Sprech- und Körperausdruck* sowie zur *Adressatenorientierung* und

zur *Strukturiertheit* angestrebt werden. Die Auswertung der offenen Begründungen der Globalbeurteilungen ermöglichen ergänzend qualitative Analysen.

Der Fragebogen für FALKE-D wurde mit der Umfrage-Software EFS-Survey von Questback erstellt. Die oben beschriebenen Vignetten wurden in der bereits erwähnten Reihenfolge (nach den Erklärprozeduren *erklären, was/ erklären, warum/ erklären, wie*) in aufsteigender Komplexität in den Fragebogen implementiert. Um Reihenfolgeeffekte zu vermeiden, wurden zwei Versionen des Fragebogens (A und B) erstellt, bei denen die Reihenfolge der präsentierten Erklärtiefe umgekehrt wurde. Das bedeutet, dass in Fragebogen A zunächst die Vignetten mit größerer Erklärtiefe zu sehen sind und danach die einfacheren Vignetten. In Fragebogen B ist das genau umgekehrt: Die Studienteilnehmer/-innen beantworteten zunächst die Fragen zu den einfacheren Vignetten und sahen anschließend die Vignetten mit größerer Erklärtiefe. Die Reihenfolge der Prozeduren ist in beiden Versionen gleich (s. o. *erklären, was/warum/wie*). Einen Überblick über das Design des Fragebogens gibt Abbildung 8. Der Fragebogen wurde wie dargelegt an zwei Messzeitpunkten erhoben, der genaue Aufbau des Fragebogens wird im Folgenden erläutert.

Ablauf Messzeitpunkt 1

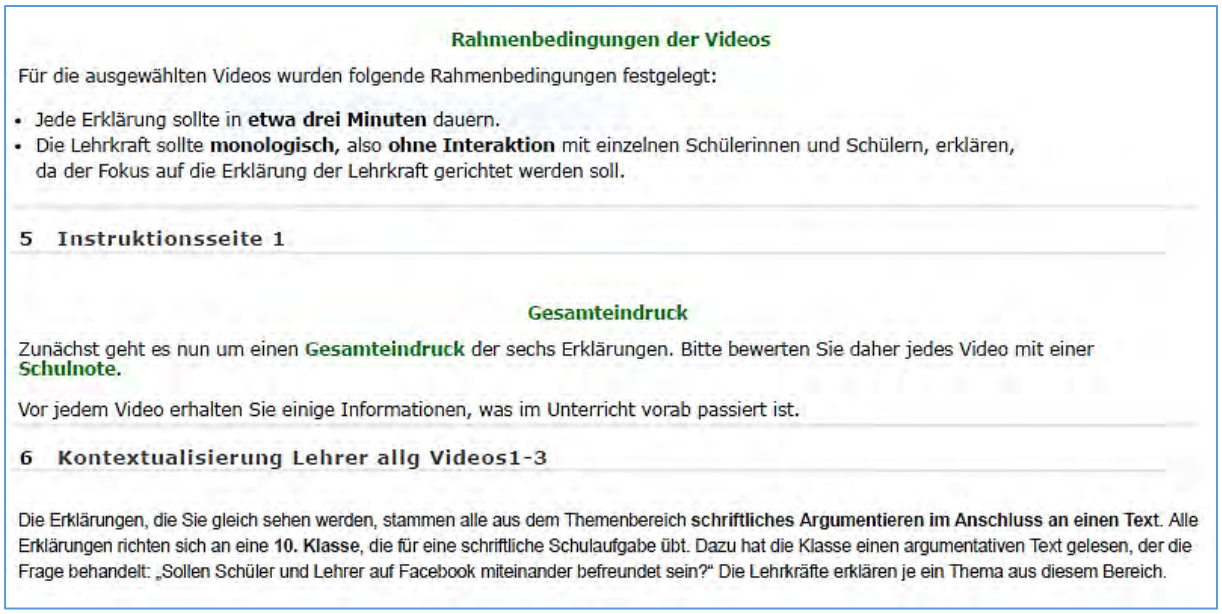
Der Fragebogen begann zum ersten Messzeitpunkt mit einem Willkommensgruß und einer allgemeinen Einleitung zur FALKE-D-Studie.

Damit Ergebnisse aus beiden Messzeitpunkten im Anschluss an die Befragung zusammengeführt werden konnten, mussten die Befragten zu Beginn beider Befragungen den gleichen Kode eintragen. Die Expertinnen und Experten generierten den Kode nach einer Anleitung (d. h. die ersten zwei Buchstaben des Vornamens der Mutter, dann des Vaters sowie die ersten zwei Ziffern der eigenen Hausnummer) selbst. Die Schüler/-innen zogen ein vorbereitetes Kodewort nach dem Zufallsprinzip (es wurden entsprechend der Anzahl der Schüler/-innen in den befragten Klassen Obst- bzw. Gemüsesorten, Tiernamen sowie Städtenamen vorbereitet).

Die Studienteilnehmer/-innen der Expertenstatusgruppen (d. h. der Studierenden, Lehrkräfte und Didaktiker/-innen, im Folgenden als SLD-Gruppe bezeichnet) wählten dann ihren Expertenstatus aus (also Student/-in, Lehrer/-in oder Hochschulmitarbeiter/-in) und wurden im Anschluss um einige spezifische biographische Angaben gebeten (zum Beispiel Studiengang der Studierenden, etwaige Funktionsstellen von Lehrkräften, Titel der Didaktiker/-innen). Bei den Schülerinnen und Schülern entfielen diese Angaben. Wie bei Befragungen dieser Art üblich wurde bei den SLD-Gruppen das Geschlecht miterhoben. Bei den Schülerinnen und Schülern entfiel dies aus datenschutzrechtlichen Gründen.

Nach der Lautstärkeanpassung folgten Erläuterungen zur Gestaltung der Erklärvignetten als monologische Lehrererklärung sowie der Auftrag, die gesehenen Erklärungen zunächst mit einer

Schulnote zu bewerten, und allgemeine Kontextinformationen zum Unterrichtsthema *Argumentieren.* Diese Situierung diente der Orientierung der Befragten und bettete die Erklärsequenzen in ein fiktives, aber möglichst authentisches Unterrichtsgeschehen ein, in dem eine 10. Klasse das schriftliche Argumentieren im Anschluss an einen Text als Vorbereitung auf eine schriftliche Leistungserhebung behandelt. Zur Veranschaulichung werden die erläuterten Kontextualisierungen in Abbildung 9 für die Statusgruppen der Expertinnen und Experten beziehungsweise in Abbildung 10 für die Schüler/-innen in der Form abgebildet, in der die Studienteilnehmer/-innen sie im Online-Fragebogen bekamen.



Rahmenbedingungen der Videos

Für die ausgewählten Videos wurden folgende Rahmenbedingungen festgelegt:

- Jede Erklärung sollte in **etwa drei Minuten** dauern.
- Die Lehrkraft sollte **monologisch,** also **ohne Interaktion** mit einzelnen Schülerinnen und Schülern, erklären, da der Fokus auf die Erklärung der Lehrkraft gerichtet werden soll.

5 Instruktionsseite 1

Gesamteindruck

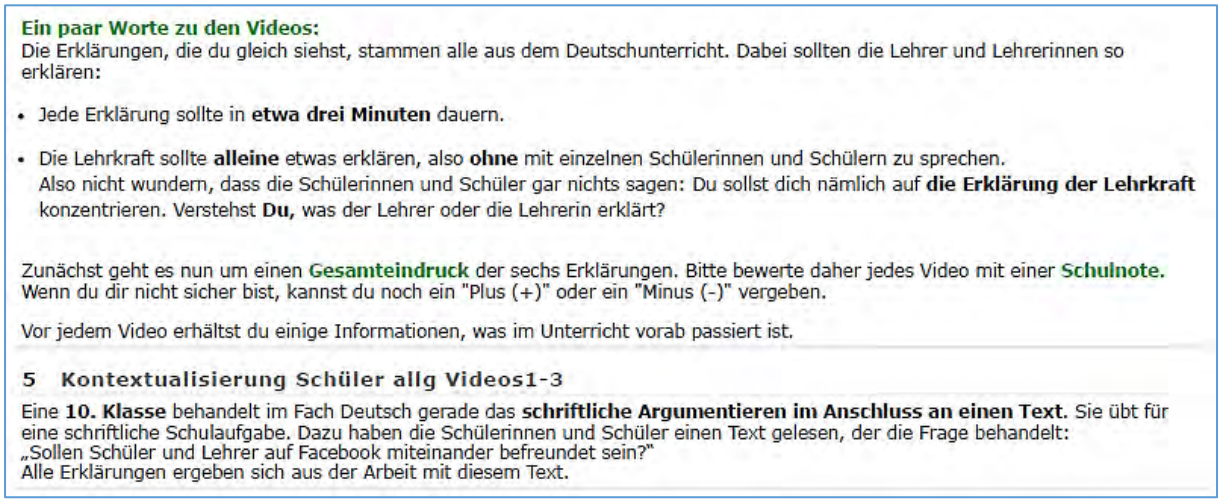
Zunächst geht es nun um einen **Gesamteindruck** der sechs Erklärungen. Bitte bewerten Sie daher jedes Video mit einer **Schulnote.**

Vor jedem Video erhalten Sie einige Informationen, was im Unterricht vorab passiert ist.

6 Kontextualisierung Lehrer allg Videos1-3

Die Erklärungen, die Sie gleich sehen werden, stammen alle aus dem Themenbereich **schriftliches Argumentieren im Anschluss an einen Text**. Alle Erklärungen richten sich an eine **10. Klasse**, die für eine schriftliche Schulaufgabe übt. Dazu hat die Klasse einen argumentativen Text gelesen, der die Frage behandelt: „Sollen Schüler und Lehrer auf Facebook miteinander befreundet sein?" Die Lehrkräfte erklären je ein Thema aus diesem Bereich.

Abbildung 9: Kontextualisierungen SLD-Gruppen



Ein paar Worte zu den Videos:
Die Erklärungen, die du gleich siehst, stammen alle aus dem Deutschunterricht. Dabei sollten die Lehrer und Lehrerinnen so erklären:

- Jede Erklärung sollte in **etwa drei Minuten** dauern.
- Die Lehrkraft sollte **alleine** etwas erklären, also **ohne** mit einzelnen Schülerinnen und Schülern zu sprechen. Also nicht wundern, dass die Schülerinnen und Schüler gar nichts sagen: Du sollst dich nämlich auf **die Erklärung der Lehrkraft** konzentrieren. Verstehst **Du,** was der Lehrer oder die Lehrerin erklärt?

Zunächst geht es nun um einen **Gesamteindruck** der sechs Erklärungen. Bitte bewerte daher jedes Video mit einer **Schulnote.** Wenn du dir nicht sicher bist, kannst du noch ein "Plus (+)" oder ein "Minus (-)" vergeben.

Vor jedem Video erhältst du einige Informationen, was im Unterricht vorab passiert ist.

5 Kontextualisierung Schüler allg Videos1-3

Eine **10. Klasse** behandelt im Fach Deutsch gerade das **schriftliche Argumentieren im Anschluss an einen Text.** Sie übt für eine schriftliche Schulaufgabe. Dazu haben die Schülerinnen und Schüler einen Text gelesen, der die Frage behandelt:
„Sollen Schüler und Lehrer auf Facebook miteinander befreundet sein?"
Alle Erklärungen ergeben sich aus der Arbeit mit diesem Text.

Abbildung 10: Kontextualisierungen Schüler/-innen

Es folgten schließlich die einzelnen Vignetten, vor denen die Teilnehmer/-innen konkrete Informationen zum jeweiligen Erklärанlass erhielten. Exemplarisch sind diese Informationen, so wie

sie im Fragebogen zu lesen waren, für die Vignette *Erklären, wie* in Abbildung 11 (für die Statusgruppen SLD) beziehungsweise in Abbildung 12 (für die Statusgruppe SuS) zu sehen.

In dem Text „Sollen Schüler und Lehrer auf Facebook miteinander befreundet sein?“ kommt unten stehende Aussage vor, mit der die Lehrerin nicht einverstanden ist. Sie widerlegt diese Aussage nun. Sie können die Aussage, auf die sich die Lehrerin bezieht, hier nachlesen:

"Die meisten meiner Kollegen engagieren sich zusätzlich zum Unterricht und versuchen, Schülern den Spaß an Schule und Lernen neu zu vermitteln. Insbesondere schwierige Schüler brauchen da starke und konstante Beziehungsarbeit. Und dafür eignen sich unter anderem die sozialen Netzwerke, da sie die Kommunikation einfach machen."

Abbildung 11: Informationen zur Vignette erklären, wie für die SLD-Gruppen

In dem Text „Sollen Schüler und Lehrer auf Facebook miteinander befreundet sein?“ kommt diese Aussage vor:

"Die meisten meiner Kollegen engagieren sich zusätzlich zum Unterricht und versuchen, Schülern den Spaß an Schule und Lernen neu zu vermitteln. Insbesondere schwierige Schüler brauchen da starke und konstante Beziehungsarbeit. Und dafür eignen sich unter anderem die sozialen Netzwerke, da sie die Kommunikation einfach machen."

Die Lehrerin ist damit nicht einverstanden. Sie widerlegt diese Aussage nun, das heißt, sie zeigt, wie man darauf reagieren kann.

Abbildung 12: Informationen zur Vignette erklären, wie für die SuS-Gruppe

Da in FALKE-D jeder Erklärgegenstand zwei Mal in jeweils unterschiedlicher Tiefe (s. Punkt *5.1.4 Variation der Erklärtiefe* bzw. Punkte *5.2-5.4* zu den *Was-/ Warum-/ Wie-Erklärungen* in dieser Arbeit) erklärt wurde, wurden die Teilnehmer/-innen nach den ersten drei Vignetten auf die Wiederholung der Themen hingewiesen.

Es war notwendig, die Befragten zunächst alle Videovignetten nacheinander ansehen zu lassen, da das Globalurteil sowie dessen Begründungen für alle Vignetten unbeeinflusst von später gestellten konkreten Fragen zu den Erklärungen gegeben werden sollte. Auf diese Weise können besonders in Verbindung mit den offenen Antworten Rückschlüsse auf Kriterien für eine gute Erklärung gezogen werden, die die befragten Personen intuitiv an eine Erklärung anlegen.

Nach der Vergabe des Globalurteils zu allen sechs Erklärungen folgte eine Situierung bezüglich des nun zu bewertenden *Sprech- und Körperausdrucks.* Um gleiche Bedingungen für alle Bewertungen zu schaffen, wurden die Items dazu an dieser Stelle bereits gezeigt. Die Befragten sahen nun wie oben dargelegt drei etwa 30 Sekunden lange Ausschnitte aus den erstgesehenen Videos (d. h. in Fragebogen A aus den vertieften Erklärungen, in Fragebogen B aus den einfachen) zur Bewertung des *Sprech- und Körperausdrucks* und der *Persönlichkeitswirkung* der erklärenden Lehrkraft. Der erste Messzeitpunkt schloss mit einem Dank an die Teilnehmer/-innen.

Ablauf Messzeitpunkt 2

Auch der zweite Messzeitpunkt begann mit einem Willkommensgruß und einer kurzen Kontextualisierung. Nach der Eingabe des Teilnehmerkodes und der Lautstärkeregulierung sowie

der Erinnerung an die Rahmenbedingungen (d. h. monologische Erklärungen) folgten Erläuterungen zur Notwendigkeit, die Aspekte mit inhaltlich ähnlichen Fragen bewerten zu lassen. Die Kontextualisierungen zu den einzelnen Vignetten waren die gleichen wie zu Messzeitpunkt 1. Erneut wurden alle Vignetten vollständig abgespielt, und zwar für jede/-n Teilnehmer/-in in der gleichen Reihenfolge wie zu Messzeitpunkt 1, das heißt, es wurde entweder zu beiden Messzeitpunkten der Fragebogen in Version A oder in Version B bearbeitet. Nach jeder Vignette wurden alle Items zu den oben beschriebenen Kriterien abgefragt. Dies geschah in der Reihenfolge *Sprachliche Verständlichkeit*, fach- und themenspezifische Aspekte in FALKE-D (aus denen sich die Aspekte *Erklärtiefe* und *Fachliche Relevanz* ergeben), *Strukturiertheit* und *Adressatenorientierung*. Die Fragen zu den einzelnen Aspekten erfolgten direkt nach dem Abspielen einer Vignette auf jeweils einer eigenen Seite, ohne den Aspekt zu benennen (s. Abb. 13 und 14 zu den Items der Aspekte *Erklärtiefe* und *Fachliche Relevanz*). Die Teilnehmer/-innen wussten demnach nur, dass sie Fragen zu fachdidaktischen (SLD-Gruppen) beziehungsweise zu weiteren Aspekten (SuS-Gruppe) erhalten würden. Auf diese Weise wurde vermieden, dass sich eigene Vorstellungen zu den einzelnen Aspekten, die allein durch zum Beispiel das Stichwort „Strukturiertheit" hervorgerufen werden, auf die Bewertung der einzelnen Punkte auswirken.

Nach den ersten drei Vignetten wurde zum zweiten Messzeitpunkt der Hinweis auf den gleichen Inhalt in den letzten drei Vignetten wiederholt.

Zur Veranschaulichung sind im Folgenden die Fragebogenseiten zu *Erklärtiefe* und *Fachlicher Relevanz* der vertieften Wie-Erklärung für die Statusgruppen SLD (Abb. 13) und für die Schüler/-innen (Abb. 14) zu sehen.

	Stimme voll zu	Stimme zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Stimme gar nicht zu
Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt.	○	○	○	○	○	○
Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen.	○	○	○	○	○	○
Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.	○	○	○	○	○	○
Die Erklärung war kognitiv aktivierend.	○	○	○	○	○	○
Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.	○	○	○	○	○	○

	Stimme voll zu	Stimme zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Stimme gar nicht zu
Das „laute Denken“, so wie es im Video gezeigt wurde, hat geholfen zu verstehen, wie man die Aussage widerlegt.	○	○	○	○	○	○
Das „laute Denken“, so wie es im Video gezeigt wurde, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt.	○	○	○	○	○	○
Das „laute Denken“, so wie es im Video gezeigt wurde, sollte im Schreibunterricht regelmäßig verwendet werden.	○	○	○	○	○	○

Abbildung 13: Fragebogenseiten zu den Aspekten Erklärtiefe und Fachliche Relevanz für die SLD-Gruppen

	Stimme voll zu	Stimme zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Stimme gar nicht zu
Das „laute Denken“, so wie es im Video gezeigt wurde, hat geholfen zu verstehen, wie man die Aussage widerlegt.	○	○	○	○	○	○
Das „laute Denken“, so wie es im Video gezeigt wurde, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt.	○	○	○	○	○	○
Das „laute Denken“, so wie es im Video gezeigt wurde, sollte im Schreibunterricht regelmäßig verwendet werden.	○	○	○	○	○	○

	Stimme voll zu	Stimme zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Stimme gar nicht zu
Die Erklärung hat geholfen, das Erklärte genau zu verstehen.	○	○	○	○	○	○
Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.	○	○	○	○	○	○
Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.	○	○	○	○	○	○
Die Erklärung hat mich zum Mitdenken angeregt.	○	○	○	○	○	○

Abbildung 14: Fragebogenseiten zu den Aspekten Erklärtiefe und Fachliche Relevanz für die SuS-Gruppe

Um möglichst gleiche Bedingungen zu schaffen, mussten sämtliche Vignetten zu beiden Messzeitpunkten vollständig abgespielt werden, bevor die Teilnehmer/-innen mit der Beantwortung

der Items beginnen konnten. Sämtliche Items zu den abgefragten Aspekten waren geschlossen und auf einer Ratingskala von 1 =*stimme voll zu* bis 6 =*stimme gar nicht zu* zu beantworten. Um möglichst vollständige Angaben zu erhalten, waren alle Items Pflichtangaben.
Der Fragebogen schließt mit einem Dank an die Teilnehmer/-innen sowie in den SLD-Gruppen mit dem Angebot, über die Ergebnisse der Studie informiert zu werden.
Im nachfolgenden Punkt werden die an der FALKE-D-Studie Teilnehmenden sämtlicher Statusgruppen sowie die Umstände der Befragungen genau beschrieben.

7.4 Statusgruppen und Erhebung

Wie oben bereits dargelegt, wird das Konstrukt des guten Erklärens in dieser Studie aus unterschiedlichen Perspektiven beleuchtet, um es möglichst ganzheitlich zu erfassen. Wie bei allen Studien, die sich mit nicht direkt mess- und sichtbaren Gegenständen wie Qualität befassen, muss die in FALKE-D untersuchte Erklärqualität als Wahrnehmung von Erklärqualität gemessen werden (s. Punkt *2.4 (Professionelle) Wahrnehmung von Unterrichtsqualität* in dieser Arbeit; einen Überblick über die Gesamtstichprobe gibt zudem Tab. 2).

7.4.1 Expertinnen und Experten

In Studien zur wahrgenommenen Qualität ist die Perspektive von Expertinnen und Experten besonders interessant. Zum Erklären steht wie erläutert eine relativ breite fachunabhängige Forschungstradition in der Lehr-/Lernpsychologie recht unterschiedlichen Forschungsständen in den einzelnen Fachdidaktiken gegenüber. In der Deutschdidaktik gibt es bislang wenig Forschung, die sich mit dem fachspezifischen Erklären beschäftigt. Obwohl die Lehr-/Lernforschung eine wesentlich höhere Expertise auf diesem Gebiet aufweist, schloss sich aufgrund des explizit deutschdidaktischen Fokus dieser Studie eine Befragung von pädagogischen Psychologinnen und Psychologen als Expertinnen und Experten für Deutsch-Erklärungen jedoch aus.
Unter Berücksichtigung oben geschilderter Grundsätze zur professionellen Wahrnehmung (s. Punkt *2.4.2 Beobachterperspektive: Wahrnehmung der Expertinnen und Experten* in dieser Arbeit) wurde entschieden, wer im Gesamtprojekt FALKE als Expertin/Experte für das Erklären anzusehen ist. Dabei wurden drei unterschiedliche Perspektiven berücksichtigt.
Für die FALKE-D-Studie wurden als erste Expertengruppe Deutschdidaktiker/-innen aufgrund ihres professionellen Wissens über deutschspezifische Lehr-/Lernprozesse ausgewählt. Obwohl sich wenige von ihnen explizit mit dem Erklären im Fach auseinandersetzen, so sind sie doch einerseits in ihrer Lehre selbst Erklärende und andererseits aufgrund ihrer Profession Analysierende und Bewertende von Unterrichtsqualität, unter die das Erklären als Lehrerhandlung fällt. Aufgrund ihrer

vertieften Ausbildung werden Didaktiker/-innen in dieser Studie als herausragende Ausübende ihrer Profession (vgl. Besser & Krauss 2009, 76) verstanden.

Neben den Didaktikerinnen und Didaktikern verfügen auch Lehrkräfte über Expertise zum Erklären. Auch sie sind, gerade beim Ausüben einer Funktionsstelle wie Seminarlehrkraft (s. Punkt *7.4.1.2 Lehrkräfte* in dieser Arbeit), mit dem Metablick auf Unterricht vertraut sowie im Ausüben ihrer Profession selbst Erklärende.

Schließlich wurden Studierende eines Lehramtes als zukünftige Expertinnen und Experten ebenfalls in die Befragung eingeschlossen. Zum einen findet während des Studiums immer wieder Reflexion von Unterricht statt (in Seminaren, aber auch „live" bei Unterrichtsversuchen der Studierenden), so dass die Studierenden mit der Metaperspektive auf Unterricht vertraut werden. Zum anderen erlaubt das Aufnehmen dieser Gruppe durch den Vergleich mit erfahrenen Lehrkräften und Fachdidaktiker/-innen einen Blick auf den Erwerbsweg der Bewertung von Unterrichtsqualität.

Somit bilden sich die Expertinnen und Experten im Gesamtprojekt FALKE aus den drei genannten Statusgruppen. Wie sich diese Statusgruppen in FALKE-D genau zusammensetzen, wird in den sich anschließenden Punkten in der Reihenfolge Didaktiker/-innen (s. Punkt *7.4.1.1* in dieser Arbeit), (Seminar-) Lehrkräfte (s. Punkt *7.4.1.2* in dieser Arbeit) sowie Studierende (s. Punkt *7.4.1.3* in dieser Arbeit) dargelegt.

Die Expertise der Statusgruppen soll wie bereits dargelegt in der Auswertung der Daten dahingehend genutzt werden, das Konstrukt des *guten Erklärens* genauer zu beschreiben und zu definieren. Über die professionelle Wahrnehmung werden Rückschlüsse auf den eigentlichen Untersuchungsgegenstand, das Erklären, gezogen. Dazu werden die Bewertungen der Expertinnen und Experten einerseits denen der Adressatinnen und Adressaten gegenübergestellt. Andererseits sind aber auch innerhalb der Expertengruppe die Aussagen der unterschiedlichen Statusgruppen getrennt voneinander zu betrachten. Hier ist vor allem die Frage nach der Validität der Bewertungen interessant, aber auch die Homogenität beziehungsweise Heterogenität untereinander und zueinander.

Für alle befragten Personen war die Teilnahme an der Studie anonym und freiwillig, sie erfolgte unentgeltlich und konnte jederzeit abgebrochen werden.

Die Zusammensetzung der einzelnen Statusgruppen bei FALKE-D wird im Folgenden genauer beschrieben.

7.4.1.1 Didaktiker/-innen

Grundlagen

Bereits die Regensburger FALKO-Gruppe einigte sich in Anlehnung an die Taxonomie Shulmans (1986, 1987) auf die Konzeptualisierung des fachdidaktischen Wissens als (1) Wissen über Erklären

und Repräsentieren von Inhalten, (2) Wissen über typische Schülerschwierigkeiten und -fehler sowie (3) Wissen über das Potenzial von Lernmaterialien (vgl. Krauss, Lindl, Schilcher, Fricke et al., 2017 bzw. Ausführungen weiter oben unter Punkt *2.1.1 Modelle zur Konzeptualisierung von qualitätvollem (Deutsch-) Unterricht*).

Auch in FALKE wird das Erklären dem fachdidaktischen Wissen zugeordnet, obwohl natürlich auch Fachwissen und pädagogisches Wissen notwendig sind, um eine gute Erklärung zu geben (s. dazu in dieser Arbeit Ausführungen in Punkt *2.3 Erklärkompetenz von Lehrkräften* sowie in Punkt *4.1 Das Konstrukt Gutes Erklären*). Aufgrund dieser Zuordnung ist es naheliegend, Didaktiker/-innen als Expertinnen/Experten zu Erklärungen heranzuziehen. Diese kommen wie angedeutet selbst immer wieder in die Lage, Lehr-/Lernsituationen, sei es an der Schule, sei es an der Universität, aus der Metaperspektive zu betrachten und im Hinblick auf ihre Qualität und/ oder Wirksamkeit zu analysieren. Dieser „geschulte" Blick auf die im Fragebogen gezeigten Erklärungen erlaubt in der Auswertung der Daten Rückschlüsse auf die Qualität von Erklärungen. Oben beschriebene professionelle Wahrnehmung beziehungsweise die „wissensgesteuerte Identifikation von Situationen und Ereignissen im Unterricht" (Seidel et al., 2010, S. 297) der befragten Expertinnen und Experten macht sich FALKE-D so zunutze.

Abgesehen davon, dass die professionelle Wahrnehmung der Didaktiker/-innen zum Beschreiben des guten Erklärens genutzt werden soll, wird natürlich auch davon ausgegangen, dass die Didaktikerinnen und Didaktiker die theoretischen Konzepte, die jeweils hinter den Videos stehen, und insbesondere die Variation, die zwischen den jeweiligen „Paaren" erzeugt wird, (er)kennen. In der Auswertung wird ein besonderes Augenmerk darauf gelegt, inwieweit die befragten Didaktiker/-innen diese zum Beispiel in den Begründungen der Globalurteile explizit nennen. In der weiteren Auswertung ist im Speziellen darauf zu achten, ob über die didaktische Qualität dieser Konzeptionen Einigkeit unter den befragten Expertinnen und Experten besteht oder nicht. Dabei ist der in dieser Studie erhobene deutschspezifische Aspekt der *Erklärtiefe* von besonderem Interesse. Entscheidend ist hier die Frage, ob die Didaktiker/-innen davon ausgehen, dass Erklärungen mit größerer Erklärtiefe gelungener sind als Erklärungen mit weniger Erklärtiefe.

Statusgruppe Didaktiker/-innen in FALKE-D

Für FALKE-D wurden Didaktikerinnen und Didaktiker in Deutschland und Österreich um Mitarbeit gebeten, die an Universitäten oder Hochschulen lehren. Zunächst wurden Professorinnen und Professoren angesprochen, die aufgrund ihres beruflichen Werdegangs sowie ihrer langjährigen Erfahrung die „professionellsten" Expertinnen und Experten in dieser Gruppe darstellen. Sowohl die Dauer der Berufstätigkeit als auch die beruflichen Erfolge zählen nach Bromme (1992) zu Indikatoren des Expertenstatus (vgl. Krauss & Bruckmaier, 2014, S. 249–250). Insbesondere

wurden, obwohl die Forschungslage im Bereich Erklärungen in den Geisteswissenschaften beziehungsweise in der Deutschdidaktik wie dargelegt defizitär ist, gezielt die (wenigen) Didaktiker/-innen angesprochen, die selbst bereits zu Erklärungen geforscht haben. Diese sind tatsächliche „Erklärexpertinnen und -experten“[45].

Um auf die erforderliche Anzahl an Teilnehmenden von N_{min}=30 zu kommen, wurde die Gruppe dann erweitert um promovierte Mitarbeiter/-innen an den Lehrstühlen für Didaktik der deutschen Sprache und Literatur sowie schließlich um (noch) nicht promovierte Mitarbeiter/-innen. Für letztere war eigene Lehrerfahrung an Schulen Voraussetzung für die Teilnahme (d. h. konkret mindestens Abschluss des Referendariats an der Schule).

Sämtliche Didaktiker/-innen wurden persönlich per E-Mail eingeladen, an der Erhebung teilzunehmen. Sie wurden gebeten, die beiden Teile der Erhebung mit einem maximalen zeitlichen Abstand von drei Wochen zu erledigen. Die Links zu den Fragebögen wurden ihnen zugeschickt. Insgesamt konnten N= 31 Didaktikerinnen und Didaktiker gewonnen werden (17 weiblich, 14 männlich), von denen die Mehrzahl über langjährige Berufserfahrung und hohe Expertise verfügen, da sie einen Doktor- (N=7) beziehungsweise Professorentitel (N=18) erwarben. Der Erhebungszeitraum erstreckte sich von Oktober 2017 bis Juni 2018.

7.4.1.2 Lehrkräfte

Grundlagen

Als zweites war es naheliegend, diejenigen als Expertengruppe zu befragen, die tagtäglich in der Schule erklären, in diesem Fall also Deutschlehrerinnen und –lehrer aus der Sekundarstufe. Ihnen sind die institutionellen Bedingungen, die die Schule den Erklärsituationen auferlegt, „in Fleisch und Blut“ übergegangen. Abgesehen davon, dass Lehrkräfte selbst Erklärende sind, bewerten sie in ihrem Alltag auch die Güte von (schriftlichen) Erklärungen, zum Beispiel aus Schulbüchern. Außerdem wurde bei der Auswahl der Studienteilnehmer/-innen versucht, gezielt Lehrer/-innen auszuwählen, die selbst oft in der Rolle der/des Bewertenden von Unterricht (und somit naturgemäß auch von Erklärungen aus diesem Unterricht) sind, damit sie mit der Metaperspektive, die sie auch beim Beurteilen der Videos dieser Studie einnehmen sollen, möglichst vertraut sind.

Dies trifft insbesondere auf Seminarlehrkräfte zu, zu deren täglicher Arbeit es gehört, diese Metaperspektive auf Unterricht einzunehmen, da sie im Zuge der Betreuung ihrer Referendarinnen und Referendare täglich Unterricht besuchen und bewerten. Sie verfügen in diesem Sinne über eine „doppelte“ Expertise, ähnlich den Didaktikerinnen und Didaktikern, nämlich die der professionellen Erklärer/-innen und die der professionellen Bewerter/-innen von (Erklärungen im) Unterricht.

[45] Aufgrund der Anonymität der Befragung kann allerdings nicht nachvollzogen werden, wie viele dieser Expertinnen/Experten tatsächlich teilgenommen haben.

Eine ähnliche Routine haben Betreuungslehrkräfte, die Referendarinnen und Referendare während des Zweigschuleinsatzes begleiten und dabei verpflichtet sind, deren Unterricht zu besuchen und zu befeedbacken.

Praktikumslehrkräfte wiederum sind mit der Betreuung von Novizinnen und Novizen betraut, das heißt von Studierenden, die während des Studiums Praktika an Schulen absolvieren und dort erste eigene Lehrversuche unternehmen. Praktikumslehrkräfte besprechen diese Versuche ausführlich mit den Studierenden; außerdem führen sie auch selbst Stunden in Anwesenheit ihrer Praktikantinnen und Praktikanten durch, die anschließend ebenso besprochen werden. Es kann also davon ausgegangen werden, dass auch Praktikumslehrkräfte mit dem für die Erklärungen der Videos in der Studie erforderlichen Metablick vertraut sind.

Auch aus den Ergebnissen der Statusgruppe Lehrkräfte sollen aufgrund deren Expertise Rückschlüsse auf die Qualität von Erklärungen gezogen werden. Ganz ähnlich wie bei der Statusgruppe der Didaktiker/-innen (s. *7.4.1.1 Didaktiker/-innen* in dieser Arbeit) soll auch die Professionalität dieser Gruppe genutzt werden, um dem guten Erklären im Deutschunterricht auf die Spur zu kommen. Daher ist es selbstverständlich, dass es auch bei der Befragung dieser Statusgruppe nicht das Ziel war, die Studienteilnehmenden zu testen. Interessant ist hierbei gerade der Vergleich der Ergebnisse der Didaktiker/-innen und der Lehrkräfte: Worin sind sich diese beiden Gruppen einig und worin unterscheiden sie sich? Während bei den Didaktikerinnen und Didaktikern davon ausgegangen wird, dass diese besonders auf die didaktischen Konzeptionen, die hinter den gezeigten Erklärungen stehen, achten, so wird bei den Lehrkräften davon ausgegangen, dass diese stärker die Tauglichkeit der Erklärungen für den Unterricht bewerten, da dies die Perspektive ist, mit der Lehrkräfte Unterricht in ihrer täglichen Arbeit beobachten, beschreiben und evaluieren.

Wie bereits in Punkt 5 zur *Konstruktion der Videos als Stimuli* in dieser Arbeit beschrieben wurde bei der Erstellung der Videos für vorliegende Studie die hohe Komplexität des Unterrichts bewusst reduziert, indem „allein" Lehrererklärungen ohne Interaktion mit Schülerinnen und Schülern aufgezeichnet wurden. Die befragten Personen konnten sich so ganz auf das Lehrerhandeln konzentrieren und erhielten auch die entsprechende Anweisung. Nichtdestotrotz liegt die Vermutung nahe, dass die Lehrkräfte beim Betrachten der Videos Unterschiedliches als relevant wahrnehmen und deswegen auch Unterschiedliches bewerten, – je nach ihrer persönlichen Erfahrung und sicherlich auch je nach ihrer persönlichen Überzeugung (vgl. Erickson, 2011). Bei der Interpretation der Daten wird besonders darauf geachtet, ob sich diese Vermutung bestätigt oder nicht.

Statusgruppe Lehrkräfte in FALKE-D

Aufgrund der Erreichbarkeit wurden Lehrkräfte befragt, die im Freistaat Bayern, vor allem im Bezirk der Oberpfalz, unterrichten. Aufgrund der Themen der Erklärungen aus der Sekundarstufe wurden Lehrkräfte der Schularten Gymnasium, Realschule beziehungsweise Mittelschule ausgewählt.

Die Erhebung der eigentlichen Daten begann nach der Pilotierung der Videos sowie der Items Anfang Oktober 2017 im Rahmen eines Seminarlehrertages an der Universität Regensburg. Mit dem Ziel, eine engere Kooperation zwischen der ersten und der zweiten Lehrerbildungsphase anzustoßen, wurden Seminarlehrer/-innen sämtlicher an FALKE beteiligten Fächer sowie sämtlicher Schularten der Bezirke Niederbayern und Oberpfalz an die Universität Regensburg eingeladen. Die Lehrkräfte wurden im Vorfeld gebeten, an der Erhebung teilzunehmen, die Bearbeitung des ersten Messzeitpunktes war Voraussetzung für die Teilnahme am Kooperationstag. Die Bearbeitung des zweiten Messzeitpunktes fand während des Kooperationstages an der Universität statt. Der Seminarlehrertag bot die Möglichkeit, den teilnehmenden Lehrkräften die Ergebnisse des ersten Messzeitpunktes vorzustellen und mit ihnen über diese Ergebnisse sowie über die Studie insgesamt ins Gespräch zu kommen. Die anwesenden Lehrkräfte zeigten sich ob der Spannweite der Ergebnisse bei dem Globalurteil der Videos überrascht: Es gab für nahezu jedes Video Bewertungen der Note 1 bis Note 5 und zwar in allen Fächern (s. dazu auch den Punkt *10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse* in dieser Arbeit).

Nahezu alle Lehrkräfte dieser Befragung haben eine der oben genannten Funktionen inne: N=23 sind Seminarlehrkräfte, N=3 Betreuungslehrkräfte und N=2 Praktikumslehrkräfte. Um auch hier auf die erforderliche Mindestanzahl von N_{min}=30 zu kommen, wurde die Statusgruppe der Lehrkräfte mit N=7 Lehrkräften im aktiven Schuldienst ohne eine der genannten Funktionen aufgefüllt; somit ergibt sich insgesamt eine Gesamtanzahl für die Statusgruppe Lehrkräfte von N=35, und zwar 18 Lehrerinnen und 17 Lehrer. Die Befragung der Lehrkräfte endete im Juni 2018.

7.4.1.3 Studierende

Grundlagen

Als dritte und letzte Statusgruppe innerhalb der Expertinnen und Experten wurden Lehramtsstudierende mit Unterrichts- oder Didaktikfach Deutsch für alle Schularten befragt. Was die Expertise anbelangt, so unterscheidet sich diese Gruppe maßgeblich von den vorherigen beiden. Ihre Urteile sind vor allem als die von Novizinnen und Novizen interessant. Durch diese können Hinweise zur Entwicklung von der Novizin zur Expertin beziehungsweise vom Novizen zum Experten erwartet werden. Es wird außerdem vermutet, dass die Urteile der Studierenden eher mit den Urteilen der Schüler/-innen denn mit denen der anderen Expertenstatusgruppen übereinstimmen, ganz einfach deswegen, weil die Studierenden vermutlich eher die Perspektive der Adressatinnen

und Adressaten von Erklärungen einnehmen, da sie noch nicht über einen geschulten Metablick auf Unterricht verfügen und die eigenen jahrelangen Erfahrungen als Schüler/-in noch nicht sehr weit zurückliegen. Zudem sind sie an der Universität wiederum Adressatinnen und Adressaten von Erklärungen. Diese Befragten stehen somit in der Rolle der/des Erklärenden sowie in der Rolle der/des Bewertenden von Unterrichtserklärungen noch am Anfang, – je nach Semesterzahl und absolvierten Praktika. Dennoch finden sich alle Studierenden auch immer wieder in der Rolle der „Erklärerin" beziehungsweise des „Erklärers": Sie müssen zum Beispiel in Referatssituationen an der Universität erklären und kommen durch Praktika in die Situation, Schülerinnen und Schülern etwas zu erklären beziehungsweise gesehene Erklärungen von Kommilitoninnen und Kommilitonen zu befeedbacken.
Bei der Interpretation der Daten ist zu überprüfen, ob die oben genannten Vermutungen zutreffen, oder ob Studierende in ihrer Bewertung eventuell viel mehr Lehrkräften oder Didaktikerinnen und Didaktikern ähneln.

Statusgruppe der Studierenden in FALKE-D

Die Studienteilnehmer/-innen dieser Statusgruppe sind Lehramtsstudierende (66 weiblich, 12 männlich) aller Schularten für das Fach Deutsch der Universität Regensburg, der Universität Augsburg sowie der PH Ludwigsburg. Die Zahl von insgesamt N=78 Studierenden teilt sich wie folgt auf die Schularten auf: 54 Studierende des Lehramtes Grundschule, 3 des Lehramtes Mittelschule, 4 des Lehramtes Realschule, 15 des Lehramtes Gymnasium sowie je 1 Studierende(r) der Sonderpädagogik beziehungsweise der Germanistik. Von den Mittelschul- und Grundschulstudierenden hatten 28 Deutsch als Didaktikfach und 29 als Unterrichtsfach gewählt. Die Studierenden befanden sich in unterschiedlichen Phasen ihres Studiums, durchschnittlich im 6,5. Fachsemester (bei einer Standardabweichung von 2,5 Semestern), mindestens jedoch im dritten Fachsemester (d. h. es wurden keine Studienanfänger/-innen befragt) und höchstens im elften Fachsemester.
Die Studierenden an der Universität Regensburg wurden im Zuge von Seminaren befragt; der zeitliche Abstand zwischen den beiden Erhebungszeitpunkten betrug hier drei Wochen. Dozierende an der Universität Augsburg sowie an der PH Ludwigsburg luden Studierende aus ihren Kursen ein, an der Erhebung teilzunehmen; diese bekamen die Links zur Erhebung per E-Mail zugeschickt, mit der Bitte zwischen den beiden Messzeitpunkten nicht mehr als drei Wochen verstreichen zu lassen. Die Statusgruppe der Studierenden wurde im Wintersemester 2017/18 und im Sommersemester 2018 erhoben.

7.4.2 Adressaten: Schüler/-innen

Grundlagen

Dass Schüler/-innen differenzierte Einschätzungen zur Unterrichtsqualität geben können, wurde oben bereits dargelegt (s. Punkt *2.4.1 Beobachterperspektive: Wahrnehmung der Schüler/-innen* in dieser Arbeit). Diese Fähigkeit macht sich FALKE für die Bewertung von Erklärungen als Mikrodimension von Unterricht zunutze. Im Unterschied zu oben dargestellten Studien wurden in der Studie FALKE-D Schüler/-innen zu fremdem Unterricht und somit auch zu fremden Erklärenden befragt. Sie unterlagen somit den gleichen Bedingungen wie die Expertengruppen und hatten als Bewertungsgrundlage nur die gezeigte dreiminütige Erklärung.

Diese Entscheidung hat auf der einen Seite den Nachteil, dass es den Schülern und Schülerinnen vermutlich schwerer fällt, einen so kurzen Ausschnitt fremden Unterrichts zu bewerten. Andererseits hat sie den Vorteil, dass Verzerrungen, die zum Beispiel aufgrund von Sympathie oder Abneigung gegen die eigene Lehrkraft bestehen und die nachweislich das Beurteilungsverhalten gerade von jüngeren Schülerinnen und Schülern (d. h. Primar- bis etwa Mittelstufe) beeinflussen können (vgl. Wisniewski et al., 2020), ausgeschlossen werden.

Während alle drei Experten-Statusgruppen die gezeigten Erklärungen aus der Metaperspektive beurteilen sollten, indem sie die Verständlichkeit für Schüler/-innen einschätzten, wurde von den Schülerinnen und Schülern diese Metaperspektive nicht verlangt. Sie sollten die Güte der Erklärungen vielmehr aus ihrer eigenen Sicht beurteilen. Schüler/-innen sind im Laufe eines Schultages Adressatinnen und Adressaten einer Vielzahl von Erklärungen in den unterschiedlichen Fächern. Diese liegen zum Teil in schriftlicher Form in Schulbüchern oder auf Arbeitsblättern vor. Zumindest im Deutschunterricht wird jedoch die Mehrzahl der Erklärungen im Unterricht mündlich präsentiert, und zwar manchmal von Mitschülerinnen und Mitschülern, meistens jedoch von der Lehrkraft (zum Vorkommen und zur Relevanz von Lehrererklärungen in der Sprachdidaktik im Deutschunterricht s. Punkt *3.4 Erklärungen im (Deutsch-)Unterricht* bzw. Punkt *5.1.1 Themenfindung* in dieser Arbeit).

Die Statusgruppe der Adressatinnen und Adressaten nimmt in dieser Studie eine besondere Rolle ein: Zunächst ist das Gelingen des Erklärprozesses immer maßgeblich von der Adressatin beziehungsweise vom Adressaten abhängig, die/der ihr/sein eigenes Wissensdefizit (an)erkennen und gewillt sein muss, dieses zu beheben. Dazu muss die/der Erklärende in dieser Rolle angenommen werden. In der „Zwangskommunikation" der Schule ist dieser Punkt etwas anders gelagert als bei „herkömmlichen", alltäglichen Erklärsituationen (s. Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit), aber dennoch zu beachten.

Aus Sicht der Adressatinnen und Adressaten ist das Ziel der Erklärung, das Verstehen, der wichtigste Schritt im Erklärprozess. Denn der beste Erklärvorgang ist hinfällig, wenn die Adressatin/der

Adressat die Erklärung nicht versteht. Um das Verstehen empirisch zu überprüfen, muss vor und nach der Erklärung ein Verständnistest durchgeführt werden. Das konnte in dieser Studie aus Zeitgründen nicht geleistet werden. Der Erklärvorgang ist allerdings auch dann schon nicht gelungen, wenn die Adressatinnen und Adressaten „nur" das Gefühl haben, nicht verstanden zu haben, und dies ist in vorliegender Studie entscheidend. Die Schülerinnen und Schüler wurden explizit aufgefordert, die Erklärungen aus ihrer eigenen Sicht zu beurteilen, also nicht zu überlegen, ob die Klasse oder Mitschüler/-innen mit der Erklärung zurechtkommen, sondern ob sie selbst diese Erklärung gelungen finden oder eben nicht, beziehungsweise ob sie selbst das Gefühl haben, die Erklärung verstanden zu haben oder nicht.

Schüler/-innen sind mit der Sprechhandlung Erklären im Kontext des Unterrichts vertraut und können deswegen in FALKE-D als „professionelle" Adressatinnen und Adressaten von schulischen Erklärungen gelten. Es wird allerdings davon ausgegangen, dass die Lernenden die gesehenen Erklärungen auf der Grundlage ihrer eigenen Erfahrungen im Deutschunterricht bewerten. Schüler/-innen vergleichen sicherlich die gesehenen Erklärungen mit dem Unterricht, den sie kennen, ganz einfach deswegen, weil das ihr vertrauter Kontext ist und weil sie die Erklärungen aus der Adressatenperspektive bewerten sollen. Wie bei den Expertenstatusgruppen muss auch bei der Interpretation der Schülerantworten mit bedacht werden, aus welchem Umfeld sie kommen und welchen Hintergrund sie vermutlich im Kopf haben. Bei den Schülerinnen und Schülern wird vermutet, dass die Bewertungen aufgrund der fehlenden Metaperspektive sehr heterogen ausfallen, – je nach persönlichen Vorlieben und Vorerfahrungen in den unterschiedlichen Klassen. Interessant ist hier besonders der Vergleich mit den Expertengruppen: Wo stimmen Urteile überein, wo zeigen sich (starke) Differenzen?

Statusgruppe der Schüler/-innen in FALKE-D

Der Kontakt zu den Lehrkräften während des oben beschriebenen Seminarlehrertages ermöglichte die Teilnahme von vier Klassen an der Studie, die von Lehrkräften, die ihrerseits an der Studie teilgenommen hatten, zur Verfügung gestellt beziehungsweise vermittelt wurden. Befragt wurden insgesamt N=113 Schülerinnen und Schüler, und zwar je aus einer 9. und einer 10. Klasse einer Realschule (insgesamt 54 Schüler/-innen) und eines Gymnasiums (insgesamt 59 Schüler/-innen) in der Oberpfalz. Aus datenschutzrechtlichen Gründen konnte wie oben bereits gesagt das Geschlecht bei dieser Statusgruppe nicht mit erhoben werden. Über die vier befragten Klassen konnte die statistisch erforderliche Anzahl an Schülerinnen und Schülern erreicht werden. Die Befragung der Schülerinnen und Schüler erstreckte sich über den Zeitraum Oktober bis Dezember 2017.

7.4.3 Überblick über Ablauf der Erhebung

Insgesamt dauerte der Erhebungszeitraum von Oktober 2017 bis Juni 2018, es wurden N=257 Studienteilnehmer/-innen in oben dargestellter Verteilung auf die vier Statusgruppen befragt. Gemäß der Anzahl der Gesamtpopulationen wurden die Statusgruppen zahlenmäßig in der Reihenfolge Schüler/-innen – Studierende – Lehrkräfte – Didaktiker/-innen erhoben. Dabei wurde darauf geachtet, dass die Teilnehmenden aus unterschiedlichen Kontexten stammen (d. h. unterschiedliche Schulen und Schularten, unterschiedliche Universitäten und unterschiedliche Studiendauer, unterschiedliche Funktionen an Schulen bzw. Universitäten). Auch die gleichmäßige Aufteilung zwischen weiblichen und männlichen Teilnehmerinnen und Teilnehmern gelang bis auf die Statusgruppe Studierende (s. hierzu auch Tab. 2 zur Übersicht über die Gesamtstichprobe in FALKE-D). Die Stichprobengültigkeit ist somit erfüllt. Die Durchführungsobjektivität ist durch das Befragungsinstrument des Online-Fragebogens gegeben. Die Befragung fand in Teilen bei den Studienteilnehmerinnen und -teilnehmern privat zu Hause am eigenen PC statt (und zwar genauer: bei einem Teil der Studierenden zu MZP1 und/oder MZP2, bei den Lehrkräften mindestens zu MZP1, zum Teil auch zu MZP2 sowie bei den Didaktikerinnen und Didaktikern zu MZP1 und MZP2). In Teilen wurde die Erhebung an der Universität Regensburg in einem Computerraum beziehungsweise an einem mitgebrachten PC (bei der Mehrzahl der Lehrkräfte zu MZP2 sowie bei den Studierenden der Universität Regensburg zu MZP1 und MZP2) durchgeführt. Die durchschnittliche Bearbeitungsdauer betrug für MZP1 35 Minuten und für MZP2 44 Minuten in den Experten-Statusgruppen.

Die Befragung der Schüler/-innen fand zur regulären Unterrichtszeit in den jeweiligen Computerräumen der Schulen statt. Sie erfolgte zu zwei Messzeitpunkten, wobei bei den beiden Klassen am Gymnasium ein zeitlicher Abstand von einer Woche zwischen den Messzeitpunkten lag, an der Realschule ein Abstand von zwei Wochen. Jeder Messzeitpunkt dauerte eine Schulstunde, also 45 Minuten. Zur Erhebung fuhren Mitarbeitende des Projekts an die Schulen und führten sie dort durch. Um gleiche Erhebungsbedingungen zu gewährleisten, wurde vorab pro Messzeitpunkt ein Manual erstellt, das den Ablauf sowie die Einführung in die Erhebung wortwörtlich festlegt. Beide Manuale sind im Anhang 8 (s. Anhang 8.1 und 8.2 Erhebungsmanuale für SuS) einzusehen.

Außerdem wurden Erläuterungen zu jedem Item aller Aspekte erstellt und schriftlich festgehalten, die gegeben wurden, falls während der Erhebung Fragen zu einem der Items auftauchen sollten. Die Erläuterungen zu den Items der fachübergreifenden Kriterien *Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit* sowie *Sprech- und Körperausdruck* und *Persönlichkeitswirkung* wurden im Gesamtprojekt FALKE erarbeitet. Die Erläuterungen zu den deutschspezifischen Aspekten der *Erklärtiefe* und der *Fachlichen Relevanz* wurden naturgemäß nur für die Studie FALKE-D erarbeitet und verwendet. Die Paraphrasierungskataloge zu Messzeitpunkt

eins und zwei sind vollständig unter Anhang 9 abgebildet. Darin werden die Paraphrasierungen in der Reihenfolge, in der sie in der Erhebung abgefragt wurden, aufgeführt (s. Anhang 9.1 und 9.2 zu den Paraphrasierungskatalogen SuS für MZP1 und MZP2).

Es kam während der Erhebung an den Schulen zu folgenden Nachfragen bezüglich einzelner Items: Zum ersten Messzeitpunkt wurde zwei Mal nach dem Item „Die Lehrkraft hatte eine abwechslungsreiche Sprechweise" gefragt (2 Mal 9. Klasse Realschule). Das wurde den Schülerinnen und Schülern so verdeutlicht: *Die Lehrerin / der Lehrer hat NICHT eintönig und NICHT langweilig oder leiernd gesprochen. Die Lehrerin / der Lehrer hat spannend gesprochen und Wichtiges betont* (s. Paraphrasierungskatalog MZP1 SuS Anhang 9.1).

Während des zweiten Messzeitpunktes wurde drei Mal bezüglich des Items „Die Erklärung hatte einen roten Faden" nachgefragt (2 Mal 9. Klasse Realschule, 1 Mal 10. Klasse Gymnasium). Dies wurde den betreffenden Schülerinnen und Schülern folgendermaßen erläutert: *Etwas hat einen roten Faden. Das bedeutet, es geht in der Erklärung von Anfang bis Ende hauptsächlich um eine Sache. Die Lehrerin / Der Lehrer ist nicht abgeschweift oder es ging nicht plötzlich um ganz andere Dinge. (Das kennst du bestimmt vom Aufsatzunterricht / Geschichten schreiben)* (s. Paraphrasierungskatalog MZP2 SuS Anhang 9.2).

Außerdem wurde zwei Mal das Item: „Die Erklärung hat nur oberflächliche Zusammenhänge erläutert" nicht verstanden (1 Mal 9. Klasse Realschule, 1 Mal 10. Klasse Realschule). Die Erläuterung hierzu lautete: *Die Erklärung war so, dass du zwar ungefähr verstehst, worum es ging, aber nicht ganz genau* (s. Paraphrasierungskatalog MZP2 SuS Anhang 9.2). Die Erläuterungen waren in allen sieben Fällen ausreichend, weitere Nachfragen kamen nicht.

Bei den Erhebungen an den Schulen und an der Universität waren jeweils zwei Projektmitarbeitende zur Aufsicht anwesend. Die Erhebungen zu Hause fanden naturgemäß ohne Aufsicht statt.

Wie oben dargelegt konnten, um möglichst gleiche Bedingungen zu schaffen, die Fragen erst nach dem vollständigen Abspielen der jeweiligen Vignette angesehen und beantwortet werden. Sämtliche Items bis auf die offenen Begründungen waren zudem Pflichtangaben. Aufgrund des Vergebens von Kodewörtern (bei Schülerinnen und Schülern) beziehungsweise des Erstellens eines anonymisierten Kodes war die Zuordnung der beiden Fragebogenteile zueinander problemlos möglich, ein Rückschluss auf die befragten Personen allerdings unmöglich.

Einen Überblick über die Gesamtstichprobe gibt Tabelle 2 aus Gaier & Schilcher (2021, in Vorb.).

Tabelle 2: Gesamtstichprobe in FALKE-D

Statusgruppe	**N**	**Geschlecht** w/m/d	**Schulart** GS/MS/RS/GY/So	**Jahrgangsstufe bzw. Erfahrung** *M (SD)*	**Funktion bzw. Akademischer Grad**
Schüler/-innen	113	---	---/---/54/59/---	Jgst.: 9 / 10	---
Studierende	78	66/12/---	54/3/4/15/2*	FS.: *6,5 (2,5)*	---
Lehrkräfte	35	18/17/---	---/6/11/18/---	---	2 PL/3 FB/23 SL
Didaktiker/-innen	31	17/14/---	---	---	7 Dr./18 Prof.
∑	257	101/43/---	54/9/69/92/2*	---	---

Bem.: GS: Grundschule, MS: Mittelschule, RS: Realschule, GY: Gymnasium; So: Sonstiges; M: Mittelwert, SD: Standardabweichung; Jgst.: Jahrgangsstufe; FS: Fachsemester, PL: Praktikumslehrkraft, FB: Fachbetreuer/-in, SL: Seminarlehrkraft. *1 Sonderschulpädagogik, 1 Germanistik; (Tabelle aus Gaier & Schilcher, 2021, in Vorb.)

7.5 Operationalisierung des Aspektes *Erklärtiefe*

Die Konzeptualisierung des FALKE-D-Aspektes *Erklärtiefe* wurde in Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit bereits ausführlich dargelegt. Von Erklärungen mit größerer Erklärtiefe wird gesprochen, wenn die Erklärungen einen Schwerpunkt auf verborgene Zusammenhänge legen und so mehr Potenzial zur kognitiven Aktivierung der Adressatinnen und Adressaten aufweisen. Um Hinweise darauf zu bekommen, inwieweit die Erklärtiefe einen Einfluss auf die Wahrnehmung von Erklärqualität hat, wurden, wie oben erläutert, in FALKE-D jeweils zwei Erklärungen zum gleichen Thema erstellt. Diese unterscheiden sich in der Erklärtiefe, da Variante eins jeweils mehr Erklärtiefe aufweist als Variante zwei. Im Folgenden wird erläutert, wie die Erklärtiefe bei FALKE-D operationalisiert wurde.

Um zu überprüfen, ob die Studienteilnehmenden die unterschiedlichen Varianten wahrnehmen und ob sie eine der beiden Varianten aufgrund der umgesetzten Erklärtiefe als besser bewerten, wurden in allen FALKE-D-Erklärungen die nun folgenden drei Items zur *Erklärtiefe* einheitlich bei jeder Vignette abgefragt.

Zunächst ging es darum zu erfassen, ob die Studienteilnehmer/-innen auf der einen Seite die unterschiedlich gestaltete Erklärtiefe auch so wahrnehmen, wie in der Umsetzung angestrebt und ob sie eine der beiden Erklärvarianten aufgrund der Erklärtiefe als besonders hilfreich einschätzen. Dazu wurden zwei gegenläufige Items gefragt, nämlich: „Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen“ (für SLD) sowie „„Die Erklärung hat geholfen, das Erklärte genau zu verstehen“ (für SuS) beziehungsweise die dazu gegengesetzte Frage „Die Erklärung hat nur oberflächliche Zusammenhänge erläutert“ (für alle Statusgruppen), deren Ergebnisse in den Analysen umkodiert wurden. Die beiden Items dienen der inhaltlichen Absicherung der Bewertungen, da sie auf das Gleiche abzielen. In den Analysen ist ein besonderes Augenmerk darauf

zu legen, ob die Wahrnehmung der beiden Vignetten zum gleichen Thema sich in Bezug auf die Erklärtiefe (deutlich) unterscheidet, das heißt konkret, ob die vertieften Varianten besser oder schlechter bewertet werden als die einfacheren Varianten.

Die wahrgenommene kognitive Aktivierung wurde schließlich bei den Expertinnen und Experten über „Die Erklärung war kognitiv aktivierend" direkt erfragt. Da dieser Fachbegriff bei der Statusgruppe der Schüler/-innen nicht vorausgesetzt werden kann, bewerteten diese stattdessen folgende Aussage: „Die Erklärung hat mich zum Mitdenken angeregt". Inhaltlich zielen diese beiden Items auf die wahrgenommene kognitive Aktivierung, in die Berechnungen sind sie deshalb als ein Item eingeflossen. Auch hier ist natürlich von besonderem Interesse, ob und welche systematischen Unterschiede sich in der Wahrnehmung der unterschiedlichen Erklärtiefe zum gleichen Thema zeigen und wie sich die Bewertung dieser Frage zu den Bewertungen der anderen Aspekte verhält.

Die drei Items, die in allen Vignetten von FALKE-D identisch abgefragt wurden, finden sich zur besseren Übersicht in Tabelle 3, eine Übersicht über sämtliche FALKE-D-Items überdies in Anhang 3.2.

Tabelle 3: Identische Items für alle Vignetten zum Aspekt Erklärtiefe

ET1	Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen. (für SLD) / Die Erklärung hat geholfen, das Erklärte genau zu verstehen. (für SuS)
ET2	Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.
ET3	Die Erklärung war kognitiv aktivierend. (für SLD) / Die Erklärung hat mich zum Mitdenken angeregt. (für SuS)

Bem.: Alle Items wurden auf einer Ratingskala von 1 = *stimme voll zu* bis 6 = *stimme gar nicht zu* beurteilt

Abgesehen von diesen drei Aussagen bewerteten die beteiligten Personen nach jeder Vignette noch zwei Items zur wahrgenommenen Erklärtiefe, die in ihrer Formulierung jeweils an den konkreten Erklärgegenstand und die konkrete Erklärmethode der Vignette angepasst wurden, inhaltlich jedoch immer auf die gleiche Fragestellung abzielten.

Das erste Item (ET4) bezog sich auf das wahrgenommene Verständnis der Erklärung. Alle Vignetten wurden so konzipiert, dass die Erklärungen für die Adressatinnen und Adressaten (d. h. Neunt- bzw. Zehntklässler/-innen) gut verständlich sein sollten. Die Auswertung der offenen Begründungen (s. Punkt *9 Qualitative Ergebnisse* in dieser Arbeit) ergab, dass dies zu großen Teilen auch von den befragten Expertinnen und Experten so eingeschätzt wurde.

Die Ergebnisse dieses Items müssen allerdings mit der gebotenen Vorsicht interpretiert werden, denn wahrgenommenes Verstehen kann trügerisch sein: „the sense of understanding is not a promising route to genuine understanding" (Trout, 2007, S. 586). Tatsächliches Verstehen müsste

wie gesagt über einen Verstehenstest im Anschluss geprüft werden. Die Ergebnisse können aber Hinweise auf (Unterschiede im) Verstehen geben, denen in weiterer Forschung nachgegangen werden muss.

Das zweite Item (ET5), das jeweils passgenau formuliert wurde, erfragt die (wahrgenommene) Merkfähigkeit, die die Erklärung auslöst. Hier kommt die Variation der Erklärtiefe ins Spiel. Es wurde wie oben dargelegt davon ausgegangen, dass die Erklärungen der Variante eins mit größerer Erklärtiefe eine bessere Merkfähigkeit generieren als die Erklärungen der Variante zwei. Auch hier ist zu prüfen, ob die befragten Personen das so wahrnehmen oder nicht (s. dazu die Ergebnisse zur *Erklärtiefe* bzw. zum *Kode Vertieft/oberflächlich*, d. h. Punkte *8.2.4, 8.3.2.3* sowie *9.4.3* in dieser Arbeit). Analog zu den Ergebnissen zum Verstehen müssen auch die Ergebnisse zum Merken vorsichtig interpretiert werden und können nur Hinweise darauf geben, in welche Richtung weiter geforscht werden muss.

Die konkrete Formulierung der einzelnen Items ist im Folgenden unter der beschriebenen Vignette zu finden, eine Übersicht findet sich an Anhang 3.2.

7.5.1 Items Was-Erklärungen

Wie unter Punkt *5.2 Was-Erklärungen* in dieser Arbeit bereits ausführlich beschrieben, hatten die beiden Vignetten der Was-Erklärung die Redewendung „Um etwas viel Aufhebens machen" zum Thema. Variante eins (mit mehr Erklärtiefe) erläuterte die Bedeutung unter Zuhilfenahme des etymologischen Hintergrundes, während Variante zwei (mit weniger Erklärtiefe) zur Bedeutungserläuterung auf synonyme Ausdrücke zurückgriff.

Um die wahrgenommene Erklärtiefe und deren Bedeutung für die Bewertung der Vignetten genauer herauszuarbeiten, wurden zu jeder Vignette noch zwei Items gefragt.

Das erste Item bezog sich auf die Bedeutung der umgesetzten Erklärtiefe auf das wahrgenommene Verständnis. Dazu war es notwendig, konkret die Art und Weise zu benennen, mit der die oben beschriebene Erklärtiefe jeweils umgesetzt wurde. Dieses Item lautete daher bei Variante eins: „Es hat beim Verstehen der Redewendung geholfen, dass erklärt wurde, wie sie entstanden ist" und bei Variante zwei analog: „Es hat beim Verstehen der Redewendung geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden". Inhaltlich beziehen sich die beiden Items trotz der unterschiedlichen Formulierung auf das Gleiche, nämlich wie gesagt darauf, inwieweit die je umgesetzte Erklärtiefe das wahrgenommene Verständnis beeinflusst.

Das zweite Item geht inhaltlich über das reine Verstehen hinaus, denn es bezieht sich auf die durch die Erklärungen ausgelöste Merkfähigkeit des neuen Inhaltes. Wie oben bereits dargelegt wird davon ausgegangen, dass höheres Potenzial zur kognitiven Aktivierung die Merkfähigkeit in Bezug auf den dargebotenen Inhalt erhöht. Das hier beschriebene Item erfasst die wahrgenommene

Merkfähigkeit, die durch die umgesetzte Erklärtiefe ausgelöst wird. Auch dazu ist es notwendig, konkret zu benennen, worauf sich das Item bezieht. Es wurde daher für Variante eins folgendermaßen formuliert: „Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass erklärt wurde, wie sie entstanden ist". Für Variante zwei wurde wiederum analog gefragt: „Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden."

Zur besseren Übersicht sind alle Items, die sich auf die *Erklärtiefe* der Was-Erklärungen beziehen (einschließlich der oben bereits erläuterten für alle Erklärungen identischen Items), noch einmal gesammelt in Tabelle 4 abgebildet.

Tabelle 4: Alle Items zur Erklärtiefe der Was-Erklärungen

ET1	Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen. (für SLD) / Die Erklärung hat geholfen, das Erklärte genau zu verstehen. (für SuS)	
ET1	Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.	
ET3	Die Erklärung war kognitiv aktivierend. (für SLD) / Die Erklärung hat mich zum Mitdenken angeregt. (für SuS)	
ET4	Vid. A (vertiefte Variante): Es hat beim Verstehen der Redewendung geholfen, dass erklärt wurde, wie sie entstanden ist.	Vid. D (einfachere Variante): Es hat beim Verstehen der Redewendung geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden.
ET5	Vid. A (vertiefe Variante): Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass erklärt wurde, wie sie entstanden ist.	Vid. D (einfachere Variante): Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden.

Bem.: Alle Items wurden auf einer Ratingskala von 1 = *stimme voll zu* bis 6 = *stimme gar nicht zu* beurteilt.

7.5.2 Items Warum-Erklärungen

Die Konzeptualisierung der Warum-Erklärungen wurde unter Punkt *5.3 Warum-Erklärungen* in dieser Arbeit bereits ausführlich dargelegt. Die umgesetzte Variation der Erklärtiefe kam in diesen Erklärungen zustande, indem eine inhaltich-funktionale einer strukturellen Begründung gegenübergestellt wurde. Erstere zeichnet sich durch das explizite Aufzeigen der Funktion eines Arguments aus, deren inhaltliche Erfüllung am konkreten Argument überprüft wird. Diese Variante weist hohes Potenzial zur kognitiven Aktivierung auf. Dahingegen wurde in Variante zwei ein eher formales Vorgehen gewählt, in dem eine (in der Schule weit verbreitete) vorgegebene Struktur am konkreten Argument überprüft wurde. Der inhaltliche Aspekt stand dabei nicht so sehr im

Vordergrund, so dass diese Variante weniger Potenzial zur kognitiven Aktivierung bietet. Inwieweit die Studienteilnehmenden das wahrnahmen, wurde zusätzlich zu den oben aufgeführten gemeinsamen Items zur *Erklärtiefe* mit den folgenden zwei Items themenspezifisch überprüft.
Analog zu den Was-Erklärungen bezog sich die erste Frage auf das wahrgenommene Verständnis und lautete bei Variante eins: „Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Inhalt konzentriert hat" und bei Variante zwei: „Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat".
Das zweite Item überprüfte (ebenso analog zu den Was-Erklärungen) die wahrgenommene Merkfähigkeit, die durch die Varianten ausgelöst wird. Es wurde für Variante eins gefragt: „Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Inhalt konzentriert hat"; und für Variante zwei wiederum analog: „Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat." Für die Schüler/-innen wurde erneut die Perspektive in den Items zu „Ich…" gewechselt, zudem wurden die Formulierungen leicht vereinfacht (s. Tab. 5 bzw. Übersicht in Anhang 3.2).
Zur besseren Übersicht sind alle Items, die sich auf die *Erklärtiefe* der Warum-Erklärungen beziehen (einschließlich der oben bereits erläuterten für alle Erklärungen identischen Items), noch einmal gesammelt in Tabelle 5 abgebildet.

Tabelle 5: Alle Items zur Erklärtiefe der Warum-Erklärungen

ET1	Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen. (für SLD) / Die Erklärung hat geholfen, das Erklärte genau zu verstehen. (für SuS)	
ET2	Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.	
ET3	Die Erklärung war kognitiv aktivierend. (für SLD) / Die Erklärung hat mich zum Mitdenken angeregt. (für SuS)	
ET4	Vid. B (vertiefte Variante): Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Inhalt konzentriert hat. (für SLD) /	Vid. E (einfachere Variante): Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat. (für SLD) /
	Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Inhalt des Arguments erklärt hat. (für SuS).	Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat. (für SuS)

ET5	Vid. B (vertiefte Variante): Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Inhalt konzentriert hat. (für SLD)	Vid. E (einfachere Variante): Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat. (für SLD)
	Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Inhalt des Arguments erklärt hat. (für SuS)	Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat. (für SuS)

Bem.: Alle Items wurden auf einer Ratingskala von 1 = *stimme voll zu* bis 6 = *stimme gar nicht zu* beurteilt; grau hinterlegt die Formulierungen für die SuS-Gruppe

7.5.3 Items Wie-Erklärungen

Wie unter Punkt *5.4 Wie-Erklärungen* in dieser Arbeit bereits ausführlich beschrieben hatten diese die Prozedur des Widerlegens eines Arguments zum Thema. Dafür wurde in Variante eins die Methode lautes Denken (mit hohem Potenzial zur kognitiven Aktivierung) in Variante zwei der Methode Modelltext (mit weniger Potenzial zur kognitiven Aktivierung, da die Schritte nicht im Entstehen erläutert werden, sondern nur ein fertiges Ergebnis präsentiert wird) gegenübergestellt.

Parallel zu den Überlegungen zu den Was- und Warum-Erklärungen wurde auch hier im ersten Item nach der Rolle der gezeigten Methode für das wahrgenommene Verständnis gefragt, und zwar für Variante eins mit: „Das „laute Denken", so wie es im Video gezeigt wurde, hat geholfen zu verstehen, wie man die Aussage widerlegt" und für Variante zwei mit: „Dass die Lehrkraft ihr vorformuliertes Gegenargument erklärt, hat geholfen zu verstehen, wie man die Aussage widerlegt".

Das zweite Item zielte, analog zu den Was- und Warum-Erklärungen, erneut auf die wahrgenommene Merkfähigkeit ab, und zwar für Variante eins mit der Frage: „Das „laute Denken", so wie es im Video gezeigt wurde, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt" beziehungsweise für Variante zwei mit der Frage: „Dass die Lehrkraft ihr vorformuliertes Gegenargument erklärt, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt."

Eine Übersicht über sämtliche Items zur *Erklärtiefe* bei den Wie-Erklärungen (einschließlich der oben bereits erläuterten für alle Erklärungen identischen Items) gibt Tabelle 6, in der die leicht abgewandelten Formulierungen für Schüler/-innen (nämlich erneut statt der Metaperspektive die direkte Ansprache „mir") zu sehen sind.

Tabelle 6: Alle Items zur Erklärtiefe der Wie-Erklärungen

ET1	Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen. (für SLD) / Die Erklärung hat geholfen, das Erklärte genau zu verstehen. (für SuS)	
ET2	Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.	
ET3	Die Erklärung war kognitiv aktivierend. (für SLD) / Die Erklärung hat mich zum Mitdenken angeregt. (für SuS)	
ET4	Vid. C (vertiefte Variante): Das „laute Denken", so wie es im Video gezeigt wurde, hat geholfen zu verstehen, wie man die Aussage widerlegt. (für SLD)	Vid. F (einfache Variante): Dass die Lehrkraft ihr vorformuliertes Gegenargument erklärt, hat geholfen zu verstehen, wie man die Aussage widerlegt. (für SLD)
	Das „laute Denken", das die Lehrerin vorgemacht hat, hat mir geholfen zu verstehen, wie man die Aussage widerlegt. (für SuS)	Dass die Lehrerin ihr vorformuliertes Gegenargument erklärt, hat mir geholfen zu verstehen, wie man die Aussage widerlegt. (für SuS)
ET5	Vid. C (vertiefte Variante): Das „laute Denken", so wie es im Video gezeigt wurde, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt. (für SLD)	Vid. F (einfache Variante): Dass die Lehrkraft ihr vorformuliertes Gegenargument erklärt, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt. (für SLD)
	Das „laute Denken", das die Lehrerin vorgemacht hat, hat mir am meisten geholfen, mir gut merken zu können, wie man die Aussage widerlegt. (für SuS)	Dass die Lehrerin ihr vorformuliertes Gegenargument erklärt, hat mir am meisten geholfen, mir gut merken zu können, wie man die Aussage widerlegt. (für SuS)

Bem.: Alle Items wurden auf einer Ratingskala von 1 = *stimme voll zu* bis 6 = *stimme gar nicht zu* beurteilt; grau hinterlegt die Formulierungen für die SuS-Gruppe

8 Quantitative Ergebnisse

Nach diesen Erläuterungen zur Forschungsmethode werden im Folgenden die quantitativen Ergebnisse der Studie FALKE-D dargestellt, wobei nach der Erläuterung der Gütekriterien zunächst auf die deskriptiven Ergebnisse und anschließend auf die inferenzstatistischen Resultate eingegangen wird. In sämtliche Berechnungen gehen die Daten der genannten Teilnehmerzahlen ein, obwohl diese nicht immer vollständig sind: Es kam zum Teil zu vorzeitigen Abbrüchen der Fragebogenbearbeitung (vor allem bei MZP2); die fehlenden Daten wurden nicht ersetzt, in den Berechnungen wurde die jeweils vorliegende Stichprobengröße verwendet.

Die deskriptiven Auswertungen und Berechnungen für FALKE-D wurden mit der Software SPSS 25 von IBM SPSS Statistics erstellt, die inferenzstatistischen Analysen wurden über Software R (Version 3.6.3) getätigt.

8.1 Qualitätssicherung

Begonnen wird mit der Darstellung der Pilotierung sowie der Beschreibung der Gütekriterien zur Qualitätssicherung der Erhebung.

8.1.1 Pilotierung

Die Auswahl der Erklärthemen sowie die sich anschließende Auswahl der Vignetten, die in den Fragebogen implementiert wurden, geschah nach vorheriger Literaturrecherche in Absprache mit Expertinnen und Experten aus der Deutschdidaktik der Universität Regensburg und der Universität Klagenfurt und wurde unter Punkt 7.2 *Erstellen und Auswahl der Videovignetten als Stimuli* in dieser Arbeit bereits dargelegt.
Im Folgenden wird die Pilotierung der Items sowie der Handhabung des Fragebogens beschrieben.

8.1.1.1 Pilotierung im Gesamtprojekt FALKE

Sämtliche in FALKE(-D) erhobenen Items zu den einzelnen Aspekten wurden mit Vertreterinnen und Vertretern aller Statusgruppen mit der Methode des *Probing* (vgl. Prüfer & Rexroth, 2005, S. 5–10) pilotiert: Die Items zu den fachübergreifenden Aspekten *Strukturiertheit, Adressatenorientierung* sowie *Persönlichkeitswirkung* wurden unter den Mitarbeiterinnen und Mitarbeitern der Fachdidaktiken aufgeteilt; jede/-r befragte zu den zugeteilten Items mindestens je eine Lehrkraft, eine/-n Schüler/-in und zwei Studierende, so dass von jeder Statusgruppe mindestens sechs Personen ausführlich befragt wurden. Die Items zu den Aspekten *Sprachliche Verständlichkeit* sowie *Sprech- und Körperausdruck* wurden von den Projektmitarbeiterinnen pilotiert (s. Thim-Mabrey & Lindl, 2021, in Vorb. und Gunga et al., 2021, in Vorb.).
Für die Pilotierungen der Items verwendete jede an FALKE beteiligte Disziplin die eigenen, für das Fach konzipierten Vignetten. Die Mitarbeiterinnen aus der Sprach- und Sprechwissenschaft griffen ebenso auf Videos aus den beteiligten Fächern zurück, da sie selbst keine Vignetten erstellt hatten. Bei den Befragungen der Statusgruppe der Schüler/-innen wurde in jedem Fach darauf geachtet, dass diese im gleichen Alter wie die später befragten Schüler/-innen waren. Da die einzelnen Fächer Vignetten für unterschiedliche Jahrgangsstufen entwickelten, etwa Musikpädagogik für die 6. Jahrgangsstufe, dahingegen Geschichte für die 11. Jahrgangsstufe, war es nötig, das gesamte Altersspektrum in der Pilotierung abzubilden. Im Fach Deutsch wurden Schüler/-innen aus der 9. beziehungsweise 10. Jahrgangsstufe für die Pilotierung befragt.

Analog zur Vorgehensweise bei Prüfer und Rexroth (2005) wurden die Teilnehmer/-innen im Pilotierungsgespräch ganz konkret nach ihrem Verständnis der in den Items verwendeten Begrifflichkeiten gefragt; also zum Beispiel für das Item „Die Erklärung hatte einen roten Faden": „Was hast du/haben Sie unter „roter Faden" verstanden?" Die Teilnehmenden antworteten daraufhin zum Beispiel: „Das heißt, dass man immer weiß, was gemeint ist" (Antwort Schülerin) oder „Bei jedem Teilaspekt ist klar, worauf es hinausläuft, dazwischen geschieht nichts, der Gegenstand wird Schritt für Schritt abgearbeitet" (Antwort Lehrkraft) (vgl. für alle Zitate aus den Pilotierungsgesprächen eigene, handschriftliche Aufzeichnungen).

Die befragten Personen wurden auch konkret nach Items gefragt, die inhaltlich ähnlich sind. Nachdem sie die Items „Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut" und „Die Erklärung war sinnvoll aufgebaut" beantwortet hatten, wurde beispielsweise gefragt: „Was ist für dich/Sie der Unterschied zwischen Schritt für Schritt und sinnvoll aufgebaut?" Die Antworten fielen unterschiedlich aus: „Schritt für Schritt kann auch Quatsch sein, sinnvoll heißt zum Beispiel, es sind auch Wiederholungen drin, es ist schlüssig, macht Sinn. Das ist eher auf den Inhalt bezogen" (Studierende/-r). Oder: „Schritt für Schritt und sinnvoll beziehen sich beide auf den Aufbau. Das überlappt sich. Die Schritte sind sinnvoll geplant" (Studierende/-r).

Durch das gemeinsame Vorgehen im FALKE-Projekt stand nach der Pilotierungsphase eine Vielzahl an Kommentaren zu jedem Item zur Verfügung, was eine Optimierung des Itemsatzes im Hinblick auf Verständlichkeit und Eindeutigkeit ermöglichte. So wurden alle Ergebnisse aus der Pilotierung im Gesamtprojekt zusammengetragen und die Items, bei denen es sich als notwendig erwiesen hatte, überarbeitet oder entfernt. Zum Beispiel wurde in Absprache mit allen beteiligten Disziplinen das Item: „Die Erklärung beschränkte sich auf das Wesentliche" aus der Erhebung des Aspektes *Strukturiertheit* ausgeschlossen. Der Begriff „das Wesentliche" wurde von einigen Befragten als zu subjektiver Begriff beschrieben. Um dennoch den Teilaspekt der Fokussierung abzufragen, wurde das Item „Die Lehrkraft hat zu viele Einzelheiten erklärt" beibehalten (s. Punkt *4.2.1 Strukturiertheit* in dieser Arbeit).

Das oben erwähnte Item „Die Erklärung war sinnvoll aufgebaut" wurde ebenfalls gestrichen und durch das Item „Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut" ersetzt, da sich bei beiden in der Pilotierung die Antworten überlappten und „sinnvoll" auf Unterschiedliches (nämlich Aufbau oder Inhalt) bezogen wurde.

Alle Ergebnisse wurden ausführlich diskutiert und Entscheidungen bezüglich einzelner Items wurden in einem gemeinsamen Einigungsprozess getroffen. Wie bereits gesagt sind sämtliche im FALKE-Gesamtprojekt umgesetzten Items zu den gemeinsamen Aspekten *Strukturiertheit, Adressatenbezug, Sprachliche Verständlichkeit, Sprech- und Körperausdruck* sowie *Persönlichkeitswirkung* im Anhang 3.1 nachzulesen.

Abgesehen von den Nachfragen zum Verständnis wurden die Pilotierungsteilnehmer/-innen außerdem noch nach der Skalierung gefragt: Die Hälfte der FALKE-Mitarbeiter/-innen pilotierte auf einer vierstufigen Ratingskala (d. h. *stimme nicht zu, stimme eher nicht zu, stimme eher zu, stimme zu*), die andere Hälfte auf einer sechsstufigen Skala (d. h. *stimme gar nicht zu, stimme nicht zu, stimme eher nicht zu, stimme eher zu, stimme zu, stimme voll zu*). Insgesamt sprachen sich die Teilnehmer/-innen für eine sechsstufige Skala aus, welche dann auch in den Fragebogen übernommen wurde.

Zusätzlich wurden außerdem die Präsentation und die Handhabung des Fragebogens pilotiert. Dabei zeigten sich in der Bedienung und dem Beantworten der Fragen bei keiner der Statusgruppen Schwierigkeiten, so dass am Layout oder an der Formatierung keine Änderungen notwendig waren.

8.1.1.2 Pilotierung in der Studie FALKE-D

Die fachspezifischen Items wurden nur in der eigenen Disziplin ebenfalls über *Probing* (vgl. Prüfer & Rexroth, 2005) pilotiert. Für FALKE-D sind das die Items zu den deutschspezifischen Aspekten *Fachliche Relevanz* und *Erklärtiefe*, auf die nun genauer eingegangen wird.

Fachliche Relevanz

Der Aspekt *Fachliche Relevanz* dient in FALKE-D wie bereits erläutert der inhaltlichen Absicherung der Vignetteninhalte. Nach der Pilotierung standen die folgenden drei Items fest, die unter Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* in dieser Arbeit bereits erläutert wurden:

(1) *Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt.*

(2) *Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.*

(3) […][46] *sollte regelmäßig im Deutschunterricht verwendet werden.*

Dieses Ergebnis wurde folgendermaßen erarbeitet: Während der Pilotierung der *Fachlichen Relevanz* wurde beim ursprünglich entwickelten Item: „Erklärungen wie die gerade gesehene sind typisch für den Deutschunterricht" deutlich, dass der Begriff „typisch" in diesem Zusammenhang negativ konnotiert sein kann: „Das kann negativ verstanden werden, im Sinne von ach ja, typisch Deutschunterricht" (Antwort Lehrkraft). Das Item wurde daraufhin aus der Befragung herausgenommen.

[46] […] steht wie oben bereits erläutert für das jeweilige Thema und die jeweilige Erklärvariante. Die Formulierung dieses Items wurde an jede Vignette einzeln angepasst. Eine Übersicht dazu findet sich in Anhang 3.2.

Entfernt wurde ebenfalls das Item: „Erklärungen wie die eben gesehene sind sinnvoll im Deutschunterricht“, denn es stellte sich heraus, dass das Item: „Erklärungen wie die eben gesehene sollten regelmäßig im Deutschunterricht vorkommen“ inhaltlich auf dasselbe abzielte, dabei aber einfacher zu verstehen war, denn es herrschte während der Pilotierung Uneinigkeit darüber, worauf sich „sinnvoll“ beziehen sollte.

Dagegen wurde das Item 2): „Erklärungen wie die eben gesehene kommen im Deutschunterricht vor“ von allen Statusgruppen in der Pilotierung als unproblematisch eingeschätzt. Dieses Item fragt danach, ob den Studienteilnehmerinnen und -teilnehmern die im Video gezeigte Erklärweise vertraut ist und ob sie Relevanz für die Teilnehmer/-innen hat. Es wurde daher im Fragebogen belassen.

Bei dem Item 1) (für die SLD-Gruppe) „Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt“ wurden die Teilnehmer/-innen gefragt: „Was verstehen Sie unter „Aufbau einer spezifischen Kompetenz“? Die Antworten reichten von „lesen, schreiben, zuhören“ (Antwort Lehrkraft) über „Texte schreiben, dazu gehört ja auch das Argumentieren“ (Antwort Lehrkraft) bis hin zu „die eigene Sprache angemessen und bewusst zu verwenden, andere überzeugen, verschiedene Sprachniveaus anwenden und auch verstehen“ (Antwort Studierende/-r). Obwohl ein/-e Studierende/-r anmerkte, dass ihr/ihm nicht ausreichend klar wäre, was genau damit gemeint sei, wurde das Item aufgrund der anderen Antworten im Fragebogen belassen. Auch dieses Item zielt auf eine inhaltliche Absicherung des Vignetteninhaltes ab: Es überprüft, ob die Teilnehmenden die Erklärung als deutschspezifisch wahrnehmen oder nicht.

Wie bereits gesagt stellte sich bei der Analyse der Ergebnisse heraus, dass diese Items ein reflexives Konstrukt mit Indexcharakter bilden, so dass sie zum Aspekt *Fachliche Relevanz* zusammengefasst wurden.

Erklärtiefe

Bei dem deutschspezifischen Aspekt der *Erklärtiefe* standen nach der Pilotierung folgende, oben erläuterte Items fest:

ET1) *Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen. bzw. Die Erklärung hat geholfen, das Erklärte genau zu verstehen.*

ET2) *Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.*

ET3) *Die Erklärung war kognitiv aktivierend* bzw. *Die Erklärung hat mich zum Mitdenken angeregt.*

ET4) *x hat dabei geholfen zu verstehen, dass y* bzw. *Die Schüler/-innen verstehen y, da x* bzw. *Ich verstehe y, da x*[47].

ET5) *x hat dabei geholfen, sich y gut merken zu können* bzw. *Die Schüler/-innen können sich y gut merken, da x* bzw. *Ich kann mir y gut merken, da x*.

Beschrieben wird nun der Prozess, in dem diese Items für die Erhebung ausgewählt wurden. Analog zu den Items zur *Fachlichen Relevanz* wurde mit den Items zum Aspekt *Erklärtiefe* verfahren. Diese Items wurden mehrmals mit Vertreterinnen und Vertretern der einzelnen Statusgruppen pilotiert. Die verschiedenen Formulierungen wurden zusätzlich Expertinnen und Experten aus der Deutschdidaktik und aus der Sprachwissenschaft zur Augenscheinvalidierung vorgelegt.

Einige Items dieses Aspektes stellten sich in der Pilotierung als unproblematisch heraus und wurden unverändert in den Fragebogen übernommen. Dies betrifft in den Expertengruppen und in der Schülergruppe sowohl das Item ET1) „Die Erklärung hat geholfen, das Erklärte vertieft (bzw. für SuS genau) zu verstehen" als auch das gegenläufige Item ET2) „Die Erklärung hat nur oberflächliche Zusammenhänge erläutert". Die Teilnehmer/-innen gaben zu ersterem an: „vertieft verstehen heißt genau verstehen" (Studierende/-r), „vertieft verstehen ist für mich etwas ganz genau erfassen" (Studierende/-r) oder „dass man es halt genau versteht" (Schüler/-in). Beim zweiten Item merkten die Teilnehmenden an: „oberflächliche Zusammenhänge heißt, es werden keine Hintergründe erklärt, das Thema wird nur schnell angerissen" (Lehrkraft) oder „oberflächlich heißt, es ist ungenau" (Schüler/-in).

Auch das Item ET3) „Die Erklärung war kognitiv aktivierend" wurde von Vertreterinnen und Vertretern der Expertengruppen richtig verstanden: „kognitiv aktivierend bedeutet, dass man nachdenkt" (Lehrkraft) beziehungsweise „kognitiv aktivierend ist geistig aktiv sein" (Studierende/-r). In der Statusgruppe der Schüler/-innen wurde letztgenanntes Item ersetzt durch „Die Erklärung hat mich zum Mitdenken angeregt". Mit dieser Formulierung hatten die befragten Schüler/-innen keine Schwierigkeiten, so gaben sie beim Probing zum Beispiel folgende Antwort: „Das heißt, dass ich dabeigeblieben bin und nicht an anderes gedacht habe."

Bei anderen Items hingegen erwies sich die Formulierung als Herausforderung, da sie zum einen konkret auf den Gegenstand (zum Beispiel das zu bewertende Argument) und die jeweilige Methode (zum Beispiel lautes Denken vs. Modelltext) zugeschnitten sein musste. In der Pilotierung zeigte sich nämlich, dass allgemeine Formulierungen, die zunächst angedacht waren, nämlich zum Beispiel für Item ET4) „Durch die Erklärung verstehen die Schüler/-innen, dass das Argument nicht gelungen ist" völlig unterschiedlich verstanden wurden und als zu wenig präzise wahrgenommen

[47] Wie oben erläutert steht x in Item ET4 und ET5 für die Erklärmethode, y steht in Item ET4 und ET5 für den Inhalt der Erklärung.

wurden: „Da weiß ich nicht, was genau gemeint ist. „Durch die Erklärung“ kann alles Mögliche sein“ (Antwort Lehrkraft).

Die notwendige Spezifität erwies sich als Schwierigkeit bei der Formulierung der Items, die nach der ersten Überarbeitung einer erneuten Pilotierung unterzogen wurden. Während dieser zeigte sich zum Beispiel, dass das folgendermaßen überarbeitete Item ET4): „Zu erklären, dass die einzelnen Teile eines Arguments nicht zueinander passen, hilft dabei zu verstehen, dass das Argument nicht gelungen ist“ vor allem aufgrund der komplexen Satzstruktur als zu schwierig erachtet wurde. Nach einer erneuten Überarbeitung mit Hilfe von mehreren Didaktikerinnen und Didaktikern wurde daraus das bereits genannte Item ET4): „Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat.“ Es wurde erneut einer/einem Studierenden vorgelegt und schließlich in den Fragebogen implementiert.

Für die Schüler/-innen wurde hier die persönliche Anrede gewählt, da sie die Frage für sich selbst beantworten sollten. Das Item ET4) heißt somit in dieser Statusgruppe leicht vereinfacht: „Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat“.

Die zweite Herausforderung war die Unterscheidung zwischen „verstehen“ und „sich gut merken können“ (d. h. Item ET4 vs. Item ET5), mit der die wahrgenommene Erklärtiefe erfasst werden sollte (s. Punkt *4.2.5 Erklärtiefe in FALKE-D* in dieser Arbeit). Interessanterweise erwies sich diese Unterscheidung als unproblematisch, denn sowohl Vertreterinnen und Vertretern der Expertengruppen als auch der Schülergruppe war der Unterschied klar, als sie beim Probing danach gefragt wurden: „Ich kann was verstehen, aber ob ich es mir merke, ist dann wieder eine andere Frage“ (Schüler/-in). „Merken ist mehr als verstehen. Merken kann ich mir ja nur etwas, was ich vorher verstanden habe“ (Lehrkraft) beziehungsweise „Etwas merken bezieht sich auf einen längeren Zeitraum als etwas verstehen“ (Studierende/-r).

Bei der Formulierung des konkreten Gegenstands sowie der konkreten Erklärmethode wurde dann analog zum gerade erläuterten „Verstehensitem“ verfahren, denn natürlich erwiesen sich auch hier zunächst angedachte Formulierungen (z. B. „Durch die Erklärung können sich die Schüler/-innen gut merken, dass das Argument nicht gelungen ist“) als zu unspezifisch, die erste Überarbeitung aber wiederum als zu komplex. Somit wurden diese Items parallel zu den gerade geschilderten formuliert. Aus: „Zu erklären, dass die einzelnen Teile eines Arguments nicht zueinander passen, hilft besonders dabei, sich gut merken zu können, dass das Argument nicht gelungen ist“ wurde in der Expertengruppe das Item ET5) „Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat.“ Für die Schüler/-innen wurde ebenso in Analogie zu oben formuliert: „Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat“.

Im Zuge der Itempilotierung wurden auch die ausgewählten Vignetten mit den Teilnehmerinnen und Teilnehmern besprochen. Dabei ergaben sich bei keiner Statusgruppe inhaltliche Verständnisschwierigkeiten. Die erklärende Lehrkraft war gut zu sehen und wirkte authentisch. Keine der befragten Personen bemerkte, dass die Erklärungen gestellt waren. Die Tonaufnahmen waren von ausreichend guter Qualität. Lediglich bezüglich der Textarbeit in den Vignetten B bis F kamen Nachfragen, da der in den Vignetten besprochene Textauszug nur vorgelesen wurde, auf der Overheadfolie im Video aber schwer zu erkennen war. Es wurden daraufhin die entsprechenden Textstellen in den Fragebogen implementiert, und zwar als zu lesender Text vor der entsprechenden Vignette (s. Punkt *7.3 Untersuchungsinstrument: Online-Fragebogen* in dieser Arbeit).
Die Pilotierungen endeten im März 2017.

8.1.2 Psychometrische Gütekriterien

Objektivität

Da in der Studie FALKE-D ein onlinebasierter Fragebogen erstellt wurde, in dem der Großteil der Daten über ein geschlossenes Antwortformat (s. Punkt *7.3 Untersuchungsinstrument: Online-Fragebogen* in dieser Arbeit) erhoben wurde, weist die Studie hohe Auswertungsobjektivität auf.
Auch die Durchführungsobjektivität (vgl. Bortz & Döring, 2015, S. 195) ist als hoch zu erachten: Die Expertengruppen (d. h. Studierende, Lehrkräfte und Didaktiker/-innen) bekamen alle relevanten Informationen zur Durchführung im Fragebogen selbst in standardisierter schriftlicher Form. Die Anweisungen zur Erhebung der Statusgruppe Schüler/-innen erfolgten nach einem vorgefertigten Skript, auf auftretende Fragen wurde ebenfalls mit vorher festgelegten, im Gesamtprojekt identischen Antworten reagiert (s. auch hierzu Punkt *7.3 Untersuchungsinstrument: Online-Fragebogen* in dieser Arbeit und Anhang 8 - 9).

Validität

Die inhaltliche Validität der in den Fragebogen aufgenommenen Items wurde durch eine ausführliche vorherige Literaturrecherche in allen beteiligten Fachdidaktiken sowie in der Lehr-/Lernpsychologie sichergestellt. Die Augenscheinvalidität wurde durch deren Diskussion in der Forschergruppe gewährleistet. Die Formulierungen der Items wurden zudem durch aufwendige Pilotierungen mit Vertreterinnen und Vertretern aller Statusgruppen (s. Punkt *8.1.1 Pilotierung* in dieser Arbeit) optimiert.
Zudem wurden die Studienteilnehmenden im Fragenbogen selbst zur inhaltlichen Validität befragt, wie unter Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* in dieser Arbeit bereits angeklungen. Die Items zur *Fachlichen Relevanz* dienten der inhaltlichen Bestätigung der implementierten Erklärungen. Die Ergebnisse zeigen insgesamt, dass die verwendeten

Methoden und Erklärthemen als fachlich relevant erachtet werden (s. Genaueres unter Punkt *8.2.2.1 Statusgruppenübergreifende Bewertungen der Aspekte* in dieser Arbeit).

Reliabilität

Die Bildung der Skalen und der ihnen zugeordneten Items erfolgte auf theoretischer Grundlage und wurde unter Punkt *4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens* in dieser Arbeit bereits dargelegt. Auf Basis dieser Grundlage wurden die einzelnen Items den fünf Skalen *Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit, Sprech-/Körperausdruck* und *Erklärtiefe* zur Erklärqualität sowie den Skalen *Persönlichkeitswirkung* (als Kontrollvariable) und *Fachliche Relevanz* (zur inhaltlichen Absicherung) zugeordnet. Bei den beiden letztgenannten war zu Beginn der Befragung unklar, ob sie ein reflexives Konstrukt induzieren, also ob die einzelnen Items tatsächlich zu einer Skala zusammengefasst werden konnten. Im Verlauf der Analysen stellte sich heraus, dass dies sowohl für die *Persönlichkeitswirkung* im Gesamtprojekt FALKE als auch (mit den unter Punkt *8.2.1 Psychometrische Gütekriterien* in dieser Arbeit genannten Einschränkungen) für die *Fachliche Relevanz* in der Studie FALKE-D möglich war.

Die Bestimmung der Reliabilität der Skalen erfolgte über Berechnung der internen Konsistenz (vgl. Bortz & Döring, 2015, S. 198–199) mithilfe des Reliabilitätsindikators Cronbachs Alpha. Negativ formulierte Items wurden für die Berechnungen umkodiert. Die Berechnung der internen Konsistenz der einzelnen Skalen erfolgte getrennt nach den Statusgruppen Schüler/-innen (= SuS) und Expertengruppen (d. h. Studierende, Lehrkräfte, Didaktiker/-innen zusammen = SLD) und nach Video (d. h. sortiert nach Erklärprozedur und Erklärtiefe) (Daten aus Gaier & Schilcher, 2021, in Vorb., s. Tab. 7 und 8; eine Auflistung der psychometrischen und deskriptiven Kennwerte auf Itemebene findet sich zudem in Anhang 10).

Auf die zum Teil niedrigen Werte in der Skala *Sprachliche Verständlichkeit* wird in dieser Arbeit nicht eingegangen. Genaueres dazu findet sich in Thim-Mabrey & Lindl (2021, in Vorb.). Ebenso wird die Skala *Sprech- und Körperausdruck* an dieser Stelle vernachlässigt, hier wird ebenfalls auf die entsprechende Publikation verwiesen (s. Gunga et al., 2021, in Vorb.).

Tabelle 7: Cronbachs alpha nach Skalen, Video und Statusgruppe SuS
n_{max}=113

Skala	Was-Erklärungen		Warum-Erklärungen		Wie-Erklärungen	
	Video D einfach	Video A vertieft	Video E einfach	Video B vertieft	Video F einfach	Video C vertieft
Strukturiertheit	*0,66*	*0,66*	*0,57*	*0,73*	*0,55*	*0,70*
Adressatenorientierung	*0,58*	*0,60*	*0,60*	*0,78*	*0,47*	*0,57*
Sprech- und Körperausdruck	*0,72*	*0,79*	*0,73*	*0,84*	*0,85*	*0,86*
Sprachliche Verständlichkeit	*0,22*	*0,50*	*0,43*	*0,69*	*0,12*	*0,51*
Persönlichkeitswirkung	*0,53*	*0,68*	*0,75*	*0,80*	*0,85*	*0,71*
Erklärtiefe	*0,71*	*0,73*	*0,72*	*0,81*	*0,74*	*0,81*
Fachliche Relevanz	*0,46*	*0,32*	*0,48*	*0,17*	*0,52*	*0,31*

Tab. angelehnt an Gaier & Schilcher, 2021, in Vorb.

Tabelle 8: Cronbachs alpha nach Skalen, Video und Statusgruppe SLD
n_{max}=144

Skala	Was-Erklärungen		Warum-Erklärungen		Wie-Erklärungen	
	Video D einfach	Video A vertieft	Video E einfach	Video B vertieft	Video F einfach	Video C vertieft
Strukturiertheit	*0,79*	*0,81*	*0,82*	*0,84*	*0,86*	*0,81*
Adressatenorientierung	*0,80*	*0,77*	*0,90*	*0,93*	*0,87*	*0,84*
Sprech- und Körperausdruck	*0,81*	*0,80*	*0,84*	*0,87*	*0,85*	*0,85*
Sprachliche Verständlichkeit	*0,69*	*0,58*	*0,77*	*0,68*	*0,65*	*0,55*
Persönlichkeitswirkung	*0,72*	*0,79*	*0,80*	*0,84*	*0,74*	*0,79*
Erklärtiefe	*0,86*	*0,85*	*0,89*	*0,88*	*0,83*	*0,87*
Fachliche Relevanz	*0,54*	*0,48*	*0,49*	*0,36*	*0,45*	*0,29*

Tab. angelehnt an Gaier & Schilcher, 2021, in Vorb.

Ausführlich dargelegt werden im Folgenden die FALKE-D-spezifischen Skalen *Erklärtiefe* und *Fachliche Relevanz*.

Reliabilität der Skala *Erklärtiefe*

Für den Aspekt der *Erklärtiefe* ergeben sich für alle Videos und in beiden Statusgruppen (also SuS bzw. SLD) Cronbachs Alpha in Höhe von 0,71 bis 0,89 (vgl. Tab. 7 und 8). Die Werte sind alle akzeptabel bis hoch, die interne Konsistenz dieser Skala ist somit psychometrisch gegeben.

Allein bei einem Item (nämlich „Die Erklärung hat nur oberflächliche Zusammenhänge erläutert") ergeben sich Werte, die statistisch nicht den Gütekriterien entsprechen. Hier wurde inhaltlichen Kriterien vor statistischen Ausschlusskriterien der Vorzug gegeben. Die Vorgehensweise bei

statistisch „problematischen" Items in der Studie FALKE-D ist somit kohärent mit der des Gesamtprojektes FALKE. Obiges Item wurde in der Skala *Erklärtiefe* belassen, obwohl es in der Statusgruppe der Schüler/-innen geringe Trennschärfe, das heißt geringe Werte in der Item-Skala-Korrelation, aufweist, und die interne Konsistenz dementsprechend steigt, wenn es aus der Skala genommen würde (exemplarisch seien die Werte für das Video D (einfache Was-Erklärung) erwähnt: Die Trennschärfe des oben genannten Items liegt bei 0,13[48], so dass Cronbachs Alpha von 0,71 (s. Tab. 7) auf 0,78 ansteigt, wenn das Item aus der Skala genommen wird). Die insgesamt niedrigen Werte bei diesem Item liegen vermutlich an der negativen Formulierung und der damit verbundenen Rekodierung. Es ist anzunehmen, dass einige Schüler/-innen analog zu den übrigen positiven Formulierungen geantwortet haben und nicht bedacht haben, dass die Skalierung bei negativ formulierten Items anders herum „funktioniert".

In der Statusgruppe der Expertinnen und Experten weist dieses Item zwar auch zum Teil niedrige Trennschärfen auf, die Auswirkungen sind jedoch rechnerisch zu vernachlässigen (so steigt zum Beispiel bei einer Trennschärfe von 0,46 für oben genanntes Item bei Video B (vertiefte Warum-Erklärung) Cronbachs Alpha der Gesamtskala von 0,88 auf 0,91, wenn das Item weggelassen wird, vgl. Fußnote 48). Da das Item aber einen spezifischen Aussagewert für die Skala *Erklärtiefe* hat (nämlich als „Gegenfrage" zu „Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen"), wurde es in der Skala belassen[49].

Reliabilität der Skala *Fachliche Relevanz*

Für den Aspekt *Fachliche Relevanz* zeigen sich insbesondere bei den Schülerinnen und Schülern, aber auch bei den Expertinnen und Experten geringe Werte, was damit zu begründen ist, dass die zwei Items (bei SuS) beziehungsweise drei Items (bei SLD) das Konzept der *Fachlichen Relevanz* sehr breit abdecken (s. Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* in dieser Arbeit). Die geringe Konsistenz ist deswegen nicht verwunderlich. Interessant ist allerdings die Tatsache, dass in beiden Gruppen die interne Konsistenz für diese Skala bei den einfachen Videos höher ausfällt als bei den komplexen. Die Werte der Skalen für *Fachliche Relevanz* sind der Vollständigkeit halber abgebildet. In der Diskussion der Ergebnisse wird für die *Fachliche Relevanz* wegen der geringen Konsistenz auf die Werte der einzelnen Items eingegangen (s. *Punkt 8.2.2.1 Statusgruppenübergreifende Bewertungen der Aspekte* in dieser Arbeit).

[48] Die Berechnungen liegen vor, werden der Übersichtlichkeit halber aber nicht abgebildet; außerdem bringen die Werte keine neuen Erkenntnisse.

[49] Auch die Reliabilitäten, die in Anhang 10 nach allen Statusgruppen getrennt berichtet werden, befinden sich auf ähnlichem Niveau wie die gerade beschriebenen. Auch diese Ergebnisse zeigen die interne Konsistenz der Skala Erklärtiefe (s. Anhang 10).

Zusammenfassend kann gesagt werden, dass durch die ausführlichen Pilotierungen im Gesamtprojekt FALKE sowie in der Studie FALKE-D und durch die Bewertungen der *Fachlichen Relevanz* durch die befragten Personen die Augenscheinvalidität und die inhaltliche Validität sowohl der Stimuli (d. h. der Erklärvignetten) als auch der implementierten Fragen (d. h. der Items) hinreichend gewährleistet ist.

Die psychometrischen Gütekriterien Objektivität, Validität und Reliabilität können somit als erfüllt betrachtet werden.

Nachdem die Voraussetzungen der Analysen geklärt wurden, werden als nächstes die deskriptiven Ergebnisse von FALKE-D dargelegt.

8.2 Deskriptive Ergebnisse

Zunächst werden die deskriptiven Ergebnisse der quantitativen Studie berichtet. Die Skalierungen gestalteten sich wie bereits erwähnt folgendermaßen: Für das Globalurteil wurde auf jede der gezeigten Erklärungen eine Schulnote von 1 bis 6 vergeben. Die Studienteilnehmer/-innen hatten die Möglichkeit, dieses Globalurteil mit einer positiven (z. B. 2+) oder negativen (z. B. 2-) Tendenz zu versehen. Diese Tendenzen flossen in die Berechnungen der Globalurteile in Höhe von 0,3 mit ein, so dass sich personenspezifische Ergebnisse von 0,7 (bei der Note 1+) bis 6,3 (bei der Note 6-) ergeben.

Alle abgefragten Items wurden auf einer sechsstufigen Antwortskala zwischen 1 (=„stimme voll zu“) bis 6 (=„stimme gar nicht zu“) erhoben. Für weiterführende Analysen wurden daher zunächst personenspezifische Mittelwerte pro Skala ermittelt, indem alle Werte der Items zusammengezählt und durch die Anzahl der Items in der Skala geteilt wurden. Negative Items wurden vor den Berechnungen umkodiert.

Einen Überblick über sämtliche arithmetische Mittelwerte und Verteilungsgrößen der Globalurteile über alle vier Statusgruppen hinweg findet sich in Tabelle 9.

Tabelle 9: Mittelwerte, Standardabweichungen und Spannweiten (über alle vier Statusgruppen) zum Globalurteil

N_{max}=257

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Globalurteil	2,63 (1,01) [0,70; 5,30]	2,21 (0,99) [0,70; 5,30]	2,33 (1,12) [0,70; 6,30]	2,52 (1,05) [0,70; 5,00]	2,40 (1,03) [0,70; 6,30]	2,25 (1,11) [0,70; 6,30]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Globalurteil: 0,7 bis 6,3.
(Tabelle gekürzt aus Gaier & Schilcher, 2021, in Vorb.)

8.2.1 Globalurteile pro Videovignette

Begonnen wird mit der Darstellung der Ergebnisse der Globalurteile der Erklärqualität, die zum ersten Messzeitpunkt gegeben wurden.

8.2.1.1 Statusgruppenübergreifende Globalurteile der Erklärqualität

Berichtet wird zunächst das Globalurteil über alle Statusgruppen hinweg (Forschungsfrage 1a). Im Mittel werden alle Erklärungen zwischen *gut* (2,21) und *befriedigend* (2,63) bewertet, keine Erklärung sticht besonders heraus. Auffällig ist zum einen die doch hohe Standardabweichung von etwa einer Schulnote in allen Vignetten. Das bedeutet, dass die Noten, die die Bewerter/-innen auf eine Vignette gaben, im Durchschnitt um eine Notenstufe voneinander abweichen. Dies deutet auf eine relativ hohe Uneinigkeit der Bewerter/-innen in Bezug auf die Gesamtqualität der Erklärungen hin, wobei allerdings zu sagen ist, dass eine Standardabweichung in dieser Höhe, wenn es um Wahrnehmung von Qualität geht, nicht ungewöhnlich ist (vgl. Clausen, 2002, S. 131). Zum anderen ist die Streuung bemerkenswert: Bei drei der sechs Erklärungen wird die gesamte Notenskala (von 1+ bis 6-) genutzt, bei den anderen drei Erklärungen wird sie nahezu ausgeschöpft (von 1+ bis 5 bzw. 5-). Interessanterweise wird sowohl bei dem am besten bewerteten Video (= A, vertiefte Was-Erklärung) als auch bei dem am schlechtesten bewerteten Video (= D, einfache Was-Erklärung) die Spannweite bei den negativen Noten nicht ganz ausgeschöpft. Auch diese Streuung deutet darauf hin, dass die Bewerter/-innen in der Beurteilung der Erklärqualität nicht übereinstimmen.

Dieses Ergebnis ist wenig überraschend, da die Globalurteile über alle Statusgruppen hinweg betrachtet wurden. Aufgrund der unterschiedlichen Perspektiven, die die Statusgruppe einnehmen (z.B. Schüler/-innen als Adressatinnen und Adressaten, Fachdidaktiker/-innen dagegen als Expertinnen und Experten für Lehr-/Lernprozesse usw.) kann man nicht davon ausgehen, dass sie in den Urteilen übereinstimmen.

Betrachtet man die Globalurteile im Hinblick auf die Variation der Erklärtiefe, so schneiden bei der Was- und bei der Wie-Erklärung die vertieften Varianten besser ab als die einfachen; bei der Warum-Erklärung ist das genau andersherum: Hier schneidet die einfache Variante insgesamt besser ab (vgl. Tab 9). Dieses Ergebnis bestätigt somit die in der Theorie angenommene höhere Qualität der vertieften Was- und Wie-Erklärung. Bei den Warum-Erklärungen geben die statusgruppenspezifischen Ergebnisse eventuell einen Einblick dahingehend, welche der Statusgruppen hier die einfache Variante vorzieht (s. Punkt *8.2.1.2 Statusgruppenspezifische Globalurteile der Erklärqualität* in dieser Arbeit).

Insgesamt muss man von hoher Uneinigkeit in den Globalurteilen aller Erklärungen sprechen, der Blick in die Statusgruppen ist also lohnend. Deshalb wird im Folgenden über das Globalurteil

innerhalb der einzelnen Statusgruppen berichtet und dieses jeweils mit den Ergebnissen der anderen Statusgruppen in Beziehung gesetzt (Forschungsfrage 2a). Wie in Tabelle 10 zu erkennen ist, ist die Streuung innerhalb der Statusgruppen ebenfalls groß.

8.2.1.2 Statusgruppenspezifische Globalurteile der Erklärqualität

Tabelle 10: Mittelwerte, Standardabweichungen und Spannweiten zum Globalurteil nach Statusgruppen getrennt

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Statusgruppe						
Schüler/-innen (n_{max} = 113)	2,55 (1,05) [0,70; 5,30]	2,37 (1,09) [0,70; 5,30]	2,19 (1,12) [0,70; 6,30]	2,31 (1,06) [0,70; 5,00]	2,76 (1,07) [0,70; 6,30]	2,59 (1,12) [0,70; 6,30]
Studierende (n_{max} = 78)	2,50 (0,93) [0,70;5,30]	2,03 (0,91) [0,70; 4,30]	2,15 (1,00) [0,70; 5,30]	2,45 (0,98) [0,70; 5,00]	2,19 (0,83) [0,70; 4,00]	1,99 (0,99) [0,70; 5,00]
Lehrkräfte (n_{max} = 35)	2,92 (1,03) [1,00; 5,00]	2,29 (0,81) [1,00; 4,30]	2,80 (1,15) [1,00; 5,30]	3,26 (0,93) [1,70; 5,00]	2,05 (0,91) [1,00; 5,00]	2,24 (1,08) [1,00; 5,00]
Didaktiker/-innen (n_{max} = 31)	2,92 (0,95) [1,00; 5,00]	1,98 (0,93) [1,00; 5,00]	2,75 (1,17) [1,00; 5,30]	2,63 (0,94) [1,00; 4,30]	1,95 (1,05) [0,70; 5,00]	1,57 (0,93) [1,00; 5,00]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Globalurteil: 0,7 bis 6,3.
(Ergebnisse aus Gaier & Schilcher, 2021, in Vorb.)

Schüler/-innen

Die Globalurteile der Statusgruppe Schüler/-innen bewegen sich zwischen den Notenstufen *gut* (2,19) und *befriedigend* (2,76), bei einer durchgehend hohen Standardabweichung von etwas über einer Notenstufe (vgl. Tab. 10).

Auffällig ist bei den Schülerinnen und Schülern, dass sie das gesamte Spektrum der Noten mit Tendenz ausnutzen: Bei jedem Video wird mindestens einmal die Note 1+, bei drei der sechs Vignetten wird mindestens einmal die schlechteste Note 6- vergeben. Im unteren Bewertungsbereich unterscheiden sich damit die Schüler/-innen von den anderen Statusgruppen, bei denen maximal die Note 5- vergeben wird. Schüler/-innen scheuen sich also offensichtlich nicht, ihr Missfallen durch das Vergeben der schlechtmöglichsten Note auszudrücken.

Trotz dieser Spannweite liegen die Schüler/-innen mit ihren Bewertungen insgesamt im Vergleich mit den anderen Statusgruppen im Mittelfeld, sie zeigen im Mittel weder eine besonders milde noch eine besonders strenge Tendenz. Auffallend ist die ähnliche Bewertung der Schüler/-innen und der Studierenden bei beiden Varianten der Was- und der Warum-Erklärung, wohingegen sich die

Bewertungen der beiden Wie-Erklärungen deutlich unterscheiden, da hier die Schüler/-innen erheblich strenger bewerten als die Studierenden. Eine mögliche Erklärung ist, dass Studierende aufgrund ihres Alters und ihrer langen Schulerfahrung doch eher eine Adressatenperspektive in der Bewertung einnehmen und daher ähnliche Ergebnisse wie die Schüler/-innen aufweisen. Bei den Wie-Erklärungen wird vermutet, dass die Studierenden beide Varianten (sowohl Arbeit mit Modelltext, besonders aber auch die Methode lautes Denken) aus den Seminaren an der Universität kennen und wissen, dass gerade letztere gute Ergebnisse in der Strategievermittlung aufweist. Es ist möglich, dass sich dieses Wissen in der guten Bewertung der vertieften Variante (lautes Denken) bei den Studierenden niederschlägt; wohingegen kaum ein/-e Schüler/-in mit dieser Methode vertraut sein dürfte, da sie im Unterricht wenig etabliert ist. Die schlechte Bewertung dieser Variante durch die Schüler/-innen kann schlicht auf das Nichtkennen der Methode zurückzuführen sein.

Interessant sind weiterhin zum Teil deutliche Diskrepanzen zwischen den Globalurteilen der Schüler/-innen und denen der Lehrkräfte: Bei der einfachen sowie der vertieften Warum-Erklärung und bei der einfachen Wie-Erklärung differieren die Globalurteile im Mittel um mehr als eine halbe Notenstufe. Dabei bewerten die Schüler/-innen die Warum-Erklärungen deutlich besser, die einfache Wie-Erklärung dagegen deutlich schlechter als die Lehrkräfte. Aufgrund des gemeinsamen Kontextes im Schulalltag der Schüler/-innen und der Lehrkräfte sind diese Differenzen überraschend und widersprechen zum Teil den Ergebnissen aus anderen Studien[50]. Eine Vermutung, die bei der Bewertung der Skalen noch überprüft werden muss, ist, dass die Schüler/-innen die Persönlichkeit der erklärenden Lehrkraft in die Bewertung mit einfließen lassen. Die Warum-Erklärungen werden von einer männlichen, sehr klar und selbstbewusst wirkenden Lehrkraft gegeben. Auch hier liegt die Annahme nahe, dass die Schüler/-innen derartige Lehrerpersönlichkeiten aus dem eigenen Unterricht kennen (und schätzen) und daher die Erklärungen dieser Lehrkraft positiv bewerten.

Eine weitere Auffälligkeit, die sich in den Globalurteilen der Schüler/-innen zeigt, ist die durchgehend schlechtere Bewertung bei beiden Varianten der Wie-Erklärung im Vergleich zu allen anderen Statusgruppen. Besonders ausgeprägt ist diese abweichende Bewertung beim Vergleich mit den Globalurteilen der Didaktiker/-innen (vgl. Tab. 10). Auch hier könnten die Sprechweise (eher langsam) und das Auftreten der erklärenden Lehrkraft (eher ruhig) einen Ausschlag bei den Bewertungen der Schüler/-innen gegeben haben.

Interessant ist weiterhin die Tatsache, dass sich die Mittelwerte innerhalb eines einzelnen Vignettenpaares (d. h. je zwischen den beiden Was-/ Warum- und Wie-Erklärungen) stärker als bei den anderen Statusgruppen ähneln (sie unterscheiden sich um höchstens 0,18). Dies könnte

[50] So stellen z. B. Wisniewski, Zierer, Dresel und Daumiller (2020) eine hohe Übereinstimmung der Urteile von Lehrkräften und Schülerinnen und Schülern fest, allerdings bei Merkmalen, die sich auf der Oberflächenstruktur von Unterricht befinden.

wiederum ein Anzeichen dafür sein, dass die erklärende Person bei den Schülerinnen und Schülern einen größeren Einfluss auf die Bewertung hat als in den anderen Statusgruppen, selbst wenn die hohe Standardabweichung dieser Vermutung wiederum entgegensteht. Bemerkenswert ist in diesem Zusammenhang, dass sich auch die Spannweiten innerhalb der gleichen Prozedur ähneln: Bei beiden Was-Erklärungen reicht sie von 1+ bis 5-, bei beiden Wie-Erklärungen von 1+ bis 6-, bei den Warum-Erklärungen unterscheiden sich die Spannweiten leicht (1+ bis 6- bei der einfachen Variante, 1+ bis 5 bei der vertieften Variante). Auch diese Ähnlichkeiten könnten bei den Schülerinnen und Schülern darauf zurückzuführen sein, dass ihre Bewertungen stärker von der erklärenden Person abhängen (s. hierzu Punkt *10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse* in dieser Arbeit).

Stellt man die Globalurteile der einfachen denjenigen der vertieften Varianten gegenüber, so erkennt man, dass bei den Was- und den Wie-Erklärungen die vertiefte Variante besser abschneidet, bei den Warum-Erklärungen jedoch die einfache Variante bevorzugt wird, wenn auch die Unterschiede, wie gerade angedeutet, nicht groß sind. Das könnte daran liegen, dass die Schüler/-innen der Erklärtiefe keine besondere Bedeutung für die Erklärqualität beimessen oder daran, dass sie die Unterschiede in der Erklärtiefe wenig wahrnehmen.

Den deskriptiven Ergebnissen der Schüler/-innen und damit der Perspektive der Adressatinnen und Adressaten wird im Folgenden die Expertenperspektive gegenübergestellt, wobei die drei Statusgruppen Studierende, Lehrkräfte sowie Didaktiker/-innen getrennt voneinander betrachtet werden.

Studierende

Die Mittelwerte der Bewertungen sämtlicher Vignetten sind in der Statusgruppe der Studierenden auf der Notenstufe *gut* und bewegen sich zwischen 1,99 und 2,50. Die Standardabweichungen sind etwas niedriger als bei den Schülerinnen und Schülern, aber mit Werten meist knapp unter einer Notenstufe immer noch hoch (vgl. Tab. 10). Bei den Mittelwerten ist bemerkenswert, dass die beste (1,99) auf eine vertiefte Erklärung (C) und die schlechteste (2,50) Note auf eine einfache Erklärung (D) fällt.

Auch die Globalurteile der Studierenden streuen weit, wenn auch bei den schlechten Bewertungen weniger stark als bei den Schülerinnen und Schülern: Die Studierenden vergeben Globalurteile von 1+ bis 5-. Es fällt auf, dass auch bei den Studierenden jedes Video mindestens einmal mit der bestmöglichen Bewertung 1+ versehen wurde und dass bei zwei Videos die schlechteste Note 4 (einfache Wie-Erklärung) beziehungsweise 4- (vertiefte Was-Erklärung) ist (vgl. Tab. 10). Gerade im unteren Bereich unterscheiden sie sich in ihren Bewertungen deutlich von denen der Schüler/-innen.

Die Statusgruppe der Studierenden ist diejenige, deren Mittelwerte sich insgesamt über alle Vignetten hinweg am ähnlichsten sind. Stellt man die Statusgruppen einander gegenüber, so bewerten die Studierenden insgesamt milder als die Schüler/-innen und als die Lehrkräfte (die Studierenden bewerten jeweils fünf von sechs Vignetten besser, vgl. Tab. 10). Im Vergleich mit den Didaktikerinnen und Didaktikern zeigt sich keine solche Tendenz (hier bewerten die Studierenden drei Mal besser, drei Mal schlechter, vgl. Tab. 10). Die Studierenden scheinen also einen eigenen – bereits vom Studium geprägten – Blick auf die Erklärungen zu haben.

Betrachtet man die Vignettenpaare (einfach vs. vertieft) so zeigt sich das gleiche Ergebnis wie bei den Schülerinnen und Schülern: Bei den Was- und Wie-Erklärungen wird die vertiefte Variante besser bewertet, bei den Warum-Erklärungen hingegen die einfache. Die Unterschiede zwischen den beiden Erklärungen der gleichen Prozedur fallen stärker aus als bei den Schülerinnen und Schülern; Studierende nehmen also die unterschiedliche Erklärtiefe als bedeutender für die Erklärqualität wahr als die Schülerinnen und Schüler. Auffällig bleibt weiterhin, dass die vertiefte Variante der Warum-Erklärung entgegen der theoretischen Annahme schlechter bewertet wird als die einfache Variante.

Als nächstes werden die Ergebnisse der Lehrkräfte dargelegt. Zu beachten sind die deutlich geringeren Stichprobengrößen bei dieser (N=35) sowie bei der Statusgruppe der Didaktiker/-innen (N=31) im Vergleich zu den Stichprobengrößen der Schüler/-innen (N=113) und der Studierenden (N=78). Die Ergebnisse der Lehrkräfte beziehungsweise der Didaktiker/-innen zeigen daher Tendenzen an.

Lehrkräfte

Die Mittelwerte der Globalurteile der Lehrkräfte bewegen sich zwischen den Notenstufen *gut* (2,05) und *befriedigend* (3,26), und liegen damit im gleichen Spektrum wie die Mittelwerte über alle Statusgruppen hinweg. Die Standardabweichungen schwanken von knapp einer Notenstufe bis etwas über eine Notenstufe, sind also ähnlich hoch wie in den vorherigen Statusgruppen der Schüler/-innen und der Studierenden.

Eine hohe Streuung zeigt sich ebenso in den Globalurteilen der Lehrkräfte, sie vergaben Noten von 1 bis 5-, das bedeutet, dass die Lehrkräfte im unteren Bewertungsbereich ähnlich bewerten wie die Studierenden. Auffällig ist, dass im Gegensatz zu den gerade besprochenen Statusgruppen die Note 1+ kein einziges Mal vergeben wurde, was sicherlich daran liegt, dass die Note im Schulunterricht für die Lehrkräfte nicht existiert. Die Uneinigkeit innerhalb dieser Statusgruppe bezüglich der Globalurteile zu den Erklärungen überrascht, denn aufgrund des Expertisestatus' der Lehrkräfte war von einer stärkeren Einigkeit auszugehen.

Ins Auge fallen die Bewertungen der vertieften Was-Erklärung, weil hier die schlechteste Note die 4- ist (im Gegensatz zu allen anderen, bei denen die schlechteste Note 5 bzw. 5- ist) und der vertieften Warum-Erklärung, weil hier die beste Note, die vergeben wurde, die 2+ ist; bei keinem anderen Video und von keiner anderen Statusgruppe wurde so eine „niedrige" beste Bewertung abgegeben. Keine der befragten Lehrkräfte fand also die Methode, eine Redewendung mit etymologischem Hintergrund zu erklären, mangelhaft, was bedeutet, dass den Lehrkräften der Mehrwert dieser Informationen bewusst ist. Im Gegensatz dazu bewertete keine der Lehrkräfte die inhaltlich-funktionale Begründung der Warum-Erklärung als sehr gut, – ein erneuter Hinweis darauf, dass die Bewertungen der Warum-Erklärungen nicht so ausfallen, wie erwartet wurde.
Zum Vergleich der Mittelwerte der Lehrkräfte mit denen der anderen Expertenstatusgruppen ist zu sagen, dass die Lehrkräfte mit einer Bewertung von *befriedigend* bei drei der sechs Vignetten (vgl. Tab.10) diejenige Statusgruppe sind, die insgesamt am strengsten bewertet. Das schlechteste Globalurteil insgesamt findet sich bei den Bewertungen der vertieften Warum-Erklärung der Lehrkräfte: Mit einem Mittelwert von 3,26 hebt es sich doch deutlich von den anderen Noten ab. Außerdem werden im Vergleich mit den Studierenden (s. o.) und den Didaktikern und Didaktikerinnen jeweils fünf der sechs Vignetten von den Lehrkräften schlechter bewertet, wenn auch jeweils unterschiedliche, nämlich bei den Studierenden alle bis auf die einfache Wie-Erklärung und bei den Didaktikerinnen und Didaktikern alle bis auf die einfache Was-Erklärung (letztere bewerten beide Statusgruppen gleich mit 2,92, vgl. Tab. 10). Die Diskrepanzen zu den Urteilen der Schüler/-innen wurden oben unter den Erläuterungen zu dieser Statusgruppe bereits dargelegt.
Die Bewertungen nach der Variation der Erklärtiefe stellen sich bei den Lehrkräften anders dar: Sie bewerten lediglich die vertiefte Was-Erklärung besser, bei den Warum- und Wie-Erklärungen präferieren sie die einfache Erklärung. Damit fallen sie bei der Wie-Erklärung als einzige Statusgruppe auf, die der einfachen Erklärung den Vorzug gibt. Hier liegt die Vermutung nahe, dass die Lehrkräfte im Schreibunterricht selbst eher mit der Methode Modelltext, denn mit der Methode lautes Denken arbeiten und sie deswegen die ihnen vertraute Methode präferieren.
Als letztes werden wie angekündigt die Ergebnisse der Statusgruppe der Didaktiker/-innen erläutert.

Didaktiker/-innen

Die Mittelwerte der Didaktiker/-innen liegen zwischen *gut* (1,57) und *befriedigend* (2,92) und somit in einer ähnlichen Spannweite wie die der Lehrkräfte, wenn sie auch im Mittel besser sind als die Globalurteile der Lehrkräfte. Bemerkenswert ist der Mittelwert 1,57 für die vertiefte Wie-Erklärung (gleichzeitig der beste Mittelwert insgesamt), der sich schon nah an der Bewertung *sehr gut* bewegt und im Vergleich zu sämtlichen Mittelwerten der anderen Statusgruppen deutlich heraussticht (vgl. Tab. 10).

Die Standardabweichungen in der Gruppe der Didaktiker/-innen sind wie bei den anderen Statusgruppen hoch (vgl. Tab.10). Es zeigt sich außerdem eine ähnliche Streuung wie bei den Urteilen der Lehrkräfte, nämlich von 1 bis 5-. Auch bei den Didaktikerinnen und Didaktikern wird die Note 1+ nicht vergeben. In unteren Bewertungsbereich wird die Note 5- nur bei einer Vignette vergeben (bei der einfachen Warum-Erklärung), in den anderen Fällen ist die Note 4- (bei der vertieften Warum-Erklärung) beziehungsweise die Note 5 (bei allen anderen Vignetten) die schlechteste vergebene Note. Obwohl die Didaktiker/-innen damit weniger stark als die anderen Statusgruppen in ihren Globalurteilen streuen, ist diese Streuung dennoch als hoch anzusehen und überrascht aufgrund der hohen Expertise bei der Bewertung von Unterricht, die Didaktiker/-innen aufweisen.

Auffälligkeiten bei den Globalurteilen der Didaktiker/-innen in Bezug auf die Mittelwerte der anderen Statusgruppen sind folgende: Im Vergleich mit den Urteilen der Schüler/-innen und der Studierenden bewerten die Didaktiker/-innen beide Warum-Erklärungen schlechter, beide Wie-Erklärungen hingegen besser; dahingegen wird die einfache Was-Erklärung schlechter und die vertiefte Was-Erklärung besser bewertet. Der Vergleich mit den Bewertungen der Lehrkräfte wurde oben unter den Beschreibungen der Ergebnisse aus dieser Statusgruppe bereits dargelegt.

Interessant ist in dieser Statusgruppe der Blick auf die Bewertungen der Erklärtiefe: Die Didaktiker/-innen sind die einzige Statusgruppe die durchgehend die vertiefte Variante der Erklärungen bevorzugen. Bei den Was- und Wie-Erklärungen tritt diese Präferenz mit Unterschieden von 0,38 (bei den Wie-Erklärungen) beziehungsweise von 0,94 (bei den Was-Erklärungen) sehr deutlich zu Tage, bei den Warum-Erklärungen ist eher von einer Tendenz zu sprechen (2,63 vs. 2,75, vgl. Tab. 10). Auffällig ist das Ergebnis dennoch, da die anderen drei Statusgruppen bei den Warum-Erklärungen die einfache Variante bevorzugen. Hier zeigt sich, dass die Didaktiker/-innen die Erklärtiefe als relevant für die Erklärqualität einschätzen, und zwar unabhängig von der Art der Erklärung. Sie bewerten als einzige Statusgruppe „theoriekonform".

8.2.1.3 Zusammenfassung der Ergebnisse in den Globalurteilen zur Erklärqualität

Insgesamt zeigen sich über die Statusgruppen hinweg und statusgruppenspezifisch folgende Tendenzen: Die Erklärungen werden im Mittel zwischen *gut* und *befriedigend* bewertet. Keine der befragten Statusgruppen zeigt eine ausgeprägte Tendenz zur Milde oder Strenge. Zu erwähnen ist aber, dass die Didaktiker/-innen immerhin drei der sechs Erklärungen besser bewerten als die anderen Gruppen, und die Schüler/-innen und die Lehrkräfte drei der sechs Erklärungen strenger bewerten als die übrigen Gruppen.

Dennoch kann man auch hier nicht von einer klaren Tendenz zu großzügigeren (bei den Didaktikerinnen und Didaktikern) oder strengeren (bei den Schülerinnen und Schülern sowie den

Lehrkräften) Bewertungen sprechen, denn die Standardabweichungen sind als hoch und die Streuungen als groß anzusehen. Damit sind weiterhin sehr heterogene Bewertungen sowohl über die Statusgruppen hinweg als auch innerhalb jeder der Statusgruppen festzuhalten. Keine der befragten Gruppen ist sich einig über die globale Bewertung einer Erklärung, was gerade bei den Gruppen mit hoher Expertise überrascht.

Bezogen auf die Wahrnehmung der Erklärtiefe wird in der Tendenz (mit gerade geschilderten Ausnahmen) die vertiefte Was- und die vertiefte Wie-Erklärung besser eingeschätzt, bei der Warum-Erklärung dagegen die einfache Variante.

Festzuhalten ist darüber hinaus, dass die Heterogenität in den Bewertungen nicht auf etwaige Reihenfolgeeffekte in der Präsentation der einzelnen Vignetten zurückzuführen ist. Wie unter Punkt *7.3 Untersuchungsinstrument: Online-Fragebogen* in dieser Arbeit bereits erläutert wurde die Reihenfolge systematisch variiert, indem zwei Versionen (A und B) des Fragebogens eingesetzt wurden, um Reihenfolgeeffekte auszuschließen. In Version A wurden die vertieften Varianten vor den einfachen bewertet, in Version B genau umgekehrt. Statistisch wurde überprüft, ob die Reihenfolge der gesehenen Erklärungen einen Einfluss auf die Bewertungen hatte: Eine Varianzanalyse mit gemischtem Design (bei der die sechs Globalurteile pro Teilnehmer/-in als sechs Messzeitpunkte und die beiden Versionen A und B in einem zweistufigen Gruppierungsfaktor berücksichtigt wurden) ergab keine signifikanten Reihenfolgeeffekte[51] (vgl. Gaier & Schilcher, 2021, in Vorb.). In der Tendenz wurden allerdings die jeweils erst gesehenen Vignetten etwas schlechter bewertet, unabhängig davon, ob zuerst die einfache oder die vertiefte Vignette gesehen wurde (vgl. Gaier & Schilcher, 2021, in Vorb.)[52].

8.2.2 Skalenebene: Bewertungen der Aspekte guten Erklärens

Nachdem die Ergebnisse der Globalurteile dargelegt wurden, werden im Folgenden die deskriptiven Ergebnisse der einzelnen Skalen berichtet (Forschungsfragen 1b und 2b).

[51] Für Hilfe bei statistischen Fragen und Berechnungen sei Alfred Lindl gedankt.

[52] Im Datensatz befinden sich keine Ausreißer. Alle sechs Globalurteile waren gemäß dem Shapiro-Wilk Test nicht normalverteilt ($p < .001$), was jedoch aufgrund der Stichprobengröße von N=257 vernachlässigt werden kann. Der Haupteffekt bezüglich der Videoreihenfolge beträgt $F(1,235) = 0,43$; $p = 0,51$ und ist somit nicht signifikant; bei Auflösung des signifikanten Interaktionseffekts zwischen der Videoreihenfolge und dem Messwiederholungsfaktor $F(5,1175) = 5,94$; $p \leq 0,01$ zeigt sich obige Tendenz (vgl. Gaier & Schilcher, 2021, in Vorb.).

8.2.2.1 Statusgruppenübergreifende Bewertungen der Aspekte

Tabelle 11: Mittelwerte, Standardabweichungen und Spannweiten (über alle vier Statusgruppen) zu den fünf Aspekten guten Erklärens, zur Persönlichkeitswirkung und zu Fachliche Relevanz*
N_{max}=257

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Aspekt						
Strukturiertheit (4 Items)	2,84 (0,86) [1,00; 5,50]	2,43 (0,84) [1,00; 4,75]	2,64 (0,80) [1,00; 4,75]	2,80 (0,98) [1,00; 6,00]	2,58 (0,89) [1,00; 5,00]	2,34 (0,95) [1,00; 5,00]
Adressatenorientierung (4/5 Items)	2,93 (0,83) [1,20; 5,75]	2,77 (0,81) [1,00; 5,25]	3,06 (0,92) [1,00; 5,60]	3,08 (1,10) [1,00; 6,00]	2,87 (0,85) [1,00; 5,25]	2,63 (0,90) [1,00; 5,50]
Sprech-/ Körperausdruck (7 Items)	2,03 (0,65) [1,00; 3,71]	1,95 (0,65) [1,00; 4,57]	2,53 (0,78) [1,00; 5,00]	2,60 (0,90) [1,00; 5,71]	2,54 (0,81) [1,00; 5,00]	2,32 (0,81) [1,00; 5,00]
Sprachliche Verständlichkeit (4 Items)	2,79 (0,78) [1,00; 4,75]	2,69 (0,77) [1,00; 5,25]	3,00 (0,94) [1,00; 5,75]	3,20 (0,97) [1,00; 6,00]	2,68 (0,72) [1,00; 4,50]	2,63 (0,81) [1,00; 6,00]
Persönlichkeitswirkung (3 Items)	2,37 (0,76) [1,00; 5,00]	2,36 (0,92) [1,00; 5,00]	2,46 (0,89) [1,00; 5,67]	2,65 (1,02) [1,00; 5,33]	2,46 (0,98) [1,00; 5,33]	2,33 (0,91) [1,00; 5,33]
Erklärtiefe (5 Items)	3,05 (0,93) [1,20; 6,00]	2,38 (0,91) [1,00; 5,20]	2,89 (0,93) [1,00; 5,00]	3,05 (1,01) [1,00; 6,00]	2,81 (0,90) [1,00; 6,00]	2,61 (1,03) [1,00; 5,60]
Fachliche Relevanz (2/3 Items)	2,80 (0,97) [1,00; 5,50]	2,82 (0,95) [1,00; 6,00]	2,50 (0,90) [1,00; 6,00]	2,64 (0,86) [1,00; 6,00]	2,70 (0,97) [1,00; 6, 00]	2,67 (1,08) [1,00; 6,00]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6.
*Die angegebenen Spannweiten der einzelnen Aspekte sind die Spannweiten der Skalenmittelwerte, nicht der Bewertungen der einzelnen Items; vgl. Erläuterungen unten im Fließtext. (Tabelle abgeändert aus Gaier & Schilcher, 2021, in Vorb.)

Zunächst werden deskriptiv die Ergebnisse (Mittelwerte, Standardabweichungen sowie Streuung) auf Skalenebene über alle Statusgruppen hinweg dargestellt (Forschungsfrage 1b), dazu wird mit den Aspekten, die im Gesamtprojekt FALKE gemeinsam erhoben wurden, begonnen. Alle Werte sind Tabelle 11 zu den Gesamtergebnissen zu entnehmen.

Zu Erinnerung: Es wurden alle Aspekte auf einer sechsstufigen Skala erhoben, wobei 1,00 die beste Bewertung („stimme voll zu“) und 6,00 die schlechteste Bewertung („stimme gar nicht zu“) darstellte. Die Spannweite der einzelnen Aspekte, die in Tabelle 11 angegeben ist, bezieht sich auf den gesamten Aspekt, es ist also je der kleinste und der größte Mittelwert der jeweiligen Skala und nicht der einzelnen Items angegeben. Eine Spannweite von 1,00 bis 6,00 bedeutet somit, dass bei der betreffenden Skala mindestens eine Person bei jedem Item die höchste Bewertung („stimme

voll zu“ = 1,00) gegeben hat und gleichzeitig mindestens eine Person bei jedem Item die schlechteste Bewertung („stimme gar nicht zu“ = 6,00) erteilt hat.

Strukturiertheit

Die Mittelwerte zum Aspekt *Strukturiertheit* liegen mit Werten zwischen 2,34 und 2,84 im mittleren Bereich (vgl. Tab. 11). Die Standardabweichung ist relativ hoch, sie bewegt sich zwischen 0,80 und 0,98 und auch die Streuungen (die von 1,00; 4,75 bis 1,00; 6,00 reichen) weisen auf heterogene Bewertungen dieses Aspektes hin. Bezogen auf die Variation der Erklärtiefe schneiden bei der *Strukturiertheit* die vertiefte Was- und die vertiefte Wie-Erklärung besser ab, bei den Warum-Erklärungen wird die *Strukturiertheit* der einfachen Variante besser bewertet.

Adressatenorientierung

Beim Aspekt der *Adressatenorientierung* liegen die Mittelwerte alle in einem etwas niedrigeren Bereich (von 2,63 bis 3,08, vgl. Tab. 11). Die *Adressatenorientierung* ist insgesamt der am schlechtesten bewertete Aspekt. Die Standardabweichung ist mit Werten zwischen 0,81 und 1,10 hoch und auch die Streuung ist groß (sie reicht von 1,00; 5,25 bis 1,00; 6,00). Auch dieser Aspekt wird also insgesamt sehr unterschiedlich bewertet. Betrachtet man diese Bewertungen differenziert nach der Variation einfach versus vertieft, so zeigt sich ein analoges Ergebnis zu dem Aspekt der *Strukturiertheit*: Bei der Was- und der Wie-Erklärung schneiden die vertieften Varianten besser ab, bei der Warum-Erklärung schneidet die einfache Variante ganz leicht besser ab.

Sprech- und Körperausdruck

Der *Sprech- und Körperausdruck* wird mit Werten zwischen 1,95 und 2,60 besser bewertet (vgl. Tab. 11) und ist insgesamt der Aspekt, der die besten Bewertungen aufweist. Gleichzeitig herrscht hier etwas mehr Einigkeit als bei den anderen Aspekten: Die Standardabweichung reicht von 0,65 bis 0,90 und ist für die vorliegenden Ergebnisse relativ niedrig. Das ist gerade deswegen interessant, weil drei Sprecher/-innen die Erklärungen gegeben haben, die doch ganz unterschiedliche Sprechweisen haben. Vermutlich sind die guten Werte aber darauf zurückzuführen, dass alle drei Lehrkräfte verständlich und klar gesprochen haben. Störende sprecherische Auffälligkeiten wie Lispeln oder Nuscheln hätten vermutlich zu schlechteren Ergebnissen geführt. Die Streuung (mit Werten von 1,00; 3,71 bis 1,00; 5,71) ist dennoch als hoch anzusehen. Beim Vergleich der einfachen und vertieften Varianten lässt sich feststellen, dass auch bei diesem Aspekt die vertiefte Was- und Wie-Erklärung besser abschneiden, bei der Warum-Erklärung dagegen erneut die einfache Variante leicht besser bewertet wird. Erwähnt werden soll in diesem Zusammenhang kurz der Vergleich mit den Urteilen aus der sprechwissenschaftlichen Fachperspektive: Gunga (2022, in

Vorb.) wertete die Skala aus sprechwissenschaftlicher Expertensicht aus. Die Ergebnisse dort sind alle schlechter als die Ergebnisse der Studienteilnehmer/-innen. Expertinnen und Experten bewerten den Sprech-/Körperausdruck also insgesamt strenger.

Sprachliche Verständlichkeit

Die *Sprachliche Verständlichkeit* liegt mit Bewertungen zwischen 2,63 und 3,20 in einem ähnlichen Bereich wie die Aspekte *Strukturiertheit* und *Adressatenorientierung* (vgl. Tab. 11). Die Standardabweichung ist mit Werten (knapp) unter 1 etwas niedriger als bei den anderen Aspekten. Die Streuung ist jedoch ähnlich hoch wie bei den anderen Aspekten (von 1,00; 4,50 bis 1,00; 6,00). Insgesamt muss man auch beim Aspekt *Sprachliche Verständlichkeit* von heterogenen Bewertungen sprechen. Die Differenz der einfachen und vertieften Vignetten stellt sich wie bei den übrigen Aspekten dar: Bei Was- und Wie-Erklärungen wird die vertiefte Variante, bei den Warum-Erklärungen hingegen die einfache Variante bevorzugt (vgl. Tab. 11).

Erklärtiefe

Schließlich ist der FALKE-D-spezifische Aspekt der *Erklärtiefe* von Interesse. Die Mittelwerte bewegen sich hier analog zu den übrigen Aspekten im Bereich von 2,38 bis 3,05. Insgesamt befindet sich die *Erklärtiefe* jedoch mit vier von sechs Vignetten, die mit Werten um 3 bewertet werden, eher im unteren Bewertungsbereich (vgl. Tab. 11). Auch bei diesem Aspekt bewegt sich die Standardabweichung bei Werten um 1, die Streuung ist gewohnt hoch (von 1,00; 5,00 bis 1,00; 6,00). Bezogen auf die Differenz der Vignetten untereinander zeigen sich erneut analoge Ergebnisse wie bei den anderen Aspekten: Die vertiefte Was- und die vertiefte Wie-Erklärung schneiden beim Aspekt *Erklärtiefe* besser ab als die einfachen Varianten. Das bedeutet, dass die in den Vignetten umgesetzte größere Erklärtiefe auch so wahrgenommen wird. Bei der Warum-Erklärung wird dagegen der einfachen Variante in den Bewertungen eine größere *Erklärtiefe* zugesprochen. Bemerkenswert ist in diesem Zusammenhang die ungewöhnlich deutliche Differenz bei den Was-Erklärungen, hier wird die Erklärtiefe der vertieften Variante mit 2,38 deutlich besser bewertet als die der einfachen Variante mit dem Mittelwert 3,05 (vgl. Tab. 11). Es ist daher davon auszugehen, dass die „höhere" Erklärtiefe bei der Erklärung mit etymologischem Hintergrund im Vergleich zur einfachen Variante (über das Geben von Synonymen) über die Statusgruppen hinweg deutlich wahrgenommen wird und als besser empfunden wird. Gleiches gilt für die inhaltlich-funktionale Begründung der Widerlegung des Arguments (wenn auch nicht mit der gleichen Deutlichkeit). Beide Methoden werden also theoriekonform als besser wahrgenommen als die Methoden, die weniger kognitive Aktivierung auslösen.

Persönlichkeitswirkung

Der Aspekt der *Persönlichkeitswirkung*, der als Kontrollvariable eingeführt wurde, weist etwas bessere Bewertungen von 2,33 bis 2,65 auf (vgl. Tab. 11). Die Standardabweichung beträgt (mit Ausnahme des Werts 0,76 bei der einfachen Was-Erklärung) Werte um 1 (vgl. Tab. 11), und bewegt sich somit in ähnlichem Rahmen wie bei den anderen Aspekten. Gleiches gilt für die Streuung, die ebenfalls weit ausfällt (sie reicht von 1,00; 5,00 bis 1,00; 5,67). Betrachtet man die Vignetten nach ihrer Variation so wiederholt sich das Ergebnis der anderen Aspekte: Die vertiefte Was- und Wie-Erklärung werden besser bewertet als die einfache Variante (wenn auch der Unterschied bei den Was-Erklärungen verschwindend gering ist). Bei der Warum-Erklärung wird dagegen die einfache Variante präferiert.

Deskriptive Bewertungen der Fachlichen Relevanz

Der Aspekt *Fachliche Relevanz* ist wie gesagt ebenfalls FALKE-D-spezifisch. Er wurde wie bereits geschildert zur Kontrolle der inhaltlichen Validität erhoben und ist kein Kriterium für die Erklärqualität. Die Mittelwerte von *Fachliche Relevanz* bewegen sich ebenfalls im mittleren Bereich, mit Werten von 2,50 bis 2,82 sind sie sich insgesamt recht ähnlich (vgl. Tab. 11). Auch hier weist die Standardabweichung durchgehend in etwa eine Höhe von 1 auf (vgl. Tab. 11). Die Streuung der Mittelwerte bewegt sich von 1,00; 5,50 bis 1,00; 6,00 in einem hohen Bereich (wie bei den anderen Aspekten ebenfalls). Bezogen auf die Variation der Erklärtiefe fällt der Aspekt *Fachliche Relevanz* als einziger aus dem Schema der anderen Aspekte heraus: Hier zeigen sich bei der vertieften Was- und Warum-Erklärung schlechtere Ergebnisse als bei den einfachen Varianten. Dahingegen wird die einfache Wie-Erklärung schlechter bewertet als die vertiefte Variante.

Der Aspekt *Fachliche Relevanz* hat wie unter Punkt *4.2.6 Persönlichkeitswirkung und Fachliche Relevanz als Kontrollvariablen* (in dieser Arbeit) dargelegt formativen Charakter, die interne Konsistenz ist schwach. Die Ergebnisse sind entsprechend unter diesen Vorgaben zu deuten und nicht als Teil der Erklärqualität aufzufassen. Daher wird zur genaueren Interpretation auf die Ergebnisse der einzelnen Items ausgewichen. In Tabelle 12 sind die Mittelwerte für die einzelnen Items dargestellt, aufgeteilt in die drei vertieften (Video A-C) und die drei einfachen (Video D-E) Erklärungen.

Tabelle 12: Mittelwerte, Standardabweichungen und Spannweiten (über alle vier Statusgruppen) für die einzelnen Items von Fachliche Relevanz
aufgeteilt in einfache und vertiefte Erklärungen; N_{max}=257

Fachliche Relevanz

Erklärtiefe **Item**	**einfache Erklärungen** (Video D-F) *M (SD)* *[range]*	**vertiefte Erklärungen** (Video A-C) *M (SD)* *[range]*
FR1) Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt. (nur SLD)	2,67 (0,63) [1,00; 4,67]	2,58 (0,74) [1,00; 4,67]
FR2) Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.	2,70 (0,88) [1,00; 6,00]	2,94 (0,92) [1,00; 6,00]
FR3) […]* sollte regelmäßig im Deutschunterricht verwendet werden.	2,53 (0,87) [1,00; 5,67]	2,38 (0,89) [1,00; 6,00]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Items: 1 bis 6.
*[…] steht für das jeweilige Thema und die jeweilige Erklärvariante.

Aus diesen Ergebnissen zu den einzelnen Items der *Fachlichen Relevanz* lassen sich folgende Schlüsse zur inhaltlichen Validität ziehen. Begonnen wird mit Item FR2), das nach der Vertrautheit der gesehenen Erklärung im Kontext des Deutschunterrichts fragt. Der Wert über alle Statusgruppen hinweg ist für Item FR2) bei den einfachen Erklärungen leicht besser als bei den vertieften. Offensichtlich erkennen die Studienteilnehmer/-innen also die Methoden, die in FALKE-D als einfache Varianten bezeichnet werden, als typisch für den Deutschunterricht. Dieses Ergebnis ist zwar aufgrund des geringen Mittelwertunterschiedes sowie der hohen Standardabweichungen und der großen Streuungen (vgl. Tab. 12) nur eine Tendenz, entspricht allerdings den Erwartungen: Bei den einfachen Varianten wurden Methoden gewählt, von denen angenommen wird, dass sie im Unterricht häufig vorkommen (z. B. der Rückgriff auf die Struktur bei der Bewertung eines Arguments in Video E). Bei den vertieften Varianten handelt es sich dagegen um Methoden, von denen angenommen wird, dass sie im Deutschunterricht noch nicht so etabliert sind (z. B. die Methode lautes Denken in Video C).

Vergleicht man alle drei Items untereinander, so fällt auf, dass Item FR1) (zum deutschspezifischen Kompetenzaufbau) und Item FR3) (zur wünschenswerten regelmäßigen Verwendung der Methode) insgesamt besser bewertet werden als Item FR2), das „nur" nach dem Bekanntheitsgrad der verwendeten Erklärmethode fragt. Das bedeutet, dass die verwendeten Methoden und Themen als sinnvoll und deutschspezifisch erachtet werden, obwohl sie zum Teil aus dem Unterricht nicht so bekannt zu sein scheinen.

Beim Vergleich der Bewertungen hinsichtlich der Erklärtiefe ist bemerkenswert, dass bei Item FR1) (das wie dargelegt nur von den Expertengruppen Studierende, Lehrkräfte und Didaktiker/-innen beantwortet wurde) und bei Item FR3) die vertieften Varianten leicht besser beurteilt werden. Auch wenn hier ebenso nur von Tendenzen gesprochen werden kann (vgl. erneut die geringen Mittelwertunterschiede sowie die hohen Standardabweichungen und die weiten Streuungen in Tab. 12), so zeigen die Ergebnisse doch, dass die verwendeten vertieften Methoden als sinnvoll für den Deutschunterricht beurteilt werden. Dieses Ergebnis ist besonders bemerkenswert, da sie gleichzeitig als weniger gebräuchlich als die einfachen Varianten erachtet werden.
Wie bereits mehrmals erläutert wird die *Fachliche Relevanz* in die folgenden Berechnungen nicht mit einbezogen, da sie kein Qualitätskriterium für die Erklärungen darstellt. Der Aussagekraft der Ergebnisse dieses Aspektes ist mit der erfolgten Erläuterung der deskriptiven Ergebnisse genüge getan, in weiteren Analysen findet er daher keine Berücksichtigung mehr.

Zusammenfassung

Beim Vergleich der Ergebnisse einer Skala über alle Erklärungen hinweg fällt auf, dass (mit Ausnahme der *Fachlichen Relevanz*, die allerdings den gerade geschilderten Sonderfall darstellt, vgl. Tab. 12) in jeder Skala eine vertiefte Erklärung am besten abschneidet: Dabei handelt es sich bei den Aspekten *Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit* und *Persönlichkeitswirkung* um die vertiefte Wie-Erklärung, bei den Aspekten *Sprech-/Körperausdruck* sowie *Erklärtiefe* um die vertiefte Was-Erklärung (vgl. Tab. 11 sowie Gaier & Schilcher, 2021, in Vorb.). Das ist bemerkenswert, denn die umgesetzte Erklärtiefe schlägt sich offensichtlich in allen Aspekten nieder. Nach den Ergebnissen von FALKE-D ist davon auszugehen, dass eine höhere Erklärtiefe in einer Erklärung auch dazu führt, dass die Erklärung als strukturierter, adressatenorientierter usw. wahrgenommen wird.
Vergleicht man die Bewertungen der Aspekte innerhalb der Varianten der Was-/Warum- und Wie-Erklärungen, so ist zu sagen, dass die Bewertungen der Skalen die Ergebnisse der Globalurteile in etwa widerspiegeln: Bei den Was- und Wie-Erklärungen wird in den Skalen die vertiefe Variante, bei den Warum-Erklärungen die einfache bevorzugt (vgl. Punkt *8.2.1.1 Statusgruppenübergreifende Globalurteile der Erklärqualität* in dieser Arbeit). Die Studienteilnehmer/-innen bewerten also kohärent, und der erste Eindruck ist nicht völlig anders als die differenzierte Bewertung der Aspekte.
Berichtenswert sind als letztes noch die Unterschiede zwischen den Aspektmittelwerten (d. h. die Bewertungen von *Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit, Sprech-/Körperausdruck, Persönlichkeitswirkung und Erklärtiefe*) der jeweils einfachen und vertieften Erklärung derselben Erklärprozedur (d. h. der Was-/ Wie- oder Warum-Erklärungen). Auffällig sind

hier die Aspekte *Strukturiertheit* und *Erklärtiefe*, bei denen die Mittelwerte zwischen den Varianten relativ deutlich differieren, bei *Strukturiertheit* z. B. bei den Was-Erklärungen um 0,41, bei *Erklärtiefe* z. B. bei den Was-Erklärungen um 0, 67 (vgl. Tab. 11). Das zeigt erneut, dass die unterschiedliche Umsetzung der Was-Erklärung besonders deutlich wahrgenommen wurde. Das verwundert im Nachhinein gar nicht so sehr, denn tatsächlich sind die Methoden etymologischer Hintergrund versus Geben von Synonymen grundlegend unterschiedlich, und so wurden sie auch wahrgenommen, wohingegen das Geben der strukturellen versus der inhaltlich-funktionalen Begründung (auch aufgrund der Umsetzung) nicht so deutlich unterschiedlich sind.
Bei den anderen Aspekten zeigen sich eher mittlere Abweichungen. Die Werte zeigen also, dass die einzelnen Aspekte in den beiden Varianten derselben Prozedur, wenn auch in verschiedener Ausprägung (s. o.), unterschiedlich wahrgenommen werden.

8.2.2.2 Statusgruppenspezifische Bewertungen der Aspekte

Im Folgenden werden analog zu den Globalurteilen auch die statusgruppenspezifischen Mittelwerte, Standardabweichungen und Spannweiten der Aspektbewertungen berichtet (Forschungsfrage 2b). Dabei dienen die Tabellen als Gesamtüberblick, in denen diese Ergebnisse nach Statusgruppe getrennt angegeben sind. Im Fließtext werden die Ergebnisse zusammenfassend berichtet, wobei Auffälliges genauer dargestellt wird.

Schüler/-innen

Tabelle 13: Mittelwerte, Standardabweichungen und Spannweiten zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Schüler/-innen*
N_{max}=113

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Aspekt						
Strukturiertheit (4 Items)	2,76 (0,80) [1,25; 5,25]	2,66 (0,80) [1,00; 4,75]	2,62 (0,79) [1,00; 4,75]	2,68 (0,96) [1,00; 6,00]	2,98 (0,81) [1,25; 5,00]	2,74 (0,90) [1,00; 5,00]
Adressatenorientierung (4 Items)	2,95 (0,85) [1,25; 5,75]	2,96 (0,82) [1,50; 5,25]	2,96 (0,84) [1,25; 5,50]	3,03 (1,12) [1,00; 6,00]	3,20 (0,86) [1,25; 5,25]	3,07 (0,89) [1,25; 5,50]
Sprech-/ Körperausdruck (7 Items)	2,26 (0,63) [1,14; 3,71]	2,10 (0,67) [1,14; 4,57]	2,25 (0,67) [1,00; 4,14]	2,38 (0,86) [1,00; 5,71]	2,70 (0,87) [1,43; 5,00]	2,51 (0,86) [1,00; 5,00]
Sprachliche Verständlichkeit (4 Items)	2,82 (0,67) [1,25; 4,50]	2,77 (0,76) [1,25; 5,25]	2,72 (0,83) [1,00; 4,50]	2,95 (0,98) [1,00; 6,00]	2,94 (0,64) [1,25; 4,50]	2,93 (0,81) [1,00; 6,00]
Persönlichkeitswirkung (3 Items)	2,53 (0,70) [1,00; 5,00]	2,46 (0,89) [1,00; 4,67]	2,20 (0,76) [1,00; 4,33]	2,32 (1,00) [1,00; 5,33]	2,71 (1,18) [1,00; 5,33]	2,53 (0,94) [1,00; 5,33]
Erklärtiefe (5 Items)	2,88 (0,84) [1,20; 5,40]	2,61 (0,87) [1,00; 4,80]	2,83 (0,87) [1,00; 5,00]	2,90 (1,00) [1,00; 6,00]	3,13 (0,88) [1,20; 6,00]	3,09 (0,99) [1,40; 5,60]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6.
*Angegeben sind die Spannweiten der Mittelwerte der Skala.

Bei der *Strukturiertheit*, der *Adressatenorientierung* und der *Sprachlichen Verständlichkeit* liegen die Bewertungen der Schüler/-innen im mittleren Bereich um den Wert 3. Die Standardabweichungen sind hoch (etwa 1) und die Streuungen groß. Obwohl die gesamte Spannweite selten ausgenutzt wird, beträgt sie bei den meisten Videovignetten mindestens vier Punkte Differenz. Angemerkt wird an dieser Stelle, dass die Bewertungen der Aspekte *Strukturiertheit, Adressatenorientierung* und *Sprachliche Verständlichkeit* in etwa den Globalurteilen, die für die einzelnen Aspekte erhoben wurden, nämlich den Werten der Items *Die Erklärung war gut strukturiert, Die Erklärung passte für mich* sowie *Die Lehrkraft hat sich gut verständlich ausgedrückt* pro Video, entsprechen (s. Übersicht über sämtliche Werte für die einzelnen Items in Anhang 10). Lediglich bei *Sprachliche Verständlichkeit* fallen die Werte für das Globalurteil insgesamt etwas besser aus als für die Skala. Damit ist noch einmal die Gültigkeit der Skalen bestätigt; sie treffen also tatsächlich das Verständnis von *Strukturiertheit, Adressatenorientierung* und *Sprachliche Verständlichkeit* der Schüler/-innen.

Die Schülerinnen und Schüler bewerten dagegen insgesamt den *Sprech-/Körperausdruck* sowie die *Persönlichkeitswirkung* etwas besser mit Werten um 2,5. Beim *Sprech-/Körperausdruck* fallen erneut die relativ geringen Standardabweichungen (mit Werten von 0,63 bis 0,87) auf; die Bewertungen stimmen hier also etwas mehr überein als bei den anderen Aspekten. Das Ergebnis ist gerade deshalb interessant, weil in FALKE-D drei unterschiedliche Sprecher/-innen gewählt wurden. Dass sich die Schüler/-innen dennoch insgesamt recht einig über die Bewertung des *Sprech-/Körperausdrucks* sind, zeigt, dass die Sprechweise und die Körpersprache der Erklärenden ähnlich wahrgenommen wird. Das Item, das als Globalurteil erhoben wurde (nämlich *Die Art und Weise, wie die Lehrkraft spricht, gefällt mir*) entspricht in etwa der Gesamtbewertung der Skala (vgl. Anhang 10). Bei der *Persönlichkeitswirkung* deuten die hohen Standardabweichungen (bei der einfachen Wie-Erklärung mit dem Wert von 1,18 sogar die höchste Standardabweichung innerhalb der Bewertungen der Schüler/-innen insgesamt) auf sehr unterschiedliche Wahrnehmungen dieses Aspektes hin und zeigt, dass die *Persönlichkeitswirkung* von Schüler/-in zu Schüler/-in unterschiedlich wahrgenommen wird, was nicht weiter überrascht, denn ob jemand als *sympathisch* wahrgenommen wird, liegt im Auge der/des Betrachterin/Betrachters.

Die *Erklärtiefe* wird von den Schülerinnen und Schülern mit Werten von 2,61 bis 3,13 versehen. Diese Bewertungen liegen in etwa auf der gleichen Ebene wie die Bewertungen von *Strukturiertheit, Adressatenorientierung* und *Sprachliche Verständlichkeit*. Die Standardabweichungen (von 0,84 bis 1,00) sind hoch und die Spannweiten groß, vier Mal beträgt die Differenz mehr als 4 Punkte. Bei der Bewertung dieses Aspekts herrscht also ebenso Uneinigkeit, das bedeutet, dass die Schüler/-innen die umgesetzte *Erklärtiefe* in den Vignetten nur zum Teil so bewerten, wie sie aus der Theorie abgeleitet wurden. Da die Schüler/-innen aber aus der Adressatenperspektive ohne didaktisches Hintergrundwissen bewerten, ist dieses Ergebnis nicht weiter überraschend.

Vergleicht man abschließend die Mittelwerte der einfachen Vignetten mit denen der vertieften Vignetten, so lässt sich die bis hier gezeigte Tendenz weiter bestätigen: Mit einzelnen kleinen Abweichungen werden die vertiefte Was- und die vertiefte Wie-Erklärung in allen Skalen besser bewertet, bei der Warum-Erklärung wird dagegen die einfache Erklärung bevorzugt. Im Vergleich zum Gesamtergebnis erkennt man in den Daten der Schüler/-innen, dass sie die gleichen Aspekte innerhalb eines Vignettenpaares im Mittelwert zwar unterschiedlich, aber doch ähnlich bewerten, denn die Differenzen betragen oft weniger als 0,1 Punkte (z. B. bei der *Erklärtiefe* der beiden Wie-Erklärungen oder bei der *Adressatenorientierung* der beiden Was-Erklärungen) und selten mehr als 0,2 Punkte (z. B. bei der *Erklärtiefe* der beiden Was-Erklärungen) (vgl. Tab. 13). Diese Tendenz wurde auch bei den Globalurteilen der Schüler/-innen festgestellt (vgl. Punkt *8.2.1.2 Statusgruppenspezifische Globalurteile der Erklärqualität* in dieser Arbeit). Offensichtlich werden also Unterschiede in der *Erklärtiefe* in den zwei Erklärungen derselben Prozedur von den

Schülerinnen und Schülern weniger deutlich wahrgenommen oder als weniger wichtig erachtet als von den anderen Statusgruppen. Das gilt für alle Erklärungen.

Als nächstes werden die Ergebnisse der Expertengruppen dargestellt, wobei mit denen der Studierenden begonnen wird.

Studierende

Tabelle 14: Mittelwerte, Standardabweichungen und Spannweiten zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Studierende*

N_{max}=78

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Aspekt						
Strukturiertheit (4 Items)	2,78 (0,86) [1,00; 5,00]	2,35 (0,84) [1,00; 4,25]	2,49 (0,84) [1,00; 4,75]	2,74 (0,86) [1,00; 4,25]	2,30 (0,65) [1,00; 3,50]	2,17 (0,89) [1,00; 4,75]
Adressatenorientierung (5 Items)	2,80 (0,81) [1,20; 5,20]	2,56 (0,77) [1,00; 4,60]	2,88 (0,92) [1,00; 5,60]	2,84 (0,80) [1,40; 5,00]	2,61 (0,62) [1,00; 4,00]	2,35 (0,56) [1,20; 3,60]
Sprech-/ Körperausdruck (7 Items)	1,89 (0,63) [1,00; 3,29]	1,84 (0,62) [1,00; 3,14]	2,68 (0,87) [1,00; 5,00]	2,47 (0,76) [1,00; 4,14]	2,63 (0,72) [1,00; 4,00]	2,35 (0,73) [1,14; 3,86]
Sprachliche Verständlichkeit (4 Items)	2,75 (0,92) [1,00; 4,75]	2,58 (0,82) [1,25; 5,25]	3,06 (1,01) [1,00; 5,50]	3,17 (0,73) [2,00; 4,75]	2,57 (0,72) [1,00; 4,00]	2,42 (0,69) [1,00; 4,50]
Persönlichkeitswirkung (3 Items)	2,12 (0,82) [1,00; 4,00]	2,29 (1,00) [1,00; 5,00]	2,50 (0,90) [1,00; 4,00]	2,61 (0,89) [1,00; 4,33]	2,41 (0,65) [1,33; 4,00]	2,40 (0,92) [1,00; 4,33]
Erklärtiefe (5 Items)	3,04 (0,92) [1,20; 5,40]	2,19 (0,82) [1,00; 4,40]	2,74 (0,93) [1,00; 5,00]	3,13 (0,88) [1,40; 5,20]	2,64 (0,76) [1,20; 4,60]	2,22 (0,82) [1,00; 4,40]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6.
*Angegeben sind die Spannweiten der Mittelwerte der Skala.

Die Studierenden bewerten die einzelnen Aspekte insgesamt besser als die Schüler/-innen, die Mittelwerte liegen insgesamt im Bereich von gut 2 bis etwa 3 Punkten. *Strukturiertheit* und *Adressatenorientierung* werden im Mittel recht ähnlich bewertet. Auch die Standardabweichungen sind mit Werten unter 1 für *Strukturiertheit* und *Adressatenorientierung* relativ parallel. Auffällig ist die *Adressatenorientierung* der vertieften Wie-Erklärung mit einer außergewöhnlich niedrigen Standardabweichung von 0,56 und einer niedrigen Spannweite mit Werten von 1,20 bis 3,60. Hier herrscht innerhalb der Bewertungen der Studierenden hohe Einigkeit. Diese Wie-Erklärung wird in der Skala *Adressatenorientierung* von allen Erklärungen als die beste eingeschätzt; auch das Globalurteil *Die Erklärung war adressatengerecht* ist mit dem Itemwert 2,20 das beste, das

vergeben wurde (vgl. Anhang 10). Das laute Denken, das die Lehrkraft in dieser Vignette zur Erklärung der Widerlegung eines Arguments einsetzt, ist eine Methode, deren Nutzen in der Deutschdidaktik immer wieder betont wird. Es ist davon auszugehen, dass die Studierenden diese Methode aus Seminaren kennen und daher als gut ansehen. Dass die Adressatenorientierung als so gut eingeschätzt wird, zeigt, dass die Passung an die Zielgruppe Neunt-/Zehntklässler/-innen aus Studierendensicht gut gelungen ist. Vergleicht man alle erhobenen Globalurteile für diese Aspekte, so stimmen die Werte mit den Skalenwerten erneut gut überein (vgl. die Werte der Items *Die Erklärung war gut strukturiert, Die Erklärung war adressatengerecht* sowie *Die Lehrkraft hat sich gut verständlich ausgedrückt* in Anhang 10). Damit sind auch für die Studierenden diese beiden Aspekte gut abgebildet.

Die Mittelwerte der *Persönlichkeitswirkung* weisen zwar ähnliche Werte auf wie die vorherigen Aspekte, hier zeigt sich allerdings insgesamt eine größere Standardabweichung und Streuung. Diese wird von den Studierenden also genau wie von den Schülerinnen und Schülern heterogen wahrgenommen.

Der *Sprech-/Körperausdruck* wird dagegen im Vergleich zu den bisher berichteten Aspekten im Mittel besser bewertet, die *Sprachliche Verständlichkeit* etwas schlechter. Bei ersterem fallen insbesondere bei den Was- und Wie-Erklärungen geringere Standardabweichungen und Streuungen auf, bei der Bewertung des *Sprech-/Körperausdrucks* dieser Erklärungen sind sich die Studierenden wie zuvor schon die Schüler/-innen ein wenig einiger als bei den anderen Skalen; erneut ein interessantes Ergebnis. Die Globalurteile zu diesem Aspekt (nämlich *Die Art und Weise, wie die Lehrkraft spricht, gefällt mir*) fallen alle etwas niedriger aus als die Skalenmittelwerte (vgl. Anhang 10).

Die *Erklärtiefe* wird mit zwei Mittelwerten unter 3 (ähnlich wie die *Sprachliche Verständlichkeit*) etwas schlechter bewertet als die anderen Aspekte, die Standardabweichungen und die Spannweiten zeigen aber im Vergleich zu den anderen Bewertungen keine Besonderheiten.

Betrachtet man allerdings jeweils die beiden Vignetten der gleichen Prozedur, so zeigen sich gerade in den Bewertungen der *Erklärtiefe* die deutlichsten Unterschiede im Vergleich zu den anderen Skalen, am stärksten bei den Was-Erklärungen, wo die Mittelwerte um 0,85 Punkte differieren (vgl. Tab. 14). Die Studierenden nehmen also das unterschiedliche Potential zur kognitiven Aktivierung bei den Methoden Geben von Synonymen versus lautes Denken deutlich und theoriekonform wahr. Bei den anderen Aspekten werden die beiden Varianten zwar auch unterschiedlich wahrgenommen, zum Teil sind sich die Bewertungen allerdings recht ähnlich (z. B. *Adressatenorientierung* der Warum-Erklärungen, *Sprech-/Körperausdruck* der Was-Erklärungen oder *Persönlichkeitswirkung* der Wie-Erklärungen, vgl. Tab. 14). Insgesamt unterscheiden die Studierenden die beiden Varianten in ihren Bewertungen etwas deutlicher als die Schüler/-innen.

Auch hier bestätigen sich die vorherigen Tendenzen: Die Studierenden präferieren im Allgemeinen bei der Bewertung der einzelnen Aspekte die vertiefte Was-Erklärung (mit kleineren Ausnahmen allerdings, vgl. Tab. 14) und die vertiefte Wie-Erklärung. Bei den Warum-Erklärungen wird dagegen (ebenfalls mit kleineren Ausnahmen, vgl. Tab.14) die einfache Variante bevorzugt.
Die nächsten Bewertungen, die dargestellt und mit den gerade geschilderten verglichen werden, sind die der Statusgruppe der Lehrkräfte.

Lehrkräfte

Tabelle 15: Mittelwerte, Standardabweichungen und Spannweiten zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Lehrkräfte*
N_{max}=35

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Aspekt						
Strukturiertheit (4 Items)	3,14 (0,94) [1,25; 5,50]	2,14 (0,75) [1,00; 3,50]	2,82 (0,69) [1,00; 4,00]	3,20 (1,12) [1,25; 6,00]	1,91 (0,76) [1,00; 3,50]	1,81 (0,71) [1,00; 3,50]
Adressatenorientierung (5 Items)	3,06 (0,80) [1,60; 4,60]	2,69 (0,76) [1,20; 4,60]	3,35 (0,85) [1,80; 4,80]	3,45 (1,11) [1,00; 5,80]	2,34 (0,65) [1,00; 3,80]	2,13 (0,73) [1,00; 4,00]
Sprech-/ Körperausdruck (7 Items)	1,82 (0,57) [1,00; 3,29]	1,90 (0,69) [1,00; 3,86]	2,98 (0,82) [2,00; 4,86]	3,19 (0,56) [2,00; 4,00]	2,31 (0,81) [1,00; 4,29]	1,76 (0,53) [1,00; 2,86]
Sprachliche Verständlichkeit (4 Items)	2,74 (0,71) [1,00; 4,50]	2,71 (0,83) [1,00; 4,25]	3,56 (0,95) [1,75; 5,75]	3,77 (0,89) [2,25; 5,25]	2,34 (0,68) [1,00; 3,75]	2,27 (0,69) [1,25; 4,00]
Persönlichkeitswirkung (3 Items)	2,37 (0,73) [1,33; 3,33]	2,33 (1,02) [1,00; 5,00]	2,90 (1,05) [1,67; 5,67]	3,33 (0,80) [2,33; 4,67]	2,08 (0,79) [1,00; 3,33]	1,81 (0,53) [1,00; 2,67]
Erklärtiefe (5 Items)	3,28 (1,05) [1,40; 6,00]	2,05 (0,89) [1,00; 5,20]	2,99 (0,90) [1,00; 4,60]	3,30 (1,08) [1,00; 5,20]	2,23 (0,74) [1,00; 3,80]	2,17 (0,83) [1,00; 4,40]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6.
*Angegeben sind die Spannweiten der Mittelwerte der Skala.

Im Vergleich zu den vorherigen Statusgruppen fällt bei den Skalenbewertungen der Lehrkräfte auf, dass diese insgesamt strenger ausfallen: Jeder Aspekt weist mindestens einen, oft zwei oder drei (z. B. bei *Sprachliche Verständlichkeit*, *Erklärtiefe* oder *Adressatenorientierung)* Mittelwert/-e schlechter als 3 auf. Auch bei den Globalurteilen für die einzelnen Aspekte ist diese Tendenz zu sehen (vgl. Anhang 10, z. B. die Werte für die Warum-Erklärungen bei *Die Erklärung war gut strukturiert)*.

Zu erwähnen sind weiterhin Besonderheiten bei beiden Wie-Erklärungen, bei denen sowohl die Standardabweichungen (keine ist höher als 0,83, viele Werte bewegen sich z. T. deutlich unter 0,7) als auch die Spannweiten (besonders auffällig beim *Sprech-/Körperausdruck* der vertieften Wie-Erklärung mit einer Spannweite von 1,00 bis 2,86) über alle Aspekte hinweg niedrig ausfallen (vgl. Tab. 15). Dies deutet darauf hin, dass sich die Lehrkräfte bei beiden Wie-Erklärungen einiger sind als bei den anderen Erklärungen und einiger als die zuvor schon berichteten Statusgruppen der Studierenden und der Schüler/-innen. Diese Einigkeit könnte darauf zurückzuführen sein, dass die Lehrkräfte die Methode „Modelltext" aus ihrem Unterricht kennen und selbst anwenden. Bemerkenswert ist die Einigkeit bei der vertieften Variante „lautes Denken", denn hier ist gerade nicht davon auszugehen, dass die Lehrkräfte diese Methode anwenden, sie ist im Deutschunterricht – mit Ausnahme bei der Einführung von Lesestrategien – wenig etabliert. Weder nannten die Seminarlehrkräfte diese Methode beim Vorgespräch zu dieser Studie noch kommt sie in Unterrichtsvideographien vor (vgl. z.B. das Metavideoportal für Lehrkräftebildung an der Universität Münster oder das Portal UnterrichtOnline.org der LMU München).

Die Bewertungen des Aspektes *Erklärtiefe* stechen bei den Lehrkräften im Vergleich zu denen der anderen Aspekte nicht heraus, die Mittelwerte liegen im mittleren Bereich, die Standardabweichungen bewegen sich um den Wert 1, die Spannweiten sind (mit Ausnahme der Bewertungen die Wie-Erklärungen) relativ hoch (vgl. Tab. 15).

Was die Vergleiche der einfachen mit den vertieften Erklärungen derselben Erklärprozedur anbelangt, so ist festzustellen, dass die Lehrkräfte hier deutlichere Unterschiede zwischen den Varianten machen als die vorherigen Statusgruppen: Wenige Mittelwerte des gleichen Aspekts bei der einfachen versus der vertieften Variante unterscheiden sich um einen geringeren Wert als 0,1. Dafür zeigen sich im Gegensatz dazu zum Teil sehr hohe Unterschiede zwischen den beiden Varianten, zum Beispiel bei der *Strukturiertheit* der Was-Erklärungen, dem *Sprech-/Körperausdruck* der Wie-Erklärungen oder der *Erklärtiefe* der Was-Erklärungen (vgl. Tab. 15). Lehrkräfte nehmen also diese Aspekte in den beiden Varianten der Erklärung als deutlich unterschiedlich wahr beziehungsweise messen dieser Unterschiedlichkeit eine gewichtige Bedeutung bei, trotz der jeweils gleichen Sprecherin der Was- und Wie-Erklärungen beziehungsweise des gleichen Sprechers der Warum-Erklärungen. Hierin unterscheiden sich ihre Bewertungen von denen der Schüler/-innen und der Studierenden. Die differenzierte Bewertung ist vermutlich auf die höhere Expertise im Bewerten von Unterricht der Lehrkräfte zurückzuführen; sie nehmen zum Beispiel die Struktur der Erklärung etymologischer Hintergrund als deutlich positiver wahr als die Struktur beim Geben von Synonymen. Die Struktur Redewendung – Etymologie – Bedeutungsübertragung - aktuelle Bedeutung, die in der Vignette eingehalten wird, nehmen die Lehrkräfte vermutlich als deutlich klarer wahr als die Struktur Redewendung - Synonym 1 - Synonym 2 usw. – aktuelle Bedeutung.

Berichtenswert sind weiterhin die deutlich größeren Unterschiede der Mittelwerte der gleichen Skala über alle Erklärungen (also Videos A-F) hinweg im Vergleich zu den Bewertungen der Schüler/-innen und Studierenden. Unterschiede von fast 1,5 Punkten sind hier die Regel (z. B. wird die *Strukturiertheit* der vertieften Wie-Erklärung mit 1,81 bewertet, die der vertieften Warum-Erklärung mit 3,20; der *Sprech-/Körperausdruck* der vertieften Wie-Erklärung wird mit 1,76, der der vertieften Warum-Erklärung dagegen mit 3,19 bewertet) (vgl. Tab. 15). Auch das bedeutet, dass die Lehrkräfte dieselben Aspekte in den verschiedenen Erklärungen unterschiedlich wahrnehmen und bewerten. Das gleiche Bewertungsverhalten lässt sich in den Globalurteilen der Aspekte feststellen (vgl. die Werte der Lehrkräfte für die Items *Die Erklärung war gut strukturiert, Die Erklärung war adressatengerecht, Die Art und Weise, wie die Lehrkraft spricht, gefällt mir* sowie *Die Lehrkraft hat sich gut verständlich ausgedrückt* in Anhang 10).

Im Folgenden werden diese Bewertungen noch mit den Bewertungen der Statusgruppe der Didaktiker/-innen in Verbindung gesetzt.

Didaktiker/-innen

Tabelle 16: Mittelwerte, Standardabweichungen und Spannweiten zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Didaktiker/-innen*

N_{max}=31

Erklärprozedur	**Was-Erklärungen**		**Warum-Erklärungen**		**Wie-Erklärungen**	
Erklärtiefe	D: einfach	A: vertieft	E: einfach	B: vertieft	F: einfach	C: vertieft
	M (SD) [range]	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*	*M (SD) [range]*
Aspekt						
Strukturiertheit (4 Items)	2,90 (0,98) [1,00; 4,75]	2,00 (0,86) [1,00; 4,00]	2,78 (0,87) [1,00; 3,75]	2,94 (1,02) [1,00; 5,50]	2,26 (1,05) [1,00; 4,75]	1,54 (0,61) [1,00; 3,00]
Adressatenorientierung (5 Items)	3,01 (0,82) [1,40; 4,60]	2,51 (0,77) [1,40; 3,80]	3,53 (1,15) [1,60; 5,20]	3,36 (1,48) [1,20; 6,00]	2,71 (0,96) [1,20; 4,40]	1,93 (0,78) [1,00; 4,60]
Sprech-/ Körperausdruck (7 Items)	1,67 (0,60) [1,00; 3,14]	1,62 (0,49) [1,00; 2,29]	2,76 (0,61) [1,43; 3,71]	3,29 (1,27) [1,14; 5,43]	2,03 (0,50) [1,29; 3,00]	2,08 (0,77) [1,14; 4,00]
Sprachliche Verständlichkeit (4 Items)	2,77 (1,01) [1,00; 4,75]	2,52 (0,64) [1,50; 4,25]	3,48 (0,79) [2,00; 5,00]	3,69 (1,05) [1,75; 5,50]	2,25 (0,63) [1,00; 3,25]	2,13 (0,65) [1,00; 3,75]
Persönlichkeitswirkung (3 Items)	2,27 (0,83) [1,33; 4,00]	2,12 (0,54) [1,00; 3,00]	2,84 (0,89) [1,67; 4,67]	3,45 (0,98) [1,67; 4,67]	2,07 (0,62) [1,33; 3,33]	1,81 (0,75) [1,00; 3,67]
Erklärtiefe (5 Items)	3,51 (0,96) [1,80; 5,40]	2,19 (1,05) [1,00; 4,00]	3,34 (1,12) [1,00; 5,00]	3,21 (1,19) [1,20; 5,40]	2,51 (0,87) [1,00; 4,40]	1,91 (0,92) [1,00; 4,60]

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6.
*Angegeben sind die Spannweiten der Mittelwerte der Skala.

Die Statusgruppe der Didaktiker/-innen bewertet die einzelnen Aspekte im Mittel ähnlich wie die Lehrkräfte. Zu berichten ist an dieser Stelle die relativ schlechte Bewertung der *Adressatenorientierung* mit Werten von zum Teil deutlich unter 3 (eine Ausnahme bildet einzig die *Adressatenorientierung* der vertieften Wie-Erklärung mit dem Wert von 1,93). Diese Bewertungen sind ähnlich denen der Lehrkräfte (vgl. Tab. 15). Allerdings sind gerade bei den niedrigen Bewertungen die Standardabweichungen sehr hoch und auch die Spannweiten fallen sehr groß aus. Die Standardabweichung von 1,48 bei der *Adressatenorientierung* der vertieften Warum-Erklärung ist die höchste Standardabweichung in der Gesamtstudie. Bemerkenswert ist in diesem Zusammenhang, dass die globale Bewertung der *Adressatenorientierung* (mit *Die Erklärung war adressatengerecht*) für alle Vignetten besser ausfällt als die Skalenmittelwerte (vgl. Werte der Didaktiker/-innen für dieses Item in Anhang 10). Dennoch ist festzuhalten, dass die Didaktiker/-innen die Passung aller Vignetten als mäßig gelungen erachten.

Auffällig sind in den Ergebnissen der Didaktiker/-innen außerdem die hohen Standardabweichungen aller Skalen der beiden Warum-Erklärungen, bei diesen beiden sind sich die Didaktiker/-innen also offensichtlich uneins in den Bewertungen (vgl. Tab. 16). Dies ist sicherlich auch der Umsetzung der Vignetten geschuldet.

Niedrige Standardabweichungen und geringe Streuung weisen dagegen erneut die Bewertungen des Aspektes *Sprech-/Körperausdruck* auf: Die Standardabweichung von 0,49 bei der vertieften Was-Erklärung ist sogar die niedrigste überhaupt in der FALKE-D-Studie, genauso wie deren Streuung von 1,00 bis 2,29. Hier sind sich die Didaktiker/-innen offensichtlich einig. Interessant ist bei diesem Aspekt noch, dass sich die Mittelwerte der beiden Warum-Erklärungen deutlich unterscheiden, obwohl es sich um den gleichen Sprecher handelt. Die Didaktiker/-innen nehmen also offensichtlich den *Sprech-/Körperausdruck* situativ unterschiedlich wahr, wenn auch bei der strengen Bewertung von 3,29 bei der vertieften Warum-Erklärung die hohe Standardabweichung mit bedacht werden muss (vgl. Tab. 16). Diese Unterschiedlichkeit entspricht der professionellen Bewertung des *Sprech-/Körperausdrucks*, da Sprechwissenschaftler/-innen von einem situativen *Sprech-/Körperausdruck* und nicht von einem personenbezogenen *Sprech-/Körperausdruck* ausgehen (s. Gunga, 2022, in Vorb.). Den Didaktikerinnen und Didaktikern gelingt es, von der persönlichen Sprechweise zu abstrahieren und wahrzunehmen, dass der Ausdruck je nach Prozedur unterschiedlich ist. Dies ist auch insofern bemerkenswert, als sich die Globalurteile (d. h. die Ergebnisse für *Die Art und Weise, wie die Lehrkraft spricht, gefällt mir*) für die gleiche Sprecherin oder den gleichen Sprecher viel stärker ähneln (vgl. Werte der Didaktiker/-innen für dieses Item in Anhang 10).

Der FALKE-D-spezifische Aspekt der *Erklärtiefe* wird mit Mittelwerten von 1,91 bei der vertieften Wie-Erklärung und 3,51 bei der einfachen Was-Erklärung sehr unterschiedlich bewertet. Auffallend

sind hier erneut die durchgehend hohen Standardabweichungen von 0,87 bis 1,19 sowie relativ große Streuungen (vgl. Tab. 16). Dennoch ist an den Ergebnissen zu sehen, dass die Statusgruppe Didaktiker/-innen theoriekonform das kognitive Aktivierungspotenzial der Methode lautes Denken als sehr viel höher einschätzt als die Methode Modelltext.

Betrachtet man die beiden Varianten der gleichen Erklärprozedur, lässt sich Folgendes berichten: Sowohl bei den Was- als auch bei den Wie-Erklärungen (mit Ausnahme des *Sprech-/Körperausdrucks*, wo der Mittelwertunterschied allerdings gering ist, vgl. Tab. 16) bewerten die Didaktiker/-innen alle Aspekte der vertieften Erklärungen besser; die Ergebnisse entsprechen also den Bewertungen der anderen Statusgruppen. Die Unterschiede sind bei den Didaktikerinnen und Didaktikern zum Teil bemerkenswert hoch, etwa bei der *Erklärtiefe* der Was-Erklärungen oder bei *Adressatenorientierung* der Wie-Erklärungen (vgl. Tab. 16). Auch hier zeigt sich der Expertenstatus deutlich: Didaktiker/-innen nehmen die Erklärtiefe in der Tiefenstruktur der Erklärungen deutlich wahr und bewerten sie entsprechend. Bei den Warum-Erklärungen wird bei den meisten Aspekten die einfache Variante besser bewertet, auch wenn die Unterschiede weniger deutlich ausfallen. Bei der *Erklärtiefe* und bei der *Adressatenorientierung* wird dagegen die vertiefte Variante der Warum-Erklärungen präferiert (vgl. Tab. 16). Mit diesen deutlich anderen Ergebnissen unterscheiden sich die Didaktiker/-innen von den drei anderen Statusgruppen.

Festzuhalten ist insgesamt, dass auch die Didaktiker/-innen mit einzelnen Ausnahmen hohe Standardabweichungen und große Streuungen in ihren Bewertungen aufweisen. Das bedeutet, dass auch diese Statusgruppe die Erklärungen und die einzelnen Aspekte sehr heterogen wahrnimmt und den einzelnen Aspekten eine je nach Thema und Erklärmethode unterschiedliche Bedeutung beimisst.

8.2.3 Zusammenfassung der deskriptiven Ergebnisse

Zusammenfassend ist zu den Forschungsfragen 1a-2b zu sagen, dass sowohl beim Globalurteil als auch bei der Bewertung der spezifischen Aspekte hohe Heterogenität innerhalb der und zwischen den bei FALKE-D befragten Gruppen der Schüler/-innen, Studierenden, Lehrkräfte sowie Didaktiker/-innen herrscht. Einigkeit besteht am ehesten beim Aspekt *Sprech-/Körperausdruck*, interessanterweise in allen Statusgruppen (s. Gunga et al., 2021, in Vorb. sowie Gunga, 2022, in Vorb.).

Bezüglich der FALKE-D-spezifischen Variation in einfache und vertiefte Erklärungen lässt sich feststellen, dass die einzelnen Aspekte in der jeweils einfachen beziehungsweise vertieften Erklärung in allen Statusgruppen unterschiedlich wahrgenommen werden. Hier zeigen sich allerdings Differenzen: Didaktiker/-innen und Lehrkräfte nehmen die Unterschiede zwischen der einfachen und der vertieften Erklärung deutlicher wahr beziehungsweise bewerten sie als

entscheidender als die Schüler/-innen und die Studierenden, denn die Bewertungen derselben Skala bei beiden Varianten sind bei den Lehrkräften und Didaktikerinnen und Didaktikern unterschiedlicher. Der professionelle Metablick auf die Erklärungen zeigt sich in diesen beiden Gruppen deutlich.

Es zeigen sich außerdem Differenzen in Bezug auf die Erklärprozedur. Sowohl in den Was- als auch in den Wie-Erklärungen werden dieselben Aspekte in den beiden Varianten deutlicher unterschieden. Dabei werden theoriekonform die Methoden etymologischer Hintergrund und lautes Denken von den Expertinnen und Experten als kognitiv aktivierender wahrgenommen und bewertet. Besonders auffällige Unterschiede zeigen sich beim FALKE-D-spezifischen Aspekt der *Erklärtiefe* (vgl. Tab. 11 sowie 13-16).

Bei den Warum-Erklärungen wird meist die einfache Variante bevorzugt, aber die Präferenzen sind hier weniger klar. So sind die Unterschiede in den Bewertungen der einzelnen Aspekte oft verschwindend gering. Die Studierenden geben der vertieften Variante in den Skalen *Adressatenorientierung* und *Sprech-/Körperausdruck* bessere Bewertungen. Die Didaktiker/-innen bevorzugen die vertiefte Variante im Globalurteil sowie in den Aspekten *Adressatenorientierung* und *Erklärtiefe*. Die Tatsache, dass diese beiden Varianten als weniger unterschiedlich wahrgenommen werden als die Was- und die Wie-Erklärungen ist sicherlich zum Teil der Umsetzung der Konzepte geschuldet. Genauere Erläuterungen dazu finden sich in Punkt *10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse* in dieser Arbeit.

8.2.4 Bedeutung der variierten Erklärtiefe für die Statusgruppen

Tabelle 17: Varianzanalyse mit gemischtem Design pro Erklärprozedur
Faktoren: Statusgruppe, Erklärtiefe

Erklärprozedur	**Was-Erklärungen**			**Warum-Erklärungen**			**Wie-Erklärungen**		
Erklärtiefe	D: einfach		A: vertieft	E: einfach		B: vertieft	F: einfach		C: vertieft
ANOVA	*Df*	*F*	η^2	*df*	*F*	η^2	*df*	*F*	η^2
Erklärtiefe	1	31,86**	0,12	1	4,91*	0,02	1	2,86	0,01
Statusgruppe	3	1,68	0,02	3	9,41**	0,11	3	13,00**	0,14
Erklärtiefe x Statusgr.	3	2,81*	0,03	3	1,32	0,02	3	1,25	0,02
Residuen	242			238			232		

Bem.: Df: Freiheitsgrade; F: F-Wert; η^2: partielles Eta-Quadrat; *: $p \leq 0,05$; **: $p \leq 0,01$.
(Tab. leicht abgewandelt aus Gaier & Schilcher, 2021, in Vorb.)

Nun wird über die Ergebnisse berichtet, die sich aus der unterschiedlichen Umsetzung der Erklärtiefe in den Vignettenpaaren ergeben (s. Punkt *5.1.4 Variation der Erklärtiefe* in dieser Arbeit), und

dargestellt, ob die umgesetzte Erklärtiefe jeweils einen statistisch nachweisbaren Einfluss auf die Bewertung der Erklärungen hat (Forschungsfrage 2c). Mithilfe der mixed ANOVA wird dazu überprüft, welche dieser Unterschiede statistisch signifikant sind. Zu diesem Zweck wurden für jede Erklärprozedur (also jeweils für die Was-, Warum- und Wie-Erklärungen) Varianzanalysen mit gemischtem Design durchgeführt. Dabei wurde die Erklärtiefe als Messwiederholungsfaktor (einfach vs. vertieft) und die Statusgruppe als vierstufiger Zwischensubjektfaktor (Schüler/-innen, Studierende, Lehrkräfte und Didaktiker/-innen) angenommen und deren Interaktion (Erklärtiefe x Statusgruppe) mit einbezogen[53]. Deskriptiv wurde bereits beschrieben, dass die Erklärtiefe in der Wahrnehmung und Bewertung der Erklärungen einen Unterschied macht, und zwar einen größeren mit steigender Expertise. Mit der mixed ANOVA wird nun überprüft, ob die Erklärtiefe auch statistisch einen bedeutsamen Unterschied für die Bewertung macht.

Die Ergebnisse werden getrennt nach der jeweiligen Prozedur dargestellt, begonnen wird mit *erklären, was*[54]*.* Bei dieser Prozedur ist die Erklärtiefe für die Beurteilung statistisch signifikant; das heißt, dass die besseren Bewertungen der vertieften Was-Erklärung auf die umgesetzte Erklärtiefe zurückgehen. Die oben erläuterten deskriptiven Ergebnisse sind somit bestätigt. Während für die Bewertungen zwischen den einzelnen Statusgruppen kein signifikanter Haupteffekt vorliegt, wird bei *erklären, was* der Interaktionseffekt von Erklärtiefe und Statusgruppe signifikant. Dies ist damit zu erklären, dass die Schüler/-innen die einfache und die vertiefte Version der Was-Erklärung recht ähnlich bewerten. Somit nehmen sie wie bereits gesagt den Unterschied in der Erklärtiefe weniger wahr und/oder sie erachten ihn für weniger wichtig. Dies könnte daran liegen, dass die Schüler/-innen in diesen Erklärungen vor allem an der Bedeutung der Redewendung interessiert sind und den Mehrwert, der durch die Etymologie entsteht, für weniger bedeutsam halten. Dahingegen könnte die größere Erklärtiefe, die durch das Erläutern des etymologischen Hintergrundes entsteht, bei den Expertinnen und Experten zu einer besseren Bewertung führen. Sie empfinden vermutlich und theoriekonform diese zusätzliche Information als aktivierend oder erleben die erklärende Lehrkraft als fachlich kompetenter (vgl. Tab. 17 sowie Gaier & Schilcher, 2021, in Vorb.).

Bei der Prozedur *erklären, warum* sind folgende Ergebnisse zu berichten[55]: Für *erklären, warum* wird die einfache Variante bevorzugt (vgl. Punkt *8.2.3 Zusammenfassung der deskriptiven Ergebnisse* in dieser Arbeit), dabei ist der Unterschied zwischen den Bewertungen der einfachen

[53] Die Sphärizität als Voraussetzung für die mixed ANOVA ist gegeben, da der Innersubjektfaktor Erklärtiefe nur zwei Stufen hat.

[54] Die Voraussetzung der Homogenität der Fehlervarianzen zwischen den Gruppen ist gemäß dem Levene-Test für die Variablen der *Erklärtiefe* erfüllt ($p > .05$); die Homogenität der Kovarianzenmatrizen ist gemäß dem Box-Test gegeben ($p = .501$).

[55] Auch hier ist die Voraussetzung der Homogenität der Fehlervarianzen zwischen den Gruppen gemäß Levene-Test für die Variablen der Erklärtiefe erfüllt ($p > .05$); die Homogenität der Kovarianzenmatrizen ist gemäß dem Box-Test gegeben ($p = .851$).

und der vertieften Erklärung wie bei den Was-Erklärungen signifikant. Die deskriptiven Ergebnisse bestätigen sich hier also statistisch. Außerdem zeigen sich signifikante Unterschiede zwischen den Statusgruppen, die also die Warum-Erklärungen nicht nur deskriptiv, sondern auch statistisch signifikant unterschiedlich bewerten (vgl. Tab. 17). Bei der Überprüfung dieses Haupteffektes durch den Post-Hoc-Test Games-Howell ergaben sich signifikante Unterschiede bei der Statusgruppe Lehrkräfte im Vergleich zu den Studierenden und im Vergleich zu den Schülerinnen und Schülern (in beiden Vergleichen jeweils $p < .001$). Das bedeutet, dass bei den Warum-Erklärungen die Lehrkräfte deutlich anders urteilen als die beiden gerade genannten Statusgruppen.

Schließlich zeigt sich Folgendes für die Prozedur *erklären, wie*[56]: Im Gegensatz zu den vorherigen beiden Prozeduren ist der Unterschied zwischen der Bewertung der einfachen und der besseren Bewertung der vertieften Wie-Erklärung nicht signifikant (auch deskriptiv fällt er geringer aus als bei den anderen Erklärprozeduren, vgl. Tab. 9 zu den Globalurteilen über alle Statusgruppen hinweg). Dagegen werden die Unterschiede der Bewertungen zwischen Statusgruppen genau wie bei *erklären, warum* signifikant und spiegeln somit die deskriptiven Ergebnisse. Gerade die Schüler/-innen und die Didaktiker/-innen bewerten ja die beiden Wie-Erklärungen sehr unterschiedlich (vgl. Tab. 10). Bei der Überprüfung dieses Haupteffektes durch den Post-Hoc-Test Games-Howell ergaben sich signifikante Unterschiede bei der Statusgruppe Schüler/-innen im Vergleich zu jeder der anderen Statusgruppen (dabei bewegen sich die Signifikanzniveaus zwischen den Schüler/-innen und den Studierenden sowie zwischen den Schüler/-innen und den Didaktiker/-innen auf $p < .001$, das Signifikanzniveau zwischen den Schüler/-innen und den Lehrkräften auf $p < .005$).

Insgesamt bestätigten sich somit die deskriptiven Unterschiede in der Varianzanalyse mit den Faktoren Statusgruppe und Erklärtiefe. Die variierte Erklärtiefe macht bei zwei der drei Prozeduren einen signifikanten Unterschied für die Bewertung der Globalurteile aus. Das Beurteilungsverhalten der Schüler/-innen ist bezogen auf die unterschiedlich umgesetzte Erklärtiefe statistisch insgesamt am deutlichsten von den Beurteilungen aller anderen Statusgruppen zu unterscheiden.

8.3 Inferenzstatistische Ergebnisse

Im nächsten Schritt werden die inferenzstatistischen Ergebnisse präsentiert.

Dies dient der Beantwortung der Forschungsfragen 3, 4 und 5, welche Kriterien bei den Bewertungen für das Globalurteil den Ausschlag geben und wo sich hier Gemeinsamkeiten und Unterschiede bei den Statusgruppen zeigen. Die deskriptiven Ergebnisse haben bereits gezeigt, dass

[56] Wie bei den anderen beiden Prozeduren ist die Voraussetzung der Homogenität der Fehlervarianzen zwischen den Gruppen gemäß Levene-Test für die Variablen der Erklärtiefe erfüllt ($p > .05$); dagegen ist die Homogenität der Kovarianzenmatrizen gemäß dem Box-Test bei dieser Prozedur verletzt ($p = .047$*). Da die Homogenitätsverletzung nur sehr gering ausfällt, werden die Ergebnisse der ANOVA dennoch auch bei dieser Prozedur berichtet und interpretiert.

die in FALKE-D untersuchten Aspekte bei der Bewertung von Erklärungen eine Rolle spielen. Mit den inferenzstatistischen Berechnungen wird nun dargelegt, welche der Aspekte einen statistisch nachweisbaren Einfluss auf die Globalbewertung der Erklärungen haben.
Da die erhobenen Aspekte im Hinblick auf ihre Bedeutung für die Gesamtbewertung der Erklärqualität (= das Globalurteil) untersucht werden, werden sie als Prädiktoren für die Qualität einer Erklärung verstanden und im Folgenden auch so benannt.

8.3.1 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil

Zunächst werden wie üblich die Interskalenkorrelationen des Globalurteils und der einzelnen abgefragten Aspekte dargestellt. Dies geschieht getrennt nach zwei Statusgruppen, den Schülerinnen und Schülern auf der einen Seite und den Expertinnen und Experten (d. h. den Studierenden, Lehrkräften und Didaktikerinnen und Didaktikern[57]) auf der anderen Seite. Eine Übersicht über die Interskalenkorrelationen bietet Tabelle 18. Zur Erinnerung: Bei der Kontrollvariable *Persönlichkeitswirkung* wird wie allgemein anerkannt davon ausgegangen, dass sie die Bewertung der Erklärqualität beeinflusst. Deswegen ist sie in die Interskalenkorrelation mit aufgenommen.
In Tabelle 18 sind die Zusammenhänge zwischen den Aspekten untereinander und zwischen den Aspekten und dem Globalurteil abgebildet.
Wie man erkennen kann, korrelieren alle Prädiktoren sowohl untereinander als auch mit dem Globalurteil stark positiv miteinander. Alle Korrelationen sind signifikant.

[57] Die Studierenden werden in dieser Berechnung zur Expertengruppe zugeordnet, obwohl ihr Alter und ihre fehlende Expertise im Unterrichten wie dargelegt für eine Zuordnung zur Adressatengruppe der Schüler/-innen sprechen könnten. Allerdings befinden sie sich, wie bereits erläutert, auf dem Weg zum Expertenstatus und bewerten die gesehenen Vignetten nicht aus der Perspektive von Adressatinnen und Adressaten. Zudem sprechen auch die Unterschiede in ihren Bewertungen im Vergleich zu denen der Schüler/-innen dafür, die Studierenden zu der Expertengruppe zu zählen. Auf die Darstellung der Interskalenkorrelationen und deren Zusammenhang mit dem Globalurteil aufgeschlüsselt nach Statusgruppen wird der Übersichtlichkeit halber verzichtet, da die Ergebnisse keine neuen Erkenntnisse bringen.

Tabelle 18: Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos
Pearson Produkt-Moment-Korrelation, nach SuS- und SLD-Gruppe getrennt

Aspekt	**Struktu-riertheit**	**Adressa-tenorien-tierung**	**Sprech-/ Körper-ausdruck**	**Sprachliche Verständ-lichkeit**	**Persönlich-keitswirkung**	**Erklärtiefe**	**Global-urteil**
Strukturiertheit		0,58**	0,42**	0,62**	0,40**	0,67**	0,35**
Adressaten-orientierung	0,72**		0,35**	0,57**	0,41**	0,62**	0,34**
Sprech-/ Körper-ausdruck	0,43**	0,51**		0,39**	0,71**	0,52**	0,49**
Spr. Verständ-lichkeit	0,61**	0,69**	0,58**		0,34**	0,58**	0,28**
Persönlichkeits-wirkung	0,50**	0,48**	0,74**	0,52**		0,46**	0,41**
Erklärtiefe	0,77**	0,75**	0,53**	0,59**	0,52**		0,40**
Globalurteil	0,57**	0,55**	0,43**	0,42**	0,47**	0,60**	

Bem.: Korrelationswerte oberhalb der Diagonale für Schüler/-innen (=SuS), unterhalb der Diagonale für Expertinnen/Experten (d. h. Studierende, Lehrkräfte und Didaktiker/-innen (=SLD)); *: $p \leq 0{,}05$; **: $p \leq 0{,}01$. Tabelle leicht abgewandelt (aufgrund der anderen Skala bei *Sprachliche Verständlichkeit*, s. Erläuterungen unter *4.2.3 Sprachliche Verständlichkeit* in dieser Arbeit) aus Gaier & Schilcher, 2021, in Vorb.

Es fällt auf, dass die Werte bei der Adressatengruppe, den Schülerinnen und Schülern, geringer ausfallen als bei der Expertengruppe Studierende, Lehrkräfte und Didaktiker/-innen.

Die deskriptiv höchste Korrelation in Tabelle 18 liegt bei den Schülerinnen und Schülern zwischen dem *Sprech-/Körperausdruck* und der *Persönlichkeitswirkung* vor. Dieses Ergebnis ist nach den Erkenntnissen der Teacher-Enthusiasm-Forschung nicht überraschend. Eine ähnlich hohe Korrelation weist bei den Schülerinnen und Schülern die *Strukturiertheit* mit der *Erklärtiefe* auf.

Die hohe Korrelation dieser beiden Prädiktoren (*Strukturiertheit* und *Erklärtiefe*) findet sich auch bei der Expertengruppe, wo sie sogar die deskriptiv höchste Korrelation darstellt (vgl. Tab. 18). Abgesehen davon fällt auf, dass drei weitere Korrelationen in der Expertengruppe ähnlich hoch ausfallen, nämlich *Adressatenorientierung* mit *Erklärtiefe, Adressatenorientierung* mit *Strukturiertheit* sowie (analog zu den Ergebnissen der Schüler/-innen) *Sprech-/Körperausdruck* mit *Persönlichkeitswirkung* (vgl. Tab. 18).

Die Tatsache, dass alle Aspekte sowohl untereinander als auch mit dem Globalurteil substanziell und statistisch signifikant korrelieren, ist nicht überraschend, da alle Aspekte Teil der Erklärqualität sind, die mit dem Globalurteil bewertet wird.

8.3.2 Zusammenhänge zwischen dem Globalurteil und den Aspekten guten Erklärens

Ein Forschungsinteresse von FALKE-D liegt darin, herauszufinden, welche der abgefragten Aspekte einen entscheidenden Einfluss auf die Bewertung einer Erklärung haben (s. Forschungsfragen 3 und 4). Dabei ist die FALKE-D-spezifische Variation der Erklärungen in einfache versus vertiefte Varianten von besonderer Bedeutung (s. Forschungsfrage 5).

Um dieser Frage nachzugehen, werden hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und den fünf Prädiktoren (*Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit, Sprech-/Körperausdruck* sowie *Erklärtiefe*) an die nach Personen und Videos strukturierten Daten angelegt (vgl. Gaier & Schilcher, 2021, in Vorb.). Die Ergebnisse für den Aspekt der *Persönlichkeitswirkung* werden außerdem als Kontrollvariable ebenfalls in die Regressionen aufgenommen. Da die Bewertungen aller Prädiktoren wie dargelegt auf einer sechsstufigen Skala basieren und das Globalurteil mit (ebenfalls sechsstufigen) Schulnoten evaluiert wurde, sind die unstandardisierten Regressionskoeffizienten in ihrer originalen Metrik interpretierbar[58] und in den Tabellen 19-21 sowie 22-25 abgebildet.

Analog zu den deskriptiven Ergebnissen werden auch diese inferenzstatistischen Verfahren statusgruppenübergreifend sowie statusgruppenspezifisch dargestellt und interpretiert.

8.3.2.1 Statusgruppenübergreifende Ergebnisse des hierarchischen linearen Regressionsmodells

Dargestellt werden als erstes die Ergebnisse für die Gesamtstichprobe (vgl. Tab. 19).

[58] Das bedeutet, dass keine „Anpassungen" der Koeffizienten nötig sind, da die Skalierung der Bewertungen der abhängigen Variable (= Globalurteil) und der Prädiktoren (= Aspekte des guten Erklärens) übereinstimmen. Der Note 2 auf eine Erklärung entspricht somit die mit 2 umgerechnete Antwort „stimme zu" auf ein Item.

Tabelle 19: Hierarchisches lineares Regressionsmodell mit der abhängigen Variable Globalurteil, Gesamtstichprobe
(unter Berücksichtigung der nach Studienteilnehmer/-innen (N_{max} = 257) und den sechs Videos geordneten Datenstruktur)

		Gesamtstichprobe ($N_{Beob.}$ = 603)
ICC	%	8,4
Konstante	*β*	0,22
Strukturiertheit	*β*	0,22**
Adressatenorientierung	*β*	0,18**
Sprech-/Körperausdruck	*β*	0,22**
Sprachliche Verständlichkeit	*β*	-0,08
Persönlichkeitswirkung	*β*	0,10
Erklärtiefe	*β*	0,22**
R^2 (marginal | konditional)		0,38 | 0,44

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N* x 6 Videos); *ICC*: Intraklassenkorrelation; *β*: unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.

Wie Tabelle 19 zu entnehmen ist, erweisen sich über alle Statusgruppen hinweg sämtliche Prädiktoren mit Ausnahme der *Sprachlichen Verständlichkeit* als hoch signifikant[59]. Das Kontrollmerkmal *Persönlichkeitswirkung* hat kein signifikantes Regressionsgewicht.

Interessant ist an diesen Ergebnissen, dass die positiven Regressionsgewichte aller prädiktiven Aspekte (also *Strukturiertheit, Sprech-/Körperausdruck, Adressatenorientierung* und *Erklärtiefe*) ähnlich hoch ausfallen (und zwar von 0,18-0,22, vgl. Tab. 19). Das heißt, dass die abgefragten Aspekte (mit Ausnahme der *Sprachlichen Verständlichkeit*) über die vier Statusgruppen hinweg alle eine ähnlich wichtige Rolle für die globalen Bewertungen einer Erklärung besitzen, *Strukturiertheit, Sprech-/Körperausdruck* sowie *Erklärtiefe* rechnerisch sogar die exakt gleiche.

Ganz konkret bedeutet das Regressionsgewicht von 0,22, dass das Globalurteil einer Erklärung um etwas mehr als eine Fünftelnote besser beurteilt wird, wenn die mittlere Bewertung der

[59] Das negative Regressionsgewicht der *Sprachlichen Verständlichkeit* ist zunächst überraschend, liegt jedoch daran, dass hier ein breites Konstrukt über sehr wenige Items (nämlich 4, s. Punkt *4.2.3 Sprachliche Verständlichkeit* in dieser Arbeit) abgedeckt wurde. Die Ergebnisse zur *Sprachlichen Verständlichkeit* werden hier nicht interpretiert, zur ausführlichen Analyse der *Sprachlichen Verständlichkeit* s. Thim-Mabrey & Lindl (2021, in Vorb.).

Strukturiertheit, des *Sprech-/Körperausdrucks* oder der *Erklärtiefe* um eine Stufe ansteigt (vgl. Tab. 19).

Aufschlussreich ist in diesem Zusammenhang das Ergebnis für das deutschspezifische Merkmal der *Erklärtiefe*: Es hat in FALKE-D den gleichen oder mehr Einfluss auf die Erklärqualität wie die aus der Lehr-/Lernforschung bereits bekannten Merkmale. Somit bestätigt FALKE-D einerseits die Bedeutung der besagten Merkmale (i. e. *Strukturiertheit, Adressatenorientierung* und *Sprech- und Körperausdruck*), weist aber andererseits der fachdidaktischen Gestaltung einen wichtigen Beitrag zur Erklärqualität nach (Genaueres hierzu s. Punkte *10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse* und *10.2 Verallgemeinerbarkeit der Ergebnisse aus FALKE-D* in dieser Arbeit).

Die Varianzaufklärung der Erklärqualität, die mit diesem Modell erzielt wird, ist mit 38 Prozent für die festen und 44 Prozent für die festen und variablen Effekte insgesamt beachtlich. Das heißt, dass die in FALKE-D gewählten Aspekte insgesamt einen hohen Anteil der Erklärqualität vorhersagen können, die kriteriale Validität der Aspekte ist somit unterstrichen. *Strukturiertheit, Adressatenorientierung, Sprech- und Körperausdruck* sowie *Erklärtiefe* sind demnach wichtige Aspekte bei der Wahrnehmung und Bewertung von Erklärqualität[60].

Nachdem die Bedeutung der Prädiktoren über alle Statusgruppen beschrieben wurde, ist gemäß Forschungsfrage 4 von Interesse, welche Gemeinsamkeiten und Unterschiede sich in der Bedeutung für die einzelnen Statusgruppen feststellen lassen.

8.3.2.2 Statusgruppenspezifische Ergebnisse der hierarchischen linearen Regressionsmodelle

Betrachtet wird daher zunächst der Vergleich zwischen den Schülerinnen und Schülern und der Gesamtexpertengruppe (d. h. analog zur Vorgehensweise bei der Interskalenkorrelation die Gruppen Studierende, Lehrkräfte und Didaktiker/-innen zusammengenommen zu SLD, zur Begründung s. *8.3.1 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil* in dieser Arbeit).

[60] Die Intraklassenkorrelation liegt bei 8,4 Prozent. Das bedeutet, dass für die Gesamtstichprobe etwa acht Prozent der Merkmalsvarianz auf die Teilnehmer/-innen und die Videos zurückzuführen ist.

Tabelle 20: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, Schüler/-innen und SLD-Gruppe
(unter Berücksichtigung der nach Studienteilnehmer/-innen (N_{max} = 257) und den sechs Videos geordneten Datenstruktur)

		Schüler/-innen ($N_{Beob.}$ = 294)	**SLD-Gruppe** ($N_{Beob.}$ = 309)
ICC	%	4,03	31,09
Konstante	*β*	0,48*	0,07
Strukturiertheit	*β*	0,08	0,30**
Adressatenorientierung	*β*	0,11	0,25**
Sprech-/Körperausdruck	*β*	0,39**	0,07
Sprachliche Verständlichkeit	*β*	-0,02	-0,08
Persönlichkeitswirkung	*β*	0,05	0,12
Erklärtiefe	*β*	0,17*	0,24**
R^2 (marginal | konditional)		0,29 | 0,29	0,47 | 0,59

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N* x 6 Videos); *ICC*: Intraklassenkorrelation; *β:* unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.
(Tabelle leicht abgewandelt aus Gaier & Schilcher, 2021, in Vorb.)

Schüler/-innen

Wie in Tabelle 20 zu erkennen ist, erweisen sich bei den Schülerinnen und Schülern zwei der Prädiktoren als signifikant: Den entscheidendsten Einfluss auf die Globalnote einer Erklärung hat bei ihnen der *Sprech- und Körperausdruck*, der mit einem Regressionsgewicht von 0,39 hoch signifikant ist. Dies ist ein weiterer Hinweis auf die oben genannten Vermutungen aus den deskriptiven Analysen: Die erklärende Person, die sich in der Performanz der Erklärung deutlicher zeigt als zum Beispiel in der Struktur einer Erklärung, hat bei den Schülerinnen und Schülern offensichtlich einen großen Einfluss auf die Bewertung der Erklärung[61]. Bemerkenswert ist jedoch, dass für die *Persönlichkeitswirkung*, die die Person selbst in den Blick nimmt, kein vergleichbarer Einfluss festgestellt werden kann (vgl. Tab. 20).

Abgesehen vom *Sprech- und Körperausdruck* lässt sich in der Statusgruppe der Schüler/-innen bei dem deutschspezifischen Aspekt der *Erklärtiefe* ein entscheidender Einfluss auf die

[61] Diese Vermutung wird durch empirische Ergebnisse gestützt, in denen sich eine gewisse Abhängigkeit der Beurteilungen von der Person der Lehrkraft zeigt: „There are strong similiarities between responses to different courses offered by the same teacher, but litte to no similarities between responses to the same courses offered by different teachers, suggesting that SPIQ [students' perceptions of instructional quality] depend more on a teacher's behavior in a given course than on particular subjects […]" (Wisniewski et al., 2020, vorab online).

Gesamtbewertung einer Erklärung nachweisen, mit einem positiven Regressionsgewicht von 0,17 ist dieser zwar nicht ganz so stark (d. h. $p \leq 0{,}05$), aussagekräftig ist der Aspekt dennoch, gerade vor dem Hintergrund, dass alle anderen Prädiktoren keinen statistisch signifikanten Einfluss auf die Bewertungen der Schüler/-innen nehmen.
Für die Schüler/-innen sind die anderen Aspekte der Erklärung in FALKE-D interessanterweise nicht so entscheidend für die Gesamtqualität.
Die Varianzaufklärung der Erklärqualität, die mit diesem Modell erzielt wird, ist mit 29 Prozent für die festen und 29 Prozent für die festen und variablen Effekte insgesamt relativ hoch. Im Vergleich zur Gesamtstichprobe (s. Punkt *8.3.2.1 Statusgruppenübergreifende Ergebnisse des hierarchischen linearen Regressionsmodells* in dieser Arbeit) sowie zu den anderen Statusgruppen (s. folgende Ausführungen) erweist sich die aufgeklärte Gesamtvarianz der Erklärqualität bei den Schülerinnen und Schülern jedoch als (zum Teil deutlich) niedriger. Das heißt, dass die Schülerinnen und Schüler zusätzlich zu den abgefragten Merkmalen Kriterien zur Bewertung einer Erklärung heranziehen, die im Modell von FALKE-D nicht berücksichtigt wurden, und zwar mehr als alle anderen Statusgruppen[62].

SLD-Gruppe

Für die Expertengruppe insgesamt (d. h. Studierende, Lehrkräfte und Didaktiker/-innen zusammen) ergibt sich ein anderes Bild (s. Tab. 20): Einen hoch signifikanten Einfluss auf die globale Bewertung einer Erklärung haben hier die Aspekte *Strukturiertheit, Adressatenorientierung* sowie *Erklärtiefe.* Dabei weist die *Strukturiertheit* das höchste Regressionsgewicht auf, die *Adressatenorientierung* und die *Erklärtiefe* haben nahezu das gleiche Gewicht.
Das Ergebnis ist aus zwei Gründen interessant: Zum einen ist es bemerkenswert, dass die Aspekte, die sich auf die Performanz der Erklärung beziehen, nämlich *Sprech-/Körperausdruck* und *Persönlichkeitswirkung* keinen (statistisch) bedeutenden Einfluss auf die Bewertung der Erklärungen nehmen. Offensichtlich abstrahieren die Expertinnen und Experten also von der erklärenden Person und fokussieren stärker den Inhalt und den Aufbau einer Erklärung. Bemerkenswert ist zum anderen, dass sich im Vergleich zu den Regressionen für die Schüler/-innen ein nahezu komplementäres Bild ergibt: Während für die Schüler/-innen lediglich der *Sprech-/Körperausdruck* hoch signifikant wird, wird er das für die Expertengruppe gar nicht. Dagegen sind hier die zwei Aspekte *Strukturiertheit* und *Adressatenorientierung* hoch signifikant, die bei den Schülerinnen und Schülern keine entscheidende Rolle spielen. „Einigkeit" herrscht lediglich bei der *Sprachlichen*

[62] Die Intraklassenkorrelation (s. Erläuterungen oben) für diese Statusgruppe beträgt 4,03 Prozent.

Verständlichkeit sowie dem Kontrollmerkmal der *Persönlichkeitswirkung*, die in keiner der Gruppen signifikant werden.

Die *Erklärtiefe* hat dagegen in beiden Gruppen einen signifikanten Einfluss, wenn auch einen stärkeren in der Expertengruppe, hier ist er auf dem gleichen Niveau wie in der Gesamtstichprobe (s. Tab. 19 und 20).

Die Varianzaufklärung der Erklärqualität, die mit diesem Modell erzielt wird, ist mit 47 Prozent für die festen und 59 Prozent für die festen und variablen Effekte insgesamt beachtlich. Das heißt, dass das Modell des guten Erklärens von FALKE-D für die Expertengruppen einen hohen Anteil an der Varianz der Erklärqualität abdeckt; das Modell „trifft" also gut die Kriterien, die die Expertinnen und Experten an die Bewertung der Erklärqualität anlegen[63].

Da die Ergebnisse so deutlich von den Schülerinnen und Schülern abweichen, lohnt der Blick in die einzelnen Expertenstatusgruppen. Abgebildet und erläutert werden im Folgenden die Regressionsgewichte der Prädiktoren auf das Globalurteil für die Statusgruppen der Studierenden, der Lehrkräfte und der Didaktiker/-innen getrennt. Dies erlaubt einen differenzierteren Blick und eine genauere Beantwortung der Frage, welche der Aspekte für die einzelnen Statusgruppen entscheidend sind (s. Forschungsfrage 4). Einen Überblick über die Ergebnisse für alle Statusgruppen gibt Tabelle 21, in der der Übersichtlichkeit halber die Daten für die Schüler/-innen aus Tabelle 20 noch einmal abgebildet sind. Im Fließtext erläutert werden nicht alle Ergebnisse, sondern nur Auffälligkeiten.

[63] Die Intraklassenkorrelation liegt für die SLD-Gruppe bei 31,09 Prozent.

Tabelle 21: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, Schüler/-innen, Studierende, Lehrkräfte und Didaktiker/-innen
(unter Berücksichtigung der nach Studienteilnehmer/-innen (N_{max} = 257) und den sechs Videos geordneten Datenstruktur)

		Schüler/-innen ($N_{Beob.}$ = 294)	**Studierende** ($N_{Beob.}$ = 152)	**Lehrkräfte** ($N_{Beob.}$ = 93)	**Didakt./-innen** ($N_{Beob.}$ = 64)
ICC	%	4,03	13,19	40,86	34,16
Konstante	*β*	0,48*	-0,09	0,40	-0,32
Strukturiertheit	*β*	0,08	0,41**	0,04	0,27*
Adressatenorientierung	*β*	0,11	0,37**	0,16	0,18
Sprech-/Körperausdruck	*β*	0,39**	0,21	-0,11	-0,02
Sprachliche Verständlichkeit	*β*	-0,02	-0,30**	0,12	-0,10
Persönlichkeitswirkung	*β*	0,05	0,23*	-0,01	0,37*
Erklärtiefe	*β*	0,17*	0,01	0,62**	0,38**
R^2 (marginal | konditional)		0,29 | 0,28	0,43 | 0,55	0,53 | 0,58	0,69 | 0,86

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N x* 6 Videos); *ICC*: Intraklassenkorrelation; *β:* unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.

Studierende, Lehrkräfte, Didaktiker/-innen einzeln und im Vergleich

Während sich bei dem Vergleich Schüler/-innen versus SLD-Gruppe eine komplementäre Verteilung der signifikanten Prädiktoren ergibt (s. Tab. 20 und Ausführungen oben), differenziert sich die Bedeutung der Prädiktoren für die einzelnen Statusgruppen wie folgt aus.

Für die Studierenden sind auf der einen Seite vier Aspekte (nämlich *Strukturiertheit, Adressatenorientierung, Sprachliche Verständlichkeit*[64] und *Persönlichkeitswirkung*) signifikant, wenn auch in unterschiedlichem Ausmaß (s. Tab. 21). Damit sind sie die Statusgruppe, die die meisten der abgefragten Prädiktoren als bedeutend wahrnimmt. Auf der anderen Seite sind die Studierenden aber auch die Statusgruppe, die als einzige die *Erklärtiefe* nicht als entscheidend bewertet und somit dem deutschspezifischen Aspekt keine so wichtige Rolle für die Qualität einer Erklärung zuspricht wie den aus der Lehr-/Lernforschung bekannten Merkmalen. Erwähnenswert ist weiterhin, dass die Kontrollvariable *Persönlichkeitswirkung* signifikant wird, der *Sprech-/Körperausdruck* aber nicht, obwohl sich beides auf die Performanz der Erklärung bezieht.

[64] Zu den negativen Regressionsgewichten der *Sprachlichen Verständlichkeit* siehe Erläuterungen unter Punkt *8.3.2.1 Statusgruppenübergreifende Ergebnisse des hierarchischen linearen Regressionsmodells* in dieser Arbeit bzw. Thim-Mabrey & Lindl (2021, in Vorb.).

Die aufgeklärte Gesamtvarianz beträgt bei den Studierenden 43 Prozent für die festen sowie 55 Prozent für die festen und variablen Effekte und ist somit erneut bemerkenswert hoch[65].

Bevor die Ergebnisse für die letzten beiden Statusgruppen, die Lehrkräfte und die Didaktiker/-innen, vorgestellt werden, ist erneut auf die geringe Stichprobengröße in diesen beiden Gruppen hinzuweisen, die sich durch fehlende Angaben noch einmal verringert. Die folgenden Ergebnisse sind also als Tendenzen zu sehen genau wie die Ausführungen oben zu den deskriptiven Resultaten.

Die Bewertungen der Lehrkräfte zeigen folgende Auffälligkeiten in den Regressionen (vgl. Tab. 21): Sie sind die einzige Statusgruppe, für die nur ein einziger Aspekt statistisch hoch signifikant das Globalurteil beeinflusst, und zwar interessanterweise der deutschspezifische Aspekt der *Erklärtiefe.* Alle anderen Aspekte haben keinen statistisch signifikanten Einfluss auf die Bewertungen der Lehrkräfte. Kurios an ihren Bewertungen ist das (wenn auch leichte und statistisch nicht signifikante) negative Regressionsgewicht des Aspektes *Sprech-/Körperausdruck*. Dies würde bedeuten, dass ein unangemessener *Sprech-/Körperausdruck* zu einer besseren Bewertung einer Erklärung führt. Aufgrund der sehr schwachen Ausprägung dieses Ergebnisses sowie der kleinen Stichprobenzahl ist das negative Gewicht vermutlich Zufall und sagt nur aus, dass dieser Aspekt keine Rolle für die Bewertungen der Lehrkräfte spielt.

Die aufgeklärte Gesamtvarianz ist für die Lehrkräfte mit 53 Prozent für die festen und mit 58 Prozent für die festen und variablen Effekte bemerkenswert und sogar noch etwas höher als bei den Studierenden[66].

Das Ergebnis für den Aspekt des *Sprech-/Körperausdrucks* bei den Lehrkräften ist umso interessanter, als es bei diesem Aspekt auch bei den Didaktikerinnen und Didaktikern zu einem negativen (aber ebenfalls leichten und statistisch nicht signifikanten) Regressionsgewicht kommt. Auch hier gilt gerade genannte Zufallsvermutung.

Abgesehen davon sind bei den Didaktikerinnen und Didaktikern die Aspekte *Strukturiertheit, Persönlichkeitswirkung* und *Erklärtiefe* statistisch signifikant für die Globalurteile der Erklärungen. Überraschend ist der signifikante Einfluss der Kontrollvariable *Persönlichkeitswirkung* ausgerechnet bei der Statusgruppe mit der höchsten Expertise. Obwohl die Didaktiker/-innen wie in den deskriptiven Ergebnissen ersichtlich die Vignetten am differenziertesten wahrnehmen, sprechen sie anscheinend der *Persönlichkeitswirkung* der/des Erklärenden dennoch einen wichtigen Einfluss auf die Qualität einer Erklärung zu. Die *Erklärtiefe* hat den wichtigsten Einfluss mit einem Signifikanzniveau von $p \leq 0{,}01$.

[65] Die Intraklassenkorrelation liegt bei 13,19 Prozent.
[66] Die Intraklassenkorrelation liegt bei 40,86 Prozent.

Die aufgeklärte Gesamtvarianz beträgt bei den Didaktikerinnen und Didaktikern 69 Prozent für die festen sowie 86 Prozent für die festen und variablen Effekte und ist somit außerordentlich hoch, deutlich höher als bei allen anderen Statusgruppen[67].

Die *Erklärtiefe* ist damit der einzige Aspekt, der in drei der vier Statusgruppen ein signifikantes Regressionsgewicht für die Globalnote erhält. Bei diesem Ergebnis unterscheiden sich die Studierenden deutlich von den anderen beiden Expertengruppen sowie von den Schülerinnen und Schülern. Zur Bedeutung der *Erklärtiefe* siehe auch den Punkt *10 Diskussion der Ergebnisse zum Konstrukt des Guten Erklärens* in dieser Arbeit.

Vergleicht man abschließend noch die aufgeklärte Gesamtvarianz, so fällt auf, dass sie mit zunehmender Expertise ansteigt. Für die Schüler/-innen ergeben sich hier wie gesagt die niedrigsten Werte, für die Didaktiker/-innen mit 69 Prozent für die festen und 86 Prozent für die festen und variablen Effekte die höchsten und tatsächlich außergewöhnlich hohen Werte (s. Tab. 21). Da bedeutet, dass das Modell des guten Erklärens in FALKE-D in dieser Statusgruppe trotz der wenigen signifikanten Prädiktoren am besten abdeckt, was die Studienteilnehmer/-innen unter gutem Erklären verstehen. Das Modell trifft somit die Expertenvorstellungen von gutem Erklären deutlich besser als die Vorstellung der Adressatinnen und Adressaten (s. dazu auch den Punkt *10.1 Zusammenfassung der quantitativen und qualitativen Ergebnisse* in dieser Arbeit).

Da die in den Erklärungen umgesetzte Erklärtiefe außer dem gerade erläuterten Aspekt für die Erklärqualität auch zur Variation in der Konzeptualisierung und Operationalisierung der Vignetten diente (s. Erläuterungen in dieser Arbeit unter Punkt *5.1.4 Variation der Erklärtiefe*), werden nach der Darstellung der hierarchischen linearen Regressionsmodelle differenziert nach Statusgruppen im Folgenden die hierarchischen linearen Regressionsmodelle für die vertieften Videos (A-C) denen für die einfachen Videos (D-F) gegenübergestellt. Damit wird der Fragestellung nachgegangen, welche Prädiktoren für die Bewertung der Erklärqualität von vertieften gegenüber einfachen Erklärungen ausschlaggebend sind, und ob und wo sich diese unterscheiden (Forschungsfrage 5). In den Tabellen sind sämtliche Ergebnisse abgebildet, berichtet werden im Folgenden die Auffälligkeiten.

8.3.2.3 Ergebnisse der hierarchischen linearen Regressionsmodelle unterschieden nach Erklärtiefe

Gesamtstichprobe

Begonnen wird mit dem hierarchischen linearen Regressionsmodell, das über die Gesamtstichprobe gerechnet wird, aufgeteilt nach den vertieften (Videos A-C) und den einfachen (Videos D-F) Varianten der Erklärungen.

[67] Die Intraklassenkorrelation liegt bei 34,16 Prozent.

Tabelle 22: Hierarchisches lineares Regressionsmodell mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Gesamtstichprobe
(unter Berücksichtigung der nach Studienteilnehmer/-innen (N_{max} = 257) und je drei Videos geordneten Datenstruktur)

Gesamtstichprobe		**Videos D-F (einfach)** ($N_{Beob.}$ = 293)	**Videos A-C (vertieft)** ($N_{Beob.}$ = 310)
ICC	%	2,52	17,24
Konstante	*β*	0,00	0,33
Strukturiertheit	*β*	0,35**	0,12
Adressatenorientierung	*β*	0,27**	0,09
Sprech-/Körperausdruck	*β*	0,01	0,39**
Sprachliche Verständlichkeit	*β*	-0,08	-0,05
Persönlichkeitswirkung	*β*	0,18*	0,04
Erklärtiefe	*β*	0,17*	0,26**
R^2 (marginal | konditional)		0,41 | 0,49	0,39 | 0,44

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N x* 3 Videos); *ICC*: Intraklassenkorrelation; *β:* unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.

Interessanterweise unterscheiden sich bei den einfachen und vertieften Videos die Prädiktoren für die Erklärqualität: Während bei den einfachen Varianten die *Strukturiertheit,* die *Adressatenorientierung* sowie (wenn auch schwächer) die *Persönlichkeitswirkung* und die *Erklärtiefe* ausschlaggebend sind, werden bei den vertieften der *Sprech-/Körperausdruck* und die *Erklärtiefe* hoch signifikant (vgl. Tab. 22).

Das Ergebnis zeigt, dass für die zwei Varianten jeweils unterschiedliche Aspekte der Erklärung als wichtig erachtet werden. Interessant ist, dass sich die Ergebnisse wieder komplementär ergänzen, mit Ausnahme der *Erklärtiefe*, die in beiden Varianten ein signifikantes Regressionsgewicht aufweist. Ein ähnlich komplementäres Ergebnis zeigt sich beim Vergleich der Regressionen der Schüler- und der SLD-Gruppe (vgl. Tab. 20).

Die aufgeklärte Gesamtvarianz liegt bei 41 Prozent für die festen und bei 49 Prozent für die festen und variablen Effekte bei den einfachen Erklärungen beziehungsweise bei 39 Prozent für die festen und bei 44 Prozent für die festen und variablen Effekte bei den vertieften Erklärungen. Für die Gesamtstichprobe unterscheidet sich die aufgeklärte Varianz somit nicht sehr für die einfachen und vertieften Vignetten. Das Modell deckt auch hier einen beachtenswerten Teil der Gesamtvarianz des guten Erklärens ab.

Auch diese Ergebnisse werden nun differenziert nach Statusgruppen betrachtet, allerdings anders aufgeteilt als bisher. Dargestellt werden die Regressionen für die Schüler/-innen (= SuS), die Studierenden (= Stud.) sowie für die Lehrkräfte und Didaktiker/-innen zusammen (= LD). Da sich die Fallzahlen in diesen Regressionsmodellen halbieren (weil jeweils nur die Videos A-C bzw. die Videos D-F einbezogen werden), ist es aufgrund der geringen Anzahl an Bewertungen nicht sinnvoll, die Lehrkräfte und die Didaktiker/-innen aufzuteilen. Diese Zusammennehmen von Lehrkräften mit Didaktikerinnen und Didaktikern ist nach theoretischen Überlegungen ebenfalls sinnvoll: Es unterscheidet die Perspektive der Adressatinnen und Adressaten (d. h. Schüler/-innen), von der Perspektive der Novizinnen und Novizen (d. h. Studierende) von der Perspektive der Expertinnen und Experten (d. h. Lehrkräfte und Didaktiker/-innen).

Schüler/-innen

Tabelle 23: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Schüler/-innen
(unter Berücksichtigung der nach Studienteilnehmer/-innen (N_{max} = 113) und je drei Videos geordneten Datenstruktur)

Schüler/-innen		**Videos D-F (einfach)** ($N_{Beob.}$ = 147)	**Videos A-C (vertieft)** ($N_{Beob.}$ = 147)
ICC	%	7,37	11,02
Konstante	*β*	0,23	0,69*
Strukturiertheit	*β*	0,20	0,06
Adressatenorientierung	*β*	0,23*	0,00
Sprech-/Körperausdruck	*β*	0,22	0,49**
Sprachliche Verständlichkeit	*β*	-0,07	-0,03
Persönlichkeitswirkung	*β*	0,14	0,01
Erklärtiefe	*β*	0,10	0,24
R^2 (marginal | konditional)		0,36 | 0,36	0,26 | 0,26

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N x* 3 Videos); *ICC*: Intraklassenkorrelation; *β:* unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.

Ähnlich wie in der Gesamtstichprobe so verteilen sich auch die Prädiktoren für die Qualität einer Erklärung bei den Schülerinnen und Schülern anders, je nachdem ob die einfachen oder die vertieften Erklärungen bewertet werden. Zu beachten ist bei diesen Ergebnissen, dass in den einfachen Vignetten nur die *Adressatenorientierung* ein signifikantes Regressionsgewicht besitzt (allerdings auf einem Niveau von p ≤ 0,05), bei den vertieften Vignetten hingegen allein der *Sprech-/Körperausdruck* signifikant wird. Dies ist vor allem deswegen bemerkenswert, als dass das

Regressionsgewicht der *Adressatenorientierung* bei den vertieften Videos so gering ausfällt, dass es mit Null anzugeben ist (vgl. Tab. 23).

Obwohl die Regressionsgewichte bei den einfachen Varianten kaum statistische Signifikanz aufweisen (vgl. Tab. 23), ist die aufgeklärte Gesamtvarianz dennoch um einiges höher als bei den vertieften Varianten. Das bedeutet, dass das Modell des guten Erklärens bei FALKE-D für die Statusgruppe der Schüler/-innen die einfachen Erklärungen besser abbildet als die vertieften. Damit unterscheiden sich die Schüler/-innen erneut von allen anderen Statusgruppen (vgl. folgende Tabellen 24 und 25).

Studierende

Tabelle 24: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Studierende
(unter Berücksichtigung der nach Studienteilnehmer/-innen ($N_{max} = 78$) und je drei Videos geordneten Datenstruktur)

Studierende		**Videos D-F (einfach)** ($N_{Beob.} = 61$)	**Videos A-C (vertieft)** ($N_{Beob.} = 91$)
ICC	%	6,17	20,82
Konstante	*β*	-0,22	-0,37
Strukturiertheit	*β*	0,65**	0,21
Adressatenorientierung	*β*	0,43*	0,47**
Sprech-/Körperausdruck	*β*	-0,05	0,47**
Sprachliche Verständlichkeit	*β*	-0,41*	-0,25
Persönlichkeitswirkung	*β*	0,24	0,20
Erklärtiefe	*β*	0,06	-0,02
R^2 (marginal | konditional)		0,43 | 0,47	0,48 | 0,64

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N x* 3 Videos); *ICC*: Intraklassenkorrelation; *β:* unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.

Auch bei den Studierenden zeigen sich unterschiedliche Ergebnisse, wenn man die Regressionen nach einfachen und vertieften Varianten getrennt darstellt: Bei den einfachen Varianten wird die *Strukturiertheit* auf hohem Niveau ($p \leq 0{,}01$) signifikant, auf etwas niedrigerem Niveau ($p \leq 0{,}05$)

die *Adressatenorientierung* und außerdem (im Negativen) die *Sprachliche Verständlichkeit*[68]. Für die vertieften Varianten werden die *Adressatenorientierung* und der *Sprech-/Körperausdruck* gleich hoch signifikant (vgl. Tab. 24).

Für die leicht negativen, nicht signifikanten Regressionsgewichte des *Sprech-/Körperausdrucks* der einfachen sowie der *Erklärtiefe* der vertieften Erklärungen gilt erneut die Zufallsvermutung (s. die Ausführungen unter Punkt *8.3.2.2 Statusgruppenspezifische Ergebnisse der hierarchischen linearen Regressionsmodelle* in dieser Arbeit).

Insgesamt ist somit festzustellen, dass die *Adressatenorientierung* in diesem Modell für die Studierenden den entscheidenden Aspekt für eine Erklärung, unabhängig von deren Erklärtiefe, darstellt; die Bedeutung von *Adressatenorientierung* für die Studierenden zeigt sich auch in dem Ergebnis des Regressionsmodells über alle Erklärungen hinweg (vgl. Tab. 21).

Die aufgeklärte Gesamtvarianz der vertieften Vignetten mit 48 Prozent für die festen und 64 Prozent für die festen und variablen Effekte ist höher als für die einfachen Vignetten (bei denen die aufgeklärte Gesamtvarianz bei 43 bzw. bei 47 Prozent liegt). Insgesamt ist die abgebildete Gesamtvarianz bei den Studierenden beachtlich. Dies zeigt, dass das Modell des guten Erklärens in FALKE-D die Einschätzungen zu Erklärungen von Studierenden gut trifft.

[68] Vgl. zu diesem Ergebnis erneut Erläuterungen unter Punkt *8.3.2.1 Statusgruppenübergreifende Ergebnisse des hierarchischen linearen Regressionsmodells* in dieser Arbeit bzw. Thim-Mabrey & Lindl (2021, in Vorb.).

Lehrkräfte und Didaktiker/-innen (LD-Gruppe)

Tabelle 25: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Lehrkräfte und Didaktiker/-innen
(unter Berücksichtigung der nach Studienteilnehmer/-innen ($N_{max} = 66$) und je drei Videos geordneten Datenstruktur)

LD-Gruppe		**Videos D-F (einfach)** ($N_{Beob.} = 85$)	**Videos A-C (vertieft)** ($N_{Beob.} = 72$)
ICC	%	43,20	45,12
Konstante	*β*	-0,10	0,39
Strukturiertheit	*β*	0,35*	-0,05
Adressatenorientierung	*β*	0,23	0,14
Sprech-/Körperausdruck	*β*	-0,09	-0,02
Sprachliche Verständlichkeit	*β*	0,21	0,03
Persönlichkeitswirkung	*β*	0,05	0,06
Erklärtiefe	*β*	0,26	0,63**
R^2 (marginal | konditional)		0,56 | 0,67	0,64 | 0,71

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen (*N* x 3 Videos); *ICC*: Intraklassenkorrelation; *β:* unstandardisierter Regressionskoeffizient; R^2: Determinationskoeffizient; *: $p \leq 0{,}05$; **: $p \leq 0{,}01$.

Als letztes wird über die Ergebnisse der vertieften gegenüber den einfachen Vignetten für die Expertengruppe bestehend aus Lehrkräften und Didaktikerinnen und Didaktikern berichtet. Diese beiden Gruppen wurden aus oben genannten statistischen und theoretischen Gründen für diese Regressionen zusammengenommen.

Analog zu den bereits berichteten Ergebnissen verteilt sich die Signifikanz auch für die Expertengruppe unterschiedlich auf die vertieften beziehungsweise einfachen Vignetten: Für die einfachen Varianten wird allein der Aspekt der *Strukturiertheit* signifikant, wenn auch nicht so stark, für die vertieften Varianten hingegen wird nur der Aspekt der *Erklärtiefe* hoch signifikant (vgl. Tab. 25). Für die leicht negativen Regressionsgewichte auf die Beurteilung des Aspektes *Sprech-/Körperausdruck* bei den einfachen und vertieften Varianten gilt aufgrund der kleinen Größen sowie der Tatsache, dass keines signifikant wird, die Zufallsvermutung (s. erneut die Ausführungen unter Punkt *8.3.2.2 Statusgruppenspezifische Ergebnisse der hierarchischen linearen Regressionsmodelle* in dieser Arbeit). Festzuhalten ist, dass der deutschspezifische Aspekt der *Erklärtiefe* für die

Expertinnen und Experten der einzige ausschlaggebende Prädiktor für die Erklärqualität der vertieften Varianten ist.
Interessant ist weiterhin, dass die aufgeklärte Gesamtvarianz in dieser Expertengruppe die höchsten Ausmaße annimmt und mit 64 Prozent bei den festen und 71 Prozent für die festen und variablen Effekte bei den vertieften Vignetten erneut eine bemerkenswerte Größe erreicht.

8.3.3 Zusammenfassung der inferenzstatistischen Ergebnisse

Fasst man alle Ergebnisse der Regressionen zusammen, so ist noch einmal zu betonen, dass die ausgewählten Aspekte insgesamt einen hohen Anteil der Erklärqualität vorhersagen können. Die Merkmale *Strukturiertheit, Adressatenorientierung, Sprech-/Körperausdruck* und *Erklärtiefe* sind also wichtige und valide Aspekte für die Wahrnehmung von Erklärqualität.
Betrachtet man die Regressionen für die einzelnen Statusgruppen, so unterscheidet sich allerdings die Bedeutung der Aspekte. Vergleicht man die Bewertungen der Adressatengruppe (Schüler/-innen) mit denen der Gesamtexpertengruppe (SLD), so ist die Verteilung interessanterweise komplementär (vgl. Tab. 20).
Zu erwähnen ist weiterhin, dass einzig bei dem FALKE-D-spezifischen Aspekt der *Erklärtiefe* (mit Ausnahme der Studierenden) Einigkeit über dessen Bedeutung für die Erklärqualität über die Statusgruppen hinweg besteht (vgl. Tab. 19-21). Dieser fachspezifische Aspekt ist also bei der Gestaltung einer Erklärung unbedingt zu beachten, denn er hat einen entscheidenden Einfluss auf die wahrgenommene Erklärqualität.
Betrachtet man hingegen die Regressionen, die aufgeteilt nach der Variation der Erklärtiefe berechnet wurden (Tab. 22-25), so ist Folgendes festzustellen.
Die Prädiktoren werden nicht als gleich wichtig für die vertieften und die einfachen Vignetten wahrgenommen. Das heißt, die Studienteilnehmer/-innen passen ihre Bewertungsmaßstäbe offensichtlich an die Art, wie erklärt wird (in FALKE-D also an die Erklärtiefe), und an die konkrete Erklärung an. Während über alle Videos hinweg für die Gesamtstichprobe alle vier Prädiktoren (*Strukturiertheit, Adressatenorientierung, Sprech-/Körperausdruck* und *Erklärtiefe*) ein hoch signifikantes Regressionsgewicht auf die Globalnote der Erklärungen haben, so verteilt sich die Signifikanz in der Gesamtstichprobe auf die vertieften und einfachen Varianten komplementär, – mit Ausnahme des Aspektes *Erklärtiefe*, der für beide Varianten relevant ist (vgl. Tab. 22).
Berichtenswert ist zudem, dass der *Sprech-/Körperausdruck* bei den vertieften Varianten sowohl für die Gesamtstichprobe als auch für die Schüler/-innen und für die Studierenden hoch signifikant ist, bei den einfachen Varianten dagegen nicht. Dies könnte mit dem sehr auffälligen Einsatz von Gestik und Mimik in der vertieften Was-Erklärung zu tun haben, in der die Lehrkraft die Etymologie der Redewendung beinahe schauspielerisch darstellt.

Erwähnt werden muss zudem die *Strukturiertheit* der einfachen Varianten, die in der Gesamtstichprobe sowie bei den Studierenden und den Expertinnen und Experten (d. h. Lehrkräfte und Didaktiker/-innen) signifikant wird. Eventuell kommt der Struktur in Erklärungen, die weniger in die Tiefe gehen, eine entscheidendere Bedeutung zu.

Ein analoges Ergebnis, allerdings für andere Statusgruppen, zeigt sich für die *Adressatenorientierung* der einfachen Varianten: Diese wird sowohl in der Gesamtstichprobe als auch bei den Schülerinnen und Schülern sowie den Studierenden ein signifikanter Prädiktor für das Globalurteil auf eine Erklärung.

Der FALKE-D-spezifische Aspekt der *Erklärtiefe* erweist sich in der Gesamtstichprobe für die einfachen und die vertieften Vignetten als signifikant, für die Expertengruppe LD dagegen nur für die vertieften Varianten. Die *Erklärtiefe* ist somit für die variierten Erklärungen insgesamt ein wichtiges Kriterium bei der Bewertung von Erklärqualität. Dass dies bei den vertieften Erklärungen stärker der Fall ist, zeigt, dass die umgesetzte Erklärtiefe von den Expertinnen und Experten auch so wahrgenommen wird. Diese Variation der umgesetzten Erklärtiefe wurde in der Varianzanalyse wie dargelegt für die Was- und Warum-Erklärungen ebenfalls als entscheidend für die globalen Bewertungen nachgewiesen (vgl. Tab. 17 sowie Ausführungen unter *8.2.4 Bedeutung der variierten Erklärtiefe* für die Statusgruppen in dieser Arbeit).

Die Ergebnisse der quantitativen Analysen sind somit dargelegt. Berichtet werden als nächstes die Erkenntnisse aus dem qualitativen Teil der Befragung.

9 Qualitative Ergebnisse

Die Studienteilnehmer/-innen gaben hauptsächlich Bewertungen im geschlossenen Antwortformat ab, indem sie die soeben analysierten Kriterien wie mehrfach bereits geschildert auf einer sechsstufigen Ratingskala bewerteten. Da ein weiteres Ziel der FALKE-D-Studie ist, auch solchen Kriterien auf die Spur zu kommen, die die Teilnehmer/-innen spontan und ohne Lenkung an Erklärungen anlegen (s. Forschungsfragen 6-7), wurde ihnen im ersten Durchlauf nach dem Vergeben der Globalnote die Möglichkeit gegeben, diese in einem offenen Antworttext zu begründen. Da Schülerinnen und Schülern für das Ausfüllen der beiden Teile der Fragebögen aus organisatorischen Gründen jeweils nur eine Schulstunde von 45 Minuten Länge zur Verfügung stand, konnte diese Statusgruppe keine offene Begründung abgeben, denn das hätte den zeitlichen Rahmen gesprengt. Offene Antworten liegen also lediglich von den Expertengruppen (d. h. Studierende, Lehrkräfte, Didaktiker/-innen) vor.

Die Forschungsfragen 6a-c, die an diese Daten angelegt wurden, werden hier noch einmal zusammengefasst: Welche Kriterien für gutes Erklären äußern die Studienteilnehmer/-innen

spontan und von sich aus, ohne Kenntnis über die Kriterien zu haben, die im Forschungsprojekt angelegt wurden? Zeigen sich Gemeinsamkeiten oder Systematiken in diesen Antworten und zwar in der Gesamtstichprobe (Forschungsfrage 6a) wie in den einzelnen Statusgruppen (Forschungsfrage 6b) und im Vergleich der einfachen mit den vertieften Vignetten (Forschungsfrage 6c)?

Vor der Auswertung der Antworten wird jedoch zunächst über das Vorgehen bei der Auswertung dieser Begründungen der Note berichtet.

9.1 Vorgehen bei der qualitativen Inhaltsanalyse

Die Analyse der offenen Begründungen des Globalurteils orientiert sich an Mayring (2010). Eine Überprüfung der bei Mayring (2010) beschriebenen qualitativen Techniken der Inhaltsanalyse ergab ein Vorgehen, das sich an der qualitativen Technik der Strukturierung (vgl. Mayring, 2010, S. 92–108) und der induktiven Kategorienbildung (vgl. Mayring, 2010, S. 83–85) orientiert. Dies liegt vor allem an der Beschaffenheit der Texte, die bei FALKE-D erhoben wurden: Zum einen waren die Begründungen der Globalurteile optional, das heißt, es liegen natürlich Fragebögen vor, die keine offenen Antworten beinhalten. Gewichtiger ist aber zum anderen, dass die Teilnehmenden explizit aufgefordert wurden, die gegebene Note kurz zu begründen; wortwörtlich lautete die Anweisung im Fragebogen: „Hier können Sie Ihre Note gerne kurz begründen. Bitte beschränken Sie sich auf maximal 3 Stichpunkte." Die Entscheidung, die Antwort auf drei Stichpunkte zu begrenzen, war vor allem der Bearbeitungsdauer des ersten Teils des Fragebogens geschuldet. Die befragten Personen sollten zeitlich nicht über Gebühr belastet werden, so dass im Sinne der Zumutbarkeit kurze, schnelle Antworten erbeten wurden. Die tatsächliche Länge der Antworten reicht von einem Stichwort bis hin zu mehreren Sätzen. Allerdings beschränkt sich die Mehrzahl der Antworten auf das Nennen von Stichpunkten, so dass die Techniken der Zusammenfassung (vgl. Mayring, 2010, S. 67–85) und der Explikation (vgl. Mayring, 2010, S. 85–92) allein aufgrund der „Kürze" des zu analysierenden Materials ausscheiden.

Zur Aufbereitung des Materials wurde die im Bereich der qualitativen Forschung schon seit längerem etablierte Analysesoftware Maxqda (Version 2018) gewählt. Maxqda eignet sich für FALKE-D vor allem, da es die Aufbereitung und Organisation des zur Verfügung stehenden Textmaterials erleichtert. Dies vereinfacht die anschließende Interpretation des Materials. Oben beschriebene offene Antworten wurden als Excel-Datei aus dem Online-Fragebogen exportiert und in die Analysesoftware Maxqda eingespeist. So standen sie direkt zur Weiterverarbeitung zur Verfügung. Das nun beschriebene Kategoriensystem wurde an die Daten aus FALKE-D angelegt.

9.2 Vorgehen bei der Kategoriebildung

Was das Vorgehen bei der Kategoriebildung anbelangt so wurde bei FALKE-D ein Zwischenweg zwischen den oben erwähnten Techniken der Strukturierung und der induktiven Kategoriebildung gewählt, da sowohl deduktiv als auch induktiv vorgegangen wurde. Zunächst wird die deduktive Kategoriebildung bei FALKE-D dargelegt.

9.2.1 Deduktive Kategoriebildung

Deduktive Kategorienbildung bedeutet, dass vorab festgelegte Kategorien an das Material angelegt werden (vgl. Mayring, 2010, S. 66). Das Vorgehen in FALKE-D bei der Analyse der offenen Antworten orientierte sich an den Schritten der strukturierenden Inhaltsanalyse nach Mayring (vgl. 2010, S. 93), wie sie in Abbildung 15 zu sehen ist. Im Folgenden wird dieses Vorgehen bei FALKE-D genauer dargelegt.

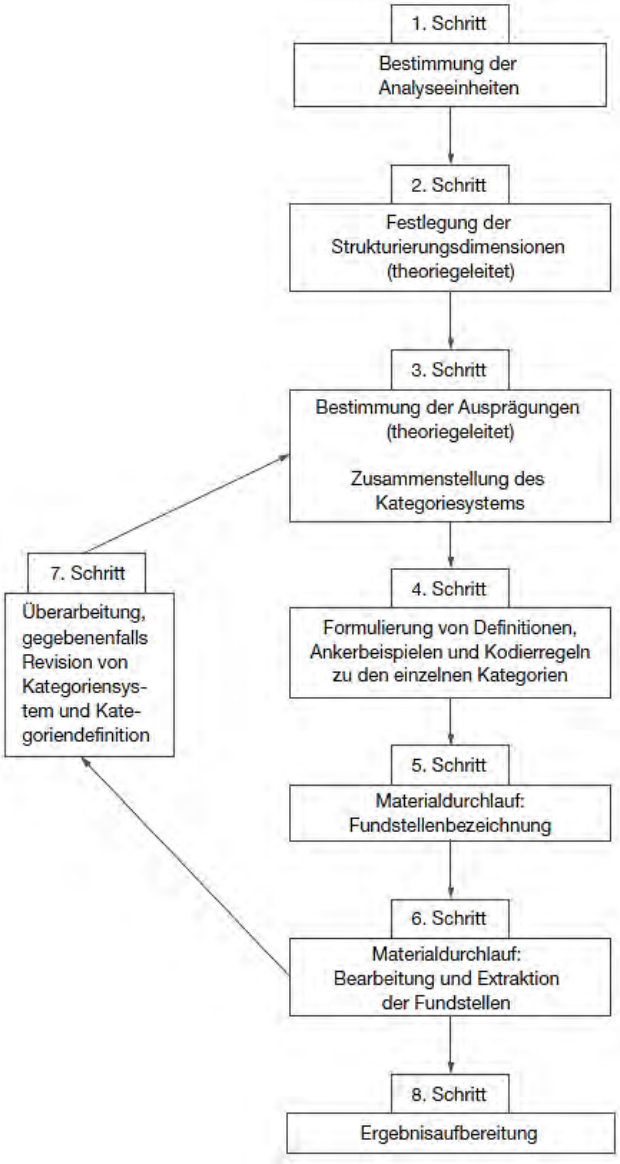


Abbildung 15: Modell zur strukturierenden Inhaltsanalyse
Mayring (2010, S. 93)

Schritt 1 bis 3 geschah bei FALKE-D bereits während der theoretischen Vorarbeit zu den Aspekten guten Erklärens, die im Fragebogen abgefragt werden sollten. Diese Vorarbeit wurde zum Teil vom Gesamtprojekt FALKE (bei den gemeinsamen Kriterien *Strukturiertheit* und *Adressatenorientierung*) geleistet beziehungsweise beim deutschspezifischen Aspekt der *Erklärtiefe* in der Teilstudie FALKE-D (s. Punkt *8.1.1 Pilotierung* in dieser Arbeit). Da in der Vorarbeit zur Erstellung des Fragebogens die Aspekte guten Erklärens, die an die Erklärungen angelegt wurden, und die entsprechenden Items dazu auf Basis der bestehenden Forschungsliteratur und des eigenen Forschungsinteresses festgelegt wurden, war es naheliegend, diese Aspekte auch in den Antworten der befragten Personen zu suchen. Somit war sowohl Schritt 1, nämlich „Bestimmung der Analyseeinheiten" (Mayring, 2010, S. 93) als auch Schritt 2, nämlich „Festlegung der Strukturierungsdimensionen"

(Mayring, 2010, S. 93) bereits erfolgt: Die Analyseeinheiten sind in FALKE-D zunächst sämtliche offene Begründungen, die zum Globalurteil pro Video abgegeben wurden; die kleinste kodierbare Einheit stellt ein Wort dar, die längste kodierte Einheit umfasst mehrere Sätze.

Bei der strukturierenden Inhaltsanalyse werden die Strukturierungsdimensionen in Anlehnung an die inhaltliche Strukturierung festgelegt, deren Ziel es ist „bestimmte Themen, Inhalte, Aspekte aus dem Material herauszufiltern und zusammenzufassen" (Mayring, 2010, S. 98). Zusammenfassungen waren aufgrund der Kürze des vorliegenden Textmaterials für FALKE-D nicht sinnvoll. Es wurden jedoch oben genannte Aspekte aus dem Material herausgefiltert. Schritt 3, also genauer die „Bestimmung der Ausprägungen [und die] Zusammenstellung des Kategoriesystems" (Mayring, 2010, S. 93) entsprach der Aufteilung der einzelnen Aspekte, die zuvor theoretisch erarbeitet wurden.

Deduktiv wurden bei FALKE-D demnach auf der einen Seite die Kriterien, die später in den geschlossenen Items abgefragt wurden, als Kategorien an die offenen Antworten angelegt. Dabei handelt es sich um die gemeinsamen Kriterien *Strukturiertheit, Adressatenorientierung, Sprech- und Körperausdruck* sowie *Sprachliche Verständlichkeit* und das deutschspezifische Kriterium der *Erklärtiefe*. Deduktiv wurden auf der anderen Seite auch *Veranschaulichungen* und *Fachlich korrekt* als Kategorien an die Antworten angelegt, die zwar aus oben genannten Gründen (s. Punkt *4.2 Konzeptualisierung und Operationalisierung der Aspekte guten Erklärens* in dieser Arbeit) nicht im Online-Fragebogen abgefragt wurden, aber in der Literatur immer wieder genannt werden. Aufgrund der theoretischen Beschäftigung mit Kriterien guten Erklärens war es sinnvoll, diese Kategorien ebenfalls vorab festzulegen (vgl. Mayring, 2010, S. 83).

Ankerbeispiele wurden im Material in Abstimmung mit der Zweitkodiererin (studentische Hilfskraft[69]) während des ersten Durchgangs festgelegt; zum Teil wurden diese in den weiteren Durchgängen noch ergänzt. So wurde für den Aspekt *Erklärtiefe* (kodiert als *Vertieft vs. oberflächlich*) als Anker „bleibt im Gedächtnis[70]" und „gut merken" festgesetzt; für den Aspekt *Fachlich korrekt* wurden „fachlich korrekt", „fachlich gut" und „klare Definition" gewählt.

Zu den Kategorien *Strukturiertheit, Adressatenorientierung* und *Veranschaulichung* wurden jeweils Unterkategorien gebildet. Sie entstanden in einer Mischung aus deduktivem und induktivem Vorgehen: Für die Aspekte *Strukturiertheit* und *Adressatenorientierung* wurden zunächst deduktiv die Unterkategorien herangezogen, die während der theoretischen Vorarbeit projektübergreifend

[69] Die Kodierungen in FALKE-D wurden von Theresa Häusl (SHK) sowie der Verfasserin der Studie durchgeführt.

[70] Alle Ankerbeispiele sind direkt dem Material entnommen. Sie werden wörtlich zitiert, genau wie alle folgenden Beispiele, die zur Illustration herangezogen werden. Etwaige Verstöße gegen die Sprachrichtigkeit werden dabei nicht korrigiert und auch nicht angemerkt. Eine Übersicht über alle Definitionen der Kodierungen und die jeweiligen Ankerbeispiele findet sich in Anhang 11.

in den entsprechenden Modellen erarbeitet wurden (s. Anhang 1 und 2). Bei der *Strukturiertheit* wurden folgende Punkte festgesetzt: *Gegenstandsbestimmung, Schritt für Schritt* und *Fokussierung*. Auch für diese Unteraspekte wurden Ankerbeispiele festgelegt. Für die Unterkategorie *Schritt für Schritt* waren das „sinnvoll aufbaut" und „Schritt für Schritt aufgebaut". Für die Unterkategorie *Gegenstandsbestimmung* „klar, worum es geht" und „Ziel klar", und schließlich für die Unterkategorie *Fokussierung* „präzise" und „zu viele Einzelheiten". Induktiv wurde die Kategorie noch um die Unterkategorie *Global Struktur* ergänzt (s. folgenden Punkt *9.2.2 Induktive Kategoriebildung* in dieser Arbeit).

Bei der *Adressatenorientierung* wurden deduktiv die Unterkategorien *Lebensweltbezug* und *Vorwissen* an das Textmaterial angelegt. Ankerbeispiele zum *Lebensweltbezug* waren: „Erklärung weicht sehr von der Lebenswelt der Schülerinnen und Schüler ab" und „für Schüler aus deren Sicht eventuell schlecht nachvollziehbar". Ankerbeispiele für *Vorwissen* waren: „gut an Vorwissen angeknüpft" und „an Vorwissen der Schüler anschließend".

Diese Unterkategorien wurden anschließend induktiv ergänzt um *Global Adressatenorientierung* und *Motivation* (s. folgenden Punkt *9.2.2 Induktive Kategoriebildung* in dieser Arbeit). Die Definitionen in Schritt 4 (vgl. Mayring, 2010, S. 94) wurden den von FALKE entwickelten Modellen (s. Punkte *4.2.1 Strukturiertheit* und *4.2.2 Adressatenorientierung* in dieser Arbeit) entnommen und ergänzt durch abstrahierte Beispiele aus den vorliegenden Begründungen. Sie sind in Anhang 11 für alle Kodes aufgelistet.

Auch für die Aspekte *Sprachliche Verständlichkeit* sowie für *Sprech- und Körperausdruck* wurden Ankerbeispiele gewählt, und zwar zu ersterem „gute Sprache" sowie „schülergemäße Sprache" und zu letzterem „schnelles Tempo" sowie „leicht stockend". Unter *Sprachliche Verständlichkeit* wurden alle Äußerungen kodiert, die sich explizit auf die Sprache und den Satzbau der erklärenden Person bezogen. Unter *Sprech- und Körperausdruck* wurden die Antworten kodiert, die sich auf die Umsetzung der Erklärung bezogen, das heißt auf Aspekte, die explizit die Performanz ansprachen. Beide Aspekte wurden im Folgenden allerdings nicht weiter ausdifferenziert. Die Kodierung diente nur dazu, sämtliche Antworten, die sich auf diese beiden Aspekte bezogen, auszusortieren. Sie wurden den Forscherinnen, die sich aus einer dezidiert fachlichen Perspektive mit diesen Aspekten beschäftigen, zur Verfügung gestellt (für genauere Ausführungen dazu s. Gunga, 2022, in Vorb. sowie Thim-Mabrey & Lindl, 2021, in Vorb.). In dieser Arbeit wird auf diese beiden Aspekte daher nicht mehr Bezug genommen.

Es erfolgte ein erster Materialdurchgang (Schritt 5 und 6) von circa zehn Prozent des Textmaterials. Währenddessen wurde das Kategoriensystem auf seine Tauglichkeit überprüft, die einzelnen Kategorien wurden durchgesprochen. Es zeigte sich, dass die vorab festgelegten Aspekte alle in den offenen Begründungen zu finden und die Ankerbeispiele leicht festzulegen waren, so dass mit

gerade beschriebener Einteilung in die Oberaspekte und die genannten Unteraspekte eine gute Basis für die Analyse geschaffen war, die insgesamt allerdings noch zu grob war. Es ergab sich somit die Notwendigkeit nach weiteren, induktiven, das heißt aus dem Material selbst abgeleiteten, Kategorien, die im Folgenden genauer erläutert werden.

9.2.2 Induktive Kategoriebildung

Im Gegensatz zu der bei Mayring beschriebenen Vorgangsweise der induktiven Kategorienbildung (vgl. 2010, S. 83–85) wurde bei FALKE-D nicht auf Zusammenfassungen des Materials zurückgegriffen. Das liegt wie erwähnt daran, dass der zur Verfügung stehende Text in den meisten Fällen so kurz ist, dass er nicht zusammengefasst werden kann. Dennoch erwies es sich nach der ersten Kodierung mit den deduktiven Kategorien als notwendig, genauer zu analysieren und Kategorien aus dem Material selbst heraus zu definieren, denn es konnte nicht der gesamte Text über die deduktiven Kategorien kodiert werden. Überdies war es Ziel der Erhebung der offenen Begründungen „intuitive" Kriterien der Statusgruppen herauszuarbeiten, die nicht zwangsläufig als Kriterien in der geschlossenen Befragung von FALKE-D erhoben wurden.

Während bei dem Vorgehen nach Mayring die einzelnen Kategorien „in einem Verallgemeinerungsprozess" (2010, S. 83) aus dem Material abgeleitet werden, wurden die induktiven Kategorien bei FALKE-D zum Teil durch das häufige Auftreten bestimmter Stichwörter in den Antworten gebildet, die dann selbst als Kategorie dienten und keinen Verallgemeinerungsprozess durchliefen.

Der andere Teil der induktiv gebildeten Kategorien entstand durch einen Abstraktionsprozess.

Induktiv war die Kategoriebildung in beiden Fällen, da diese Kategorien nicht auf Grundlage vorheriger theoretischer Überlegungen gebildet wurden, sondern erst während der Arbeit mit dem Textmaterial entstanden sind.

Das Vorgehen richtete sich nach Mayrings Prozessmodell induktiver Kategorienbildung (vgl. Mayring, 2010, S. 84; sowie in dieser Arbeit Abb. 16): Der Gegenstand und das Ziel der Analyse stand wie bei der deduktiven Kategorienbildung erläutert von Anfang an fest, nämlich Kriterien, die die unterschiedlichen Statusgruppen und Studienteilnehmer/-innen intuitiv an die Qualität von Erklärungen im Deutschunterricht anlegen, zu erfassen. Dabei wurde die gesamte Textmenge gesichtet, was aufgrund der Kürze der Antworten möglich war.

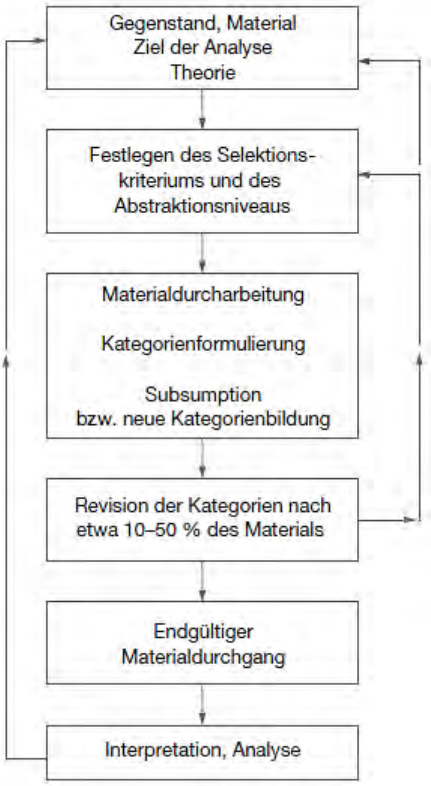


Abbildung 16: Prozessmodell induktiver Kategorienbildung
Mayring (2010, S. 84)

Statt wie Mayring (2010, S. 84) es vorschlägt, für die Forschungsfrage Irrelevantes wegzulassen und aus der Analyse auszuschließen, wurde für derartiges Textmaterial zunächst die Kategorie *Fehlende Abstraktion bezogen auf Designvorgaben* gebildet. Diese Kategorie wurde für alle Kommentare gewählt, die nicht die Qualität der gesehenen Erklärung betrafen, sondern sich auf die aus theoretischen Überlegungen notwendige Umsetzung der Erklärungen in FALKE-D bezogen. Diese betreffen vor allem die monologische Gestaltung der Erklärung und die damit fehlende Interaktion mit Schülerinnen und Schülern. Bemerkungen wie „SuS nicht wirklich mit einbezogen" oder „sehr frontal" wurden hierunter kodiert. Gerade in Bezug auf die *Fachliche Relevanz* ist es wichtig, auch diese Bemerkungen in die Analyse mit aufzunehmen. Sie geben Auskunft darüber, ob und wem die notwendige Umsetzung der Erklärungen als monologische Lehrererklärung so unnatürlich vorkam, dass eine Konzentration auf die Qualität der Erklärung nicht möglich war.

Ähnlich verfahren wurde mit Bemerkungen, die hauptsächlich beschreibender Natur waren. Bezogen sich diese Zusammenfassungen allein auf die Methodik, die in der Erklärung verwendet wurde, so wurde dies unter der (induktiven) Kategorie *Direkter Bezug zur Methode* kodiert. Entscheidend war, dass allein auf die Erklärmethode Bezug genommen wurde, ohne dass Kriterien wie Struktur oder Ähnliches genannt wurden. Ankerbeispiele dafür sind: „Herkunft des Ausdrucks erklärt" oder „Strategie mit lautem Denken ist überzeugend". Auch diese Kategorie lässt nur wenige

Rückschlüsse auf intuitive Kriterien zu Erklärqualität der Studienteilnehmer/-innen zu, da sich die Antworten wie gesagt vor allem auf deskriptiver Ebene befinden. Da jedoch in Anbetracht der Kürze des Textmaterials festgelegt wurde, alles zu kodieren, wurden auch Bemerkungen dieser Art mit aufgenommen.

Ein dritter Fall der Antworten, der nicht auf der Ebene der Qualität der Erklärungen liegt, sondern sich vielmehr aus dem Forschungsdesign ergibt, ist die Kategorie *Direkter Vergleich Variante 1 und 2*. Auch diese ergab sich induktiv während der Arbeit am Textmaterial. Es fielen Bemerkungen nach der Bewertung der jeweils an zweiter Stelle gesehenen Variante auf, die sich eindeutig aus dem Design (d. h. ein Erklärthema, das in zwei verschiedenen Varianten erklärt wird) ergeben und die explizit die beiden Varianten miteinander vergleichen. Als Ankerbeispiele dienten die Bemerkungen: „jetzt besser", „erneut" und „immer noch". Das Vergleichsmoment selbst (z. B. „wenig Schülereinbezug") wurde wiederum unter dem entsprechenden Kriterium kodiert. Auskunft gibt diese Kategorie daher darüber, bei welchen Varianten die Studienteilnehmer/-innen in ihrer Bewertung konkret auf die erst gesehene Variante zurückgreifen.

Schließlich wurde als letztes eine Kategorie *Sonstiges* induktiv angelegt, worunter all die Antworten kodiert wurden, die tatsächlich unter keine der anderen Kategorien fielen. Wie bereits erwähnt, sollte die Gänze des Materials kodiert werden; diese Entscheidung machte das Einrichten einer solchen Kategorie notwendig. Ankerbeispiele festzulegen, erwies sich ob der individuellen Natur dieser Antworten als schwierig. Aufgenommen wurden in diese Kategorie in Absprache mit der Zweitkodiererin schließlich Beschreibungen aller Art, die sich nicht auf die Methode beziehen, Unverständliches und Verbesserungsvorschläge. Beispiele aus den Begründungen sind: „Thematik Waffen", „… wäre hilfreich gewesen" oder „gebrauchsorientiert". Diese Kategorie wurde mehrmals überarbeitet.

Der zweite Schritt nach Mayring „Festlegen des Selektionskriteriums und des Abstraktionsniveaus" (2010, S. 84) wurde wie oben bereits angedeutet folgendermaßen durchgeführt: Zum Teil ergaben sich die Kategorien durch die häufige Nennung eines Stichwortes. Dies war zum Beispiel bei der gerade beschriebenen Kategorie *Direkter Vergleich Variante 1 und 2* der Fall, zu der Komparative („besser") beziehungsweise vergleichende Ausdrücke („erneut", „immer noch") kodiert wurden. Auch bei *Beispiel* (als Subkategorie von *Veranschaulichung*) trifft eine häufige Nennung zu, denn es taucht in einer Vielzahl von Antworten auf: „Beispiel genannt" oder „an vielen Beispiele erklärt". Bei anderen Kategorien war eine Abstraktion notwendig. Auf diese Weise entstand etwa die Kategorie *Direkter Bezug zur Methode*. Hier wurden vor allem Bemerkungen kodiert, die auf einer rein beschreibenden Ebene angesiedelt sind (vgl. Anmerkungen oben). Gleichermaßen vorgegangen wurde bei *Motivation* (als Subkategorie von *Adressatenorientierung*), worunter Bemerkungen wie „langweilig" und „spannend" oder „Interesse aufrechterhalten" kodiert wurden.

Wie den Beispielen zu entnehmen ist, wurden hier Einschätzungen, mit welchem Interesse und Gefühl die Adressatinnen und Adressaten der Erklärung folgen, abstrahiert.

Analog ergab sich die Kategorie *Veranschaulichung medial* (als Subkategorie von *Veranschaulichung*). Nach einem ersten Durchgang, in dem die Kategorie *Veranschaulichung* deduktiv angelegt worden war, zeigte sich, dass die Kodierungen zu unpräzise und zu allgemein waren und eine Ausdifferenzierung nötig war. Die Kategorie wurde daraufhin induktiv, das heißt nur aufgrund des vorliegenden Textmaterials, präzisiert. Dies geschah in enger Abstimmung mit der zweiten Kodiererin (studentische Hilfskraft). Unter die Kategorie *Veranschaulichung medial* fallen sämtliche Medien, die in diesem Zusammenhang erwähnt wurden, also zum Beispiel folgende Bemerkungen: „Sehr gut die Tafel", „Visualisierung auf der Folie" oder auch „Die Visualisierung ist größtenteils gut gelungen: die beiden Medien parallel und Häkchen/Fragezeichen/Blitz; jedoch wirken die beiden Spalten (anfangs) irritierend."

Ebenso ergab sich während dieses Durchgangs die Kategorie *Global.* Darunter fallen alle Antworten, die sich insgesamt auf die Erklärqualität bezogen. Ankerbeispiele sind „Erklärung überzeugt insgesamt", „plausibel" und „ordentlich erklärt".

Das Material wurde daraufhin erneut durchgegangen und mit den neuen Kategorien kodiert. Wiederum wurden diese unabhängig voneinander von den zwei Raterinnen an etwa einem Drittel des Gesamtmaterials angewandt.

Es kam vor der endgültigen Kodierung noch einmal zu Revisionen, die sich während dieses dritten Kodiervorgangs als notwendig herausstellten.

Dies betraf zwei Kategorien, die sich noch als zu grob erwiesen, nämlich die Kategorien *Global* und *Veranschaulichung.*

In Absprache mit der Zweitkodiererin wurde beschlossen, die Kategorie *Global* aufzuspalten, da sich sehr viele Bewertungen in dieser Kategorie fanden und beiden Kodiererinnen während der Kodierarbeit aufgefallen war, dass sie sich noch weiter differenzieren ließe. Während der Kodierung der Kategorie *Global* fielen zahlreiche Äußerungen auf, die einen Aspekt, zum Beispiel Struktur, erwähnten, wenn auch ohne diesen näher zu spezifizieren. Nach einigen Überlegungen, wie mit Bemerkungen dieser Art verfahren werden sollte, einigten sich die Raterinnen darauf, zu allen Aspekten, zu denen Unterkategorien gebildet wurden, zusätzlich die Kategorie *Global* einzuführen. Dies brachte folgende zusätzliche Subkategorien hervor: *Global Adressatenorientierung, Global Struktur* sowie *Global Veranschaulichung.* Darunter wurden unter anderem Bemerkungen wie folgende kodiert: „schülerangepasst" oder „über die Köpfe der Schüler hinweg, eher abstrakt" (*Global Adressatenorientierung*) beziehungsweise „gut strukturiert" oder „etwas unstrukturiert" (*Global Struktur*) beziehungsweise „Sehr anschaulich" oder „schlecht veranschaulicht" (*Global Veranschaulichung).*

Auch die Kategorie *Veranschaulichung* mit den bisherigen Subkategorien *Global, Beispiel* und *Medial* erwies sich noch als zu grob. Induktiv am Textmaterial wurden zu diesem Punkt noch die Unterkategorien *Inhalt* und *Gestik* herausgearbeitet. Bemerkungen, die zuvor unter Sprechen kodiert waren, wurden daraufhin noch einmal genauer betrachtet. Da *Veranschaulichung* eine Kategorie ist, die oft genannt wurde, wurde beschlossen, alle Bemerkungen, die die Gestik im Zusammenhang mit ihrer veranschaulichenden Funktion erwähnen, unter *Veranschaulichung Gestik* zu kodieren. Ankerbeispiele sind hierzu: „unterstützt Aussagen durch Gestik" oder „gestisch verdeutlicht". Andere Begründungen, die sich allein auf die Gestik beriefen (etwa „sehr schön die Gestik"), wurden unter *Sprech- und Körperausdruck* belassen.

Als letzte Unterkategorie zur *Veranschaulichung* ergab sich *Veranschaulichung Inhalt.* Dazu wurden vor allem Begründungen, die bis dato unter *Veranschaulichung Global* beziehungsweise *Veranschaulichung Beispiel* fielen, erneut durchgesehen. Hierunter fallen vor allem Begründungen, die sich konkret auf das in der Erklärung Gesagte beziehen, zum Beispiel die Bemerkungen „Veranschaulichung durch geschichtl. Hintergrund" oder „inhaltlich anschaulicher".

Nach diesen letzten Absprachen erfolgte die endgültige Kodierung des gesamten Materials.

Die Verfasserin der Studie kodierte das gesamte Material, die Hilfskraft kodierte insgesamt 60% des Textes.

Eine Übersicht über sämtliche Kodierungen und Unterkodierungen gibt Tabelle 26, in der auch verzeichnet ist, welche Kategorien deduktiv und welche induktiv entstanden sind. Die Definitionen und die Ankerbeispiele (aufgeteilt nach typischen und randständigen Beispielen) finden sich für alle Kodes und Subkodes in Anhang 11.

Tabelle 26: Übersicht über deduktive und induktive Kodes und Subkodes

Kode/*Subkode*	**deduktiv/induktiv**
Global	induktiv
Fachlich korrekt	induktiv
Direkter Bezug zur Methode	induktiv
Direkter Vergleich Variante 1 und 2	induktiv
Vertieft vs. oberflächlich	deduktiv
Fehlende Abstraktion bez. auf Designvorgaben	induktiv
Sprechen	deduktiv
Sprache	deduktiv
Sonstiges	induktiv
Veranschaulichung	deduktiv
Global Veranschaulichung	induktiv
Beispiel	induktiv
Medial	deduktiv
Gestik	induktiv
Inhalt	induktiv
Struktur	deduktiv
Global Struktur	induktiv
Gegenstandsbestimmung	deduktiv
Schritt für Schritt	deduktiv
Fokussierung	deduktiv
Adressatenorientierung	deduktiv
Global Adressatenorientierung	induktiv
Vorwissen	deduktiv
Motivation	induktiv
Lebensweltbezug	deduktiv

9.3 Gütekriterien

Die Diskussion der Gütekriterien richtet sich wie die Analyse der offenen Antworten nach Mayring (vgl. 2010, S. 116–123), daher werden im Folgenden Gütekriterien besprochen, die sich explizit auf die Methode der qualitativen Inhaltsanalyse beziehen (zur Diskussion der klassischen Gütekriterien im Zusammenhang mit der qualitativen Inhaltsanalyse vgl. Mayring, 2010, S. 116-123).

Validität

In Anlehnung an Krippendorff (1980) unterteilt Mayring die Validität als Gütekriterium für die qualitative Inhaltsanalyse in diese Aspekte: Semantische Gültigkeit, Stichprobengültigkeit,

Korrelative Gültigkeit, Vorhersagegültigkeit und Konstruktgültigkeit (vgl. Mayring, 2010, S. 119). Diese Kriterien erfüllt die Studie FALKE-D wie folgt:

Die Semantische Gültigkeit ist dann gegeben, wenn die Kategorien, die an das Material angelegt werden, dieses richtig erfassen (d. h. „Angemessenheit der Kategoriendefinitionen" Mayring, 2010, S. 119). Nach Krippendorff (1980) gibt es zwei Indikatoren für semantische Gültigkeit: „Semantical validity is indicated by a substantial agreement between two different partitions of the same set of units of analysis. One is obtained from the way the analytical procedure assigns units to its categories. The other is obtained according to an external criterion, for example, involving judges familiar with the symbolic nature of the same language who group these units according to their semantic similarity" (Krippendorff, 1980, S. 161). Beides ist bei FALKE-D geschehen: Die Kategorien, die deduktiv an die offenen Antworten in FALKE-D angelegt wurden, fußen auf der theoretischen Vorarbeit im Gesamtprojekt beziehungsweise in der Teilstudie FALKE-D. Ihre Definitionen sind aufgrund der ausführlichen Diskussion innerhalb der gesamten Forschungsgruppe inhaltlich als ausreichend abgesichert anzusehen. Ankerbeispiele und Kodierregeln wurden, sofern sie Aspekte des Gesamtprojektes betrafen, ebenfalls mit Mitarbeitenden aus FALKE diskutiert. Die Aspekte (*Erklärtiefe* sowie ein Teil der induktiven Kategorien, s. obige Erläuterungen), die nur in der Studie FALKE-D von Belang waren, wurden in Absprache mit der Zweitkodiererin inhaltlich abgesichert; dazu wurden immer in mehreren Durchgängen (s. o.) sämtliche Stellen, die unter einem Kode gefasst wurden, von den zwei Kodiererinnen verglichen.

Das Kriterium der Stichprobengültigkeit wurde unter Punkt *7.4.3 Überblick über Ablauf der Erhebung* in dieser Arbeit bereits erläutert.

Die Korrelative Gültigkeit (d. h. der Vergleich der Ergebnisse mit bereits vorliegenden Studienergebnissen, die sich mit dem gleichen Thema beschäftigt haben, vgl. Mayring, 2010, S. 119) kann wegen fehlender Studien zur Wahrnehmung der Qualität von Erklärungen aus dem Deutschunterricht für FALKE-D nicht angegeben werden. Es sei jedoch darauf hingewiesen, dass die abgefragten Kategorien zur Erklärqualität sich natürlich aus theoretischer Vorarbeit ergaben (s. Punkte *3 Theoretische Sichtweisen auf Erklären* und *4 Erklären in FALKE-(D)* in dieser Arbeit) und dass zum Beispiel Handreichungen zum guten Erklären ebenfalls diese oder ähnliche Kriterien angeben (vgl. z. B. Schopf et al., 2015; Lehner, 2018; Scheffel, 2019).

Die Vorhersagegültigkeit bezeichnet die Möglichkeit, Prognosen aus den Ergebnissen zu erstellen (vgl. Mayring, 2010, S. 120). Die kann nur gelingen, wenn diese Ergebnisse sichtbar, das heißt in der Wirklichkeit erkennbar, gemacht werden können: „To test predictive validity, the factuality of the criterion is essential" (Krippendorff, 1980, S. 166). Die Vorhersagegültigkeit lässt sich in vorliegender Studie nicht überprüfen, da die Wahrnehmung von Erklärqualität im Fokus steht.

Folgestudien müssen zum Beispiel klären, ob die als wichtig erachteten Kriterien für eine Erklärung im Deutschunterricht auch zu höheren Verstehensleistungen führen.

Die Konstruktvalidität ist bei qualitativer Inhaltsanalyse dann gegeben, wenn die Operationalisierung des Konstruktes auf theoretischer Vorarbeit beruht und die Ergebnisse vor dem Hintergrund der Theorie plausibel sind (vgl. Mayring, 2010, S. 117). Dabei können laut Krippendorf folgende Aspekte herangezogen werden: „past successes with similar constructs and/or similar situations; experiences with the context of available data; established theories and models of the contextual dependencies of data; representative interpretes and experts" (Krippendorff, 1980, S. 167). Im Gegensatz zur Vorhersagegültigkeit ist die Konstruktvalidität durch die bereits ausführlich dargelegte theoretische Vorarbeit zum Erklären (d. h. *established theories*) und zu Kriterien guten Erklärens sowie die Diskussionen in der Gesamtforschergruppe FALKE (d. h. *representative experts*) gewährleistet.

Reliabilität

Das Kriterium der Stabilität nach Mayring (vgl. 2010, S. 120) ist durch das wiederholte Anwenden des Kategoriensystems erfüllt. Die Übereinstimmung der beiden Kodiererinnen wird über die Interkoderreliabilität gemessen. Sie gibt Auskunft darüber, wie gut die Kodierungen der einzelnen Kategorien übereinstimmen. Wie bereits geschildert wurde durch wiederholte Absprachen und Vergleiche während des Kodierprozesses versucht, Klarheit über die angelegten Kategorien zu schaffen und so eine möglichst hohe Übereinstimmung zu erzielen. Dazu wurde 60 Prozent des Gesamtmaterials von zwei Raterinnen kodiert (Vorgehen s. oben); der Rest wurde nur noch von einer Raterin bearbeitet.

Zunächst wird die prozentuale Übereinstimmung der einzelnen kodierten Segmente angegeben[71]. Diese liegt insgesamt bei FALKE-D bei 78,35 Prozent. Die prozentuale Übereinstimmung variiert von Kode zu Kode; die Werte liegen zwischen 66,67 Prozent für *Adressatenorientierung, Subkode Lebensweltbezug* und 87,50 Prozent für *Adressatenorientierung, Subkode Motivation*. Da der erste Wert sich aus nur drei Kodierungen zusammensetzt, ist er akzeptabel. Die anderen Werte liegen mit über 70 Prozent alle in einem hohen Bereich.

Zusätzlich wurde in FALKE-D die Interkoderreliabilität über Kappa berechnet. Der Kappa-Wert gibt einen zufallskorrigierten Koeffizienten an. Das bedeutet, dass aus der prozentualen Übereinstimmung „herausgerechnet" wird, wie viel dieser Übereinstimmung allein durch Zufall entstanden sein kann, der Wert ist somit zufallsbereinigt. Maxqda gibt den Kappa-Wert nach Brennan und Prediger (1981) an, er beträgt bei FALKE-D $\kappa_n = 0{,}77$ und ist somit in Anlehnung an

[71] Als Kodeüberlappung gelten Übereinstimmungen von ≥ 90 Prozent.

Landis und Koch (1977) als beachtlich zu bezeichnen. Diese Übereinstimmung zwischen den Raterinnen zeigte zum einen, dass das Kategoriensystem anwendbar ist und funktioniert, und machte es zum anderen überflüssig, das gesamte Material von zwei Raterinnen zu kodieren.

9.4 Ergebnisse der qualitativen Analysen

Nachdem das Vorgehen bei den qualitativen Analysen ausführlich dargelegt wurde, werden im Folgenden die Ergebnisse dieser Analysen vorgestellt. Diese werden für die Gesamtstichprobe sowie nach Statusgruppen und nach Erklärtiefe gesondert interpretiert, damit wird analog zur Interpretation der quantitativen Analysen vorgegangen.

9.4.1 Statusgruppenübergreifende Ergebnisse der qualitativen Analysen

Begonnen wird mit den Antworten für die Gesamtstichprobe (Forschungsfrage 6a). Zu den Ergebnissen der qualitativen Analysen werden Abbildungen mit den Häufigkeiten der Nennungen aller vergebenen Kodes abgebildet. Es wird im Fließtext jedoch nicht auf jedes einzelne Ergebnis eingegangen, berichtet und interpretiert werden Auffälligkeiten in diesen Daten[72]. Weitere Analysen liegen vor, werden aber aufgrund der Lesbarkeit nicht dargelegt. Sie können bei der Verfasserin erfragt werden.

[72] Nicht berichtet werden die Kodierungen zu *Sprache* und *Sprechen*; hier sei auf die entsprechenden Publikationen verwiesen: Thim-Mabrey & Lindl (2021, in Vorb.) sowie Gunga (2022, in Vorb.). Auffällig sind jedoch insgesamt sehr wenige Antworten zu *Sprache*, obwohl der Sprachstil in den FALKE-D-Vignetten mit sehr langen und langen Sätzen sowie hochkomplexen und komplexen Gesamtsatzkonstruktionen auffällige bildungssprachliche Merkmale aufweist (s. Thim-Mabrey & Lindl, 2021, in Vorb. sowie Thim-Mabrey, in Vorb.).

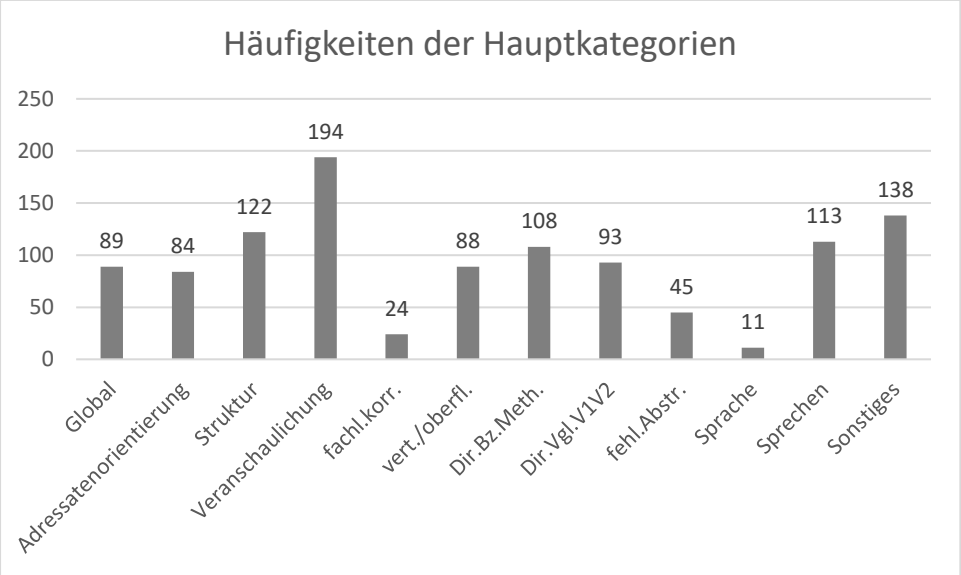


Abbildung 17: Häufigkeiten der Kodes für die Gesamtstichprobe
(dargestellt sind die absoluten Häufigkeiten, d. h. es wird angegeben, wie oft ein Kode im Gesamtdokument vergeben wurde)

Kode Veranschaulichung

Interessant an den offenen Antworten ist zunächst, dass die meisten Nennungen insgesamt die *Veranschaulichung* betreffen (s. Abb. 17). Wie oben beschrieben sind Veranschaulichungen ein bekanntes und anerkanntes Merkmal von Erklärqualität, das allerdings in FALKE-D nicht erhoben wurde. Die sehr zahlreichen offenen Antworten dazu zeigen, dass dieser Aspekt in der Beurteilung einer Erklärung sehr präsent ist. Betrachtet man die Antworten zu *Veranschaulichung* genauer, so fällt auf, dass sich die überwiegende Anzahl auf die Medien bezieht, die zur Veranschaulichung genutzt werden (s. Abb. 18, *Veranschaulichung Medial*). Unter diesem Subkode werden auf der einen Seite die in den Erklärungen genutzten Medien entweder als positiv hervorgehoben, zum Beispiel bei Video A „Sehr gut die Tafel", bei Video C „grafisch sehr schön dargestellt" oder bei Video F: „farbliche Markierungen gut"; oder die verwendeten Medien werden kritisiert, zum Beispiel Video D: „zu viel Text an der Tafel", Video E: „Folie unübersichtlich" oder Video F: „zu viele Farben". Einige Anmerkungen sind eher beschreibender denn wertender Natur: Video B „Visualisierung auf der Folie" oder Video F: „durch Farben markiert". Die überwiegenden Äußerungen zum Medieneinsatz sind positiv, indem sie die verwendeten Medien als hilfreich, sinnvoll oder gut bezeichnen (insgesamt wurden 49 eindeutig positive Anmerkungen dazu gezählt; sämtliche Auszählungen erfolgten per Hand). Kritisiert wurde 17 Mal, meistens weil der Medieneinsatz als unübersichtlich oder die Schrift/Folie als zu klein beschreiben wurde. Bei 18 Anmerkungen wird gar keine Wertung vorgenommen, sondern die Medien werden „nur" erwähnt (s. Beispiele oben).

Interessant ist in diesem Zusammenhang noch die Verteilung der Bemerkungen pro Video: Dabei zeigt sich ganz klar, dass in den Erklärungen, bei denen ein stärkerer Medieneinsatz stattfindet (das sind v.a. Video C mit der Tafelanschrift, die während des lauten Denkens entsteht, sowie Video F mit der Arbeit mit mehreren Farben auf der Folie), dieser Medieneinsatz sehr oft Erwähnung findet (nämlich 39 Mal bei Video C und 31 Mal Video F), häufig sogar als einzige Anmerkung. Gleiches gilt für den Subkode *Veranschaulichung Gestik,* bei dem nahezu alle Bemerkungen (nämlich 38 von 41) auf Video A fallen, bei dem die sehr anschauliche Gestik der Lehrkraft als positiv herausgestellt wird (zur Erinnerung: Video A ist die Erklärung, in der die Lehrkraft die etymologische Herkunft der Redewendung fast schauspielerisch inszeniert).
Bemerkenswert ist auf der anderen Seite bei diesem Kode *Veranschaulichung*, dass auch das Fehlen dieser an einigen Stellen angemerkt wird. Dies ist ein noch größeres Indiz für die Bedeutung, die die Studienteilnehmenden diesem Aspekt beimessen, als die Kommentierung des Medieneinsatzes in den einzelnen Erklärungen. Wird in den offenen Antworten darauf hingewiesen, dass eine Veranschaulichung hilfreich gewesen wäre (z. B. Video A: „ein illustrierendes Bild wäre hilfreich gewesen" oder Video F: „besser wäre eine Skizze gewesen, die den Gedankengang kurz erläutert"), so bedeutet das, dass die befragten Personen Veranschaulichungen als ein Merkmal der Erklärqualität im Kopf haben und sie (ohne in ihrer Einschätzung gelenkt worden zu sein) dieses heranziehen, um die Qualität einer Erklärung einzuschätzen.
Die anderen Subkodes spielen nur eine untergeordnete Rolle.

Kode *Struktur*
Ein weiterer Aspekt, der sehr häufig in den Begründungen zu den Globalurteilen angemerkt wurde, ist die *Strukturiertheit* der Erklärungen, kodiert unter *Struktur*. Auch dieser Kode wurde in Subkodes unterteilt, so wie unter Punkt *9.2 Vorgehen bei der Kategoriebildung* in dieser Arbeit beschrieben. Betrachtet man die vier Subkodes (*Global, Gegenstandsbestimmung, Schritt für Schritt* sowie *Fokussierung*) näher, so sticht lediglich die *Fokussierung* (mit 43 Kodierungen) ein wenig heraus, bei den anderen drei Subkodes lässt sich eine relativ gleichmäßige Verteilung der Antworten feststellen (22 bis 32 Antworten pro Kode). Die *Fokussierung* ist eine Unterdimension im Modell zur *Strukturiertheit* von FALKE, die aber aus oben dargelegten Gründen nicht in die Kriterien der quantitativen Untersuchung aufgenommen wurde (s. Ausführungen unter Punkt *4.2.1 Strukturiertheit* in dieser Arbeit). In der Auswertung der offenen Antworten wurde die *Fokussierung* aber als Subkode der *Strukturiertheit* untergeordnet.
Unter *Fokussierung* (vgl. Abb. 18) wurden alle Kommentare gefasst, die sich auf die Ausführlichkeit oder Knappheit beziehungsweise auf die inhaltliche Präzision beziehen. Das umfasst zum Beispiel folgende Bemerkungen für Video A: „präzise erklärt", für Video B: „genaue Erklärung" oder für

Video F: „Zu ausschweifende Erklärung“. Bei diesem Subkode überwiegen die negativen Kommentare, wobei vor allem eine unnötig lange Erklärung kritisiert wird (hier insbesondere bei Video D: „zu viel geredet“, „Zu viel um den heißen Brei geredet“ oder „hätte schneller gesagt werden können“) oder eine unstrukturierte Erklärung moniert wird (hier vor allem bei Video F: „unstrukturiert erklärt“, „Zu ausschweifende Erklärung“). Es sei der Vollständigkeit halber aber erwähnt, dass bei beiden Videos auch (vereinzelte) Kommentare zu finden sind, die genau das Gegenteil loben, zum Beispiel bei Video D „sehr präzise und knapp formuliert“ oder bei Video F „präzise Erklärung ohne Umwege“. Man sieht also auch in den offenen Antworten eine relativ hohe Uneinigkeit, wenn auch nicht ganz so stark wie bei den geschlossenen Antworten. Das liegt sicherlich daran, dass die offenen Antworten optional gegeben wurden, so dass die befragten Personen nur das anmerkten, was ihnen besonders auffiel. Das ist normalerweise das, was in den Videos besonders deutlich hervortritt.

Zu anderen Subkodes von *Struktur* sind noch folgende Auffälligkeiten zu berichten. Die Bemerkungen, die unter *Global* kodiert wurden, sind überwiegend positiv und lauten „gute Strukturierung“ (Video C), „strukturierte Erklärung“ (Video E) oder „Gut strukturiert“ (Video F). Nur drei Bemerkungen (der insgesamt 32) sind negativ und beziehen sich auf eine fehlende Struktur. Offensichtlich fällt den Studienteilnehmer/-innen also vor allem eine gelungene Struktur auf.

Unter *Gegenstandsbestimmung* fallen Bemerkungen, die das inhaltliche Ziel der Erklärung betreffen. Hier finden sich positive Kommentare (z. B. Video C: „Das Ziel der LK wird klar“ oder Video F: „erläutert zu Beginn worauf es ankommt >gut!“) und negative Anmerkungen (Video B: „Ziel der Erklärung?“ oder Video D:„eigentliches Sprichwort nicht im Fokus“) in etwa in gleicher Anzahl (12 positiven stehen 10 negative gegenüber). Auffällt hier eine Häufung der negativen Kommentare bei Video D. Dies liegt vermutlich an der Umsetzung der Erklärung: „Hier wird die Redewendung am Anfang von der Lehrerin nur an die Tafel geschrieben, allerdings nie ausgesprochen. […] Die ausbleibende Verbalisierung wird vor allem in Gruppe B (8x) [gemeint sind Studienteilnehmende mit Fragebogenversion B] sehr negativ bewertet, hier herrscht in manchen Fällen sogar Unsicherheit darüber, welche Redewendung genau erklärt werden soll“ (Häusl, 2019, S. 38). Interessanterweise fällt das den Teilnehmenden, die zuvor Video A gesehen hatten (also Studienteilnehmende an Fragebogenversion A), nicht auf, hier finden sich nur zwei negative Anmerkungen dazu. Sie kannten die Redewendung bereits aus der erst gesehenen Erklärung.

Für die Kodierungen unter *Schritt für Schritt* ist dagegen festzustellen, dass diese überwiegend positiv ausfallen. Unter diesen Subkode fallen alle Bemerkungen, die sich explizit auf den (schrittweisen) Aufbau der Erklärung beziehen: „Argument wird Schritt für Schritt durchgegangen“ (Video B), „Sehr gute schrittweise Entwicklung“ (Video C) oder „Prinzip/ Bedeutung am Ende zusammengefasst -> gut“ (Video D). Auch für diesen Subkode gilt also, dass Bemerkungen zum

schrittweisen Aufbau einer Erklärung dann gemacht werden, wenn dieser als gelungen empfunden wird.

Kode *Adressatenorientierung*

Zahlenmäßig geringer fallen die Kommentare aus, die unter die letzte Oberkategorie *Adressatenorientierung* zählen und unter die insgesamt 84 Bemerkungen kodiert wurden (s. Abb. 17). Davon entfallen in etwa gleich viele auf die Subkodes *Adressatenorientierung Global* (37, s. Abb. 18) beziehungsweise *Vorwissen* (34, s. Abb. 18). Zum ersten Subkode zählen positive (z. B. Video A: „schülerangepasst" oder Video C: „sehr schülernah") sowie negative Anmerkungen (z. B. Video B: „dürfte für schwächere Schüler schwerer nachvollziehbar sein" oder Video E: „aber Schülerinnen und Schüler können m.E. nicht folgen"). Auffällig ist die Häufung der negativen Kommentare zur *Adressatenorientierung Global* bei Video B und bei Video E, also den beiden Warum-Erklärungen, auf die insgesamt 15 der 25 negativen Kommentare fallen. Die Teilnehmenden nehmen diese Erklärungen als zu überfrachtet und als zu schnell wahr: „kein Achten darauf, ob die Schüler der Erklärung folgen können" (Video B) beziehungsweise „zu schnell angesichts der Komplexität" (Video E). Festzuhalten ist insgesamt zu diesem Subkode, dass die negativen Kommentare überwiegen (25 zu 12), was bedeutet, dass die *Adressatenorientierung* vor allem dann angemerkt wird, wenn sie bemängelt wird.

Beim Subkode *Vorwissen* sieht die Verteilung anders aus: Hier überwiegen die positiven Kommentare (22). Eine genaue Aufteilung nach Videos findet sich in der im Zuge der FALKE-D-Studie entstandenen, nicht veröffentlichten Zulassungsarbeit (Häusl, 2019). Sie stellt fest, dass bei Vignetten E und C ausschließlich positive Kommentare zum Vorwissen gemacht werden (vgl. Häusl, 2019, S. 30). Diese ergeben sich allerdings fast ausnahmslos durch eine Kontextuierung, die die erklärenden Lehrkräfte im Video simulieren. So sagt zum Beispiel die Lehrkraft in der einfachen Warum-Erklärung (Video E): „Das habe ich euch ja auch beigebracht…"(s. auch Transkript zu Vignette E, Anhang 6.4). Dies wird dann von den Studienteilnehmerinnen und -teilnehmern zum Anlass genommen zu kommentieren: „Bezug auf die vorhergehenden Lerninhalte" oder „Vor der Erklärung fand eine kurze Wiederholung statt". Ähnliches geschieht in der einfachen Was-Erklärung (Video D), in der die Lehrkraft die Schüler/-innen fragt, ob sie eine Redewendung kennen (s. auch Transkript zu Vignette D, Anhang 6.2). Dies wird ebenfalls von den befragten Personen aufgenommen und kommentiert: „fragt nach Vorwissen der SuS und küpft daran an". Diese Kommentare sind bemerkenswert, da sie zeigen, wie wichtig der Aspekt des *Vorwissens* für die befragten Teilnehmer/-innen ist: Es genügt ein kurzer Hinweis auf die Vorstunde zur positiven Kommentierung dieses Kriteriums. Dementsprechend findet sich bei diesen Videos auch gehäuft eine Aussage zum Vorwissen, gerade bei der einfachen Warum-Erklärung (Video E) wird der Bezug

zur Vorstunde/zum Vorwissen ganze zwanzig Mal positiv erwähnt. Anders herum gilt das allerdings nicht: „ Obwohl dem Bezug auf das Vorwissen in der fachdidaktischen Forschung große Bedeutung bezüglich des Anknüpfens und der Vernetzung von Wissen zugeschrieben wird, wird ein fehlender Bezug auf das Vorwissen nur selten als Begründung für eine schlechte Bewertung herangezogen“ (Häusl, 2019, S. 31).

Die übrigen Subkodes von Adressatenorientierung spielen nur eine untergeordnete Rolle.

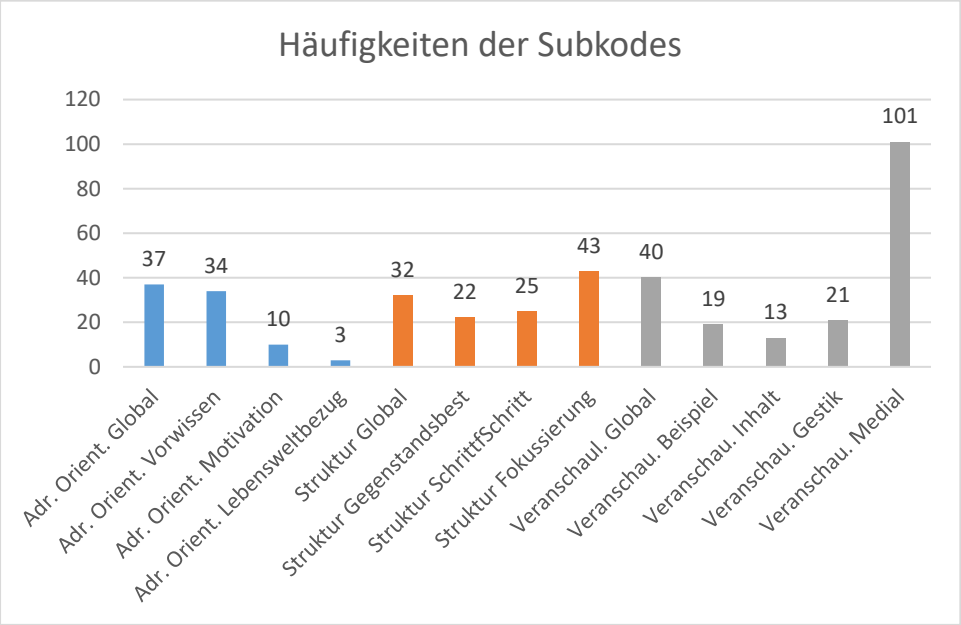


Abbildung 18: Häufigkeiten der Subkodes für die Gesamtstichprobe
(dargestellt sind die absoluten Häufigkeiten, d. h. es wird angegeben, wie oft ein Kode im Gesamtdokument vergeben wurde)

Zu den anderen Hauptkategorien wurden keine Subkodes gebildet. Hier ist die Verteilung folgendermaßen:

Kode *Sonstiges*

Obwohl das Kategoriensystem ausführlich diskutiert und in mehreren Schritten überarbeitet wurde (vgl. Punkt *9.2 Vorgehen bei der Kategorienbildung* in dieser Arbeit), blieb eine ganze Reihe an Bemerkungen „unkodierbar“, diese wurden unter die Kategorie *Sonstiges* zusammengefasst. Durch diese Kategorie war es wie gesagt möglich, den gesamten für FALKE-D vorliegenden Text zu kodieren. Bei der Bildung dieser Kategorie wurde schon erläutert, dass unter diesen Kode vor allem Verbesserungsvorschläge oder nicht zuordenbare Beschreibungen gefasst wurden. Betrachtet man die Verteilung auf die einzelnen Videos, so stellt man fest, dass relativ wenige Bemerkungen auf die

Was-Erklärungen fallen (nur 20 der 138); die anderen Bemerkungen, die sich nur mit *Sonstiges* kodieren lassen, fallen auf die Warum- und Wie-Erklärungen.

Kodes *Direkter Bezug zur Methode, Direkter Vergleich Variante 1 und 2* sowie *Fehlende Abstraktion in Bezug auf Designvorgaben*

Berichtet wird als nächstes über die Kategorien, die alle aufgrund der konkreten Gestaltung der Erklärungen und der Erhebung entstanden sind (s. hierzu Erläuterungen oben). Unter *Direkter Bezug zur Methode* wurden wie erläutert alle Bemerkungen zusammengefasst, die die in der Vignette angewandte Erklärweise kommentieren. Dabei werden Methoden, die klar zu erkennen sind (z. B. das laute Denken in Video C) oder als bekannt wahrgenommen werden (z. B. die Struktur eines Arguments in Video E) besonders oft erwähnt. Das laute Denken wird dabei oft „nur" benannt (z. B. „laut gedacht" oder „lautes Denken als solches benannt") oder als positiv herausgestellt (z. B. „Strategie mit dem lauten Denken ist überzeugend" oder „lautes Denken vermitteln den SchülerInnen ein Muster für das Formulieren von Gegenargumenten"). Negative Kommentare zum lauten Denken finden sich gar nicht.

Bei der einfachen Warum-Erklärung, bei der auf die schulbekannte Struktur eines Arguments eingegangen wird, sind auch die meisten Bemerkungen positiv: „positiv: Struktur und Aufbau einer Argumentation erklärt, teils auch mit Verbesserungsvorschlägen" oder „Strukturierung des Arguments wird klar". Hier finden sich vereinzelt auch negative Anmerkungen, zum Beispiel „Allerdings sind die Ausführungen sehr auf formale Aspekte fokussiert (ist die richtige Abfolge eingehalten, entspricht es dem was wir in der Schule gelernt haben …). Die Funktionalität des Textabschnittes wird kaum thematisiert: Es wird lediglich darauf verwiesen, dass die Behauptung zu allgemein ist, die Begründung nicht zur Behauptung passt usw. Begründungen dafür werden - über die schulische Normierung hinaus - den Schüler/-innen nicht geliefert". Die negativen Kommentare sind jedoch deutlich seltener (nur drei von insgesamt 20).

Auch bei den Was-Erklärungen finden sich zahlreiche Bemerkungen zu den Methoden etymologischer Hintergrund und synonyme Ausdrücke. Bei ersterem sind viele (sieben von 19) beschreibender Natur (z. B. „Herkunft des Ausdrucks erklärt"), nur einmal wird etwas Negatives angemerkt, die übrigen sind positive Kommentare (z. B. „gut war der Rückgriff auf die Vergangenheit, um die Bedeutung der Redewendung noch deutlicher zu machen"). Bei der einfachen Was-Erklärung stellt sich das etwas anders dar, das Nennen synonymer Ausdrücke wird öfter kritisiert (nämlich 10 von 21): „Beispielschlacht" oder „umständlich erklärt: Versuch Redewendung mit anderer Redewendung zu erklären". Die übrigen Kommentare sind positiv oder beschreibender Natur. Zu den anderen Vignetten (nämlich Video B, vertiefte Warum-Erklärung

(Funktional-inhaltliche Begründung) sowie Video F, einfache Wie-Erklärung (Modelltext)) finden sich weniger Kommentare zur Methode.

Unter den Kode *Direkter Vergleich Variante 1 und 2* wurden all diejenigen Bemerkungen kodiert, die beim Ansehen der zweiten Variante direkt auf die erste Variante Bezug nahmen. Auch wenn von den Teilnehmenden kein Vergleich zwischen den beiden Vignetten zum gleichen Thema verlangt wurde, so nahmen sie diesen doch häufig intuitiv vor. Dabei verglichen sie zum überwiegenden Teil die beiden Varianten (z. B. „besser als die erste Erklärung"). Interessant sind zwei Fälle, bei denen die befragte Person die Bewertung, die sie auf die zuvor gesehene Vignette vergeben hatte, korrigierte (z. B. „auch hier würde ich bei der 1. Erklärung im Vergleich jetzt auf 3 gehen"). Das bedeutet, dass die Teilnehmer/-innen nicht nur die Erklärungen zum gleichen Thema einander gegenüberstellen, sondern in Einzelfällen auch das eigene Bewertungsverhalten reflektieren und korrigieren (und zwar „nur" aufgrund des Ansehens einer anderen Erklärung zum selben Thema).

Schließlich finden sich unter dem Kode *Fehlende Abstraktion in Bezug auf Designvorgaben* all diejenigen Kommentare, die die konkrete Umsetzung der Erklärungen in den Vignetten kritisieren, und zwar obwohl die Vorgaben zur Umsetzung im Fragebogen selbst erläutert und begründet werden. Dies bezieht sich meistens auf die (aus forschungstheoretischen Gründen notwendige) monologische Umsetzung der Erklärungen: „Schüler nicht wirklich einbezogen" oder „deduktiv und wenig induktiv". Bemerkungen dieser Art finden sich bei allen sechs Vignetten. Sie zeigen, dass einige Teilnehmer/-innen die Erklärungen als so unnatürlich wahrnahmen, dass sie das anmerken wollten, trotz der geschilderten Erläuterungen dazu.

Insgesamt ist bei allen soeben beschriebenen Kategorisierungen erneut klar zu erkennen, dass die Teilnehmenden das anmerken, was ihnen in den Videos besonders auffällt, entweder weil es ihnen bekannt ist oder weil die Methode ungewöhnlich ist und daher „ins Auge sticht."

Kode *Vertieft/oberflächlich*

Ähnlich viele Kommentare wie bei den gerade erläuterten Kodes, nämlich 88, bezogen sich auf die in den Vignetten umgesetzte Erklärtiefe und wurden daher mit dem Kode *Vertieft/oberflächlich* kategorisiert. Interessant war für FALKE-D besonders, welche der Begründungen sich auf das spezifische Forschungsinteresse der Erklärtiefe beziehen, ohne dass die befragten Personen das Design oder die Kriterien kannten. Die Tatsache, dass viele Bemerkungen dazu gemacht wurden, zeigt, dass die Erklärtiefe durchaus ein Kriterium ist, dass die beteiligten Personen an Erklärungen anlegen, wenn auch nicht so häufig wie zum Beispiel das Kriterium *Veranschaulichung* (s. Abb. 17). Interessant ist bei dieser Kategorie die Verteilung, es zeigt sich nämlich, dass sich positive Kommentare (d. h. in diesem Fall Kommentare, die die Tiefe der Erklärung hervorheben) und

negative Kommentare (d. h. in diesem Fall Kommentare, die die Erklärungen als zu oberflächlich beschreiben) vignettenweise verteilen: So fallen auf die vertiefte Was-Erklärung fast ausnahmslos positive Kommentare, die das Erklären des etymologischen Hintergrundes als Möglichkeit beschreiben, das Gelernte zu verknüpfen beziehungsweise zu behalten: „geschichtlicher Kontext sorgt für gutes Merken“ oder „Bild bleibt im Gedächtnis, kann mit der Redewendung verküpft werden“. Demgegenüber stehen nahezu ausnahmslos negative Kommentare (in oben beschriebenem Sinne) bei der einfachen Was-Erklärung, bei der vor allem moniert wird, dass die Redewendung nicht in einen Kontext eingebettet wird, was die Merkfähigkeit verringert: „Die Redewendung kann mental nicht verankert werden“ oder „Erklärung isoliert, kein Kontext“. Viele Bemerkungen fallen zudem noch auf die vertiefte Wie-Erklärung (Methode lautes Denken); zu dieser wiederum finden sich fast nur positive Kommentare, die sich vor allem auf die Nachvollziehbarkeit des Vorgehens der Lehrkraft in der Erklärung beziehen: „gut nachvollziehbar und bleibt meiner Meinung nach besser im Gedächtnis hängen“ oder „Gedanken sind nachvollziehbar erläutert“. Auf die anderen Vignetten entfallen weniger Kommentare zu dieser Kategorie.

Kode *Global*

Als letzte Kategorie, auf die viele Kodierungen entfallen, ist der Kode *Global* zu nennen. Wie oben dargelegt wurde diese Kategorie induktiv erarbeitet und mehrmals überarbeitet. Darunter fallen Bemerkungen, die keinen Aspekt der Erklärung explizit erwähnen, sondern ein Gesamturteil abgeben, dass sich auf die Qualität der Erklärung insgesamt bezieht. Dabei ist bemerkenswert, dass nur zehn (der 89) Kodierungen negativ sind, der überwiegende Teil der Bemerkungen also positiv ausfällt. Sehr viele der Kommentare bestehen dabei nur aus einem Wort („klar“, „plausibel“) oder aus wenigen Stichworten („sehr gelungene Erklärung“, „klar und deutlich“).

Kode *Fachlich korrekt*

Der letzte Kode *Fachlich korrekt* umfasst deutlich weniger Aussagen als die bisher genannten (24, s. Abb. 17). Hier sind einige Bemerkungen zu finden, die die fachliche Korrektheit bestätigen (z. B. „fachlich sehr gut“ oder „völlig korrekt“). Es finden sich aber zu den Vignetten auch einzelne Kommentare, die den Inhalt der Vignetten als fachlich falsch kritisieren (z. B. Video D, einfache Was-Erklärung: „Wesentliches wird nicht thematisiert“ oder Video B, vertiefte Warum-Erklärung: „das eigentliche Argument ist nicht deshalb wenig überzeugend, weil es eine Hierarchie vergisst“). Bei der Erläuterung der theoretischen Vorarbeit wurde bereits geschildert, dass und wie die Inhalte der Vignetten fachlich abgesichert wurden. Dass es dennoch Widerspruch dazu gibt, ist der Natur des Faches Deutsch mit seinen typischen divergenten Lösungen geschuldet und deswegen nichts

Schlechtes. Die Kommentare, die die fachliche Korrektheit einer der Erklärungen in Frage stellen, sind zudem zahlenmäßig nur sehr wenige (insgesamt 11), so dass sie vernachlässigt werden können. Interessant ist im Folgenden der Blick in die Statusgruppen, mit dem dargelegt wird, wie sich die Kommentare auf die drei Expertenstatusgruppen verteilen (Forschungsfrage 6b). In den folgenden vier Abbildungen (Abb. 19-22) werden die relativen Häufigkeiten abgebildet (s. auch die entsprechenden Fußnoten). Aufgrund der unterschiedlichen Teilnehmerzahlen der einzelnen Statusgruppen erlauben die relativen Häufigkeiten einen aussagekräftigen Vergleich.

9.4.2 Statusgruppenspezifische Ergebnisse der qualitativen Analysen

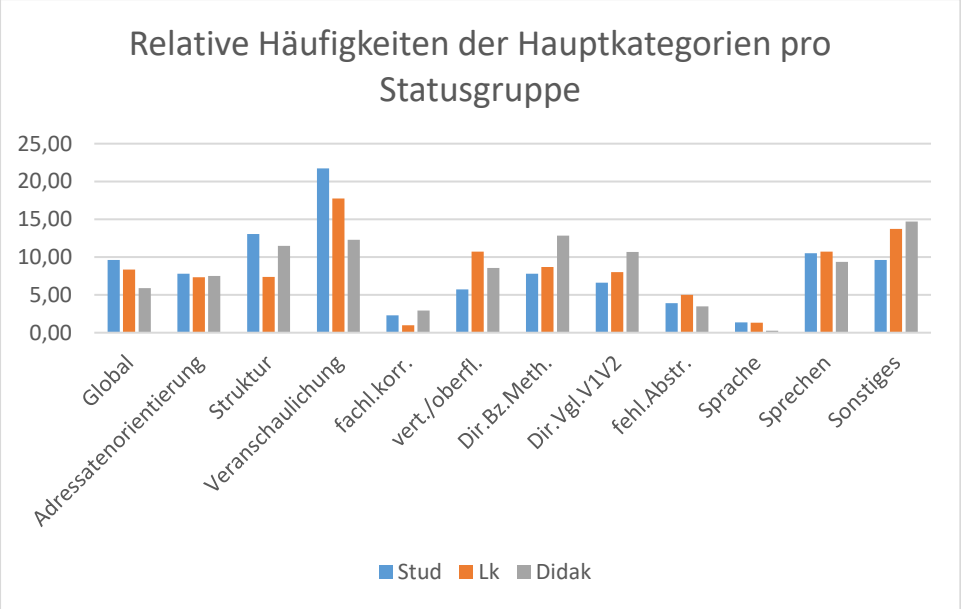


Abbildung 19: Quantitative Kontrastierung der Kodes nach Statusgruppen
(angegeben sind die relativen Häufigkeiten, d. h. pro Säule wird angegeben, wie oft der Kode in einer Statusgruppe prozentual zur Gesamtanzahl der in der Statusgruppe erteilten Kodes vergeben wurde. Somit ist der Vergleich zwischen den Säulen der Statusgruppen aussagekräftig.)

Analog zur Darlegung der quantitativen Ergebnisse werden auch die qualitativen Ergebnisse nach Statusgruppen getrennt betrachte. Berichtet werden wiederum nur Auffälligkeiten. Detaillierte Auswertungen können bei der Verfasserin angefragt werden.

Studierende

Als erstes werden die offenen Begründungen der Statusgruppe Studierende betrachtet, die mit $N_{max} = 78$ die größte Statusgruppe bildet. Analog zu den Ergebnissen der Gesamtstichprobe entfallen bei den Studierenden die meisten Bemerkungen auf die Kategorie *Veranschaulichung*.
Am häufigsten wird die Unterkategorie *Medial* vergeben, wobei hier vor allem bei Vignette C (vertiefe Wie-Erklärung) das Tafelbild erwähnt wird („das Tafelbild hat sehr gut gezeigt, wie die

Lehrerin denkt“ oder „Anschaulich durch Tafelbild“) und bei Vignette F (einfache Warum-Erklärung) die Arbeit an der Overhead-Folie („Gut mit Farben eingekreist“ oder „guter Farbeinsatz“) (vgl. Abb. 20). Darauf folgt zahlenmäßig die Unterkategorie *Global,* die gehäuft auf Vignette A (vertiefte Was-Erklärung) vergeben wird, mit elf der insgesamt 22 Kodierungen. Diese Vignette wird vor allem mit „anschaulich“ bewertet. Auffallend an den Bemerkungen zu diesem Kode ist, dass die Studierenden im Vergleich zu den Lehrkräften und Didaktikerinnen und Didaktikern besonders viele Bemerkungen zum Unterkode *Beispiel* machen (s. Abb. 20). Dabei sind die Bemerkungen alle entweder positiv („gute Beispiele“, „passende Beispiele“) oder beschreibend („an Beispielen erklärt“). Interessant ist, dass sich keine negative Anmerkung zu *Beispiel* finden lässt und dass die Studierenden für jede der Vignetten mindestens einmal diese Kategorie vergeben, das heißt also, dass diese Kategorie unabhängig vom Thema und von der Erklärweise herangezogen wird. Die anderen Unterkodes von *Veranschaulichung,* nämlich *Inhalt* und *Gestik* werden selten erteilt.

Am zweithäufigsten folgt auf *Veranschaulichung* die Kategorie *Struktur*, bei der die auffallend oft vertretene Unterkategorie *Fokussierung* im Vergleich zu den anderen Gruppen ins Auge sticht (vgl. Abb. 20), bei der 29 von insgesamt 43 auf die Statusgruppe der Studierenden entfällt. Sie merken hier entweder an, dass die Erklärung „präzise erklärt“ oder „kurz und bündig genug“ war, oder dass sie als zu ungeordnet wahrgenommen wurde („unstrukturiert erklärt“ oder „zu sprunghaft und durcheinander“). Bei der Verteilung auf die einzelnen Vignetten fallen eine Konzentration von negativen Kommentaren bei der Vignette D (einfache Was-Erklärung), die als zu ausschweifend beurteilt wird, und von positiven Kommentaren zur Vignette A (vertiefte Was-Erklärung) auf, die als präzise wahrgenommen wird. Die anderen Unterkategorien werden erheblich seltener genannt.

Die letzte Kategorie, für die Unterkategorien gebildet wurden, die *Adressatenorientierung,* wird weniger häufig vergeben, dafür in etwa gleich oft in den einzelnen Statusgruppen (s. Abb. 19). Die häufigsten Subkodes, die sich den Bemerkungen der Studierenden zuweisen lassen, sind *Global* und *Vorwissen*, mit je zwölf Kodierungen. Bei *Global* überwiegen negative Anmerkungen, die die Erklärungen als zum Beispiel „fern von den Schülern“ bezeichnen. Bei *Vorwissen* ist auffällig, dass zehn der zwölf Kodierungen auf die einfache Warum-Erklärung (Vignette E) fallen, bei der das Anknüpfen an das Vorwissen positiv erwähnt wird (zur Erinnerung: Dies liegt daran, dass der Lehrer in der Einleitung zu seiner Erklärung einen direkten Bezug zu Vorstunden anspricht).

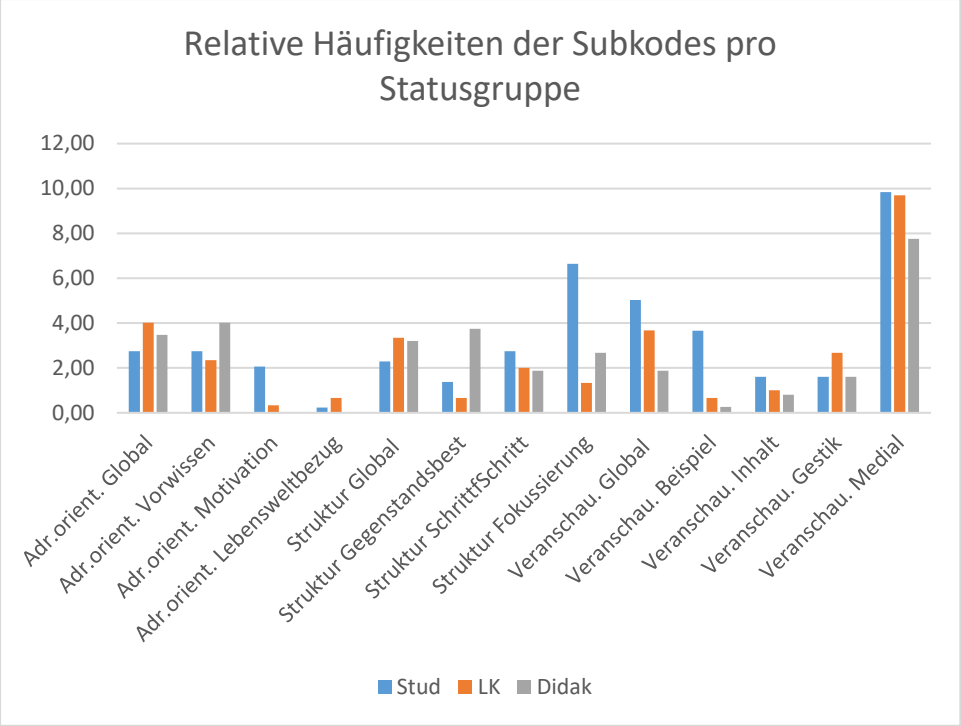


Abbildung 20: Quantitative Kontrastierung der Subkodes nach Statusgruppen
(angegeben sind die relativen Häufigkeiten, d. h. pro Säule wird angegeben, wie oft der Subkode in einer Statusgruppe prozentual zur Gesamtanzahl der in der Statusgruppe erteilten Kodes vergeben wurde. Somit ist der Vergleich zwischen den Säulen der Statusgruppen aussagekräftig.)

Bei den anderen Kategorien sieht die Verteilung für die Studierenden wie folgt aus: Ähnlich häufig vergeben werden die Kategorien *Global* und *Sonstiges*. Nahezu alle Anmerkungen zu *Global* sind positiv (es gibt nur zwei negative, die auf die vertiefte Wie-Erklärung (Video F) fallen), sie verteilen sich in etwa über alle Videos gleich. Die Kategorie *Sonstiges*, die wie dargelegt alles umfasst, was keiner anderen Kategorie zugeordnet werden kann, umfasst inhaltlich sehr unterschiedliche Kommentare. Die meisten davon (22 von 42) fallen interessanterweise auf die beiden Warum-Erklärungen.

Die Kategorien *Direkter Bezug zur Methode, Direkter Vergleich Variante 1 und 2* und *Vertieft/oberflächlich* weisen eine ähnliche Anzahl (zwischen 25 und 34) an Kodierungen auf. Es fällt auf, dass alle drei Kategorien von den Studierenden im Vergleich weniger oft vergeben werden als von den anderen Statusgruppen. Bei *Direkter Bezug zur Methode* bezieht sich interessanterweise fast die Hälfte der Kodierungen, nämlich 16 von 34 auf die Was-Erklärungen und damit auf die Methoden Erläutern des etymologischen Hintergrundes und Verwenden synonymer Ausdrücke. Nur vier Kommentare beziehen sich bei den Studierenden auf das laute Denken, obwohl das in der Gesamtgruppe eine Methode ist, auf die oft Bezug genommen wird (s. Ausführungen unter Punkt *9.4.1 Statusgruppenübergreifende Ergebnisse der qualitativen Analysen* in dieser Arbeit). Bei den Kodierungen unter *Direkter Vergleich Variante 1 und 2* ist zu berichten, dass die Vergleiche sich zu

großen Teilen auf die Was- und die Wie-Erklärungen beziehen. Zwischen den Vignetten dieser beiden Prozeduren nehmen die Studierenden also vor allem Vergleiche vor. Das könnte daran liegen, dass sie die Vignetten deutlicher voneinander unterscheiden als die beiden Warum-Erklärungen. Bei *Vertieft/oberflächlich* fällt auf, dass hier vor allem negative Kommentare bei beiden Warum-Erklärungen zu finden sind, die die fehlenden Anknüpfungsmöglichkeiten (z. B. „aus dem Kontext gerissen") monieren, während vor allem die vertiefte Was- und die vertiefte Wie-Erklärung als „nachvollziehbar" bezeichnet werden.

Auf die Kategorien *Fachlich korrekt* und *Fehlende Abstraktion in Bezug auf Designvorgaben* entfallen nur sehr wenige Kommentare. Die Bemerkungen zu *Fehlende Abstraktion in Bezug auf Designvorgaben* beziehen sich alle auf die fehlende Schülerbeteiligung, die wie bereits erläutert aus theoretischen Gründen notwendig war. Hier zeigt sich, dass die Studierenden trotz der vorgeschalteten Erläuterungen der notwendigen Designvorgaben diese aus dem Blick verlieren und als negativ anmerken. Da Bemerkungen dieser Art aber nur vereinzelt fallen, ist insgesamt davon auszugehen, dass die Studierenden die oben erläuterten Merkmale der Authentizität, nämlich Immersion und Resonanz, wahrnehmen (vgl. Ausführungen unter *7.1 Design der Studie* in dieser Arbeit sowie Friesen et al., 2018).

Es folgt der Vergleich mit den Bemerkungen der Lehrkräfte.

Lehrkräfte

Wie in den Abbildungen 17 und 18 zu erkennen ist, beziehen sich die Anmerkungen der Lehrkräfte ebenfalls am häufigsten auf den Kode *Veranschaulichung*, und hier auf die Unterkategorie *Medial.* Dabei wird erneut vor allem (mit 14 von insgesamt 29 Kodierungen) auf das Tafelbild Bezug genommen, das die Lehrkraft in Vignette C parallel zum lauten Denken erstellt, das mit Ausnahme einer Begründung positiv hervorgehoben wird als „anschaulich" oder „hilfreich". Abgesehen davon wird noch häufig der Folieneinsatz in Vignette F (einfache Wie-Erklärung) erwähnt, hier finden sich neun Kommentare, von denen sechs den Medieneinsatz positiv anmerken (z. B. „schöne Veranschaulichung mit Farbe und Visualisierung auf der Folie"). Drei Bemerkungen allerdings kritisieren den Medieneinsatz, zum Beispiel „unsaubere Folienarbeit (Vorbildfunktion)". Zu den anderen Vignetten finden sich nur vereinzelte Bemerkungen aus dieser Kategorie.

Die zweithäufigste Unterkategorie ist *Global* (auch hier antworten die Lehrkräfte, was die Häufigkeit anbelangt, genau wie die Studierenden). Hier findet sich vor allem die Begründung „anschaulich" ohne weitere Kommentierung. Dies ist für alle Vignetten in etwa gleich verteilt, nur bei Vignette B (vertiefte Warum-Erklärung) findet sich ein negativer Kommentar („keine Anschaulichkeit bei der Erklärung").

Ab dieser Kategorie unterscheiden sich die Verteilungen nach Häufigkeit von denen der Studierenden: Bei den Lehrkräften ist als nächstes die Unterkategorie *Gestik* zu finden. Hier ist sehr auffällig, dass sich alle Kodierungen auf die Was-Erklärungen beziehen und dort alle bis auf eine auf die vertiefte Vignette. In dieser ist der Einsatz der Körpersprache bereits als besonders auffällig beschrieben worden. Dies empfinden die Lehrkräfte ebenso. Die Begründungen sind hier alle positiv. Zu den verbleibenden Unterkategorien *Inhalt* und *Beispiel* finden sich nur noch vereinzelt Kodierungen bei den Lehrkräften.

Zu *Struktur* und *Adressatenorientierung* machen die Lehrkräfte nahezu gleich viele Anmerkungen. In den Unterkategorien zu *Struktur* zeigen sich keine nennenswerten Auffälligkeiten.

Für die *Adressatenorientierung* ist Folgendes festzustellen: Wie bei den Studierenden ist hier die vorherrschende Unterkategorie *Global*, worunter Anmerkungen fallen, die die Angemessenheit an die Zielgruppe allgemein betreffen. Auffallend sind hier die ausschließlich negativen Kommentare zu den beiden Warum-Erklärungen, die als nicht schülerorientiert kritisiert werden: „über die Köpfe der Schüler hinweg" (Vignette B) oder „-zu schnell angesichts der Komplexität - Schüler fehlt Zeit zum Mitdenken" (Vignette E). Auch bei den Kommentaren zu den anderen Vignetten fällt auf, dass diese hauptsächlich die Nichtangemessenheit der Erklärung an die Adressatengruppe kritisieren. *Adressatenorientierung* scheint also vor allem dann ein Bewertungskriterium für die Lehrkräfte zu sein, wenn sie sie als nicht angemessen empfinden. Die anderen Unterkategorien, nämlich *Vorwissen, Lebensweltbezug* und *Motivation,* werden nur in Ausnahmefällen angemerkt.

Als nächstes werden die Kategorien dargelegt, die nicht in Unterkategorien unterteilt wurden. Dabei ist als erstes die Kategorie *Sonstiges* bei den Lehrkräften zu nennen, darunter fallen Begründungen wie „zu theoretisch" (Vignette A) oder „zu Beginn verwirrend" (Vignette E). Bei dieser Kategorie zeigen sich erneut auffällige Häufungen bei den beiden Warum-Erklärungen (Vignetten B und E), auf die 22 der insgesamt 40 Kodierungen entfallen und die häufig negativer Natur sind („aber auch hier werden subjektive Floskeln eingeflochten" Vignette B oder „suggerieren eine dann auch tendenziell subjektive Auslegungsweise, was eine gute oder eine schlechte Argumenatation darstellt" Vignette E). Die Bemerkungen zu den anderen Vignetten verteilen sich in etwa gleich, auch hier gibt es neben einigen positiven Bemerkungen vor allem negative.

Die nächstzahlreiche Kategorie bei den Lehrkräften ist *Vertieft/oberflächlich,* die sehr gleichmäßig auf alle Vignetten verteilt ist (auf jede entfallen zwischen vier und sechs Bemerkungen). Sehr interessant ist bei dieser Kategorie, dass mit Ausnahme der Vignette C bei allen Vignetten ungefähr im Verhältnis halb-halb sowohl die Erklärtiefe gelobt wird (z. B. „Bezug zu vorhandenem Wissen wird aktiviert und dadurch erschließt sich das Ganze" Vignette E) als auch die fehlende Erklärtiefe kritisiert wird (z. B. „nur in Ansätzen hilfreich für Eigentätigkeit der Schüler bei der Planung einer eigenen Widerlegung" Vignette F). Einzig bei Vignette C (lautes Denken) sind alle fünf Bemerkungen

positiv. Das heißt aber insgesamt, dass keine Einigkeit unter den Lehrkräften darüber besteht, ob eine Erklärung genügend in die Tiefe geht oder nicht.

Bei der Kategorie *Direkter Bezug zur Methode* lässt sich folgende Verteilung berichten: Viele Bemerkungen fallen auf die beiden Was-Erklärungen (zwölf von 26). Dabei wird die Methode (etymologischer Hintergrund bzw. synonyme Ausdrücke) im Normalfall positiv erwähnt (z. B. „gut war der Rückgriff auf die Vergangenheit, um die Bedeutung der Redewendung noch deutlicher zu machen" Vignette A oder „Die Bedeutung wurde durch die Synonyme klar" Vignette D), lediglich bei letzterer finden sich auch drei kritische Anmerkungen (z. B. „nur anhand anderer Sprichwörter erklärt"). Immerhin sechs Bemerkungen finden sich für Vignette C zum lauten Denken, das positiv erwähnt wird. Auf die anderen Methoden wird weniger oft Bezug genommen.

Bei der Kategorie *Global* lassen sich keine Häufungen bei einzelnen Vignetten feststellen. Die Bemerkungen sind hier fast alle positiv („klare Aussagen" bei Vignette E oder „Die Erklärung ist gut verständlich" bei Vignette F). Lediglich zwei Kommentare finden sich, die negativ sind (z. B. „nicht immer ganz logisch" bei Vignette B).

Ähnlich häufig finden sich Bemerkungen, die unter die Kategorie *Direkter Vergleich Variante 1 und 2* fallen. Dabei ist bemerkenswert, dass im Gegensatz zu den Studierenden die meisten Vergleiche zwischen den beiden Warum-Erklärungen angestellt werden, wobei auch hier Uneinigkeit herrscht: Sowohl Variante 1 (Vignette B) als auch Variante 2 (Vignette E) werden als „besser" bezeichnet. Das zeigt, dass die Lehrkräfte zwischen den beiden Varianten einen deutlichen Unterschied bemerken, der sie zu diesen Vergleichen veranlasst. Zwischen den anderen beiden Vignettenpaaren finden auch Vergleiche statt, aber seltener.

Zu den beiden verbleibenden Kategorien *Fehlende Abstraktion in Bezug auf Designvorgaben* sowie *Fachlich korrekt* machen die Lehrkräfte nur noch sehr vereinzelt Kommentare.

Als letztes werden nun die Ergebnisse der offenen Begründungen der Didaktiker/-innen betrachtet.

Didaktiker/-innen

Analysiert man die offenen Antworten der Didaktiker/-innen, so fällt als erstes auf, dass sich ihre Begründungen gleichmäßiger auf die einzelnen Kategorien verteilen. Folgende sechs Kategorien werden ähnlich oft vergeben (vgl. Abb. 19): *Sonstiges, Direkter Bezug zur Methode, Veranschaulichungen, Struktur, Vergleich Variante 1 und 2* sowie *Vertieft/oberflächlich*. Das bedeutet, dass diese Statusgruppe unterschiedlichere Begründungen für die Globalnoten angibt als die anderen Statusgruppen.

Analog zu den Ausführungen zu den Ergebnissen der Studierenden und der Lehrkräfte wird auch hier mit den Kategorien begonnen, zu denen noch Unterkodes gebildet wurden, auch wenn diese, was die Häufigkeiten betrifft, nicht ganz vorne liegen.

Von diesen drei Kodes wird auch bei den Didaktikerinnen und Didaktikern der Kode *Veranschaulichung* am häufigsten vergeben. Erneut führt der Unterkode *Medial* diese Kategorie an, mit deutlichem Abstand zu den übrigen Unterkodes (s. Abb. 20). Auffällig ist hier, dass zu allen Vignetten (mit Ausnahme der Vignette D, einfache Was-Erklärung) Bemerkungen zum Einsatz der Medien gemacht werden: Dabei wird erneut vor allem die Tafelanschrift der vertieften Wie-Erklärung (Vignette C) lobend erwähnt (z. B. „- Visualisierung der zerlegten Erklärung = funktionaler Medieneinsatz“) und auch die Folienarbeit der einfachen Wie-Erklärung (Vignette F) wird positiv hervorgehoben (z. B. „zeigt dies im Text mit verschiedenen Farben“). Zu den anderen Vignetten finden sich positive und negative Kommentare in Bezug auf den Medieneinsatz.

Mit deutlichem Abstand folgen der Unterkode *Global* und fast gleichauf *Gestik*. Bei ersterem wird das Globalurteil „anschaulich“ vergeben (auf alle Vignetten außer auf D, einfache Was-Erklärung, die einmal mit der Bemerkung „schlecht veranschaulicht“ bewertet wird). Von den insgesamt sechs Kommentaren unter *Gestik* fallen fünf auf die vertiefte Was-Erklärung (s. oben) und eine auf die vertiefte Wie-Erklärung, alle Kommentare sind positiv. Zu den übrigen Unterkategorien *Inhalt* und *Beispiel* finden sich nur vereinzelte Anmerkungen.

Der nächste Kode, der betrachtet wird, ist die *Strukturiertheit*. Hier zeigt sich eine deutlich andere Verteilung als bei den anderen beiden Statusgruppen. Die Didaktiker/-innen machen die häufigsten Bemerkungen zur Unterkategorie *Gegenstandsbestimmung*, wobei in etwa gleich viele Anmerkungen auf die einzelnen Vignetten entfallen. Die meisten dieser Bemerkungen sind positiv (z. B. „- Das Ziel der LK wird klar“ (Vignette C) oder „Nennt bereits zu Beginn die Zielsetzung: Was soll im folgenden gelernt werden?“ (Vignette F)), lediglich bei Vignette D (einfache Was-Erklärung) finden sich drei negative Kommentare. Die nächsthäufige Kategorie ist *Global*, dabei entfallen die meisten Kommentare wie „gut strukturiert“ auf die vertiefte Was-Erklärung. Zu den anderen Vignetten finden sich hier nur vereinzelt Anmerkungen, die mit einer Ausnahme bei Vignette E (einfache Warum-Erklärung: „fehlende Struktur“) alle positiv ausfallen. Die Kommentare zu *Fokussierung* verteilen sich wiederum relativ gleichmäßig auf die Vignetten und sind bei den Was- und Wie-Erklärungen positiv: „Konzentration auf wichtige Aspekte“ (Vignette A) oder „keine Überfrachtung mit Information“ (Vignette F). Bei den Warum-Erklärungen sind die Kommentare dagegen negativ: „Wortschwall, Struktur der Argumentation wird nicht klar“ (Vignette B) oder „überlädt die Schüler mit Information“ (Vignette E).

Auf die *Adressatenorientierung* entfallen insgesamt weniger Kommentare, und dabei nur auf die Unterkodes *Vorwissen* und *Global*. Für die anderen beiden Unterkategorien finden sich bei den Didaktikerinnen und Didaktikern keine Beispiele. Die Begründungen unter *Vorwissen* verteilen sich sehr deutlich auf zwei Vignetten: Die einfache Was-Erklärung, in der die Lehrkraft wie erläutert die Frage stellt „schon einmal gehört?“ (s. auch Transkript zu Vignette D, Anhang 6.2). Dies wird in

zahlreichen Begründungen als positiv herausgestellt (z. B. „an Vorwissen der Schüler anschließend"). Es finden sich aber vier (der sieben) Bemerkungen, die die synonymen Ausdrücke als möglicherweise problematisch für den Vorwissensbezug der Schüler/-innen ansehen: „wenn Schüler die Bedeutung der anderen Redewendungen nicht kennen, dann ist das keine Hilfestellung". Die andere Vignette, bei der Begründungen unter Bezugnahme auf *Vorwissen* gemacht werden, ist die einfache Warum-Erklärung, in der der Lehrer wie schon dargelegt den Bezug zu vorherigen Stunden explizit macht (s. auch Transkript zu Vignette E, Anhang 6.4). Auch dies wird positiv angemerkt: „Bezug auf vorhergehende Lerninhalte". Negative Kommentare finden sich nicht. Zu den anderen Vignetten gibt es kaum Anmerkungen. Bei der zweiten Unterkategorie, *Global*, ist das Bild gemischter. Sie beziehen sich auf die Vignetten B-E und sind sowohl positiv (z. B. „nimmt die Schüler an die Hand" Vignette C) als auch negativ (z. B. „insofern überkomplex" Vignette D).

Als nächstes werden die Kodes berichtet, die nicht weiter untergliedert wurden. Am häufigsten wurde der Kode *Sonstiges* vergeben, bei dem sich keine Häufung auf einzelne Vignetten findet. Wie bereits erläutert finden sich hier Anmerkungen, die keiner anderen Kategorie zugeordnet werden können. Zum Teil ist nicht klar, ob diese Anmerkungen positiv oder negativ zu verstehen sind, da sie vielmehr das Gesehene beschreiben (z. B. „Aufstellen von Fragen und Beantworten dieser" Vignette E oder „die Lehrerin zeigt hier etwas" Vignette F).

Fast genauso viele Begründungen fallen unter den Kode *Direkter Bezug zur Methode*. Auch hier finden sich zu allen Vignetten Anmerkungen, jedoch in unterschiedlicher Anzahl. Die meisten (nämlich 16 von 48) Begründungen werden zu Vignette C (lautes Denken) gemacht, gefolgt von Vignette D (synonyme Ausdrücke, mit elf Kodierungen). Dabei sind sämtliche Anmerkungen zum lauten Denken positiv (z. B. „- Schülern wird Vorgehen der Lehrkraft transparent gemacht (lautes Denken) - gelungene Entwicklung eines (kurzen) Schreibplans vom Ausgangstext ausgehend" oder „Modellierung eines expertenhaften Umgangs/ Prozesscharakter"). Ungewöhnlich hohe Einigkeit herrscht also bei den Didaktikerinnen und Didaktikern über die Tauglichkeit der Methode des lauten Denkens. Das Verwenden synonymer Ausdrücke wird unterschiedlich bewertet, sieben der insgesamt elf Anmerkungen fallen negativ aus: „so wurde sie nur durch andere Redewendungen paraphrasiert" oder „Beispielschlacht". Immerhin vier Mal wird aber genau diese Methode gelobt, zum Beispiel: „gut über Analogien gelöst". Hier sind sich also die Didaktiker/-innen eher uneins, ob die Methode geeignet ist oder nicht. Bei den Bemerkungen zum Erklären des etymologischen Hintergrunds fallen neben positiven Kommentaren auch eher beschreibende auf (z. B. „Stattdessen wird die heutige übertragene Bedeutung aus der (fast) wörtlichen historischen Bedeutung abgeleitet"). Die Bemerkungen zu den anderen Vignetten sind deutlich seltener, hier finden sich sowohl positive als auch negative Kommentare.

Ähnlich viele Bemerkungen wie zu *Direkter Bezug zur Methode* fallen auf *Direkter Vergleich Variante 1 und 2.* Dabei finden die meisten Vergleiche zwischen den Warum-Erklärungen und den Wie-Erklärungen (nämlich je 16 von insgesamt 40) statt. Die Didaktiker/-innen haben also bei diesen Prozeduren stärker das Bedürfnis, auf die zuerst gesehene Erklärung Bezug zu nehmen. Sie teilen somit die Vergleiche der Warum-Erklärungen mit den Lehrkräften, die Vergleiche der Wie-Erklärungen dagegen mit den Studierenden. Dabei finden allgemeine Vergleiche zwischen den Vignetten statt (z. B. „besser" oder „anschaulicher", beides Vignette B). Auch hier wird in einigen Begründungen das eigene Bewertungsverhalten revidiert („auch hier würde ich bei der 1. Erklärung im Vergleich jetzt auf 3 gehen", Vignette B). In Einzelfällen wird auch die Lehrkraft verglichen („L. nervöser", Vignette B) oder die Vergleichbarkeit der beiden Erklärungen in Frage gestellt („Problem: diese Erklärung ist nur schlecht mit der 1. vergleichbar, weil sie anderes fokussiert", Vignette C).
Als letztes Kriterium, das in etwa gleich viele Begründungen anführt, ist der Kode *Vertieft/oberflächlich* zu nennen. Hier finden sich Kommentare zu allen Vignetten, allerdings unterschiedlichen Inhaltes. Während die Begründungen zu den Vignetten C (lautes Denken) und F (Modelltext) alle die Erklärtiefe positiv hervorheben (z. B. „das hat wirklch Modellcharakter für die Textarbeit der Lernenden" Vignette C bzw. „Die Denkprozesse eines Experten werden dadurch nachvollziehbar und unterstützen die Schüler/-innen in der Konstruktion eigener Strategien" Vignette F), werden die Erklärungen in den Vignetten D (synonyme Ausdrücke) und E (strukturelle Begründung) als zu oberflächlich und zu wenig anschlussfähig kritisiert (z. B. „deren Vorwissen oder Mitdenken und deren eigene Bedeutungsfindung nicht angeregt werden" Vignette D bzw. „Begründungen, warum das Argument nicht so gefiel, reichen meiner Ansicht nach nicht aus, um die Schüler so anzuleiten, dass sie es besser machen können" Vignette F). Bei den verbleibenden Vignetten (also A etymologischer Hintergrund und B funktional-inhaltliche Begründung) finden sich positive wie negative Kommentare. Auch diese Ergebnisse sind interessant, da sie eine Einigkeit zwischen den Didaktikerinnen und Didaktikern andeuten: Sie nehmen eine hohe Erklärtiefe bei den Methoden lautes Denken und Modelltext an, dagegen eine eher oberflächliche Erklärung bei den Methoden synonyme Ausdrücke sowie strukturelle Begründung.
Die übrigen Kategorien werden in den Begründungen dieser Statusgruppe deutlich weniger genannt.

9.4.3 Ergebnisse der qualitativen Analysen unterschieden nach Erklärtiefe

Nachdem die offenen Antworten aufgeteilt nach den einzelnen Statusgruppen ausgewertet wurden, folgt als letztes noch die Aufteilung in die vertieften (Vignetten A-C) und einfachen (Vignetten D-F) Varianten (Forschungsfrage 6c). Auch dieses Vorgehen erfolgt analog zu den quantitativen Analysen (s. Punkte *8.2.4 Bedeutung der variierten Erklärtiefe für die Statusgruppen*

sowie *8.3.2.3 Ergebnisse der hierarchischen linearen Regressionsmodelle unterschieden nach Erklärtiefe* in dieser Arbeit) und dient dazu herauszufinden, ob und wie sich die Kriterien, die die Studienteilnehmer/-innen von sich aus an die Erklärungen anlegen, je nach umgesetzter Erklärtiefe unterscheiden. Diese Analysen erfolgen für die Gesamtstichprobe, die nochmalige Sortierung nach Statusgruppen wäre zu kleinschrittig. Dort, wo es Auffälligkeiten bezogen auf einzelne Statusgruppen gibt, werden diese berichtet. Da es sich um das gleiche Textmaterial wie bei den vorherigen Analysen handelt, das „nur" unter anderen Vorzeichen betrachtet wird, wird auf Zitate aus dem Textmaterial im Folgenden verzichtet. Zahlreiche Zitate wurden oben bereits als Beispiele angeführt, so dass sich zu viele Wiederholungen ergeben, wenn das in den folgenden Analysen auch geschehen würde.

Die Reihenfolge der Darlegung erfolgt nach Häufigkeiten. Begonnen wird mit den Hauptkategorien und ihren entsprechenden Unterkodes.

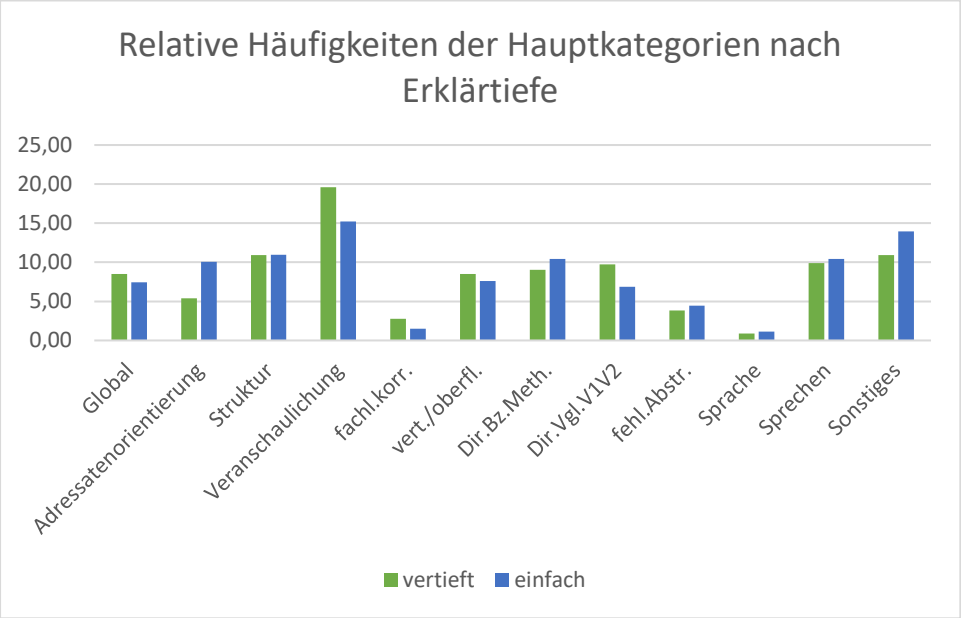


Abbildung 21: Quantitative Kontrastierung der Kodes nach Erklärtiefe
(angegeben sind die relativen Häufigkeiten, d. h. pro Säule wird angegeben, wie oft der Kode in der einfachen bzw. vertieften Variante prozentual zur Gesamtanzahl der in der Variante erteilten Kodes vergeben wurde. Somit ist der Vergleich zwischen den Säulen *vertieft* und *einfach* aussagekräftig.)

Kode *Veranschaulichung*

Wie Abbildung 21 zu entnehmen ist, fallen beim Vergleich der am häufigsten vergebenen Kategorie *Veranschaulichung* in Relation mehr Kommentare auf die vertieften Vignetten. Zerlegt man dies noch genauer nach den Unterkategorien, ergibt sich folgendes Bild (vgl. Abb. 22): Bei der Unterkategorie *Medial* werden bei den vertieften Vignetten mehr Kommentare als bei den einfachen gemacht. Bei ersteren beziehen sich, wie oben bereits herausgearbeitet, die Anmerkungen vor allem auf die Vignette C, in der das Tafelbild als positive Visualisierung

festgehalten wird. Bei den einfachen Varianten fallen die meisten Bemerkungen auf Vignette F, bei der die Arbeit mit der Folie positiv angemerkt wird.
Am zweithäufigsten wird der Unterkode *Global* vergeben, und zwar auch hier wieder öfter bei den vertieften Vignetten, bei der vor allem Vignette A mit dem Urteil „anschaulich" versehen wird. Bei den einfachen Varianten lässt sich keine Häufung feststellen, hier wird zu den drei Erklärungen das Urteil „anschaulich" erteilt. Interessant ist der nächsthäufige Subkode *Gestik*, bei dem die Kodierungen bei den vertieften Vignetten bei weitem überwiegen. Dies liegt vor allem, wie in den obigen Ausführungen bereits erläutert, an der sehr auffälligen Gestik, die in Vignette A eingesetzt wird. Auf diese entfallen nahezu alle Bemerkungen (18 von insgesamt 19), während bei den einfachen nur sehr vereinzelte Bemerkungen zur Gestik gemacht werden.

Kode *Adressatenorientierung*

Die zweithäufigste Kategorie, die im Folgenden betrachtet wird, ist die *Adressatenorientierung*, bei der der häufigste Subkode *Vorwissen* ist, mit weitem Vorsprung bei den einfachen Varianten (vgl. Abb. 22). Dabei beziehen sich die Anmerkungen alle auf Vignette D, in der die Lehrkraft zu Beginn fragt, ob jemand die Redewendung schon einmal gehört hat (s. oben) und auf Vignette E, in der die Lehrkraft fiktiven Bezug zu den vorherigen Stunden herstellt und mit einer „Wiederholung" einsteigt. Keine Anmerkungen finden sich zur dritten einfachen Vignette F und nur sehr vereinzelt Anmerkungen zu den drei vertieften Versionen. Bei diesem Vergleich wird noch einmal deutlich, dass das Vorwissen nur dann aufgeführt wird, wenn es explizit erwähnt wird. Der folgende Unterkode ist *Global*, der wiederum bei den vertieften etwas häufiger vorkommt. Hier verteilt er sich auf die drei Vignetten, wobei auffällt, dass sowohl bei der vertieften Warum-Erklärung die fehlende Adressatenorientierung global angemerkt wird, bei der vertieften Wie-Erklärung dagegen die Schülerorientierung gelobt wird. Bei der Was-Erklärung finden sich Kommentare in beide Richtungen. Bei den einfachen Varianten finden sich vor allem Kommentare zur Was-Erklärung und zur Warum-Erklärung. Bei beiden wird vor allem fehlende Schülerorientierung bemängelt, zum Beispiel aufgrund des zu schnellen Tempos.

Kode *Struktur*

Bei der letzten Hauptkategorie *Struktur* sieht die Verteilung gleichmäßiger aus, leicht angeführt wird sie von der Unterkategorie *Fokussierung*, und zwar etwas mehr bei den einfachen Varianten, bei denen auffällig ist, dass die Kommentare fast alle negativ sind (s. o.), also mangelnde Fokussierung feststellen, und zwar bei allen drei einfachen Vignetten. Bei den vertieften Vignetten ist das Bild differenzierter: Hier finden sich bei der Was- und der Wie- Erklärung positive und negative Kommentare, bei der Warum-Erklärung vor allem negative.

Es folgt erneut die Unterkategorie *Global*, die bei beiden Varianten nahezu gleich oft vergeben wird. Bei den einfachen Varianten werden vor allem die Warum- und die Wie-Erklärung als „strukturiert" bezeichnet, die Was-Erklärung deutlich seltener. Bei den vertieften Varianten werden dagegen die Was- und die Wie-Erklärung als „strukturiert" bewertet, die Warum-Erklärung dagegen kaum. Beim Unterkode *Schritt für Schritt* fällt eine klare Mehrheit auf die vertieften Varianten, wobei die Kommentare fast alle positiv sind und sich relativ gleichmäßig auf die drei Vignetten verteilen. Bei den einfachen Erklärungen sind die Bemerkungen auch positiv, verteilen sich aber vor allem auf die Warum- und die Wie-Erklärung. Der letzte Unterkode *Gegenstandsbestimmung* findet sich wieder etwas häufiger bei den einfachen Vignetten. Dabei sind die Bemerkungen zur Was-Erklärung vor allem negativ, weil kritisiert wird, dass nicht klar sei, welche Redewendung eigentlich erklärt werde. Die anderen Bemerkungen sind in der Mehrzahl positiv, wobei sich die meisten auf die Wie-Erklärung beziehen. Bei den vertieften Vignetten sind die Kommentare zur Was-Erklärung eher negativ, weil kritisiert wird, dass die Redewendung nur einmal genannt wird. Zu den anderen beiden Vignetten finden sich vereinzelt positive Kommentare.

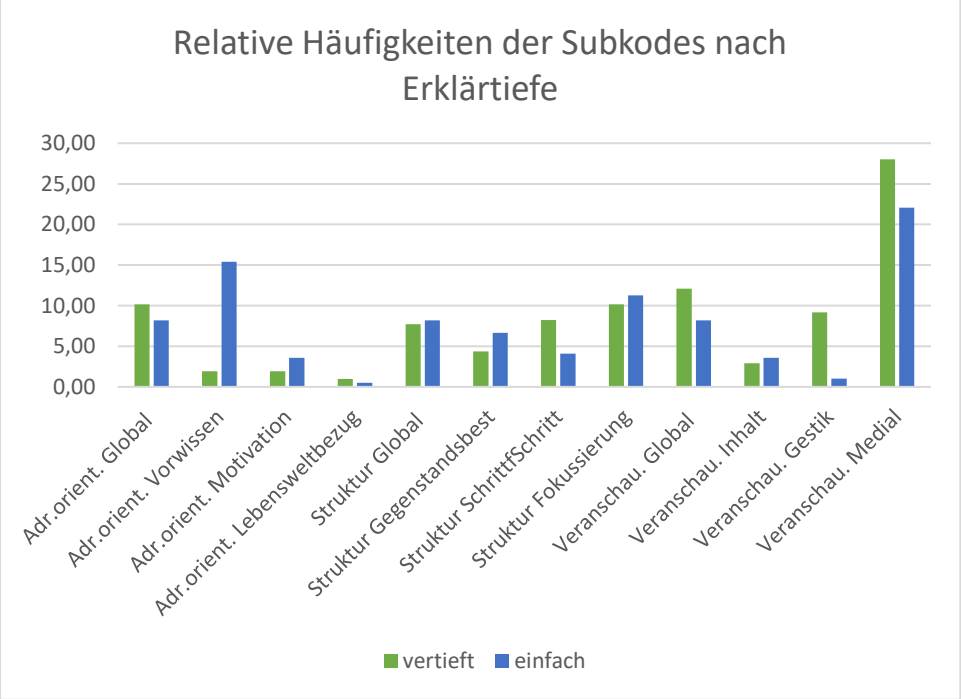


Abbildung 22: Quantitative Kontrastierung der Subkodes nach Erklärtiefe
(angegeben sind die relativen Häufigkeiten, d. h. pro Säule wird angegeben, wie oft der Kode in der einfachen bzw. vertieften Variante prozentual zur Gesamtanzahl der in der Variante erteilten Kodes vergeben wurde. Somit ist der Vergleich zwischen den Säulen vertieft und einfach aussagekräftig.)

Im Folgenden werden nun noch die Kodes erläutert, die keine weitere Unterteilung erfahren haben. Hier wird nach Relevanz der Ergebnisse vorgegangen.

Kode *Direkter Bezug zur Methode*

Der Kode *Direkter Bezug zur Methode* findet sich bei den einfachen Varianten ein wenig häufiger als bei den vertieften. Bei ersteren wird vor allem das Nennen synonymer Ausdrücke zur Redewendung negativ angemerkt. Bei der Warum- und der Wie-Erklärung finden sich vor allem positive Bemerkungen zu den verwendeten Methoden.

Bei den vertieften Vignetten finden sich vor allem Kommentare zur Was-Erklärung und zur Wie-Erklärung, die für beide Vignetten überwiegend positiv ausfallen.

Kode *Direkter Vergleich Variante 1 und 2*

Ähnlich viele Bemerkungen werden zum Unterkode *Direkter Vergleich Variante 1 und 2* gemacht, hier allerdings öfter bei den vertieften Varianten. Dieser Vergleich ist allerdings nur dahingehend aussagekräftig, dass die Studienteilnehmer/-innen öfter einen Vergleich zwischen den Vignetten anstrebten, wenn sie zuerst die einfache Variante gesehen hatten (nur dann konnten sie bei der vertieften Variante darauf zurückgreifen). Die Unterschiede sind aber nicht sehr groß (s. Abb. 21). Betrachtet man die Erklärprozeduren, so zeigt sich, dass mit 36 beziehungsweise 35 Bezugnahmen die Warum- und die Wie-Erklärungen ähnlich oft verglichen werden, die Was-Erklärungen mit insgesamt 22 Kommentaren etwas seltener.

Kode *Vertieft/oberflächlich*

Interessant ist unter diesem Blickwinkel natürlich besonders der Kode *Vertieft/oberflächlich*, der die umgesetzte Erklärtiefe selbst in den Blick nimmt. Es finden sich bei den vertieften Varianten hier etwas mehr Anmerkungen. Sehr deutlich ist hier die Verteilung der positiven Anmerkungen (z. B. Anknüpfungspunkte, gute Merkfähigkeit) bei den Vignetten A (Was-Erklärung) und C (Wie-Erklärung), insbesondere bei den Didaktikerinnen und Didaktikern. Die höhere umgesetzte Erklärtiefe wird von den Studienteilnehmenden hier erkannt und für gut befunden. Dagegen finden sich bei der Vignette B (Warum-Erklärung) nur wenige positive Kommentare. Die Mehrzahl der Kommentare zieht hier vielmehr die Merkfähigkeit und Anknüpfungsfähigkeit in Zweifel.

Bei den einfachen Vignetten sieht die Verteilung anders aus: Während nahezu alle Kommentare die Was-Erklärung als zu oberflächlich und als zu isoliert bezeichnen, finden sich bei der Warum-Erklärung sowohl Kommentare, die die Erklärtiefe loben als auch solche, die sie als fehlend kritisieren. Die Wie-Erklärung dagegen empfinden viele Studienteilnehmer/-innen als kognitiv aktivierend.

Kode *Fachlich korrekt*

Der Kode *Fachlich korrekt* wird nurmehr selten und bei den vertieften Varianten etwas häufiger als bei den einfachen vergeben. Bei den vertieften Vignetten beziehen sich die Kommentare vor allem auf die Was-Erklärung, bei der die fachliche Korrektheit gelobt, aber auch in Zweifel gezogen wird, und auf die Warum-Erklärung, die ebenso als fachlich sehr gut und als inhaltlich verwirrend bezeichnet wird.

Die wenigen Kommentare zu den einfachen Varianten beziehen sich gleichmäßig verteilt auf alle drei Vignetten. Hier wird vor allem die fachliche Korrektheit herausgestellt.

9.4.4 Zusammenfassung der qualitativen Ergebnisse

Gesamtstichprobe und Statusgruppen

Betrachtet man die Ergebnisse der qualitativen Analysen insgesamt, so zeigen die Antworten zunächst einmal Folgendes: „Personen unterschiedlicher Statusgruppen sind in der Lage, Erklärungen ohne Vorgaben von Kriterien zu bewerten […]" (Häusl, 2019, S. 43) Dabei beziehen sie sich auch auf Kriterien, die in der Lehr-/Lernforschung und in den Didaktiken als bedeutend für die Qualität einer Erklärung gelten. Zur Gewichtung der intuitiven Kriterien kann man Nachstehendes feststellen: Ein bedeutsames Kriterium, das die Studienteilnehmer/-innen von sich aus an die gesehenen Erklärungen anlegen, ist *Veranschaulichung* (vgl. Abb. 19 und 20). Auffällig ist, dass das Kriterium auch dann angemerkt wird, wenn es fehlt. Für alle drei Statusgruppen konnte überdies festgestellt werden, dass hier der Subkode *Medial* eine wichtige Rolle spielt. Dabei wird in FALKE-D vor allem die mediale Unterstützung in Vignette C (Tafelanschrift beim lauten Denken) und Vignette F (Folienarbeit mit unterschiedlichen Farben) positiv angemerkt. Auch hier ist aber bezeichnend, dass die Teilnehmenden ebenfalls anmerken, wenn ihnen eine mediale Veranschaulichung fehlt. Häufig wird außerdem unter dem Subkode *Gestik* die auffällige gestische und mimische Gestaltung in Vignette A angemerkt. Auch hier ist also festzuhalten, dass Besonderheiten in den Vignetten auch besonders häufig in den offenen Antworten angemerkt werden.

Beim Kode *Struktur* werden insgesamt mehr positive Kommentare gemacht, das heißt die Strukturiertheit einer Erklärung wird dann angemerkt, wenn sie als gelungen empfunden wird. Uneinigkeit herrschen bei den Subkodes *Fokussierung* und *Gegenstandsbestimmung*. Die mangelnde Fokussierung wird vor allem von den Studierenden bei der einfachen Was-Erklärung (Nennen synonymer Ausdrücke) und von den Didaktikerinnen und Didaktikern bei beiden Warum-Erklärungen kritisiert.

Bei der *Adressatenorientierung* verhält es sich genau gegenteilig. Diese wird vorrangig dann angemerkt, wenn sie als fehlend wahrgenommen wird. Dies geschieht in allen Statusgruppen

häufig unter dem Subkode *Global.* Davon abgesehen ist noch das Antwortverhalten für den Subkode *Vorwissen* interessant. Dieses wird vor allem von den Studierenden und den Didaktikerinnen und Didaktikern positiv bei Vignette E (einfache Warum-Erklärung) erwähnt, in der der Lehrer wie erläutert auf fiktive Vorstunden explizit Bezug nimmt. Das bloße Erwähnen eines Zusammenhangs mit bereits Gelerntem reicht also aus, den Vorwissensbezug als positiv wahrzunehmen.

Bei dem Kode *Direkter Bezug zur Methode* sind sich die Studienteilnehmer/-innen aller Statusgruppen neben zahlreichen beschreibenden Kommentaren einig, dass die Methode lautes Denken lernförderlich ist. Bei der Verwendung synonymer Ausdrücke herrscht große Einigkeit darüber, dass das keine geeignete Methode zum Erlernen einer Redewendung ist. Bei den anderen Methoden finden sich sowohl positive als auch negative Kommentare.

Bei der Kategorie *Direkter Vergleich Variante 1 und 2* verteilt sich das Antwortverhalten auf die einzelnen Prozeduren unterschiedlich je nach Statusgruppe: Während die Studierenden vor allem die beiden Was- und die beiden Wie-Erklärungen miteinander vergleichen, fällt bei den Lehrkräften der Vergleich vor allem auf die Warum-Erklärungen. Die Didaktiker/-innen wiederum vergleichen vor allem die beiden Warum- und die beiden Wie-Erklärungen. Offensichtlich ist es also so, dass den Statusgruppen hier zwischen den Varianten der gleichen Prozedur anderes auffällt, was sie zum direkten Vergleich anregt.

Bei der letzten häufig vergebenen Kategorie *Vertieft/oberflächlich* ist bemerkenswert, dass diese vor allem von Lehrkräften und Didaktiker/-innen vergeben wird. Dabei herrscht Einigkeit, dass die Methode lautes Denken vertiefte Denkprozesse auslöst, die Methode synonyme Ausdrücke für die Redewendung dagegen zu oberflächlich erscheint; das Bild passt also zu den Bemerkungen unter *Direkter Bezug zur Methode*. Bei den anderen Methoden herrscht Uneinigkeit.

Insgesamt ist zu sagen, dass die heterogenen Bewertungen der quantitativen Antworten sich in den offenen Antworten spiegeln, aber nicht so stark. Es ist bei den offenen Antworten zumindest bei einzelnen Kodes eine stärkere Einigkeit festzustellen.

Einfache versus vertiefte Varianten

Betrachtet man nun die Ergebnisse, die sich aus der unterschiedlichen Umsetzung der Erklärtiefe ergeben, so bestätigt sich das Ergebnis der Analysen nach Statusgruppen: Angemerkt wird vor allem das, was besonders auffällig ist. Daher kommt die häufige Nennung von *Veranschaulichung Medial* und von *Veranschaulichung Gestik* bei den vertieften Varianten, die auf Vignette C mit der gelungenen Tafelanschrift und auf Vignette A und die darin enthaltene auffällige Gestik zurückgehen. Genauso erklärt sich die häufige Nennung von *Adressatenorientierung Vorwissen* bei den einfachen Vignetten, da diese in zwei Vignetten fiktiv hergestellt wird.

Bei dem Subkode *Fokussierung* (von *Struktur*) ist zu berichten, dass dieser vor allem bei den einfachen Varianten genannt wird, und zwar weil dort die Fokussierung nach Meinung der Studienteilnehmenden fehlt.
Schließlich könnte die häufigere Nennung von *Direkter Bezug zur Methode* bei den einfachen Varianten ein Hinweis darauf sein, dass die befragten Personen diese Methoden besser aus dem Schulalltag kennen als die Methoden der vertieften Varianten.
Die Kriterien, die die Studienteilnehmer/-innen von sich aus an die Erklärungen anlegen, entsprechen somit den in FALKE-D erhobenen Aspekten guten Erklärens, müssen allerdings um einige erweitert werden, nämlich um *Veranschaulichung* mit den genannten Subkodes sowie um *Fachlich korrekt* und oben beschriebene induktiv erstellte Subkodes zu *Strukturiertheit* und *Adressatenorientierung*. Zudem veranlasst die konkrete Umsetzung der Erklärungen in FALKE-D die befragten Personen zu Bemerkungen. Hierunter fallen die genannten Kategorien *Direkter Bezug zur Methode, Vergleich Variante 1 und 2* sowie *Fehlende Abstraktion in Bezug auf Designvorgaben*.
Hervorzuheben ist noch einmal, dass die *Erklärtiefe*, die in FALKE-D als fachspezifischer Prädiktor konzeptualisiert wurde, auch in den Begründungen der Studienteilnehmer/-innen spontan als Kriterium für die Bewertung von Erklärqualität herangezogen wird.

10 Diskussion der Ergebnisse zum Konstrukt des *Guten Erklärens*

Es folgt nun die Auswertung und Diskussion der Gesamtergebnisse zum Konstrukt des guten Erklärens, bei der die quantitativen mit den qualitativen Ergebnissen in Bezug gesetzt werden. In der Studie FALKE-D wurde erstmals das Konstrukt des guten Erklärens im Deutschunterricht konzeptualisiert und operationalisiert. Die Konzeptualisierung berücksichtigt fachübergreifende Aspekte, die sich im Gesamtprojekt FALKE ergaben, genauso wie den fachspezifischen Aspekt der *Erklärtiefe* und die Unterscheidung in die Erklärprozeduren *erklären, was/warum/wie*. Die Operationalisierung erfolgte über entsprechende Videovignetten und dazugehörige Items in einem Online-Fragebogen.

10.1 Zusammenfassung der quantitativen und der qualitativen Ergebnisse

Im Folgenden werden die soeben erläuterten Ergebnisse der quantitativen und qualitativen Analysen zusammengefasst, so dass sich dem Konstrukt *Gutes Erklären* aus beiden Forschungsperspektiven genähert wird und aus den Ergebnissen insgesamt Schlüsse gezogen werden können. Dabei interessieren die Fragestellungen, welche dieser Schlüsse für Erklärungen im Deutschunterricht verallgemeinerbar sind und wo die Grenzen der Reichweite der Studienergebnisse liegen.

10.1.1 Erklären als Teil der Lehrerkompetenz

Wie unter Punkt *1.2 Erkenntnisinteresse und Verortung* sowie unter Punkt *2 Aspekte professionellen (Deutsch-) Unterrichts* in dieser Arbeit bereits dargelegt wurde, wird das Erklären in FALKE-D als Teil der Lehrerkompetenz aufgefasst.

Insbesondere die Ergebnisse für die Gesamtstichprobe bei FALKE-D zeigen, dass die abgedeckte Gesamtvarianz der Erklärqualität für das FALKE-Modell des guten Erklärens hoch ist (vgl. Tab. 19). Hinzu kommt, dass für die Gesamtstichprobe nahezu alle Prädiktoren (mit Ausnahme der *Sprachlichen Verständlichkeit* und der Kontrollvariable *Persönlichkeitswirkung*) signifikant für die Qualität einer Erklärung sind (vgl. Tab. 19). Die abgefragten Kriterien sind somit nahezu alle bedeutsam für die Bewertung von Erklärqualität. Das bedeutet, dass die Konzeptualisierung des guten Erklärens in FALKE für die Studienteilnehmer/-innen relevante Aspekte des Erklärens, nämlich *Strukturiertheit, Adressatenorientierung* und *Sprech-/Körperausdruck* abdeckt. Dabei ist das deutschspezifische Kriterium der *Erklärtiefe* genauso wichtig wie die aus der Lehr-/Lernforschung bekannten Merkmale. Die Ergebnisse von FALKE-D zeigen, dass die fachspezifische Gestaltung der Erklärtiefe bei der Gestaltung von Erklärungen im Deutschunterricht mit bedacht werden muss.

Lehrkräfte, die diese Aspekte berücksichtigen und im Prozess des Erklärens tatsächlich verwirklichen, weisen vermutlich hohe Erklärkompetenz auf (s. hierzu auch Ausführungen unter Punkt *10.3 Grenzen der Studie FALKE-D* sowie Punkt *11 Fazit und Ausblick* in dieser Arbeit). Auch wenn diese Vermutung empirisch noch zu überprüfen ist, so bietet das Modell des guten Erklärens von FALKE-D doch eine theoretisch und (zumindest was die Wahrnehmung von Erklärqualität betrifft) auch empirisch gesicherte Grundlage der Erforschung von Erklärkompetenz von Lehrkräften. Damit leistet FALKE-D einen Baustein im weiten Feld der Professionsforschung von Lehrkräften.

10.1.2 Multiperspektivität: Statusgruppenübergreifende und statusgruppenspezifische Ergebnisse

Die Erfassung der Wahrnehmung von Erklärqualität aus der Sicht unterschiedlicher Statusgruppen erlaubt Schlussfolgerungen über die Wahrnehmung von Erklären insgesamt, aber auch über die spezifische Wahrnehmung der einzelnen Gruppen.

Statusgruppenübergreifende Ergebnisse der Aspektbewertungen

Die folgenden Ausführungen beziehen sich zunächst auf die Forschungsfragen, die die Gesamtstichprobe betreffen und daher die Wahrnehmung des Erklärens insgesamt betreffen. Zusammengefasst wird, wie die Erklärungen global und in ihren Aspekten bewertet werden (Forschungsfragen 1a und 1b), welchen Einfluss dabei die abgefragten Aspekte auf die Gesamtbewertung einnehmen (Forschungsfrage 3) und wie diese Bewertungen mit den offenen Begründungen, die nach dem Globalurteil gegeben werden konnten, zusammenhängen (Forschungsfrage 6a). Wie dargelegt zeigen sich bei der deskriptiven Analyse der Bewertungen durchgehend recht heterogene Urteile, und zwar sowohl bei den Bewertungen, die in geschlossenem Antwortformat gegeben wurden (also Globalurteile und die einzelnen Aspektbewertungen), als auch bei den offenen Begründungen des Globalurteils. Diese Heterogenität besteht in der Gesamtstichprobe mit allen Studienteilnehmenden ebenso wie innerhalb der vier einzelnen Statusgruppen, obwohl bei den offenen Begründungen innerhalb der Statusgruppen zum Teil eine höhere Einigkeit besteht (s. folgende Ausführungen zu den Statusgruppenspezifischen Ergebnissen*)*. Es herrscht somit in der Gesamterhebung von FALKE-D keine Einigkeit über die Qualität der gesehenen Erklärungen (vgl. Tab. 9-16).

Dennoch erweist sich das Modell des guten Erklärens insgesamt als tragfähig, wie die Regressionen zeigen, die nahezu allen Prädiktoren einen signifikanten Einfluss auf die Erklärqualität nachweisen, und das insgesamt eine hohe aufgeklärte Gesamtvarianz zeigt (s. gerade erläuterte Ausführungen sowie Tab. 9).

Statusgruppenübergreifende Ergebnisse der offenen Antworten

Stellt man nun dem Antwortverhalten der Gesamtstichprobe die offenen Begründungen in den Expertengruppen gegenüber, so ist dort, wie oben angedeutet, zum Teil eine höhere Homogenität festzustellen. Es ist zu vermuten, dass besonders Auffälliges in den Erklärungen von allen Statusgruppen bemerkt und oft in den offenen Begründungen erwähnt wird. Dies tritt in FALKE-D besonders deutlich bei der schauspielerischen Darstellung des etymologischen Hintergrundes in Vignette A und dem parallel zum lauten Denken entstehenden Tafelbild in Vignette C zu Tage. Beides sind klar hervortretende beziehungsweise ungewöhnliche Erklärweisen, was sich in den offenen Antworten stark niederschlägt. Interessant ist, dass beide in überwiegender Zahl positiv vermerkt werden. Diese beiden Vignetten (A und C) sind sowohl insgesamt als auch bei den Expertenstatusgruppen die am besten beziehungsweise mit am besten bewerteten Vignetten. Die

Methoden Erklären einer Redewendung mit etymologischem Hintergrund und die Methode lautes Denken beim Widerlegen eines Argumentes werden insgesamt also als gutes Erklären anerkannt. Außerdem finden sich die Aspekte, die in den Regressionen ein signifikantes Gewicht auf das Globalurteil aufweisen (d. h. *Strukturiertheit, Adressatenorientierung, Sprech-/Körperausdruck* und *Erklärtiefe*) auch in den offenen Antworten zumindest in ähnlicher Häufigkeit (vgl. Abb. 17). Diese Aspekte spielen also definitiv auch bei der intuitiven Bewertung einer Erklärung eine Rolle und bilden das Erklären gut ab. Es ist aber auch zu betonen, dass die befragten Personen noch andere, in FALKE-D nicht abgefragte Kriterien zur Bewertung von Erklärungen heranziehen, und zwar zum Teil deutlich häufiger als die abgefragten Kriterien (v. a. *Veranschaulichung*, wie schon mehrmals herausgearbeitet, s. Abb. 17 und 18).

Statusgruppenspezifische Ergebnisse der Aspektbewertungen

Der Blick in die Ergebnisse nach Statusgruppen getrennt erlaubt außerdem genauere Aussagen darüber, welches Merkmal in welcher Gruppe als bedeutsam erachtet wird. Zusätzlich können mit diesen Ergebnissen Aussagen über das Beurteilungsverhalten der einzelnen Gruppen und im Vergleich gemacht werden.

Bei den statusgruppenspezifischen Ergebnissen interessieren zunächst die Fragen, wie die einzelnen Statusgruppen der Schüler/-innen, Studierenden, Lehrkräfte und Didaktiker/-innen die Erklärungen global und in den abgefragten Aspekten bewerten, sowie insbesondere die Frage, ob und wo sich Einigkeit innerhalb und zwischen den Gruppen feststellen lässt (Forschungsfrage 2a und 2b). Dabei ist auch von Belang, welche Kriterien für welche Statusgruppe/-n einen entscheidenden Einfluss auf die Globalnote haben (Forschungsfrage 4). Diese Ergebnisse werden wiederum mit den offenen Antworten der einzelnen Statusgruppen verglichen (Forschungsfrage 6b).

Zunächst ist festzustellen, dass sich die Teilnehmer/-innen innerhalb derselben Statusgruppe auch nicht einig über die Qualität der gesehenen Erklärungen sind, wie die hohen Standardabweichungen und weiten Streuungen in den Globalurteilen und den Aspektbewertungen zeigen (s. Tab. 10 sowie Tab. 13-16 und Ausführungen oben). Übereinstimmungen in den Urteilen zwischen und innerhalb der Statusgruppen finden sich vor allem bei Auffälligkeiten (s. oben), die in den einzelnen Erklärungen besonders hervortreten.

Eine interessante Feststellung ist über das Beurteilungsverhalten der zwei Vignetten der je gleichen Prozedur zu machen. Die beiden Was-/ Warum- und Wie-Erklärungen werden mit steigender Expertise unterschiedlicher bewertet: Während sich die Bewertungen der Schüler/-innen zwischen der je einfachen und vertieften Variante ähnlich sind, werden sie von Statusgruppe zu Statusgruppe immer unterschiedlicher. Die Bewertungen der Didaktiker/-innen weisen die höchsten Unterschiede auf. Dies gilt sowohl für die Globalnoten als auch für die Aspektbewertungen (s. Tab. 10 und 16).

Statusgruppenspezifische Ergebnisse der offenen Antworten

Diese nach Expertisegrad zunehmende Differenzierung lässt sich zudem auch in den offenen Antworten feststellen: Während sich bei den Studierenden in auffälliger Weise und (weniger deutlich) bei den Lehrkräften die Bemerkungen auf bestimmte Merkmale konzentrieren (besonders auffällig ist das bei *Veranschaulichung*, s. Abb. 19), so verteilen sie sich bei den Didaktikerinnen und Didaktikern gleichmäßiger auf die unterschiedlichen Merkmale (s. die nahezu gleiche Anzahl an Nennungen von *Struktur, Veranschaulichung, Direkter Bezug zur Methode* sowie *Direkter Vergleich Variante 1 und 2*). Damit ist angedeutet, was aus der Expertiseforschung und der Forschung zu *professional vision* bekannt ist, nämlich, dass der Blick auf die Erklärungen (genau wie der Blick auf Unterricht) mit höherer Expertise differenzierter wird und Unterschiede deutlicher wahrgenommen werden (vgl. Jahn et al., 2014; Sherin et al., 2011b).

Auch die oben angedeutete Vermutung, dass Didaktiker/-innen sich in besonderem Maße auf die hinter den Erklärungen stehenden didaktischen Prinzipien konzentrieren, hat sich im Großen und Ganzen bestätigt; bei einzelnen Aspekten herrscht obendrein eine relativ hohe Einigkeit über deren Tauglichkeit, auffällig ist das zum Beispiel in den positiven offenen Antworten zur Methode lautes Denken (Vignette C) oder in den negativen offenen Antworten zur Erklärtiefe des Nennens synonymer Ausdrücke (Vignette D) (vgl. Punkt *7.4.1.1. Didaktiker/-innen* und Punkt *9.4.2 Statusgruppenspezifische Ergebnisse der qualitativen Analysen* in dieser Arbeit sowie die häufige Nennung unter *Direkter Bezug zur Methode* in Abb. 19).

Interferenzstatistische statusgruppenspezifische Ergebnisse

Während die Bewertungen mit „ansteigendem" Expertisegrad immer differenzierter werden, so steigt mit dem Expertisegrad interessanterweise auch die aufgeklärte Gesamtvarianz (vgl. Tab. 19-21). Wie oben dargelegt ist die Gesamtvarianz durchgehend erfreulich hoch, am höchsten jedoch bei den Didaktikerinnen und Didaktikern. Das Modell des *Guten Erklärens* von FALKE-D bildet offensichtlich Expertenvorstellungen von gutem Erklären am besten ab.

Es zeigt sich insgesamt genau wie für die Gesamtstichprobe ein heterogenes Bewertungsverhalten für jede der Statusgruppen. Dennoch bewerten die Gruppen anders, und zwar mit steigender Expertise sowohl in den geschlossenen Bewertungen als auch in den offenen Begründungen differenzierter.

Auch in den Regressionen ergibt sich für die einzelnen Statusgruppen ein gemischtes Bild. Genauso wenig wie sich die deskriptiven Ergebnisse zwischen den einzelnen Statusgruppen ähneln, genauso wenig entsprechen sich die signifikanten Prädiktoren für die Erklärqualität, die jeweils unterschiedlich und auch in verschiedener Anzahl für die einzelnen Statusgruppen relevant sind (s.

Tab. 20 und 21). Einzig der Aspekt der *Erklärtiefe* bildet hier eine Ausnahme, da er immerhin in drei Statusgruppen (nämlich Schüler/-innen, Lehrkräfte sowie Didaktiker/-innen) statistisch signifikant für das Globalurteil der Erklärungen wird. In den offenen Antworten finden sich denn auch Begründungen, die auf die *Erklärtiefe* (Kode *Vertieft/oberflächlich*) Bezug nehmen, bei den Lehrkräften und Didaktikerinnen und Didaktikern häufiger als bei den Studierenden (s. Abb. 19). Diese Ergebnisse unterstreichen noch einmal die Bedeutung des speziell in FALKE-D konzipierten Aspektes der *Erklärtiefe* sowie die Tatsache, dass eine fachspezifische Ausgestaltung von Erklärungen im Deutschunterricht notwendig ist.

Bei den Schülerinnen und Schülern hat vor allem der *Sprech-/Körperausdruck* einen signifikanten Einfluss auf die Gesamtbewertung einer Erklärung, was die Vermutung aus den deskriptiven Analysen, nämlich dass die Performanz der Erklärung bei den Schülerinnen und Schülern entscheidend für die Bewertung ist, noch einmal bestätigt. Auch die recht ähnlichen Bewertungen innerhalb der gleichen Prozedur weisen in diese Richtung, da jeweils die/der gleiche Sprecher/-in beide Erklärungen einer Prozedur gesprochen hat. Somit bezeugen die oben bereits erwähnten ähnlichen Bewertungen der beiden Erklärungen derselben Prozedur bei den Schülerinnen und Schülern wahrscheinlich nicht nur eine weniger differenzierte Wahrnehmung der Erklärungen, sondern auch, dass die Schüler/-innen sich in ihren Bewertungen stärker von der Person der/des Erklärenden lenken lassen. So ist es zum Beispiel denkbar, dass die deutlich besseren Bewertungen der beiden Warum-Erklärungen der Schüler/-innen im Vergleich zu allen anderen Statusgruppen auch auf die Person der Lehrkraft zurückzuführen sind. Es wird wie oben bereits angedeutet vermutet, dass die erklärende Person, ein männlicher Lehrer aus dem aktiven Schuldienst, als typisch und klar wahrgenommen wird, was die Schüler/-innen mit guten Bewertungen honorieren (s. auch Punkt *7.4.2 Adressaten: Schüler/-innen* in dieser Arbeit). Abgesehen von dieser Besonderheit im Antwortverhalten der Schüler/-innen ist noch zu erwähnen, dass sie im Mittel sowohl die Erklärungen global als auch die einzelnen Aspekte etwas strenger bewerten, die Didaktiker/-innen hingegen etwas milder als die anderen Statusgruppen.

Abschließender Vergleich des Bewertungsverhaltens der Statusgruppen

Zwischen einzelnen Statusgruppen zeigen sich keine nennenswerten Ähnlichkeiten, es ist also weder so, dass zum Beispiel Schüler/-innen und Lehrkräfte aufgrund des gemeinsamen Kontextes der Schule ähnlich bewerten, noch dass etwa Schüler/-innen und Studierende aufgrund ihres Alters und ihrer Rolle als (vornehmlich) Adressatinnen und Adressaten von Erklärungen ein ähnliches Bewertungsverhalten aufweisen. Oben geäußerte Vermutung, dass die Studierenden in den Beurteilungen vermutlich eher die Rolle von Adressatinnen und Adressaten einnehmen, bestätigt sich somit nicht (vgl. Punkt *7.4.1.3 Studierende* in dieser Arbeit). Die Antworten der Studierenden,

gerade auch die offenen Antworten, deuten vielmehr darauf hin, dass sie bereits eine Metaperspektive auf die Erklärqualität einnehmen. An den weniger spezifischen und präzisen Antworten zeigt sich aber ihr Novizenstatus.

Auch die Bewertungen zwischen Lehrkräften und Didaktikerinnen/Didaktikern ähneln sich nicht besonders, was man aufgrund der höheren Expertise eigentlich annehmen könnte. Ebenso wenig sind sich Didaktiker/-innen und Studierende einig, obwohl sie an den Universitäten beide mit den neusten Forschungsergebnissen vertraut sind. Es ist im Gegenteil nach den in FALKE-D vorliegenden Daten so, dass die Statusgruppen alle ein eigenes Bewertungsverhalten aufweisen, bei dem sich gerade in den offenen Antworten relativ deutlich die jeweilige „Brille", das heißt der berufliche Hintergrund sowie die damit verbundenen Erfahrungen und Überzeugungen der einzelnen Statusgruppen zeigt. Die Ergebnisse reihen sich in Studien zur Validität der Aussagen unterschiedlicher Statusgruppen ein. So wurde unter Punkt *2.4.4 Übereinstimmung der unterschiedlichen Perspektiven* in dieser Arbeit bereits dargelegt, dass zum Beispiel Lehrkräfte die Lernziele einer Stunde gut bewerten, während externe Beobachter/-innen vor allem didaktische Prinzipien beurteilen. In FALKE-D finden sich in den offenen Antworten der Lehrkräfte tatsächlich einige, die die „Tauglichkeit" der Erklärung (z. B. aufgrund von *Adressatenorientierung* und *Vorwissensbezug*) beurteilen und diese deutlich zutage treten lassen. Es ist somit deutlich, wie der eigene Kontext das Beurteilungsverhalten prägt. Die oben erwähnten Vermutungen von Erickson (2011), wonach eigene Überzeugungen das Beurteilungsverhalten lenken, zeigen sich in den offenen Antworten von FALKE-D damit bestätigt.

Interessant ist in diesem Zusammenhang vor allem das doch recht unterschiedliche Antwortverhalten der Schüler/-innen im Vergleich zu den anderen Gruppen: Sie sind die eigentlichen Adressatinnen und Adressaten der Erklärungen, daher sollten ihre Bewertungen der Qualität einer Erklärung in Überlegungen zur Gestaltung mit einbezogen werden (s. oben erläuterte Forderung von Gindele & Voss, 2017). Dass bei den Schülerinnen und Schülern außer dem *Sprech- und Körperausdruck* nur die fachspezifisch konzeptualisierte *Erklärtiefe* einen signifikanten Einfluss auf das Gesamturteil aufweist, unterstreicht noch einmal die Bedeutung dieses Aspektes für die Gestaltung einer Erklärung im Deutschunterricht.

10.1.3 Erklären im Deutschunterricht

Spezifika des Erklärens im Fach Deutsch zu erarbeiten ist das zentrale Anliegen von FALKE-D. Zusammenfassend ist noch einmal festzuhalten, dass die für das Modell guten Erklärens in FALKE-D konzipierten Aspekte *Strukturiertheit, Adressatenorientierung, Sprech-/Körperausdruck* sowie *Erklärtiefe* als bedeutsam für die Qualität von Erklärungen im Deutschunterricht wahrgenommen werden. Durch die Analyse der offenen Antworten konnte insbesondere der Aspekt der

Veranschaulichung noch zusätzlich als Merkmal einer guten Erklärung bestätigt werden. Die Aspekte müssen bei der Gestaltung von Erklärungen im Deutschunterricht demnach berücksichtigt werden (s. Ausführungen oben). Besonders die fachspezifische Ausgestaltung des Aspektes *Erklärtiefe* erwies sich als zentral und wurde in FALKE-D erstmalig konzeptualisiert und operationalisiert. Rückschlüsse auf diesen Aspekt bieten auch die Ergebnisse, die sich durch die Variation der Erklärtiefe in der Umsetzung der beiden Vignetten einer Prozedur ergaben.

10.1.4 Bedeutung der Erklärtiefe für Erklärungen im Deutschunterricht

Als letztes erfolgt daher die Zusammenfassung und Diskussion der FALKE-D-spezifischen Variation nach Erklärtiefe; dabei interessiert, ob die einfachen oder die vertieften Varianten besser bewertet werden, welche Kriterien jeweils für diese Bewertungen ausschlaggebend sind und was in den Begründungen zum Globalurteil dazu genannt wird (Forschungsfragen 5 und 6c).

Die Ergebnisse der quantitativen Analysen ergeben, dass insgesamt sowohl bei den Globalurteilen als auch bei den Bewertungen der Skalen die vertiefte Was- und die vertiefte Wie-Erklärung besser bewertet werden, bei den Warum-Erklärungen dagegen die einfache Variante. Statusgruppenspezifische Ausnahmen bilden hier die Lehrkräfte, die auch bei den Wie-Erklärungen die einfache Variante bevorzugen sowie die Didaktiker/-innen, die bei allen drei Vignettenpaaren der vertieften Variante den Vorzug geben.

Theoriekonform werden also insgesamt die vertieften Varianten der Was- und Wie-Erklärungen besser bewertet als die einfachen Varianten. Dass gerade die Gruppe mit der höchsten Expertise (Didaktiker/-innen) alle drei vertieften Varianten bevorzugt, zeigt einmal mehr das vom spezifischen Hintergrund abhängige Beurteilungsverhalten: Wie angesprochen achten die Didaktiker/-innen vermutlich besonders auf die umgesetzten fachdidaktischen Prinzipien in den Erklärungen und somit auf die umgesetzte Erklärtiefe. Folglich sind sie die einzige Gruppe, die auch bei den Warum-Erklärungen der vertieften Variante (mit der funktional-inhaltlichen Begründung) den Vorzug gibt. Bedacht werden muss hier aber einerseits der sehr kleine Unterschied in den Bewertungen der Didaktiker/-innen.

Andererseits sind sich die Bewertungen der beiden Vignetten der Warum-Erklärungen sicherlich auch aufgrund der konkreten Realisierung der beiden Varianten insgesamt ähnlicher. Wie bereits angedeutet war es schwierig, die beiden Prinzipien (strukturelle vs. funktional-inhaltliche Begründung) klar erkennbar umzusetzen, so dass sie trotz wiederholtem Aufnehmen und der Auswahl der besten Vignetten nicht deutlich genug voneinander zu unterscheiden sind. Es finden sich tatsächlich in der strukturellen Begründung auch inhaltliche Gründe für das misslungene Argument sowie in der funktional-inhaltlichen Begründung auch Hinweise auf die wenig gelungene

Struktur des Arguments. Die Ergebnisse zu den Warum-Erklärungen sind dem sicherlich zumindest zum Teil geschuldet.

Einen statistisch signifikanten Einfluss auf die Globalnote haben unterschiedliche Prädiktoren für die einfachen versus vertieften Varianten (s. Tab. 22-25). Dies zeigt erneut, dass die Aspekte des Modells in FALKE-D insgesamt differenziert wahrgenommen und an die jeweilige Erklärung angepasst werden.

Beachtlich ist dabei, dass die Bedeutung des FALKE-D-spezifischen Prädiktors der *Erklärtiefe* in den Regressionen, die getrennt für die einfachen versus die vertieften Varianten gerechnet wurden, insofern erhalten bleibt, als dass er für die Gesamtstichprobe in beiden sowie bei den vertieften Varianten für die Expertengruppe (hier = LD) signifikant wird. Somit scheint auch unter diesem Blickwinkel die *Erklärtiefe* ein maßgelblicher Faktor bei der Bewertung von Erklärqualität zu sein, und zwar bedeutsamer als alle anderen Prädiktoren.

Die Ergebnisse zeigen außerdem, dass die Variation der beiden Erklärungen der gleichen Prozedur in eine einfache und eine vertiefte Variante gelungen ist. Das Konzipieren von Erklärtiefe, als Erläutern verborgener Zusammenhänge, so dass die Erklärung ein höheres kognitives Aktivierungspotenzial für die Adressatinnen und Adressaten bietet, wird mit diesen Ergebnissen bestätigt.

In den offenen Antworten zeigt sich ebenfalls, dass die Studienteilnehmer/-innen bei den einfachen und den vertieften Erklärungen jeweils anderes gehäuft anmerken. Bei dem Subkode *Fokussierung* (von *Struktur*) ist eine Auffälligkeit zu berichten, denn dieser wird vor allem bei den einfachen Varianten als fehlend angemahnt und hier ganz besonders bei der einfachen Warum-Erklärung. Dies ist verblüffend, da bei den Warum-Erklärungen die einfache insgesamt bevorzugt wird. Die fehlende Fokussierung gibt hier also offensichtlich nicht den Ausschlag für die Bewertung.

Aufschlussreich ist bei den offenen Antworten noch der Kode *Vertieft/oberflächlich*, mit dem die umgesetzte Erklärtiefe bewertet wird. Die Antworten hier spiegeln die obigen deskriptiven Ergebnisse in etwa wider: Während bei der vertieften Was- und Wie-Erklärung deutlich Anmerkungen überwiegen, die die Erklärtiefe positiv hervorheben, sind bei der vertieften Warum-Erklärung vor allem negative Kommentare zu finden, dafür finden sich bei der einfachen Warum-Erklärung wiederum positive Kommentare. Die allgemeine Präferenz der einfachen Warum-Erklärung zeichnet sich hier also auch ab. Der deutschspezifische Aspekt der *Erklärtiefe* findet sich in den Begründungen für die Was- und Wie-Erklärungen theoriekonform, bei den Warum-Erklärungen dagegen nicht, was aber zumindest zum Teil an der oben geschilderten Umsetzung liegt.

Auch unter diesem Blickwinkel auf die Antworten der Studie wird deutlich, dass vor allem Auffälliges bemerkt und angemerkt wird, und dass der FALKE-D-spezifische Aspekt der *Erklärtiefe*

eine wichtige Rolle in den Bewertungen spielt, und zwar unabhängig von der tatsächlich umgesetzten Erklärtiefe in der konkreten Erklärung.

Somit ist insgesamt zu unterstreichen, dass dem für die Studie FALKE-D fachspezifisch konzipierten Aspekt der *Erklärtiefe* eine wichtige Rolle zukommt, da er sich in den Bewertungen als zentral erweist. Die Variation der Erklärtiefe, die durch ein unterschiedlich hohes kognitives Aktivierungspotenzial in den Erklärungen in FALKE-D umgesetzt wurde, wurde in den Beurteilungen insgesamt (mit den genannten Einschränkungen) entsprechend wahrgenommen. Die Ergebnisse von FALKE-D zeigen zudem, dass ein höheres kognitives Aktivierungspotenzial zu einer besseren Bewertung der Erklärung führt. Das bedeutet, dass das Qualitätskriterium der kognitiven Aktivierung von Unterricht tatsächlich auch bei der Bewertung von Erklärungen eine wichtige Rolle spielt. Theoriekonform erweist sich die Bedeutsamkeit je nach Expertisegrad der befragten Statusgruppe als unterschiedlich hoch, am bedeutsamsten ist sie bei den Didaktikerinnen und Didaktikern.

Inwieweit sich diese Erkenntnisse verallgemeinern lassen, wird im Folgenden diskutiert.

10.2 Verallgemeinerbarkeit der Ergebnisse aus FALKE-D

Die Frage nach der Verallgemeinerbarkeit der gewonnenen Erkenntnisse stellt sich in FALKE-D zunächst aufgrund der für eine quantitative Studie vergleichsweise kleinen Stichprobengröße von $N_{max} = 257$, die sich innerhalb der einzelnen Statusgruppen bis $N_{max} = 31$ verringert.

Dazu ist zu sagen, dass die Stichprobe als repräsentativ angesehen werden kann, insbesondere die Statusgruppe der Didaktiker/-innen, die deutschlandweit erhoben wurde, und die der Studierenden, die an drei Universitäten beziehungsweise pädagogischen Hochschulen in Bayern und in Baden-Württemberg durchgeführt wurde. Schüler/-innen und Lehrkräfte wurden ausschließlich aus Bayern befragt. Hier wurden aber unterschiedliche Schulformen und Jahrgangsstufen berücksichtigt. Dennoch wäre eine größere Anzahl an Studienteilnehmenden sowie eine deutschlandweite Befragung aller Statusgruppen wünschenswert.

Trotz der relativ kleinen Anzahl an Befragungen lässt sich die Notwendigkeit der Abfrage von unterschiedlichen Perspektiven bezogen auf die wahrgenommene Qualität von Unterrichtsmerkmalen (bei FALKE-D Lehrererklärungen) verallgemeinern. Gerade in den offenen Begründungen tritt recht deutlich zu Tage, dass die Teilnehmenden die gesehenen Erklärungen jeweils vor dem Hintergrund ihres eigenen Kontextes beurteilen. Abgesehen davon wird statusgruppenunabhängig vor allem Auffälliges angemerkt (vgl. dazu auch Häusl, 2019, S. 43–44). Diese Erkenntnis ist nicht neu, sondern aus der Wahrnehmungsforschung bekannt. Es ist dennoch bemerkenswert, dass sie sich auch bei FALKE-D deutlich zeigt. Zum Teil sind der eigene Hintergrund und die eigenen Überzeugungen so präsent, dass es den Teilnehmerinnen und Teilnehmern (und

zwar aller Statusgruppen) nicht möglich ist, die notwendige Abstraktion im Hinblick auf Vorgaben des Designs zu vollziehen; stattdessen bewerten sie die Erklärungen unter dem Aspekt der Umsetzbarkeit oder Eignung für authentischen Unterricht. Man sieht somit deutlich, dass Studienteilnehmer/-innen in einer ganz bestimmten Rolle an Erhebungen teilnehmen. Um ein komplexes Konstrukt wie das *Gute Erklären* zu erfassen, ist die Erhebung unterschiedlicher Perspektiven somit notwendig.

Festzustellen ist weiterhin, dass das in FALKE-D erarbeitete Modell des Erklärens im Deutschunterricht die Erklärqualität gut abbildet (s. aufgeklärte Gesamtvarianz in den Tabellen 19-25 und Ausführungen oben). Da dies für die Gesamtstichprobe ebenso wie für die einzelnen Statusgruppen gilt, ist anzunehmen, dass das Modell tragfähig ist (zu den niedrigeren Ergebnissen für die Statusgruppe der Schüler/-innen s. Punkt *10.3 Grenzen der Studie FALKE-D* in dieser Arbeit). Aus der Analyse der hierarchischen linearen Regressionen ergibt sich außerdem, dass nahezu alle abgefragten Prädiktoren (nämlich *Strukturiertheit, Adressatenorientierung, Sprech-/Körperausdruc*k sowie der deutschdidaktische Aspekt der *Erklärtiefe*) eine wichtige Rolle für die Bewertung von Erklärungen spielen. Nimmt man die Ergebnisse der offenen Antworten hinzu, so wird auch die Bedeutung von *Veranschaulichung* klar: Das ist der Aspekt, der spontan am häufigsten genannt wird, wobei das Fehlen von Veranschaulichungen ebenfalls angemerkt wird. Auch wenn *Veranschaulichung* viele unterschiedliche Facetten abdeckt und eine Unterteilung notwendig ist (s.o.), so zeigen die Ergebnisse dennoch, dass dieser Aspekt ebenfalls bedeutsam ist. Trotz der kleinen Stichprobegröße und trotz der Tatsache, dass die Themen der Erklärungen alle aus der Sprachdidaktik stammen, kann man die Ergebnisse mit der gebotenen Vorsicht (s. Punkt *10.3 Grenzen der Studie FALKE-D* in dieser Arbeit) für Erklärungen im Deutschunterricht verallgemeinern. Die gemeinsam in FALKE erhobenen Aspekte sind auf der einen Seite in der Lehr-/Lernforschung bereits anerkannt und werden auf der anderen Seite in FALKE-D in den geschlossenen Antworten statistisch signifikant bestätigt. Zudem werden alle Aspekte in den offenen Antworten von FALKE-D auch spontan als wichtige Kriterien von Erklärqualität genannt (und zwar über die Statusgruppen hinweg). Es ist daher anzunehmen, dass diese Aspekte beim Erklären im Deutschunterricht themenunabhängig berücksichtigt werden sollten.

Verallgemeinert werden kann auch die Bedeutung des Aspektes *Erklärtiefe*, der spezifisch für FALKE-D konzipiert und erhoben wurde. Die *Erklärtiefe* wird differenziert wahrgenommen und je nach Erklärung unterschiedlich bewertet. In den vertieften Erklärungen, in denen eine größere Erklärtiefe umgesetzt wurde, wird der Aspekt *Erklärtiefe* insgesamt auch besser bewertet, was zeigt, dass die Studienteilnehmer/-innen die unterschiedliche Umsetzung von Erklärtiefe wahrnehmen, und zwar theoriekonform die Expertinnen und Experten genauer als die anderen Teilnehmer/-innen. Die Regressionen weisen diesem Aspekt überdies ein mindestens genauso hohes Gewicht nach wie

den aus der Lehr-/Lernforschung bekannten Merkmalen, und zwar sowohl in der Gesamtstichprobe wie für einzelne Stichproben und in der Variation einfach versus vertieft (s. Tab. 19, 20-21 sowie 22-25). Da dieser Aspekt als einziger unter so vielen unterschiedlichen Blickwinkeln signifikant wird, kann davon ausgegangen werden, dass *Erklärtiefe* bei Erklärungen im Deutschunterricht insgesamt eine Rolle spielt und ihre Bedeutung verallgemeinert werden kann. Somit muss die *Erklärtiefe* bei Erklärungen im Deutschunterricht mitreflektiert werden. Neben den in der Bildungsforschung schon bekannten Kriterien leistet die fachdidaktische Gestaltung von Erklärungen offensichtlich einen wichtigen Beitrag zu deren Qualität.

10.3 Grenzen der Studie FALKE-D

Nach der Zusammenfassung der Ergebnisse soll auch angesprochen werden, was die Studie FALKE-D nicht leisten kann.

Trotz der gerade erläuterten Verallgemeinerbarkeit der Aspekte guten Erklärens im Deutschunterricht muss man sich der Heterogenität der Themen und der für die Inhalte des Faches Deutsch typischen divergenten Lösungen bewusst bleiben. Das bedeutet, dass die konkrete Ausgestaltung der Aspekte an den jeweiligen Gegenstand angepasst werden muss. Aus den Ergebnissen von FALKE-D kann abgelesen werden, dass die *Strukturiertheit* einer Erklärung im Deutschunterricht ein wichtiges Kriterium für ihre Qualität ist, wie genau diese Strukturiertheit aber aussieht, muss für jeden Gegenstand entschieden werden. Dies hat FALKE-D nur für die drei Erklärungen aus der Sprachdidaktik geleistet. Wie die Umsetzung bei anderen sprachdidaktischen Themen oder bei Themen aus der Literaturdidaktik aussehen muss, darüber kann die Studie FALKE-D keine Aussagen treffen. Analoges gilt für die anderen Aspekte, insbesondere für den fachspezifischen Aspekt der *Erklärtiefe.* Dieser wurde in FALKE-D für die oben ausführlich geschilderten drei Erklärungen variiert. Die Ergebnisse zeigen exemplarisch, dass die Variation auch so wahrgenommen wird und dass *Erklärtief*e einen entscheidenden Aspekt für die Erklärqualität darstellt. Für die konkrete Ausgestaltung von Erklärungen zu anderen Themen liefert FALKE-D den Grundsatz, nämlich dass verborgene Zusammenhänge deutlich gemacht werden müssen, die das kognitive Aktivierungspotenzial erhöhen. Die konkrete Ausgestaltung muss aber immer im Einzelfall erfolgen.

Damit einher geht zum einen das Bewusstsein, dass die für FALKE-D konzipierten Erklärungen künstliche Erklärungen sind, die monologisch gestaltet wurden, um die Konzentration auf das Erklärhandeln der Lehrkraft zu ermöglichen. Die Erklärungen sind also explizit keine Mustererklärungen, die so eins zu eins in die Schule übernommen werden sollen (die doch zahlreichen offenen Antworten, die genau diesen Punkt kritisieren, zeigen, dass diese Abstraktion trotz vorheriger Erläuterungen nicht von allen Studienteilnehmerinnen und -teilnehmern geleistet

wurde). Sie dienen vielmehr dem Herauskristallisieren von Kriterien und Prinzipien, die dann für die Gestaltung von Erklärungen zum Beispiel in die Lehreraus- und -weiterbildung übernommen werden können. Zum anderen muss klar sein, dass Erklärkompetenz eine von zahlreichen Kompetenzen ist, die eine Lehrkraft beherrschen muss, um qualitätvollen Unterricht zu erteilen. Mit der Auswahl des Erklärens als Forschungsgegenstand wird keine Hierarchie innerhalb dieser Kompetenzen aufgestellt und noch weniger ein Ausschließlichkeitsanspruch dieser Lehrerkompetenz postuliert. Genauso wenig trifft FALKE-D Aussagen darüber, an welcher Stelle im Unterricht Erklärungen notwendig sind, und wann andere Methoden vielversprechender für den Lernerfolg von Schülerinnen und Schülern sind. Insofern leistet diese Arbeit einen Beitrag zum großen Ganzen der Lehrerprofessionalität.

Das Konstrukt des Erklärens bei FALKE-D wird aufgrund der oben beschriebenen eingehaltenen Gütekriterien als valide angenommen, die Regressionen zeigen wie mehrfach betont, dass das Modell das Erklären für die Studienteilnehmer/-innen gut abdeckt. Was die Studie allerdings nicht leistet, sind Aussagen darüber, inwiefern die als für gutes Erklären bedeutsam erachteten Kriterien, zum Beispiel *Strukturiertheit* oder *Erklärtiefe,* auch tatsächlich effektiv sind. Das heißt konkret, ob Erklärungen, die diese Kriterien erfüllen, auch wirklich und theoriekonform zu besseren Verstehens- und Merkleistungen bei den Adressatinnen und Adressaten führen als einfache Erklärungen. Für diese Fragestellung müssen weitere Studien durchgeführt werden, die entsprechende Verstehenstests entwickeln (s. folgende Ausführungen *11.1 Anregungen aus FALKE-D für weitere Forschungsarbeiten*).

Weiterhin wurde zwar mit der Variation nach Erklärtiefe und der entsprechenden Konzeptualisierung und Operationalisierung in den Erklärungen ein Beitrag zur Erforschung des kognitiven Aktivierungspotenzials im Deutschunterricht geleistet. Allerdings ist die kognitive Aktivierung wie mehrfach betont ein anspruchsvolles und komplexes Prinzip und keinesfalls in FALKE-D ausreichend dargelegt, denn das hätte den Rahmen der Studie gesprengt. Hier sind weitere Arbeiten sowohl zu anderen Themen mit entsprechender Operationalisierung des kognitiven Aktivierungspotenzials als auch Studien zum Verstehen der Adressatinnen und Adressaten dringend nötig.

Ausgehend von den dargelegten Grenzen der FALKE-D Studie ergeben sich Themen für weiterführende Studien. Dieses Potenzial wird im Folgenden erläutert.

11 Fazit und Ausblick

Sowohl aus FALKE als auch aus FALKE-D ergeben sich Anregungen zu weiteren Forschungsarbeiten, und zwar zu folgenden unterschiedlichen Themenbereichen.

11.1 Anregungen aus FALKE-D für weitere Forschungsarbeiten

Die effektive Erklärung

Folgende Überlegungen sind in dieser Arbeit unter Punkt *10.3 Grenzen der Studie FALKE-D* bereits angeklungen. FALKE-D nimmt das Konstrukt des guten Erklärens in den Blick. Dabei steht die Wahrnehmung der Erklärqualität aus der Sicht der befragten Statusgruppen im Mittelpunkt. Die Bedeutung bereits bekannter Aspekte für die Qualität einer Erklärung wie *Strukturiertheit, Sprech-/Körperausdruck* oder *Adressatenorientierung* bestätigt sich auch in der Befragung von FALKE-D. Ebenso erweist sich das fachspezifisch konzeptualisierte und operationalisierte Merkmal der *Erklärtiefe* als zentral. Gleichzeitig zeigt sich durch die Variation der Erklärprozeduren (Was-/ Wie-/ Warum-Erklärung) in je eine vertiefte und eine einfache Variante insgesamt eine Präferenz für die vertieften Erklärungen. Somit wird die Forderung nach kognitiv aktivierendem Unterricht ebenfalls in den Ergebnissen der FALKE-D-Befragung offenbar. Eine gewinnbringende und notwendige nächste Fragestellung ist nun, ob die als gut wahrgenommenen Erklärungen auch effektiv sind. Effektiv meint in diesem Fall die Verstehens- und Merkleistung, die die Erklärungen hervorrufen. Über Verstehenstests muss geprüft werden, ob die als besser wahrgenommenen vertieften Erklärungen zu besseren Verstehens- und Merkleistungen führen als die einfachen. Gerade bei komplexen Anforderungen wie dem Widerlegen eines Arguments sind auch Longitudinalstudien in Betracht zu ziehen.

Ein weiterer interessanter Aspekt ist in diesem Zusammenhang die Notwendigkeit einer vertieften Erklärung im konkreten Unterrichtsgeschehen. Denn selbst wenn sich diese als effektiver im gerade geschilderten Sinne herausstellen, so sind sie doch auch zeitaufwändiger, und es gibt vermutlich Unterrichtssituationen, in denen eine einfache Erklärung ausreicht. Zu berücksichtigen sind hier besonders Situationen, in denen die Erklärung eine Nebenkommunikation ist, die die Weiterarbeit am eigentlichen Thema ermöglichen soll. Es scheint in diesem Zusammenhang wahrscheinlich, dass dann eine einfache Erklärung genügt, um möglichst schnell zum eigentlichen Unterrichtsgegenstand zurückzukehren, und damit eine geringere kognitive Aktivierung der Schüler/-innen in diesen Fällen zielführender ist. In FALKE-D könnte das zum Beispiel bei der Erklärung der Redewendung der Fall sein, da sie eine solche Nebenkommunikation darstellt.

Damit einher geht die Überlegung, dass vertiefte Erklärungen in bestimmten Kontexten sogar störend sein können, weil sie zu viel kognitive Aktivität in Anspruch nehmen (s. obige Überlegungen zur kognitiven Aktivierung nach Praetorius und anderen, 2018). Gerade im Zusammenhang mit Erklärungen ist dieser Punkt zu beachten, worauf auch Schnotz hinweist: „many commonly instructional techniques result in an unnecessarily high cognitive load on working memory that interferes with learning“ (2016, S. 114). Dass es für die Entscheidung, wie kognitiv aktivierend eine Unterrichtshandlung sein muss, um möglichst effektiv zu sein, keine Patentrezepte gibt, sondern

dass sie im Gegenteil sehr anspruchsvoll ist, nennt Schnotz ebenfalls, – nur zwei Sätze später: „However, reduction of cognitive load is not always helpful for learning either" (2016, S. 114-155). Diesem Aspekt muss genauer nachgegangen werden. Von Studien, die zum Beispiel mit authentischem Unterrichtsgeschehen arbeiten und diesen Aspekt untersuchen, sind vielversprechende Ergebnisse zu erwarten.

Erklärtiefe

Mehrfach ist in dieser Arbeit betont worden, dass die deutschdidaktische Operationalisierung von *Erklärtiefe* ein ebenso bedeutsames Merkmal von Erklärqualität darstellt wie bereits bekannte Merkmale (zum Beispiel *Strukturiertheit* oder *Veranschaulichung*). Mit *Erklärtiefe* ist in FALKE-D das kognitive Aktivierungspotenzial gemeint, das eine Erklärung in sich birgt. Es wird davon ausgegangen, dass durch hohes kognitives Aktivierungspotenzial vertieftes Verstehen möglich ist, bei niedrigem Potenzial dagegen nur oberflächliches Verständnis erzeugt werden kann. Ob dies tatsächlich so ist, muss wie eben dargelegt mit Verstehenstests für Erklärungen aus dem Fach Deutsch überprüft werden. Aber auch das kognitive Aktivierungspotenzial selbst muss noch weiter erforscht werden, denn es wurde bereits mehrfach angedeutet, wie unterschiedlich, je nach Gegenstand und Ziel, kognitive Aktivierung aussehen kann. Weitere Forschung ist daher notwendig, um zu konkretisieren, wie genau eine aktive geistige Auseinandersetzung bei den Schülerinnen und Schülern im Deutschunterricht angeregt werden kann. FALKE-D leistet hier einen Baustein, indem das kognitive Aktivierungspotenzial anhand von drei Gegenständen aus der Sprachdidaktik variiert und untersucht wurde. Die Übertragung auf andere Gegenstände kann ein nächster Schritt sein, um das Aktivierungspotenzial in Erklärungen aus dem Deutschunterricht noch genauer einzukreisen. Dazu können insbesondere die drei identischen Items zur *Erklärtiefe* (s. Tab. 3) genutzt werden, die für andere Gegenstände übernommen werden können.

Mehrperspektivische Untersuchung

Die mehrperspektivische Anlage der Studie FALKE-D erlaubt es, statusgruppenspezifische Rückschlüsse auf die Kriterien, die bei der Bewertung von Erklärqualität eine Rolle spielen, zu ziehen und die unterschiedlichen Gruppen in ihrem Bewertungsverhalten zu vergleichen. Gleichzeitig schließen sich aber an diese mehrperspektivische Anlage offene Fragestellungen an.

Wie dargelegt sticht die Statusgruppe der Schüler/-innen, also die Adressatengruppe, in mehreren Hinsichten in der Datenanalyse heraus. Zum einen fällt die aufgeklärte Gesamtvarianz des Konstruktes *gutes Erklären* bei den Schülerinnen und Schülern geringer aus als bei allen anderen Statusgruppen. Es ist offensichtlich, dass Schüler/-innen außer den abgefragten Kriterien noch andere Kriterien zur Bewertung von Erklärungen heranziehen. Diese Kriterien genauer zu

bestimmen, zum Beispiel über Nachfragen in sich anschließenden Interviews oder die Erhebung offener Begründungen bei der Statusgruppe Schüler/-innen, ist eine wichtige Aufgabe weiterer Forschung. Es sei noch einmal betont, dass gerade die Adressatenperspektive beim Erklären von besonderem Interesse ist. Die Diskrepanzen, die sich in den Wahrnehmungen der Erklärqualität aus Sicht der Schüler/-innen und Didaktiker/-innen zeigen, gibt zu denken. Offensichtlich stimmen die Qualitätskriterien, die von fachdidaktischer Seite an Erklärungen angelegt werden, nicht mit denen der Schüler/-innen überein. Die Aufnahme von Kriterien von Schülerinnen und Schülern würde somit zu einem neuen Modell des guten Erklärens (aus Schülersicht) führen, das dann wiederum mit dem in FALKE-D erarbeiteten Konstrukt in Verbindung gesetzt werden kann. Ein Ziel muss dabei sein, die aufgeklärte Gesamtvarianz von Erklärqualität bei den Schülerinnen und Schülern zu erhöhen. In Verbindung mit der Frage, welche der Erklärungen dann auch tatsächlich effektiv ist, versprechen weiterführende Studien zu diesen Aspekten wichtige und interessante Ergebnisse.

Weiterhin ist der Vergleich des Beurteilungsverhaltens der Statusgruppen, das heißt entsprechende Unterschiede und Gemeinsamkeiten ein lohnender Forschungsgegenstand. Wie oben dargelegt gibt es in der Forschung hierzu widersprüchliche Ergebnisse, was die Übereinstimmung und die Verlässlichkeit der Urteile der einzelnen Statusgruppen anbelangt. In FALKE-D lassen sich aufgrund der sehr heterogenen Bewertungen in allen Gruppen keine eindeutigen Ausprägungen im Bewertungsverhalten feststellen. Es zeigen sich aber dennoch bereits erläuterte Tendenzen dahingehend, dass die Schüler/-innen und die Lehrkräfte etwas strenger, die Didaktiker/-innen etwas milder als die anderen Statusgruppen bewerten. Weitere Forschung wird zeigen, ob sich diese Tendenzen bestätigen oder nicht. Lohnend scheint in diesem Zusammenhang auch eine genauere qualitative Befragung zur wahrgenommenen Qualität von Erklärungen in einem iterierenden Prozess. Die recht knappen Antworten bei FALKE-D erlauben nur einen ersten Einblick in die intuitiven Bewertungskriterien. Zahlreiche fehlende sowie zu allgemeine beziehungsweise unklare Begründungen bei FALKE-D könnten daher rühren, dass die Bewertung einer Erklärung zwar einfach scheint, diese Bewertung jedoch sachlich zu begründen erheblich schwerer fällt (vgl. Häusl, 2019, S. 45). Ausführlichere Begründungen, die zum Beispiel in Form von Interviews erhoben werden, können hierüber Aufschluss geben.

11.2 Ausblick auf FALKE II

In Punkt *2.2.2 Professionswissen von Lehrkräften* in dieser Arbeit wurde der Weg von COACITV über FALKO zu FALKE geschildert. Auch FALKE ist aber nicht der Endpunkt dieser Forschung, vielmehr läuft seit Januar 2020 das Nachfolgerprojekt FALKE II an der Universität Regensburg. Ziel dieses Projektes ist es nach dargelegter Modellierung von Erklären den Fokus verstärkt auf die

Reflexion von Erklärungen im Unterricht zu legen und somit Professionalisierungsprozesse zu untersuchen und zu fördern.
Im Folgenden wird an zwei Punkten konkret aufgezeigt, wie die Ergebnisse aus FALKE genutzt werden können, für FALKE II und darüber hinaus.

Erklärkompetenz in der Lehrerbildung

Nutzbar gemacht werden können die Ergebnisse von FALKE-D für die Lehreraus- und -fortbildung. Da in der Gesamtstichprobe die vier genannten Prädiktoren (*Strukturiertheit, Adressatenorientierung, Sprech-/Körperausdruck* sowie *Erklärtiefe*) entscheidend für die Qualität einer Erklärung sind, sollten alle in der deutschdidaktischen Lehrerausbildung genutzt werden. Gerade die deutschdidaktische Operationalisierung von *Erklärtiefe*, die wie mehrmals betont für die Expertinnen und Experten einen ähnlich hohen Einfluss auf die Erklärqualität hat wie jedes aus der Lehr-/Lernforschung bekannte Merkmal guten Erklärens, bietet hier Potenzial. Erklärkompetenz der Studierenden in den deutschdidaktischen Seminaren an der Universität zu schulen, ist ein wünschenswerter Bestandteil der Ausbildung der zukünftigen Lehrkräfte.
Im Zuge des FALKE-Projekts fanden dazu an der Universität Regensburg Tandemseminare zwischen zwei unterschiedlichen Fächern statt. Die Studierenden hatten dabei den Auftrag den jeweils fachfremden Studierenden ein Thema aus dem eigenen Fach zu erklären. Diese Erklärungen wurden anschließend ausführlich besprochen. Gerade die Distanz, die durch den fachfremden Gegenstand entstand, erwies sich als fruchtbar: So hatten zum einen die Studierenden bei der Erstellung ihrer Erklärung „echte" (im Sinne von unwissende) Adressatinnen und Adressaten als Zielgruppe. Die Erklärung musste also so vorbereitet werden, dass Studierende mit wenig Hintergrundwissen der Erklärung folgen konnten, ohne dass der Inhalt durch die Verkürzung fachlich falsch werden durfte; – eine Herausforderung, der sich Lehrkräfte tagtäglich stellen müssen.
Gleichzeitig erhielten die Erklärenden von diesen Adressatinnen und Adressaten Feedback zur Qualität der Erklärung, da zum anderen die Studierenden, die zuhörten, in die Rolle von „echten" (s. o.) Adressatinnen und Adressaten versetzt wurden. Dadurch, dass den fachfremden Studierenden der Gegenstand nicht bekannt war, bestand echter Erklärbedarf. Die Qualität der Erklärung konnte somit tatsächlich in der Rolle der Adressatin/des Adressaten befeedbackt werden.
Diese Tandemseminare fanden während der gesamten Laufzeit von FALKE vom Wintersemester 2015 bis zum Sommersemester 2019 in variierenden Konstellationen und in unterschiedlich intensiver Zusammenarbeit statt. Es gab Formate, bei denen es am Ende des Semesters einen gemeinsamen Erklärtag gab, beziehungsweise Kurse, die einige Sitzungen auch zur theoretischen Vorarbeit teilten und dann mit einem Erklärtag schlossen, aber auch Seminare, die jede Sitzung gemeinsam durchführten. Darüber hinaus wurde in einzelnen Sitzungen mit der deutschen

Sprachwissenschaft und der Sprechwissenschaft zusammengearbeitet, so dass den Studierenden auch sprachliche und sprecherische Grundlagen des Erklärens vermittelt wurden. Die Evaluationen, die im Nachgang erhoben wurden, fielen sehr positiv aus, wobei insbesondere die Möglichkeit der interdisziplinären Zusammenarbeit, gerade auch mit der Sprach- und Sprechwissenschaft, und die Schulung der eigenen Erklärkompetenz gelobt wurden.

Die Tandemseminare bleiben an der Universität Regensburg während des Nachfolgeprojekts FALKE II erhalten und werden breiter implementiert, um möglichst vielen Studierenden über ein Feedback „von außerhalb des Faches" die Chance zu geben, die eigene Erklärkompetenz zu verbessern. Es wäre wünschenswert und gewinnbringend, sie auch nach der Projektlaufzeit von FALKE II durchzuführen und das Konzept an anderen Universitäten vorzustellen, damit es auch dort übernommen werden kann. Ebenso wäre es wünschenswert, die Ergebnisse in Form einer Lehrerfortbildung auch Lehrkräften in der dritten Lehrerbildungsphase zur Verfügung zu stellen.

Nicht vergessen werden darf in diesem Zusammenhang ein weiterer Aspekt, der kurz angerissen werden soll, nämlich die Bedeutung von Mündlichkeit und mündlicher Ausdrucksfähigkeit für die Schüler/-innen: Unterricht findet in weiten Teilen mündlich statt, außerdem ist eine sprachbewusste mündliche Ausdrucksfähigkeit eine Schlüsselkompetenz für das spätere (Berufs-)Leben. Mündlichkeit muss daher im Deutschunterricht den gleichen Stellenwert wie Schriftlichkeit einnehmen (vgl. dazu Quasthoff & Heller, 2014). Erklären zu können ist dabei eine von vielen Sprechhandlungen, die bei Schülerinnen und Schülern gefördert werden muss (s. Ausführungen oben unter Punkt *3.4 Erklärungen im (Deutsch-) Unterricht* in dieser Arbeit). Studierende, die sich bereits im Studium intensiv mit dem Erklären und den Merkmalen des guten Erklärens auseinandergesetzt haben, können ihre Schüler/-innen auch besser darin anleiten. So kommt die Schulung der eigenen Erklärkompetenz auch der Förderung der Erklärkompetenz der Schüler/-innen zugute.

Interdisziplinäre Vergleiche

Ein weiteres Forschungsinteresse, das sich aus der interdisziplinären Anlage von FALKE ergibt, sind fachübergreifende Vergleiche darüber, welche Aspekte guten Erklärens in welchen Disziplinen als entscheidend angesehen werden und welche eine geringere Gewichtung erfahren. Dabei sind sowohl Gemeinsamkeiten und Unterschiede zwischen einzelnen Fächern von Belang als auch Gemeinsamkeiten und Unterschiede zwischen Fächergruppen, zum Beispiel zwischen naturwissenschaftlichen und geisteswissenschaftlichen Fächern.

Ein erster Blick in die Ergebnisse der anderen Fächer zeigt zum Beispiel, dass die Aspekte *Strukturiertheit* und *Adressatenorientierung*, die in FALKE-D bei den Expertengruppen SLD ein hohes signifikantes Regressionsgewicht auf die Globalnote haben, auch in den Fächern Geschichte

und Englisch hoch signifikant sind. Das Ergebnis, dass der *Sprech-/Körperausdruck* bei den Schülerinnen und Schülern einen entscheidenden Einfluss auf die Globalnote hat, teilt Deutsch mit den Fächern Musik und evangelische Religionspädagogik. Für ausführliche interdisziplinäre Analysen sei auf Schilcher, Krauss, Lindl und Hilbert (2021, in Vorb.) verwiesen. Betont werden in diesem Zusammenhang noch einmal die hohen Anforderungen, die interdisziplinäre Forschung an alle Beteiligten stellt:

> Interdisciplinary communication in the field of educational research is therefore not a trivial task. It requires the readiness to listen, to understand other perspectives, and to engage in rigorous discussions about conceptual consistency. Simultaneously, it needs showing full respect for other disciplinary views even when for the time being some issues could not be satisfactorily clarified yet, and it requires willingness to engage in common activities from the different introduced perspectives. Building bridges between disciplines and building bridges between interdisciplinary research and educational practice is certainly highly challenging. However, it is also highly rewarding by providing new insights in the field (Schnotz, 2016, S. 115).

Nach der vierjährigen Arbeit im Projekt FALKE kann diesem Zitat vollumfänglich zugestimmt werden: Interdisziplinäre Forschung benötigt ein hohes Maß an Selbstdisziplin sowie Respekt vor anderen Herangehensweisen und Schwerpunktsetzungen, was manchmal einher geht mit Kompromisslösungen, bei denen das eigene Interesse und die Besonderheiten des eigenen Faches hinten angestellt werden müssen. Auf der anderen Seite bietet sie die großartige Chance, Neues zu lernen, über den Tellerrand zu blicken und Erkenntnisse zu gewinnen, die ohne fachfremde Impulse nicht möglich gewesen wären. Vor diesem Hintergrund ist es umso erfreulicher, dass alle Disziplinen im Projekt FALKE „durchgehalten" haben und dass die interdisziplinäre Arbeit auch in FALKE II fortgeführt wird.

Aufgrund der Vielzahl an insgesamt erhobenen Daten im Projekt FALKE sind weitere Analysen möglich, zum Beispiel im Hinblick auf die erhobenen Klassenstufen und Schularten. Auch die neuen Daten aus dem Nachfolgerprojekt FALKE II können hier mit einfließen. Eine interessante Fragestellung ist dabei auch, ob es Aspekte des Erklärens gibt, die sich besonders für interdisziplinäre Vergleiche eignen und andere, die dafür eventuell weniger in Frage kommen. Die Überlegungen von Wisniewski und anderen (2020) bezogen auf die Vergleichbarkeit von Unterrichtsqualitätsmerkmalen können hier ein Ansatzpunkt sein. Sie konstatieren beispielsweise, dass sich Motivation oder Lebenswelt weniger für Vergleiche eignen, da sie stark von konkreten Bedingungen des Unterrichtsgeschehens abhängen. Besser geeignet sind dagegen Strukturierung und Klassenführung zum Vergleich zwischen Fächern und Messzeitpunkten. Es ist sicherlich interessant, unter diesem Blickwinkel die Vergleiche von Erklärungen zwischen den Fächern weiterzudenken.

Verknüpfung von Theorie und Praxis

Oben ist bereits angeklungen, wie wichtig es ist, die aus FALKE(-D) gewonnenen Erkenntnisse für die Praxis nutzbar zu machen (vgl. hierzu Seidel & Thiel, 2017). Dazu können auch die erstellten Videovignetten herangezogen werden, die so aufbereitet werden können, dass sie nicht nur für Studierende in Seminaren, sondern auch für Lehrkräfte im aktiven Schuldienst interessant sind. Denkbar ist zum Beispiel, sie zur Reflexion von monologischen Erklärungen sowie deren sprachlicher Gestaltung einzusetzen.

Aufbauend auf den Erkenntnissen von FALKE entsteht aktuell in FALKE II an der Universität Regensburg eine Vielzahl an Vignetten zu unterschiedlichen Themen aus dem Deutschunterricht. Auch diese Vignetten können für die Praxis genutzt werden.

Die in der theoretischen Auseinandersetzung und in der empirischen Forschung gewonnenen Erkenntnisse aus der Deutschdidaktik sind wichtig und spannend. Eine tatsächliche Umsetzung der Ergebnisse in der schulischen Praxis aber lässt das Herz jeder Didaktikerin und jeden Didaktikers höher schlagen. Deswegen sei am Schluss dieser Arbeit die Hoffnung formuliert, dass die in FALKE-D gewonnenen Erkenntnisse zum Erklären im Deutschunterricht manche Lehrkraft interessieren und zur Reflexion des eigenen Erklärverhaltens anregen können.

Danksagung

Seit vielen Jahren ist das Laufen ein fester Bestandteil in meinem Leben, und so lag die Analogie nahe, das Entstehen dieser Arbeit mit dem Marathonlauf meines Lebens zu vergleichen. Und wie beim „echten" Laufen so war auch der „Lauf" zu dieser Arbeit nur möglich mit Menschen, die mir auf ganz unterschiedliche Weise geholfen haben und so zum Gelingen dieser Arbeit beigetragen haben. Ihnen allen möchte ich meinen Dank aussprechen.

Dankbar bin ich dem aktiven Publikum des Laufs, also Kolleginnen und Kollegen, die die Videovignetten für mich aufgenommen haben, und den Teilnehmerinnen und Teilnehmern an der Studie, die alle zwei Mal 45 Minuten in einen Fragebogen investiert haben. Das ist keine Selbstverständlichkeit, hat aber den empirischen Teil dieser Arbeit erst möglich gemacht.

Dankbar bin ich den Wasserreichern an den Verpflegungsstellen, meinen Kolleginnen und Kollegen am Lehrstuhl für Deutschdidaktik der Universität Regensburg, die alle stets ein offenes Ohr hatten und durch viele kluge Bemerkungen diese Arbeit immer wieder vorangebracht haben.

Dankbar bin ich den Streckenposten, meinen studentischen Hilfskräften Julia Hofmann, Maximilian Lang und Theresa Häusl, auf die ich mich stets verlassen konnte und deren Elan mich über so manche schwierige Strecke hinweggebracht hat.

Dankbar bin ich dem Streckenchef Alfred Lindl für all seine Unterstützung bei statistischen Berechnungen und bei den Interpretationen der Ergebnisse und für die freundliche Geduld, die er dabei stets mit mir hatte.

Dankbar bin ich der Pacemakerin, meiner Kollegin und Freundin Michaela Haberkorn-Trenner, für das sorgfältige Korrekturlesen der ersten Fassung, aber vor allem für eine Freundschaft, die mein Leben bereichert.

Dankbar bin ich der Gesamtleiterin des Laufs, meiner Erstgutachterin Anita Schilcher, die mir diese Arbeit von Anfang an zugetraut hat und mir den Freiraum gegeben hat, selbst zu gestalten. Ich habe viel gelernt.

Ebenso dankbar bin ich dem Vizeleiter des Laufs, meinem Zweitgutachter Markus Pissarek, für die unkomplizierte Übernahme des Zweitgutachtens und die immer freundlichen und konstruktiven Kommentare.

Dankbar bin ich meinen Mitläuferinnen und Mitläufern, allen Promovierenden im FALKE-Projekt. Ich habe ungemein von unserem interdisziplinären Austausch profitiert und viel von ihnen allen gelernt. Bei so manchem Berg, den es zu überwinden galt, waren wir uns vielleicht über die Strategie nicht immer einig, aber stets getragen von dem gemeinsamen Bestreben, vorwärtszukommen. Es war mir außerdem eine Freude, mit so vielen jungen, interessierten und engagierten Menschen zusammenzuarbeiten, in einem Team, das seinesgleichen sucht. Hoch leben die FALKEN!

Dankbar bin ich meinen Fans, meinen Freundinnen und Freunden, Svenja, Alex, Ulli und Doris, die wissen, was es heißt, eine Doktorarbeit zu schreiben und mir auf meinem Weg tapfer beiseite standen. Der fachliche Austausch war mir sehr wertvoll, unsere Freundschaft ist mir unersetzlich.

Dankbar bin ich den Fans meiner Familie, meinen Eltern, Großeltern und Geschwistern, für viele interessierte Nachfragen zu diesem Projekt.

Dankbar bin ich meiner Tochter Johanna, die mich, wenn ich mich verlaufe, mit Leichtigkeit auf den Boden der Tatsachen zurückholt. Nichts könnte mich jemals mit mehr Stolz erfüllen, als zu sehen, zu was für einer wundervollen Frau mein Kind herangewachsen ist.

Y gracias, Juan Diego por acompañarme en ese camino de la vida, por haberme soportado, por seguir queriéndome y por creer en mí. Nada me hace sentir más segura que estar a tu lado.

Tabellen- und Abbildungsverzeichnis

Tabellenverzeichnis

Tabelle 1: Übersicht über Videovignetten 107
Tabelle 2: Gesamtstichprobe in FALKE-D 156
Tabelle 3: Identische Items für alle Vignetten zum Aspekt Erklärtiefe 157
Tabelle 4: Alle Items zur Erklärtiefe der Was-Erklärungen 159
Tabelle 5: Alle Items zur Erklärtiefe der Warum-Erklärungen 160
Tabelle 6: Alle Items zur Erklärtiefe der Wie-Erklärungen 162
Tabelle 7: Cronbachs alpha nach Skalen, Video und Statusgruppe SuS 171
Tabelle 8: Cronbachs alpha nach Skalen, Video und Statusgruppe SLD 171
Tabelle 9: Mittelwerte, Standardabweichungen und Spannweiten (über alle vier Statusgruppen) zum Globalurteil 173
Tabelle 10: Mittelwerte, Standardabweichungen und Spannweiten zum Globalurteil nach Statusgruppen getrennt 175
Tabelle 11: Mittelwerte, Standardabweichungen und Spannweiten* (über alle vier Statusgruppen) zu den fünf Aspekten guten Erklärens, zur Persönlichkeitswirkung und zu Fachliche Relevanz182
Tabelle 12: Mittelwerte, Standardabweichungen und Spannweiten (über alle vier Statusgruppen) für die einzelnen Items von Fachliche Relevanz 186
Tabelle 13: Mittelwerte, Standardabweichungen und Spannweiten* zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Schüler/-innen 189
Tabelle 14: Mittelwerte, Standardabweichungen und Spannweiten* zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Studierende 191
Tabelle 15: Mittelwerte, Standardabweichungen und Spannweiten* zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Lehrkräfte 193
Tabelle 16: Mittelwerte, Standardabweichungen und Spannweiten* zu den fünf Aspekten guten Erklärens und zur Persönlichkeitswirkung für die Statusgruppe Didaktiker/-innen 195
Tabelle 17: Varianzanalyse mit gemischtem Design pro Erklärprozedur 198
Tabelle 18: Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos 202
Tabelle 19: Hierarchisches lineares Regressionsmodell mit der abhängigen Variable Globalurteil, Gesamtstichprobe 204
Tabelle 20: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, Schüler/-innen und SLD-Gruppe 206

Tabelle 21: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, Schüler/-innen, Studierende, Lehrkräfte und Didaktiker/-innen 209
Tabelle 22: Hierarchisches lineares Regressionsmodell mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Gesamtstichprobe 212
Tabelle 23: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Schüler/-innen 213
Tabelle 24: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Studierende 214
Tabelle 25: Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil, aufgeteilt nach einfachen und vertieften Videos, Lehrkräfte und Didaktiker/-innen 216
Tabelle 26: Übersicht über deduktive und induktive Kodes und Subkodes 229

Abbildungsverzeichnis

Abbildung 1 Dimensionen von Unterrichtsqualität 25
Abbildung 2 Dimension der kognitiven Aktivierung in deutschdidaktischen Studien 32
Abbildung 3: Das Kaskadenmodell 34
Abbildung 4: Professionelle Kompetenz 35
Abbildung 5: Professionswissen 49
Abbildung 6: Modell zur Strukturiertheit bei FALKE 85
Abbildung 7: Modell zur Adressatenorientierung bei FALKE 88
Abbildung 8: Übersicht über Design der Studie 137
Abbildung 9: Kontextualisierungen SLD-Gruppen 141
Abbildung 10: Kontextualisierungen Schüler/-innen 141
Abbildung 11: Informationen zur Vignette erklären, wie für die SLD-Gruppen 142
Abbildung 12: Informationen zur Vignette erklären, wie für die SuS-Gruppe 142
Abbildung 13: Fragebogenseiten zu den Aspekten Erklärtiefe und Fachliche Relevanz für die SLD-Gruppen 144
Abbildung 14: Fragebogenseiten zu den Aspekten Erklärtiefe und Fachliche Relevanz für die SuS-Gruppe 144
Abbildung 15: Modell zur strukturierenden Inhaltsanalyse 221
Abbildung 16: Prozessmodell induktiver Kategorienbildung 225
Abbildung 17: Häufigkeiten der Kodes für die Gesamtstichprobe 233
Abbildung 18: Häufigkeiten der Subkodes für die Gesamtstichprobe 237
Abbildung 19: Quantitative Kontrastierung der Kodes nach Statusgruppen 241
Abbildung 20: Quantitative Kontrastierung der Subkodes nach Statusgruppen 243
Abbildung 21: Quantitative Kontrastierung der Kodes nach Erklärtiefe 250
Abbildung 22: Quantitative Kontrastierung der Subkodes nach Erklärtiefe 252

Literaturverzeichnis

Abraham, Ulf; Baurmann, Jürgen & Feilke, Helmuth. (2015). Materialgestütztes Schreiben. *Praxis Deutsch*, (251), S. 4–12.

Abraham, Ulf & Rothgangel, Martin (2017). Fachdidaktik im Spannungsfeld von "Bildungswissenschaft" und "Fachwissenschaft". In H. Bayrhuber, U. Abraham, V. Frederking, W. Jank, M. Rothgangel & H. J. Vollmer (Hrsg.), *Auf dem Weg zu einer allgemeinen Fachdidaktik* (Fachdidaktische Forschungen, S. 15–21). Münster: Waxmann.

Alber, Kerstin & Neumeister, Nicole (2009). Wortbedeutungserklärungen unter empirischer und didaktischer Perspektive. In M. Krelle & C. Spiegel (Hrsg.), *Sprechen und Kommunizieren. Entwicklungsperspektiven, Diagnosemöglichkeiten und Lernszenarien in Deutschunterricht und Deutschdidaktik* (S. 139–155). Baltmannsweiler: Schneider Hohengehren.

Allhoff, Dieter-W. & Allhoff, Waltraud. (2006). *Rhetorik & Kommunikation. Ein Lehr- und Übungsbuch* (14., völlig neu bearb. und erw. Aufl.). München u.a.: Reinhardt.

Appel, Joachim (2009). Erklären im Fremdsprachenunterricht. In J. Spreckels (Hrsg.), *Erklären im Kontext. Neue Perspektiven aus der Gesprächs- und Unterrichtsforschung* (S. 33–48). Baltmannsweiler: Schneider Hohengehren.

Bailey, Alison L.; Butler, Frances A.; LaFramenta, Charmien & Ong, Christine. (2004). *Towards the Characterization of Academic Language in Upper Elementary Science Classrooms. CSE Report 621*. Los Angeles.

Baumert, Jürgen & Kunter, Mareike. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9*(4), S. 469–520.

Baumert, Jürgen & Kunter, Mareike (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.

Bayer, Klaus. (1981). Einige Aspekte des Sprachhandlungstyps "Erklären". *Deutsche Sprache*, *9*(1), S. 25–43.

Becker-Mrotzek, Michael. (2004). *Schreibentwicklung und Textproduktion. Der Erwerb der Schreibfertigkeit am Beispiel der Bedienungsanleitung.* Radolfzell: Verlag für Gesprächsforschung.

Becker-Mrotzek, Michael; Schneider, Frank & Tetling, Klaus. (2010). *Argumentierendes Schreiben – lehren und lernen. Vorschläge für einen systematischen Kompetenzaufbau in den Stufen 5 bis 8*. Verfügbar unter https://www.schulentwicklung.nrw.de/cms/upload/netzwerk_NfUE/deutsch/argumentieren_einfuehrung_lang.pdf

Becker-Mrotzek, Michael; Schramm, Karen; Thürmann, Eike & Vollmer, Helmut J. (Hrsg.). (2013). *Sprache im Fach. Sprachlichkeit und fachliches Lernen.* Münster: Waxmann.

Becker-Mrotzek, Michael & Vogt, Rüdiger. (2009). *Unterrichtskommunikation. Linguistische Analysemethoden und Forschungsergebnisse* (Germanistische Arbeitshefte, Bd. 38, 2. Aufl.). Berlin: De Gruyter.

Begrich, Lukas; Fauth, Benjamin; Kunter, Mareike & Klieme, Eckhard. (2017). Wie informativ ist der erste Eindruck? Das Thin-Slices-Verfahren zur videobasierten Erfassung des Unterrichts. *Zeitschrift für Erziehungswissenschaft*, *20*(Supplement 1), S. 23–47.

Berkemeier, Anne (2009). „Wissen macht Ah!". SchülerInnen erklären SchülerInnen Orthographieregeln medial. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 57–66). Tübingen: Stauffenburg.

Berkemeier, Anne & Harren, Inga. (2016). "Wie kann man das besser erklären?" Erklärungen in Sachtexten überarbeiten. *Praxis Deutsch*, (256), S. 28–33.

Berliner, David C. (1987). Simple views of effective teaching and a simple theory of classroom instruction. In D. C. Berliner & B. V. Rosenshine (Hrsg.), *Talks to teachers. A Festschrift for N. L. Gage* (S. 93–110). New York: Random House.

Berliner, David C. (2005). The Near Impossibility of Testing for Teacher Quality. *Journal of Teacher Education*, *56*(3), S. 205–213.

Berliner, David C. & Rosenshine, Barak V. (Hrsg.). (1987). *Talks to teachers. A Festschrift for N. L. Gage*. New York: Random House.

Besser, Michael & Krauss, Stefan (2009). Zur Professionalität als Expertise. In O. Zlatkin-Troitschanskaia (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 71–82). Weinheim: Beltz.

Blömeke, Sigrid (2011). Teacher Education and Development Study: Learning to Teach (TEDS-LT) – Erfassung von Lehrerkompetenzen in gering strukturierten Domänen. In S. Blömeke, K. Schwippert, A. Bremerich-Vos, H. Haudeck, G. Kaiser, G. Nold et al. (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (S. 7–24). Münster: Waxmann.

Blömeke, Sigrid; Gustaffson, J.-E. & Shavelson, Richard J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, (233 1), S. 1–13.

Bortz, Jürgen & Döring, Nicola. (2015). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Mit 87 Tabellen* (4., überarb. Aufl.). Berlin: Springer.

Bräuer, Christoph & Wieser, Dorothee (Hrsg.). (2015). *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik*. Wiesbaden: Springer VS.

Bremerich-Vos, Albert. (2019). Zum Professionswissen von (zukünftigen) Deutschlehrkräften. Empirische Befunde und offene Fragen. *Didaktik Deutsch*, *24*(46), S. 47–63.

Bremerich-Vos, Albert & Dämmer, Jutta (2013). Professionelles Wissen im Studienverlauf Lehramt Deutsch. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Keßler et al. (Hrsg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerausbildung aus TEDS-LT* (S. 47–75). Münster: Waxmann.

Bremerich-Vos, Albert; Dämmer, Jutta; Willenberg, Heiner & Schwippert, Knut (2011). Professionelles Wissen von Studierenden des Lehramtes Deutsch. In S. Blömeke, K. Schwippert, A. Bremerich-Vos, H. Haudeck, G. Kaiser, G. Nold et al. (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT* (S. 47–76). Münster: Waxmann.

Brennan, Robert L. & Prediger, Dale J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, *41*(3), S. 687–699.

Brinker, Klaus; Cölfen, Hermann & Pappert, Steffen. (2018). *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden* (Grundlagen der Germanistik, Bd. 29, 9., durchgesehene Auflage). Berlin: Erich Schmidt.

Bromme, Rainer. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens*. Bern: Huber.

Bromme, Rainer; Rambow, Riklef & Nückles, Matthias. (2001). Expertise and estimating what other people know: The influence of professional experience and type of knowledge. *Journal of Experimental Psychology: Applied*, *vol. 7*(issue 4), S. 317–330.

Bromme, Rainer; Rheinberg, Falko; Minsel, Beate; Winteler, Adi & Weidenmann, Bernd (2006). Die Erziehenden und die Lehrenden. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (5., vollst. überarb. Aufl., S. 269–356). Weinheim: Beltz PVU.

Burger, Thomas. (2018). *Rhetorik für Lehrkräfte*. Bad Heilbrunn: Julius Klinkhardt.

Bußmann, Hadumod. (2008). *Lexikon der Sprachwissenschaft* (Kröners Taschenausgabe, Bd. 452, 4. Aufl.). Stuttgart: Alfred Kröner.

Charalambous, Charalambos Y.; Hill, Heather C. & Ball, Deborah L. (2011). Prospective teachers' learning to provide instructional explanations. How does it look and what might it take? *Journal of Mathematics Teacher Education*, *14*(6), S. 441–463.

Clausen, Marten. (2002). *Unterrichtsqualität. Eine Frage der Perspektive?: empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 29). Zugl.: Berlin, Freie Univ., Diss., 2000. Münster: Waxmann.

Collins, Allan; Brown, John S. & Newman, Susan E. (1989). Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics. In L. B. Resnick (Hrsg.), *Knowing, learning, and instruction. Essays in honor of Robert Glaser* (S. 453–494). Hillsdale NJ u.a.: Erlbaum.

Danielson, Charlotte. (2007). *Enhancing professional practice. A framework for teaching* (Gale virtual reference library, 2. Aufl.). Princeton, N.J.: Recording for the Blind & Dyslexic.

Dannerer, Monika (2012). Routiniert vom ersten bis zum letzten Satz? - Die Rolle von Textroutinen in der Erzählentwicklung von Jugendlichen. In H. Feilke & K. Lehnen (Hrsg.), *Schreib- und Textroutinen. Theorie, Erwerb und didaktisch-mediale Modellierung* (Forum Angewandte Linguistik - F.A.L, Bd. 52, S. 101–124). Frankfurt a. M.: Peter Lang.

Deci, Edward L. & Ryan, Richard M. (1985). *Intrinsic motivation and self-determination in human behavior* (Perspectives in social psychology). New York, NY: Plenum Press.
Deppermann, Arnulf (2018). Wissen im Gespräch. In K. Birkner & N. Janich (Hrsg.), *Handbuch Text und Gespräch* (Handbücher Sprachwissen (HSW), Bd. 5, S. 104–142). Berlin: De Gruyter.
Duffy, Gerald G. (2014). *Explaining Reading:. A Resource for Explicit Teaching of the Common Core Standards* (3. Aufl.). New York: Guilford Publications.
Dutke, Stephan & Reimer, Torsten. (2000). Evaluation of two types of online help for application software. *Journal of Computer Assisted Learning*, *16*(4), S. 307–315.
Eckert, Marcus & Sieland, Bernhard (2016). Psychologie der Lehrerpersönlichkeit. In M. K. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion: Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (Schule und Gesellschaft, Bd. 24, 3., überarb. u. aktual. Aufl. 2017, S. 147–165). Wiesbaden: Springer VS.
Ehlich, Konrad (2009). Sprechen im Deutschunterricht - didaktische Denkanstöße. In M. Krelle & C. Spiegel (Hrsg.), *Sprechen und Kommunizieren. Entwicklungsperspektiven, Diagnosemöglichkeiten und Lernszenarien in Deutschunterricht und Deutschdidaktik* (S. 8–14). Baltmannsweiler: Schneider Hohengehren.
Ehlich, Konrad & Rehbein, Jochen. (1986). *Muster und Institution. Untersuchungen zur schulischen Kommunikation* (Kommunikation und Institution, Bd. 15). Tübingen: Narr.
Erath, Kirstin. (2017). *Mathematisch diskursive Praktiken des Erklärens* (Research, Band 27). Wiesbaden: Springer Fachmedien.
Erickson, Frederick (2011). On Noticing Teacher Noticing. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Hrsg.), *Mathematics teacher noticing. Seeing through teachers' eyes* (Studies in Mathematical Thinking and Learning, S. 17–34). New York: Routledge.
Ernst-Weber, Diana. (2019). *Mündliche Erklärfähigkeiten von Schülerinnen und Schülern der Sekundarstufe I. Vergleichende Gesprächsanalysen*. Tübingen: Stauffenburg.
Fauth, Benjamin; Decristan, Jasmin; Rieser, Svenja; Klieme, Eckhard & Büttner, Gerhard. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive. Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, *28*(3), S. 127–137.
Feigenspan, Katja & Michalak, Magdalena (2018). Sprachbewusster Biologieunterricht durch Scaffolding. In U. Spörhase-Eichmann & W. Ruppert (Hrsg.), *Biologie Methodik. Handbuch für die Sekundarstufe I und II* (4., überarbeitete Auflage, S. 94–99). Berlin: Cornelsen.
Feilke, Helmuth (2010, Mai). *„Aller guten Dinge sind drei" – Überlegungen zu Textroutinen & literalen Prozeduren.* Fest-Platte für Gerd Fritz. Verfügbar unter http://www.festschrift-gerd-fritz.de/files/feilke_2010_literale-prozeduren-und-textroutinen.pdf
Feilke, Helmuth (2012). Was sind Textroutinen? - Zur Theorie und Methodik des Forschungsfeldes. In H. Feilke & K. Lehnen (Hrsg.), *Schreib- und Textroutinen. Theorie, Erwerb und didaktisch-mediale Modellierung* (Forum Angewandte Linguistik - F.A.L, Bd. 52, S. 1–31). Frankfurt a. M.: Peter Lang.
Feilke, Helmuth (2013). Bildungssprache und Schulsprache am Beispiel literal-argumentativer Kompetenzen. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (S. 113–131). Münster: Waxmann.
Feilke, Helmuth (2014). Argumente für eine Didaktik der Textprozeduren. In T. Bachmann & H. Feilke (Hrsg.), *Werkzeuge des Schreibens. Beiträge zu einer Didaktik der Textprozeduren* (S. 11–34). Stuttgart: Fillibach bei Klett.
Fend, Helmut. (1980). *Theorie der Schule*. München u.a.: Urban & Schwarzenberg.
Fiehler, Reinhard (2007). *Formen und interaktive Strukturen des Erklärens.* Unveröffentlichtes Handout eines Vortrags des Symposions „Erklären in unterrichtlichen und außerunterrichtlichen Kontexten" an der Pädagogischen Hochschule, Freiburg im Breisgau.
Findeisen, Stefanie. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen. Eine Untersuchung zum Erklären im Rechnungswesen*. Wiesbaden: Springer Fachmedien.
Fleischer, Jürgen. (1987). *Verbalisieren des Denkens als Problemlösungsstrategie. Empirische Untersuchung zu den Wirkungen des lauten Denkens auf die Lösung von Problemaufgaben*. Dissertation zur Erlangung des Grades eines Doktors der Philosophie der Universität Bremen.
Flick, Uwe. (2019). *Qualitative Sozialforschung. Eine Einführung* (Rororo Rowohlts Enzyklopädie, Bd. 55694, 9. Auflage, Originalausgabe). Reinbek bei Hamburg: Rowohlt.

Frei, Mario; Hofmann, Bernhard & Puffer, Gabriele (2021, in Vorb.). Musiktheoretische Sachverhalte im Musikunterricht erklären. In A. Schilcher, S. Krauss, A. Lindl & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären.* Weinheim und Basel: Beltz Juventa.

Friedrich Verlag (Hrsg.). (2017). *Deutsch per Smartphone* (Praxis Deutsch, Bd. 265).

Friedrich Verlag (Hrsg.). (2021). *Erklärtexte lesen und schreiben* (Praxis Deutsch, Bd. 285).

Friesen, Marita; Kuntze, Sebastian & Vogel, Markus (2018). Videos, Texte oder Comics? Die Rolle des Vignettenformats bei der Erhebung fachdidaktischer Analysekompetenz zum Umgang mit Darstellungen im Mathematikunterricht. In J. Rutsch, M. Rehm, M. Vogel, M. Seidenfuß & T. Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung. Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (S. 153–177). Wiesbaden: Springer.

Fussell, Susan R. & Krauss, Robert M. (1992). Coordination of knowledge in communication. Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), S. 378–391.

Gaier, Lisa & Schilcher, Anita (2021, in Vorb.). Gutes Erklären im Deutschunterricht – eine empirische Studie zu unterschiedlichen Beobachterperspektiven. In A. Schilcher, S. Krauss, A. Lindl & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären.* Weinheim und Basel: Beltz Juventa.

Gebhardt, Markus; Rauch, Dominique P.; Mang, Julia; Sälzer, Christine & Stanat, Petra (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 275–308). Münster: Waxmann.

Gindele, Verena & Voss, Thamar. (2017). Pädagogisch-psychologisches Wissen: Zusammenhänge mit Indikatoren des beruflichen Erfolgs angehender Lehrkräfte. *Zeitschrift für Bildungsforschung*, *7*(3), S. 255–272.

Gogolin, Ingrid & Lange, Imke (2011). Bildungssprache und Durchgängige Sprachbildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel. Mehrsprachigkeit* (S. 107–127). Wiesbaden: VS Verlag für Sozialwissenschaften.

Göllner, Richard; Praetorius, Anna-Katharina; Wagner, Wolfgang; Lenske, Gerlinde & Fauth, Benjamin. (2020). Who sees what? *Zeitschrift für Pädagogik Beiheft*, (66), S. 138–155.

Graham, Steve & Perin, Dolores. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools. A report to Carnegie Corporation of New York.* Washington, D.C.: Alliance for Excellent.

Grießhaber, Wilhelm (2010). (Fach-)Sprache im zweitsprachlichen Fachunterricht. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (S. 37–53). Tübingen: Narr.

Gruehn, Sabine. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 12). Zugl.: Berlin, Freie Univ., Diss., 1998. Münster: Waxmann.

Grundler, Elke. (2011). *Kompetent argumentieren. Ein gesprächsanalytisch fundiertes Modell* (Stauffenburg Linguistik, Bd. 56). Zugl.: Ludwigsburg, Pädag. Hochsch., Diss., 2011. Tübingen: Stauffenburg.

Gunga, Eileen. (2022, in Vorb.). *Einfluss des Sprech- und Körperausdrucks von Lehrkräften auf die wahrgenommene Qualität von schulischen Erklärungen – Eine Multilevelanalyse und auditiv-akustische Untersuchung von Lehrer*innenstimmen*.

Gunga, Eileen; Schilcher, Anita; Kranich, Wieland & Gegner, Christian (2021, in Vorb.). FALKE-SE (Sprechwissenschaft/Sprecherziehung): Einfluss des Sprech- und Körperausdrucks von Lehrkräften auf die wahrgenommene Qualität von schulischen Erklärungen. In A. Schilcher, S. Krauss, A. Lindl & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären.* Weinheim und Basel: Beltz Juventa.

Hanisch, Anna-Katharina. (2018). *Kognitive Aktivierung im Rechtschreibunterricht* (Internationale Hochschulschriften, Band 648). Dissertation. Münster: Waxmann.

Harren, Inga (2009). Formen der Begriffsarbeit – wie im Unterrichtsgespräch Inhalte und Fachtermini verknüpft werden. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 151–168). Tübingen: Stauffenburg.

Harris, Karen R.; Graham, Steve; Mason, Linda H. & Friedlaender, Barbara. (2010). *Powerful writing strategies for all students* (2nd printing). Baltimore: Paul H. Brookes.

Hatano, Giyoo & Inagaki, Kayoko (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma & K. Hakuta (Hrsg.), *Child development and education in Japan* (A Series of Books in Psychology, S. 262–272). New York: Freeman.

Hattie, John A. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Häusl, Theresa. (2019). *Gut erklärt? Auswertung einer Studie des FALKE-Projekts zur offenen Bewertung von Lehrererklärungen im Hinblick auf die Kriterien Adressatenorientierung und Struktur*. Zulassungsarbeit (unveröffentlicht). Universität Regensburg.

Heinze, Aiso; Herwartz-Emden, Leonie & Reiss, Kristina. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, *53*(4), S. 562–581.

Helmke, Andreas. (2014). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (5. aktual. Aufl., berücks. die Hattie-Studien). Seelze-Velber: Klett/Kallmeyer.

Hempel, Carl G. (1977). *Aspekte wissenschaftlicher Erklärung* (De-Gruyter-Studienbuch: Grundlagen der Kommunikation). Berlin u.a.: De Gruyter.

Hempel, Carl G. & Oppenheim, Paul. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, (15), S. 135–175.

Hohenstein, Christiane. (2006). *Erklärendes Handeln im wissenschaftlichen Vortrag. Ein Vergleich des Deutschen mit dem Japanischen* (Studien Deutsch, Bd. 36). Univ., Diss.--Hamburg, 2003. München: Iudicium.

Hohenstein, Christiane (2009). Interkulturelle Aspekte des Erklärens. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 37–55). Tübingen: Stauffenburg.

Holzberger, Doris & Kunter, Mareike (2016). Unterricht aus der Perspektive der Pädagogischen Psychologie und empirischen Unterrichtsforschung. In J. Möller, M. Köller & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung: Schule und Unterricht – Lehren und Lernen* (S. 39–52). Seelze: Klett Kallmeyer.

Jahn, Gloria; Seidel, Tina; Stürmer, Kathleen & Prenzel, Manfred. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden. Eine Scaling-up Studie des Observe-Projekts. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *46*(4), S. 171–180.

Jahr, Silke (2000). Vertextungsmuster Explikation. In K. Brinker (Hrsg.), *Text- und Gesprächslinguistik 1. Ein internationales Handbuch zeitgenössischer Forschung* (Handbücher zur Sprach- und Kommunikationswissenschaft, S. 385–397).

Josefy, Almuth (2009). Erklärprozesse im Fach Englisch. Zur Vermittlung grammatischer Inhalte in der Sekundarstufe. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 79–92). Tübingen: Stauffenburg.

Keller, Melanie M.; Hoy, Anita W.; Goetz, Thomas & Frenzel, Anne C. (2016). Teacher Enthusiasm. Reviewing and Redefining a Complex Construct. *Educational Psychology Review*, *28*(4), S. 743–769.

Keßler, Jörg-Uwe (2009). Englischdidaktik in „Erklärungsnot". Implizites und explizites Wissen und die Rolle der Bewusstmachung im schulischen Englischerwerb. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 93–107). Tübingen: Stauffenburg.

Kiel, Ewald. (1999). *Erklären als didaktisches Handeln* (Bibliotheca academica /Reihe Pädagogik, Bd. 1). Univ., Habil.-Schr.--Göttingen, 1998. Würzburg: Ergon-Verl.

Kirschner, Sophie; Szcudlek, Melanie; Tepner, Oliver; Borowski, Andreas; Fischer, Hans E.; Lenske, Gerlinde et al. (2017). Professionswissen in den Naturwissenschaften (ProwiN). In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 113–130). Wiesbaden: Springer VS.

Klauer, Karl J. & Leutner, Detlev. (2012). *Lehren und Lernen. Einführung in die Instruktionspsychologie.* Weinheim: Beltz.

Klein, Josef. (1987). *Die konklusiven Sprechhandlungen. Studien zur Pragmatik, Semantik, Syntax und Lexik von Begründen, Erklären-warum, Folgern und Rechtfertigen* (Reihe Germanistische Linguistik, Bd. 76). Berlin: De Gruyter.

Klein, Josef (2009). ERKLÄREN-WAS, ERKLÄREN-WIE, ERKLÄREN-WARUM. Typologie und Komplexität zentraler Akte der Welterschließung. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 25–36). Tübingen: Stauffenburg.

Kleinbub, Iris (2016). Ein-Blick ins Klassenzimmer: deutschdidaktische Unterrichtsforschung am Beispiel der Videostudie VERA -Gute Unterrichtspraxis. In M. Krelle & W. Senn (Hrsg.), *Qualitäten von Deutschunterricht* (S. 15–36). Stuttgart: Fillibach bei Klett.

Klieme, Eckhard; Eichler, Wolfgang; Helmke, Andreas; Lehmann, Rainer H.; Nold, Günther; Rolff, Hans-Günter et al. (Hrsg.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (Beltz Pädagogik). Weinheim: Beltz.

Klieme, Eckhard; Lipowsky, Frank; Rakoczy, Katrin & Ratzka, Nadja (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts "Pythagoras". In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.

Klieme, Eckhard; Schümer, Gundel & Knoll, Steffen (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur" und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS - Impulse für Schule und Unterricht: Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–58). Bonn: Bundesministerium für Bildung und Forschung.

Kluge, Friedrich & Seebold, Elmar. (2002). *Etymologisches Wörterbuch der deutschen Sprache. EBookPlus* (24., aktualisierte und erw. Aufl.). Berlin: De Gruyter.

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2003). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Beschluss vom 04.12.2003.*

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2004a). *Bildungsstandards im Fach Deutsch für den Hauptschulabschluss. Beschluss vom 15.10.2004.*

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2004b). *Bildungsstandards im Fach Deutsch für den Primarbereich. Beschluss vom 15.10.2004.*

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2012). *Bildungsstandards im Fach Deutsch für die allgemeine Hochschulreife. Beschluss vom 18.10.2012.*

Koole, Tom (2009). Erklären in der Mathematikklasse. Eine angewandte Konversationsanalyse. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 109–121). Tübingen: Stauffenburg.

Köster, Juliane. (2016). *Aufgaben im Deutschunterricht. Wirksame Lernangebote und Erfolgskontrollen* (Praxis Deutsch, 2. Auflage). Seelze: Klett/Kallmeyer.

Kounin, Jacob S. (1976). *Techniken der Klassenführung* (Abhandlungen zur pädagogischen Psychologie, Neue Folge, Bd. 6, 1. Aufl.). Bern: Huber.

Krauss, Stefan; Blum, Werner; Brunner, Martin; Neubrand, Michael; Baumert, Jürgen; Kunter, Mareike et al. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135–161). Münster: Waxmann.

Krauss, Stefan & Bruckmaier, Georg (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. u. erg. Aufl., S. 241–261). Münster: Waxmann.

Krauss, Stefan; Bruckmaier, Georg; Lindl, Alfred; Hilbert, Sven; Binder, Karin; Steib, Nicole et al. (2020). Competence as a continuum in the COACTIV study: the "cascade model". *ZDM*, *52*(2), 311-327.

Krauss, Stefan; Lindl, Alfred; Schilcher, Anita; Fricke, Michael; Göhring, Anja; Hofmann, Bernhard et al. (Hrsg.). (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik; mit neuen Daten aus der COACTIV-Studie*. Münster: Waxmann.

Krauss, Stefan; Lindl, Alfred; Schilcher, Anita & Tepner, Oliver (2017). Das Forschungsprojekt FALKO - ein einleitender Überblick. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann et al. (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik; mit neuen Daten aus der COACTIV-Studie* (S. 9–65). Münster: Waxmann.

Krauss, Stefan; Neubrand, Michael; Blum, Werner; Baumert, Jürgen; Brunner, Martin; Kunter, Mareike et al. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematikdidaktik*, *29*(3/4), S. 223–258.

Krippendorff, Klaus. (1980). *Content Analysis. An introduction to its methodology*. London: Sage Publications.

Kulgemeyer, Christoph. (2017). *Physik erklären. Kumulative Habilitationsschrift*. Universität Bremen.

Kulgemeyer, Christoph. (2018). A Framework of Effective Science Explanation Videos Informed by Criteria for Instructional Explanations. *Research in Science Education,* S. 1–22.

Kulgemeyer, Christoph (2019). Qualitätskriterien zur Gestaltung naturwissenschaftlicher Erklärvideos. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe.* (S. 258–288).

Kulgemeyer, Christoph & Tomczyszyn, Elisabeth. (2015). Physik erklären – Messung der Erklärensfähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, *21*(1), S. 111–126.

Kunter, Mareike & Baumert, Jürgen. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), S. 231–251.

Kunter, Mareike; Baumert, Jürgen; Blum, Werner; Klusmann, Uta & Neubrand, Michael (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.

Kunter, Mareike & Ewald, Silvia (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 9–31). Münster: Waxmann.

Kunter, Mareike; Klusmann, Uta; Dubberke, Thamar; Baumert, Jürgen; Blum, Werner; Brunner, Martin et al. (2007). Linking aspects of teacher competence to their instruction: Results from the COACTIV project. In M. Prenzel (Hrsg.), *Studies on the educational quality of schools: The final report on the DFG Priority Programme* (S. 39–59). Münster: Waxmann.

Kunter, Mareike & Trautwein, Ulrich. (2018). *Psychologie des Unterrichts* (StandardWissen Lehramt, Bd. 3895). Paderborn: Schöningh.

Kuntze, Sebastian (2016). Expertisemerkmale von Mathematiklehrkräften und anforderungshaltige Situierungen - Fragen an Untersuchungsdesigns. In F. Caluori, H. Linneweber-Lammerskitten & C. Streit (Hrsg.), *Beiträge zum Mathematikunterricht 2015. Vorträge auf der 49. Tagung für Didaktik der Mathematik vom 09.02.2015 bis 13.02.2015 in Basel* (S. 528–531). Münster: WTM-Verlag.

Landis, J. R. & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33,* S. 159–174.

Lang, Ewald (1974). Erklärungstexte. In F. Daneš (Hrsg.), *Probleme der Textgrammatik* (Studia grammatica, S. 147–181). Berlin: Akad.-Verl.

Lehnen, Katrin (2012). Erwerb wissenschaftlicher Textroutinen. In H. Feilke & K. Lehnen (Hrsg.), *Schreib- und Textroutinen. Theorie, Erwerb und didaktisch-mediale Modellierung* (Forum Angewandte Linguistik - F.A.L, Bd. 52, 33-60). Frankfurt a. M.: Peter Lang.

Lehner, Martin. (2018). *Erklären und Verstehen. Eine kleine Didaktik der Vermittlung.* Bern: Haupt.

Leinhardt, Gaea (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Hrsg.), *Handbook of research on teaching* (4. Aufl., S. 333–357). Washington, D.C.: American Educational Research Association.

Leisen, Josef. (2007). Das Erklären im Unterricht. *Der mathematische und naturwissenschaftliche Unterricht*, (60 (8)), S. 459–462.

Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich.* Münster: Waxmann.

Lindl, Alfred & Krauss, Stefan (2017). Transdisziplinäre Perspektiven auf domänenspezifische Lehrerkompetenzen. Eine Metaanalyse zentraler Resultate des Forschungsprojektes FALKO. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann et al. (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik; mit neuen Daten aus der COACTIV-Studie* (S. 381–438). Münster: Waxmann.

Lindow, Ina (2015). Systematische Schwierigkeiten in der schulischen Förderung literarischer Rezeptionskompetenz. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 35–48). Wiesbaden: Springer VS.

Lipowsky, Frank (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Lehrbuch, 2. vollständ. überarb. u. aktual. Aufl., S. 69–105). Berlin: Springer.

Lotz, Miriam. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule. Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Wiesbaden: Springer VS.

Ludwig, Otto & Spinner, Kaspar H. (2000). Mündlich und schriftlich argumentieren. *Praxis Deutsch*, *27*(160), S. 16–22.

Maag Merki, Katharina & Werner, Silke (2014). Erfassung und Bewertung professioneller Kompetenz von Lehrpersonen. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. u. erg. Aufl., 745-763). Münster: Waxmann.

Macaro, Ernesto. (2013). Students' strategies in response to teachers' second language explanations of lexical items. *The Language Learning Journal*, *45*(3), S. 352–367.

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist*, *59*(1), S. 14–19.

Mayring, Philipp. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (Pädagogik, 11., aktual. und überarb. Auflage). Weinheim: Beltz.

Michalak, Magdalena; Lemke, Valerie & Goeke, Marius. (2015). *Sprache im Fachunterricht. Eine Einführung in Deutsch als Zweitsprache und sprachbewussten Unterricht* (Narr Studienbücher). Tübingen: Narr Francke Attempto.

Miller, Kevin F. (2011). What Educators Can Learn From Video-Based Research in Other Fields. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Hrsg.), *Mathematics teacher noticing. Seeing through teachers' eyes* (Studies in Mathematical Thinking and Learning, S. 51–65). New York: Routledge.

Möller, Kornelia (2012). Konstruktion vs. Instruktion oder Konstruktion durch Instruktion? Konstruktionsfördernde Unterstützungsmaßnahmen im Sachunterricht. In H. Giest, E. Heran-Dörr & C. Archie (Hrsg.), *Lernen und Lehren im Sachunterricht. Zum Verhältnis von Konstruktion und Instruktion* (Probleme und Perspektiven des Sachunterrichts, Band 22, S. 37–50). Bad Heilbrunn: Julius Klinkhardt.

Möller, Kornelia; Kleickmann, Thilo & Lange, Kim (2013). Naturwissenschaftliches Lernen im Übergang von der Grundschule zur Sekundarstufe. In H. E. Fischer & E. Sumfleth (Hrsg.), *nwu-essen. 10 Jahre Essener Forschung zum naturwissenschaftlichen Unterricht: Elke Sumfleth* (Studien zum Physik- und Chemielernen, Bd. 100, S. 57–120). Berlin: Logos.

Morek, Miriam. (2012). *Kinder erklären – Interaktionen in Familie und Unterricht im Vergleich*. Tübingen: Stauffenburg.

Morek, Miriam (2013a). Erklären. In B. Rothstein & C. Müller (Hrsg.), *Kernbegriffe der Sprachdidaktik Deutsch. Ein Handbuch* (S. 70–72). Baltmannsweiler: Schneider Hohengehren.

Morek, Miriam (2013b). Erklärung. In B. Rothstein & C. Müller (Hrsg.), *Kernbegriffe der Sprachdidaktik Deutsch. Ein Handbuch* (S. 72–75). Baltmannsweiler: Schneider Hohengehren.

Morek, Miriam; Heller, Vivien & Quasthoff, Uta M. (2017). Erklären und Argumentieren. Modellierungen und empirische Befunde zu Strukturen und Varianzen. In I. Meißner & E. L. Wyss (Hrsg.), *Begründen - Erklären - Argumentieren. Konzepte und Modellierungen in der Angewandten Linguistik* (Stauffenburg Linguistik, Band 93, S. 11–46). Tübingen: Stauffenburg.

Naftulin, Donald H.; Ware, John E. & Donnelly, Frank A. (1973). The Doctor Fox Lecture: A Paradigm of Educational Seduction. *Journal of Medical Education*, (48), S. 630–635.

Neuland, Eva & Peschel, Corinna. (2013). *Einführung in die Sprachdidaktik*. Stuttgart: J. B. Metzler.
Neumeister, Nicole. (2011). *(Wie) Wird im Deutschunterricht erklärt? Wissensvermittelnde Handlungen im Sprachunterricht der Sekundarstufe I*. Dissertation an der Pädagogischen Hochschule Ludwigsburg.
Neumeister, Nicole (Hrsg.). (2019). *Erklären* (Deutsch 5-10, Bd. 59): Friedrich.
Neumeister, Nicole & Vogt, Rüdiger (2015). Erklären im Unterricht. In M. Becker-Mrotzek & W. Ulrich (Hrsg.), *Mündliche Kommunikation und Gesprächsdidaktik* (Deutschunterricht in Theorie und Praxis, DTP; Handbuch zur Didaktik der deutschen Sprache und Literatur in elf Bänden / hrsg. von Winfried Ulrich ; Bd. 3, 3., korrigierte Aufl., S. 562–583). Baltmannsweiler: Schneider Hohengehren.
Nickerson, Raymond S.; Baddeley, Alan & Freeman, Barbara. (1987). Are People's Estimates of what other People know influcenced by what they themselves know? *Acta Psychologica*, (64), S. 245–259.
Nückles, Matthias. (2001). *Perspektivenübernahme von Experten in der Kommunikation mit Laien. Eine Experimentalserie im Internet* (Internationale Hochschulschriften, Bd. 368). Zugl.: Münster (Westfalen), Univ., Diss., 2000. Münster: Waxmann.
Ossner, Jakob. (2008). *Sprachdidaktik Deutsch: Eine Einführung für Studierende* (2. Aufl.). Paderborn: Schöningh.
Pabst-Weinschenk, Marita. (2016). *Stimmlich stimmiger Unterricht. Professionelle Kommunikation und Rhetorik*. Göttingen: Vandenhoeck Ruprecht.
Penzel, Joachim. (2015). Anschaulich erklären. Zur Arbeit mit Iconic Concept Mapping in anderen Unterrichtsfächern - eine Transferleistung der Kunstpädagogik. *BDK Mitteilungen*, *51*(1), S. 29–32.
Pflugmacher, Torsten (2015). Verstehen verstehen -verstehen. Literaturpädagogische Professionalität als Herausforderung von Literaturdidaktik und LiteraturvermittIung. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 131–158). Wiesbaden: Springer VS.
Philipp, Maik. (2014). Einblicke in die Schreibprozesse gewähren: Modellieren. *Praxis Grundschule*, *37*(1), S. 28–31.
Pissarek, Markus & Schilcher, Anita (2015). Fachspezifische Lehrerkompetenzen im Fach Deutsch messen? Modellierung und Konstruktvalidierung eines Erhebungsinstruments im Rahmen der Projektgruppe FALKO Regensburg. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 321–340). Wiesbaden: Springer VS.
Pissarek, Markus & Schilcher, Anita (2017). FALKO-D: Die Untersuchung des Professionswissens von Deutschlehrenden. Entwicklung eines Messinstruments zur fachspezifischen Lehrerkompetenz und Ergebnisse zu dessen Validierung. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann et al. (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik; mit neuen Daten aus der COACTIV-Studie* (S. 67–111). Münster: Waxmann.
Praetorius, Anna-Katharina; Decristan, Jasmin; Hess, Miriam & Holzberger, Doris. (2020). Oberflächen- und Tiefenmerkmale. *Zeitschrift für Pädagogik Beiheft*, (66), S. 102–116.
Praetorius, Anna-Katharina; Klieme, Eckhard; Herbert, Benjamin & Pinger, Petra. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, *50*(3), S. 407–426.
Praetorius, Anna-Katharina; Lenske, Gerlinde & Helmke, Andreas. (2012). Observer ratings of instructional quality. Do they fulfill what they promise? *Learning and Instruction*, *22*(6), S. 387–400.
Praetorius, Anna-Katharina; Pauli, Christine; Reusser, Kurt; Rakoczy, Katrin & Klieme, Eckhard. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, *31,* S. 2–12.
Prediger, Susanne & Erath, Kirstin. (2014). Content or Interaction, or both? Synthesizing two German traditions in a video study on learning to explain in mathematics classroom microcultures. *Eurasia Journal of Mathematics, Science & Technology Education*, (10(4)), 313–327.
Prüfer, Peter & Rexroth, Margit. (2005). *Kognitive Interviews.* Mannheim: Zentrum für Umfragen, Methoden und Analysen.
Quasthoff, Uta M. & Hartmann, Dietrich. (1982). Bedeutungserklärungen als empirischer Zugang zur Wortbedeutung. *Deutsche Sprache*, (10), S. 97–118.
Quasthoff, Uta M. & Heller, Vivien (2014). Mündlichkeit und Schriftlichkeit aus sprachwissenschaftlicher und sprachdidaktischer Sicht. Grundlegende Ein-/Ansichten und methodischen Anregungen. In A. Neumann

(Hrsg.), *Empirische Methoden der Deutschdidaktik. Audio- und videografierende Unterrichtsforschung* (S. 6–37).
Quittenbaum, Nancy. (2016). *Training für direkte Instruktion. Die Entwicklung und Erprobung eines Kommunikationstrainings für den Unterricht mit direkter Instruktion*. Bad Heilbrunn: Julius Klinkhardt.
Reinert, Johannes (2015). Unterrichtliches Brauchtum und Kommunikationsorientierung: Spannungen bei der unterrichtlichen Arbeit mit schulischen Textformen am Beispiel der Inhaltsangabe. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 227–251). Wiesbaden: Springer VS.
Reinmann, Gabi & Mandl, Heinz (2006). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (5., vollst. überarb. Aufl., S. 613–658). Weinheim: Beltz PVU.
Reiss, Kristina; Weis, Mirjam; Klieme, Eckhard & Köller, Olaf (Hrsg.). (2019). *Pisa 2018. Grundbildung im internationalen Vergleich. Zusammenfassung*. Münster: Waxmann.
Renkl, Alexander. (1997). *Lernen durch Lehren. Zentrale Wirkmechanismen beim kooperativen Lernen* (DUV-Psychologie). Zugl.: München, Univ., Habil.-Schr., 1996. Wiesbaden: Dt. Univ.-Verl.
Renkl, Alexander & Wittwer, Jörg (2006). Instruktionale Erklärungen beim Erwerb kognitiver Fertigkeiten: sechs Thesen zu einer oft vergeblichen Bemühung. In I. Hosenfeld (Hrsg.), *Schulische Leistung. Grundlagen, Bedingungen, Perspektiven* (S. 205–223). Münster: Waxmann.
Resnick, Lauren B. & Hall, Megan W. (1998). Learning organizations for sustainable education reform. *Daedalus*, *127*(4), S. 98–118.
Reusser, Kurt (2006). Konstruktivismus - vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In M. Baer, M. Fuchs, P. Füglister, K. Reusser & H. Wyss (Hrsg.), *Didaktik auf psychologischer Grundlage. Von Hans Aeblis kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung* (S. 151–168). Bern: h.e.p. verlag ag.
Rezat, Sara. (2011). Schriftliches Argumentieren. Zur Ontogenese konzessiver Argumentationskompetenz. *Didaktik Deutsch*, (31), S. 50–67.
Röhrich, Lutz. (2003). *Lexikon der sprichwörtlichen Redensarten* (Bd. 1, 3. Auflage, 3 Bände). Freiburg im Breisgau: Herder.
Scardamalia, Marlene & Bereiter, Carl (1983). The Development of Evaluative, Diagnostic and Remedial Capabilities in Children's Composing. In M. Martlew (Hrsg.), *The Psychology of written language* (Wiley series in developmental psychology, S. 67–95). Chichester West Sussex: Wiley.
Scheffel, Lars. (2019). *Erklären im Unterricht. Taktiken und Strategien*. Weinheim: Beltz.
Scherer, Ronny; Nilsen, Trude & Jansen, Malte. (2016). Evaluating Individual Students' Perceptions of Instructional Quality: An Investigation of their Factor Structure, Measurement Invariance, and Relations to Educational Outcomes. *Frontiers in psychology*, *7,* S. 1–16.
Scherf, Daniel (2015). Von ‚ich gehe in sämtlichen Fächern (…) mit Texten anders um' bis ‚die Zeit haben wir ja gar nicht": Vier Fallskizzen zum Umgang mit lesedidaktischen Wissensangeboten. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 71–87). Wiesbaden: Springer VS.
Schilcher, Anita (Hrsg.). (2015). *Märchen* (Praxis Grundschule, Bd. 5): Westermann.
Schilcher, Anita & Krauss, Stefan (2019). Anforderungen an die Lehrerbildung aus fachdidaktischer Perspektive. In N. McElvany, W. Bos, H. G. Holtappels, A. Ohle-Peters, G. Bellenberg, A. Gröschner et al. (Hrsg.), *Bedingungen und Effekte von Lehrerbildung, Lehrkraftkompetenzen und Lehrkrafthandeln* (Dortmunder Symposium der Empirischen Bildungsforschung, Bd. 4, S. 29–49). Münster: Waxmann.
Schilcher, Anita; Krauss, Stefan; Lindl, Alfred & Hilbert, Sven (Hrsg.). (2021, in Vorb.). *Fachspezifische Lehrerkompetenzen im Erklären*. Weinheim und Basel: Beltz Juventa.
Schilcher, Anita; Krauss, Stefan; Rincke, Karsten & Hilbert, Sven (2017). Ausblick - Aus FALKO wird FALKE Fachspezifische Lehrerkompetenz im Erklären. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann et al. (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik; mit neuen Daten aus der COACTIV-Studie* (S. 439–451). Münster: Waxmann.

Schmidt-Thieme, Barbara. (2009). Erklären als fachspezifische Kompetenz in fächerübergreifender Perspektive. *Beiträge zur Mathematikdidaktik 2009 online. Beiträge auf der 43. Tagung für Didaktik der Mathematik*.

Schmölzer-Eibinger, Sabine (2013). Sprache als Medium des Lernens im Fach. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (S. 25–40). Münster: Waxmann.

Schmölzer-Eibinger, Sabine & Fanta, Johanna (2014). Erklären lernen. Ein prozedurorientiertes didaktisches Modell zur Förderung literaler Handlungskompetenz am Beispiel des Erklärens. In T. Bachmann & H. Feilke (Hrsg.), *Werkzeuge des Schreibens. Beiträge zu einer Didaktik der Textprozeduren* (S. 157–175). Stuttgart: Fillibach bei Klett.

Schnotz, Wolfgang. (2016). Learning and Instruction. A review of main research lines during recent decades. *Zeitschrift für Erziehungswissenschaft*, *19*(1), S. 101–119.

Schoenbach, Ruth. (2006). *Lesen macht schlau. Neue Lesepraxis für weiterführende Schulen*. Berlin: Cornelsen Scriptor.

Schopf, Christiane; Zwischenbrugger, Andrea & Baier, Rainer. (2015). *Handbuch verständlich erklären. Eine Heuristik mit Beispielen aus Betriebswirtschaft, Rechnungswesen, Volkswirtschaft und Wirtschaftsinformatik* (Wissen plus Wissenschaft). Wien: Manz.

Schworm, Silke & Renkl, Alexander. (2006). Computer-supported example-based learning. When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), S. 426–445.

SDD. (2018). *Mitgliederbrief 57*.

Seidel, Tina; Blomberg, Geraldine & Stürmer, Kathleen (2010). „Observer" – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 296–306). Zeitschrift für Pädagogik. Beiheft 56. Weinheim: Beltz.

Seidel, Tina & Shavelson, Richard J. (2007). Teaching Effectiveness Research in the Past Decade. The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, *77*(4), S. 454–499.

Seidel, Tina & Stürmer, Kathleen. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, *51*(4), S. 739–771.

Seidel, Tina & Thiel, Felicitas. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaft*, *20*(S1), S. 1–21.

Selting, Margret; Auer, Peter; Barden, Birgit; Bergmann, Jörg; Couper-Kuhlen, Elizabeth; Günthner, Susanne et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte*, (173), S. 91–122.

Settekorn, Wolfgang (1991). Konversationelle Erklärungen. Zur Beziehung von Erklärungsbegriffen in wissenschaftstheoretischen und konversationsanalytischen Überlegungen. In E. Gülich, U. Dausendschön-Gay & U. Krafft (Hrsg.), *Linguistische Interaktionsanalysen. Beiträge zum 20. Romanistentag 1987* (Linguistische Arbeiten, Bd. 254, S. 235–262). Tübingen Germany: Max Niemeyer.

Sevian, Hannah & Gonsalves, Lisa. (2008). Analysing how Scientists Explain their Research. A rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, *30*(11), S. 1441–1467.

Sherin, Miriam G.; Jacobs, Victoria R. & Philipp, Randolph A. (Hrsg.). (2011a). *Mathematics teacher noticing. Seeing through teachers' eyes* (Studies in Mathematical Thinking and Learning). New York: Routledge.

Sherin, Miriam G.; Jacobs, Victoria R. & Philipp, Randolph A. (2011b). Situating the Study of Teacher Noticing. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Hrsg.), *Mathematics teacher noticing. Seeing through teachers' eyes* (Studies in Mathematical Thinking and Learning, S. 3–13). New York: Routledge.

Sherin, Miriam G. & van Es, Elizabeth A. (2009). Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, *60*(1), S. 20–37.

Shulman, Lee S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), S. 4–14.

Shulman, Lee S. (1987). Knowledge and teaching: Foundations of the new Reform. *Harvard Educational Review*, *57*(1), S. 1–22.

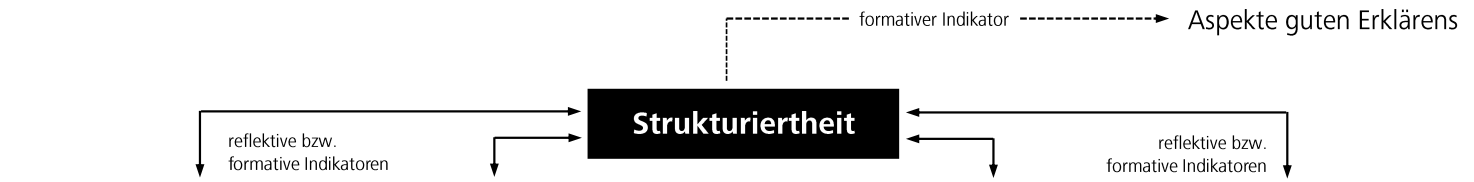
Spiegel, Carmen. (2007). „Kannst du mir das mal erklären?". Gesprächskompetenzen erwerben durch partnerschaftliches Lernen im Unterricht. *Grundschulunterricht*, (9), 4-11.
Spreckels, Janet (2009). Mündliches Erklären im Deutschunterricht. In M. Krelle & C. Spiegel (Hrsg.), *Sprechen und Kommunizieren. Entwicklungsperspektiven, Diagnosemöglichkeiten und Lernszenarien in Deutschunterricht und Deutschdidaktik* (S. 117–138). Baltmannsweiler: Schneider Hohengehren.
Spreckels, Janet & Trojahn, Tanja (2009). „*n objekt is also des is SCHWER zu erklären*". Eine empirische Untersuchung von Erklärungen im Grammatikunterricht. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 133–149). Tübingen: Stauffenburg.
Staatsinstitut für Schulqualität und Bildungsforschung (Hrsg.). *LehrplanPlus*. Verfügbar unter https://www.lehrplanplus.bayern.de/
Stahns, Ruven. (2013). *Kognitive Aktivierung im Grammatikunterricht. Videoanalysen zum Deutschunterricht* (Thema Sprache - Wissenschaft für den Unterricht, Bd. 7). Dissertation. Baltmannsweiler: Schneider Hohengehren.
Stark, Robin; Gruber, Hans; Mandl, Heinz & Hinkofer, Ludwig. (2001). Wege zur Optimierung eines beispielbasierten Instruktionsansatzes: Der Einfluss multipler Perspektiven und instruktionaler Erklärungen auf den Erwerb von Handlungskompetenz. *Unterrichtswissenschaft*, *29,* S. 26–40.
Sturm, Afra (16.09.15). *Das Schreiben denkend verlangsamen und steuern*, Zürich. Verfügbar unter https://wiki.edu-ict.zh.ch/_media/quims/fokusa/b6sturm12.pdf
Taut, S. & Rakoczy, Katrin. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction*, (46), S. 45–60.
Thim-Mabrey, Christiane. (in Vorb.). *Sprachwissenschaftliche Untersuchungen zu verstehensrelevanten Eigenschaften von unterrichtlichen Erklärungen*.
Thim-Mabrey, Christiane & Lindl, Alfred (2021, in Vorb.). Erklärqualität aus sprachwissenschaftlicher Sicht. Die Wahrnehmung der sprachlichen Verständlichkeit von Erklärvideos aus elf verschiedenen Fächern. In A. Schilcher, S. Krauss, A. Lindl & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären.* Weinheim und Basel: Beltz Juventa.
Thorndike, Edward L. (1920). A constant error in psychological rating. *Journal of Applied Psychology*, (4), S. 25–29.
Trout, J. D. (2007). The Psychology of Scientific Explanation. *Philosophy Compass*, *2*(3), S. 564–591.
Ulrich, Winfried. (2011). Wortschatzarbeit im Deutschunterricht. *ide*, (1), S. 9–21.
Van Es, Elizabeth A. (2009). Participants' Roles in the Context of a Video Club. *Journal of the Learning Sciences*, *18*(1), S. 100–137.
VanLehn, Kurt; Siler, Stephanie; Murray, Charles; Yamauchi, Takashi & Baggett, William B. (2003). Why Do Only Some Events Cause Learning During Human Tutoring? *Cognition and Instruction*, *21*(3), S. 209–249.
Vogt, Rüdiger (2009). Die Organisation von Erklärprozessen im Unterricht. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 203–225). Tübingen: Stauffenburg.
Wagner, Anke & Wörn, Claudia. (2011). *Erklären lernen - Mathematik verstehen. Ein Praxisbuch mit Lernangeboten* (1. Aufl.). Stuttgart: Klett.
Wagner, Wolfgang; Göllner, Richard; Helmke, Andreas; Trautwein, Ulrich & Lüdtke, Oliver. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and domain-generalizability of domain-independent assessments. *Learning and Instruction*, *28,* S. 1–11.
Wagner, Wolfgang; Göllner, Richard; Werth, Sarah; Voss, Thamar; Schmitz, Bernhard & Trautwein, Ulrich. (2016). Student and teacher ratings of instructional quality. Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, *108*(5), S. 705–721.
Webb, Noreen M. & Mastergeorge, Ann M. (2003). The Development of Students' Helping Behavior and Learning in Peer-Directed Small Groups. *Cognition and Instruction*, *21*(4), S. 361–428.
Weinert, Franz E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Seattle: Hogrefe und Huber.

Weinert, Franz E. (2014). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (Beltz Pädagogik, 3. Aufl., S. 17–31). Weinheim und Basel: Beltz.
Wellenreuther, Martin. (2015). *Lehren und Lernen - aber wie? Empirisch-experimentelle Forschungen zum Lehren und Lernen im Unterricht* (Grundlagen der Schulpädagogik, Bd. 50, 8. korr. u. überarb. Aufl.). Baltmannsweiler: Schneider Hohengehren.
Wenzel, Birgit (2009). Geschichte erklären. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (Stauffenburg Linguistik, Bd. 52, S. 169–187). Tübingen: Stauffenburg.
Westermann Verlag (Hrsg.). (2017). *Sprachwandel* (Deutschunterricht, Bd. 1): Westermann.
Westermann Verlag (Hrsg.). (2018). *YouTube & Co. - Mit Webvideos arbeiten* (Deutschunterricht, Bd. 4): Westermann.
Wieser, Dorothee (2015). Theorie(?)-Praxis-Konstellationen in Lehrerforschung und Lehrerbildung: Fragen an die aktuelle deutschdidaktische Lehrerforschung. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 17–34). Wiesbaden: Springer VS.
Wild, Johannes & Wildfeuer, Alfred (Hrsg.). (2019). *Sprachendidaktik: Eine Ein- und Weiterführung zur Erst- und Zweitsprachdidaktik des Deutschen*. Tübingen: Narr Francke Attempto.
Winkler, Iris (2006). Argumentierendes Schreiben im Deutschunterricht im Spiegel von Aufgaben für Lern- und Leistungssituationen. In E. Grundler & R. Vogt (Hrsg.), *Argumentieren in Schule und Hochschule. Interdisziplinäre Studien; [Symposion, 13.10. und 14.10.2005, Ludwigsburg]* (Stauffenburg-Linguistik, Bd. 42, S. 157–166). Tübingen: Stauffenburg.
Winkler, Iris. (2017). Potenzial zu kognitiver Aktivierung im Literaturunterricht. Fachspezifische Profilierung eines prominenten Konstrukts der Unterrichtsforschung. *Didaktik Deutsch*, (43), S. 78–97.
Winkler, Iris & Steinmetz, Michael (2016). Zum Spannungsverhältnis von deutschdidaktischen Fragestellungen und empirischen Erkenntnismöglicheiten am Beispiel des Projekts KoALa. In M. Krelle & W. Senn (Hrsg.), *Qualitäten von Deutschunterricht* (S. 37–56). Stuttgart: Fillibach bei Klett.
Wisniewski, Benedikt; Zierer, Klaus; Dresel, Markus & Daumiller, Martin. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction*, *66,* vorab online. Verfügbar unter https://doi.org/10.1016/j.learninstruc.2020.101303
Wisniewski, Benedikt; Zierer, Klaus & Hattie, John. (2017). *Visible Feedback. Ein Leitfaden für erfolgreiches Unterrichtsfeedback*. Baltmannsweiler: Schneider Hohengehren.
Witschel, Elfriede & Saxalber, Annemarie. (2012). Von fremden zu eigenen Texten. Folgerungen aus einer LehrerInnenbefragung für den Schreibunterricht in der Sekundarstufe II. *ide*, (1), S. 74–84.
Wittwer, Jörg & Renkl, Alexander. (2008). Why Instructional Explanations Often Do Not Work. A Framework for Understanding the Effectiveness of Instructional Explanations. *Educational Psychologist*, *43*(1), S. 49–64.
Wragg, Edward C. (1993). *Primary teaching skills*. London: Routledge.

Anhang

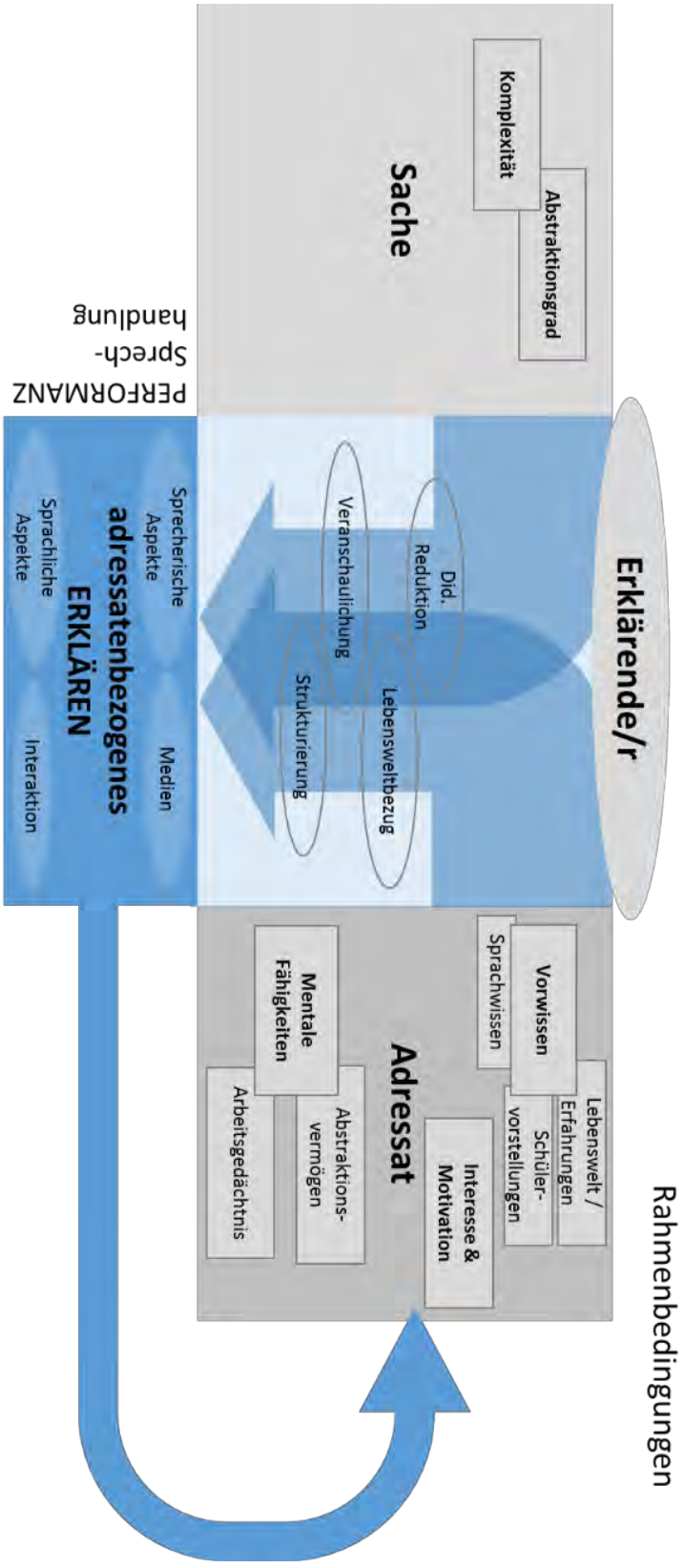
Anhang 1 Modell zur *Strukturiertheit* bei FALKE 295
Anhang 2 Modell zur *Adressatenorientierung* bei FALKE 296
Anhang 3.1 Itemübersicht FALKE 297
Anhang 3.2 Itemübersicht FALKE-D 300
Anhang 4 Interviewleitfragen für Gespräche mit Seminarlehrkräften über erklärwürdige Situationen 304
Anhang 5 Ausgangstext im Original 305
Anhang 6.1 Transkript zu Vignette A 307
Anhang 6.2 Transkript zu Vignette D 310
Anhang 6.3 Transkript zu Vignette B 312
Anhang 6.4 Transkript zu Vignette E 315
Anhang 6.5 Transkript zu Vignette C 318
Anhang 6.6 Transkript zu Vignette F 321
Anhang 7 Modelltext für Erklärung in Vignette F 323
Anhang 8.1 Erhebungsmanual Deutsch MZP1 SuS 324
Anhang 8.2 Erhebungsmanual Deutsch MZP2 SuS 325
Anhang 9.1 Paraphrasierungskatalog MZP1 SuS 326
Anhang 9.2 Paraphrasierungskatalog MZP2 SuS 328
Anhang 10 Deskriptive und psychometrische Kennwerte auf Itemebene 335
Anhang 11 Übersicht über sämtliche FALKE-D-Kodes mit Ankerbeispielen 341

Anhang 1 Modell zur *Strukturiertheit* bei FALKE



Konstrukt				
Dimensionen	**Gegenstandsbestimmung**	**Aufbau**	**Transparenz**	**Fokussierung**
Items	□ Mir ist klar, worum es in der Erklärung geht. □ Das Ziel der Erklärung ist mir klar.	□ Für mich hat die Erklärung einen roten Faden. □ Aus meiner Sicht baut die Erklärung das Thema Schritt für Schritt auf. □ Die Erklärung ist sinnvoll aufgebaut	□ Die Lehrkraft macht mir klar, was die wichtigen Punkte sind. □ Die Lehrkraft kündigt die einzelnen Schritte der Erklärung an.	□ Aus meiner Sicht erklärt die Lehrkraft zu viele Einzelheiten auf einmal. □ Die Erklärung bleibt stets beim Wesentlichen.
Beschreibung der Dimension	Der Gegenstand der Erklärung ist (von Anfang an) klar bestimmt. Dies ist eine Voraussetzung von Strukturiertheit, denn erst dadurch wird eine inhaltlich präzise und sprachlich prägnante Erklärung möglich.	Die Ordnung der Teilschritte der Erklärung ist sinnvoll. Sie kann grundsätzlich als induktiv oder deduktiv charakterisiert werden. Auch Beziehung zwischen den Teilaspekten der E. ist sinnvoll. Sie kann z.B. logischer, kausaler, chronologischer, funktionaler, exemplarischer, narrativer (etc.) Art sein. Der Aufbau gewährleistet Stringenz und bildet den Kernaspekt von Strukturiertheit.	Eine transparente Kommunikation des Aufbaus und der Gewichtung der Teilaspekte trägt zu Strukturiertheit bei. Die Transparenz unterstützt hierbei den gedanklichen Nachvollzug der Struktur.	Die Erklärung konzentriert sich auf das Wesentliche, Nötige und Wichtige. Die Fokussierung ist eine notwendige Bedingung für eine strukturierte Erklärung.
Performanz (Ausführung der Erklärung)	Der Erklärgegenstand ergibt sich entweder aus dem Kontext oder wird explizit gemacht.	Erklärende wählen und präsentieren die Beziehung der einzelnen Erklärschritte in Anpassung an den Erklärgegenstand.	Erklärende legen die Gliederung der Erklärung offen oder sie betonen bzw. gewichten die Teilschritte z.B. durch metasprachliche Einschübe, Intonation oder Gesten.	Erklärende verzichten auf unwichtige Exkurse und Ausschweifungen.
Wirkungsannahme	Die Adressaten wissen, was genau erklärt wird. So ist es ihnen möglich, die neuen Inhalte mit dem Erklärgegenstand zu vernetzen.	Die Adressaten können der Erklärung gut folgen und so die einzelnen Teilschritte aufeinander beziehen.	Die Adressaten können die Informationen leichter filtern und aufnehmen.	Die Aufmerksamkeit der Adressaten kann bei den wichtigen Aspekten bleiben.

Anhang 2 Modell zur *Adressatenorientierung* bei FALKE



Anhang 3.1 Itemübersicht FALKE

SLD-Gruppe

Strukturiertheit

1) Es ist klar, was genau die Lehrkraft erklären wollte.

2) Die Erklärung hatte einen roten Faden.

3) Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut.

4) Die Lehrkraft hat zu viele Einzelheiten erklärt.*

5) Die Erklärung war gut strukturiert.**

Adressatenorientierung

1) Für Schülerinnen und Schüler war die Erklärung zu schwierig.*

2) Die Erklärung berücksichtigte die Lebenswelt der Schülerinnen und Schüler.

3) Die Erklärung passte für Schülerinnen und Schüler dieser Jahrgangsstufe.

4) Die Erklärung war angemessen didaktisch reduziert.

5) Schülerinnen und Schüler konnten der Erklärung mit ihrem Wissen gut folgen.

6) Die Erklärung war adressatengerecht.**

Sprech- und Körperausdruck

1) Die Lehrkraft hatte eine angenehm klingende Stimme.

2) Die Lehrkraft hatte eine deutliche Aussprache.

3) Die Lehrkraft hatte eine angemessene Sprechgeschwindigkeit.

4) Die Lehrkraft hat die Sprechpausen so gesetzt, dass ich gut folgen konnte.

5) Die Lehrkraft hatte eine abwechslungsreiche Sprechweise.

6) Die Lehrkraft hatte einen Körperausdruck, der die Erklärung unterstützt hat.

7) Die Lehrkraft hat sich den Schülerinnen und Schülern zugewendet.

8) Die Art und Weise, wie die Lehrkraft spricht, gefällt mir.**

Sprachliche Verständlichkeit

1) Die Lehrkraft hat sich gut verständlich ausgedrückt.

2) Manchmal hat die Lehrkraft extra gesagt, dass etwas sehr wichtig ist.

3) Bei manchen Wörtern wissen die Schülerinnen und Schüler eventuell nicht, was sie bedeuten.*

4) Manche Sätze hat die Lehrkraft zu lang gemacht.*

Persönlichkeitswirkung

1) Die Lehrkraft wirkte in diesem Video auf mich sympathisch.

2) Die Lehrkraft wirkte in diesem Video auf mich begeistert.

3) Die Lehrkraft wirkte in diesem Video auf mich natürlich.

Bem.: *: Item wird in der zugehörigen Skala *rekodiert* berücksichtigt; **: Item wird in der zugehörigen Skala *nicht* berücksichtigt.

Anhang 3.1 Itemübersicht FALKE (SuS-Gruppe)

Strukturiertheit
1) Es ist klar, was genau die Lehrkraft erklären wollte.
2) Die Erklärung hatte einen roten Faden.
3) Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut.
4) Die Lehrkraft hat zu viele Einzelheiten erklärt.*
5) Die Erklärung war gut strukturiert.**
Adressatenorientierung
1) Die Erklärung war zu schwierig für mich.*
2) Die Inhalte der Erklärung hatten etwas mit meinem Leben zu tun.
3) Die Erklärung passte für mich.
---
5) Ich konnte der Erklärung mit meinem Wissen gut folgen.

Sprech- und Körperausdruck
1) Die Lehrkraft hatte eine angenehm klingende Stimme.
2) Die Lehrkraft hatte eine deutliche Aussprache.
3) Die Lehrkraft hatte eine angemessene Sprechgeschwindigkeit.
4) Die Lehrkraft hat die Sprechpausen so gesetzt, dass ich gut folgen konnte.
5) Die Lehrkraft hatte eine abwechslungsreiche Sprechweise.
6) Die Lehrkraft hatte einen Körperausdruck, der die Erklärung unterstützt hat.
7) Die Lehrkraft hat sich den Schülerinnen und Schülern zugewendet.
8) Die Art und Weise, wie die Lehrkraft spricht, gefällt mir.**
Sprachliche Verständlichkeit
1) Die Lehrkraft hat sich gut verständlich ausgedrückt.
2) Manchmal hat die Lehrkraft extra gesagt, dass etwas sehr wichtig ist.
3) Bei manchen Wörtern habe ich nicht gewusst, was sie bedeuten.*
4) Manche Sätze hat die Lehrkraft zu lang gemacht.*
Persönlichkeitswirkung
1) Die Lehrkraft wirkte in diesem Video auf mich sympathisch.
2) Die Lehrkraft wirkte in diesem Video auf mich begeistert.
3) Die Lehrkraft wirkte in diesem Video auf mich natürlich.

Bem.: *: Item wird in der zugehörigen Skala *rekodiert* berücksichtigt; **: Item wird in der zugehörigen Skala *nicht* berücksichtigt; --- entsprechendes Item fehlt in dieser Statusgruppe.

Anhang 3.2 Itemübersicht FALKE-D
SLD-Gruppe

Erklärtiefe

1) Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen.
2) Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.*
3) Die Erklärung war kognitiv aktivierend.

4) Vid. a: Es hat beim Verstehen der Redewendung geholfen, dass erklärt wurde, wie sie entstanden ist./
Vid. d: Es hat beim Verstehen der Redewendung geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden./
Vid. b: Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Inhalt konzentriert hat./
Vid. e: Die Schüler/-innen verstehen, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat./
Vid. c: Das „laute Denken", so wie es im Video gezeigt wurde, hat geholfen zu verstehen, wie man die Aussage widerlegt./
Vid. f: Dass die Lehrkraft ihr vorformuliertes Gegenargument erklärt, hat geholfen zu verstehen, wie man die Aussage widerlegt.

5) Vid. a: Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass erklärt wurde, wie sie entstanden ist./
Vid. d: Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden./
Vid. b: Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Inhalt konzentriert hat./
Vid. e: Die Schüler/-innen können sich gut merken, warum das Argument nicht gelungen ist, weil sich die Erklärung auf den Aufbau konzentriert hat./
Vid. c: Das „laute Denken", so wie es im Video gezeigt wurde, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt./

Vid. f: Dass die Lehrkraft ihr vorformuliertes Gegenargument erklärt, hat am meisten geholfen, sich gut merken zu können, wie man die Aussage widerlegt.

Fachliche Relevanz

1) Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt.

2) Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.

3) Vid. a: Bei Redewendungen sollte im Deutschunterricht regelmäßig erklärt werden, wie sie entstanden sind./

Vid. d: Bei Redewendungen sollten im Deutschunterricht regelmäßig andere Ausdrücke mit der gleichen Bedeutung genannt werden./

Vid. b: Es sollte regelmäßig erklärt werden, warum ein Argument inhaltlich nicht überzeugt./

Vid. e: Es sollte regelmäßig erklärt werden, warum der Aufbau eines Arguments nicht überzeugt./

Vid. c: Das „laute Denken", so wie es im Video gezeigt wurde, sollte im Schreibunterricht regelmäßig verwendet werden./

Vid. f: Vorformulierte Gegenargumente sollten im Schreibunterricht regelmäßig verwendet werden.

Bem.: *: Item wird in der zugehörigen Skala *rekodiert* berücksichtigt.

Anhang 3.2 Itemübersicht FALKE-D (SuS-Gruppe)

Erklärtiefe

1) Die Erklärung hat geholfen, das Erklärte genau zu verstehen.

2) Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.*

3) Die Erklärung hat mich zum Mitdenken angeregt.

4) Vid. a: Es hat beim Verstehen der Redewendung geholfen, dass erklärt wurde, wie sie entstanden ist./

Vid. d: Es hat beim Verstehen der Redewendung geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden./

Vid. b: Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Inhalt des Arguments erklärt hat./

Vid. e: Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat./

Vid. c: Das „laute Denken", das die Lehrerin vorgemacht hat, hat mir geholfen zu verstehen, wie man die Aussage widerlegt./

Vid. f: Dass die Lehrerin ihr vorformuliertes Gegenargument erklärt, hat mir geholfen zu verstehen, wie man die Aussage widerlegt.

5) Vid. a: Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass erklärt wurde, wie sie entstanden ist./

Vid. d: Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden./

Vid. b: Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Inhalt erklärt hat./

Vid. e: Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat./

Vid. c: Das „laute Denken, das die Lehrerin vorgemacht hat, hat mir am meisten geholfen, mir gut merken zu können, wie man die Aussage widerlegt./

Vid. f: Dass die Lehrerin ihr vorformuliertes Gegenargument erklärt, hat mir am meisten geholfen, mir gut merken zu können, wie man die Aussage widerlegt.

Fachliche Relevanz

1) ---

2) Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.

3) Vid. a: Bei Redewendungen sollte im Deutschunterricht regelmäßig erklärt werden, wie sie entstanden sind./

Vid. d: Bei Redewendungen sollten im Deutschunterricht regelmäßig andere Ausdrücke mit der gleichen Bedeutung genannt werden./

Vid. b: Es sollte regelmäßig erklärt werden, warum ein Argument inhaltlich nicht überzeugt./

Vid. e: Es sollte regelmäßig erklärt werden, warum die einzelnen Bestandteile eines Arguments sich nicht aufeinander beziehen./

Vid. c: Das „laute Denken", so wie es die Lehrerin vorgemacht hat, sollte im Schreibunterricht regelmäßig verwendet werden./

Vid. f: Vorformulierte Gegenargumente sollten im Schreibunterricht regelmäßig verwendet werden.

Bem.: *: Item wird in der zugehörigen Skala *rekodiert* berücksichtigt; --- entsprechendes Item fehlt in dieser Statusgruppe.

Anhang 4 Interviewleitfragen für Gespräche mit Seminarlehrkräften über erklärwürdige Situationen

1) Welche Themen fallen Ihnen ein, für die selbstentdeckendes Lernen im Deutschunterricht nicht funktioniert?

2) In welchen Lernbereichen erklären Sie besonders viel oder wenig?

3) Bei welchen Themen / in welchen Situationen erklären Berufsanfänger
zu wenig
zu viel
falsch?

4) Welche Themen aus dem Deutschunterricht fallen Ihnen ein zu
erklären, was
erklären, wie
erklären, warum?

5) Wann und wie erklären Sie die Bedeutung von Wörtern? Was ist Ihnen dabei wichtig?

6) Welchen Stellenwert weisen Sie folgenden Aussagen zu?
Erklärungen müssen den Zweck / die Funktion des Erklärgegenstandes deutlich machen.
Erklärungen müssen in einen größeren Zusammenhang eingebettet werden.
Erklärungen sollen tieferes Verstehen ermöglichen.

7) Welche Situationen /Themen fallen Ihnen ein, in denen Sie eigentlich immer kurz und knapp etwas erklären?

Anhang 5 Ausgangstext im Original

Kolumne Streit
Like den Lehrer?!
Sollten Schüler und Lehrer auf Facebook befreundet sein?
Eine Schülerin und eine Lehrerin sagen, was sie davon halten – und sie sind sehr unterschiedlicher Meinung

An sozialen Netzwerken kommt man nicht mehr vorbei, heute sind selbst die Lehrer auf Facebook. Aber sich deswegen gleich dort mit ihnen anfreunden? Auch wenn's vielleicht sinnvoll klingt – das geht doch ein bisschen zu weit. Finden zumindest die Ministerien einiger Bundesländer und haben diese Form der Schüler-Lehrer-Interaktion gleich mal untersagt. Aber zugegeben: Ganz unpraktisch ist es auch wieder nicht, eigenverantwortlich muss das doch wohl in Ordnung sein. Oder? Wir lassen mal wieder streiten!

Pro: Warum denn bitte nicht? Dorothea Kleffner

Die heutige Beziehung zwischen Schülern und Lehrern lässt es zu, in sozialen Netzwerken befreundet zu sein, meint Lehrerin Dorothea Kleffner. Außerdem: Wo sonst kann sie für ihre Schüler derart gut erreichbar sein? Auch deshalb nutzt sie Facebook seit Jahren. Zu Problemen habe das noch nie geführt

Ich arbeite nun schon seit zwölf Jahren als Lehrerin in Hamburg. Meine Schüler sind hauptsächlich Jugendliche zwischen 10 und 18 Jahren. Seit ungefähr sechs Jahren bin ich bei Facebook registriert und kommuniziere dort auch mit einigen Schülern.
Ich habe mir seitdem eigentlich nie die Frage gestellt, ob ich womöglich die nötige Distanz zu meinen Schülern aufgebe, wenn ich mit ihnen auf Facebook befreundet oder über WhatsApp mit ihnen in Kontakt bin. Wahrscheinlich liegt das daran, dass ich mit manchen meiner Schüler mehr Zeit verbringe als deren Eltern.
Natürlich weiß ich um die Kritik, die es an Facebook gibt. Das thematisieren wir auch in der Medienerziehung an unserer Schule und halten die Schüler dazu an, sich gut zu überlegen, was sie auf Facebook an persönlichen Details über sich bekannt geben. Aber es wäre auch weltfremd, Jugendlichen heute die Benutzung der großen sozialen Netzwerke ausreden zu wollen. Wenn ich mit ihnen dort kommuniziere und eine behutsame Nutzung vorlebe, erreiche ich mehr.
Da ich viele Projekte mit Schülern koordiniere, etwa Filmprojekte, Malaktionen und eine Schülerfirma, die Caterings veranstaltet, sind Facebook oder WhatsApp schon sehr praktische Hilfsmittel, um Informationen möglichst schnell weiterzuleiten. Innerhalb der geschlossenen Gruppen können dann Fragen und Wünsche ausgetauscht werden, und man muss nicht jeden einzeln anrufen oder in der Schule suchen und ansprechen. Zudem bietet Facebook sich an, wenn es um Werbung für eine Veranstaltung oder für unsere Schülerfirma geht. Man erreicht über diese Medien viel mehr Menschen und kann sich die Kosten für Flyer sparen. […]

Ich mache kein Geheimnis aus meinem Leben
Ich persönlich habe noch nie die Erfahrung gemacht, dass Schüler die Gelegenheit nutzen, via Facebook Böses über mich zu veröffentlichen oder mich in meiner Freizeit zu belästigen. Ohnehin bin ich dem ja nicht wehrlos ausgeliefert: Ich kann selbst entscheiden, was ich von mir persönlich

preisgebe. Ich kann Kontakte löschen, wenn ich kein gutes Gefühl bei der Sache habe. Und ich bin selbst dafür verantwortlich, wie oft ich meinen Computer oder mein Handy nach Nachrichten überprüfe. Ich bin eben nur dann erreichbar, wenn ich erreichbar sein will.
Meine Schüler wissen alle, dass ich ein Privatleben habe. Veröffentliche ich wirklich mal Bilder von der Familie oder von privaten Erlebnissen, interessiert sie das kaum – zumindest haben sie mich noch nie darauf angesprochen oder derartige Postings kommentiert. Wahrscheinlich ist es auch weniger interessant, da ich kein Geheimnis aus meinem Leben mache.
Wozu auch, ich bin weder besonders wichtig, noch habe ich brisante Informationen. Da ich glaube, über einen gesunden Menschenverstand zu verfügen, ist mir klar, dass ich nur ausgewählte Bilder oder Informationen verbreiten kann, aber dabei geht es mir weniger um meine Schüler als um die Tatsache, dass diese Informationen im Netz gespeichert bleiben.

Die Lehrerrolle verändert sich

Um die Frage nach Nähe oder Distanz von Lehrern und Schülern wird meines Erachtens zu viel Aufhebens gemacht. Da sich die Lehrerrolle sowieso stetig verändert, man heute nicht mehr der reine Wissensvermittler ist, sondern Begleiter und oft auch Bezugsperson, gibt es meiner Meinung nach immer weniger Lehrer, die ihren Job nur im Schulgebäude erledigen und danach ein anderer Mensch sind. Die meisten meiner Kollegen engagieren sich zusätzlich zum Unterricht und versuchen, Schülern den Spaß an Schule und Lernen neu zu vermitteln. Insbesondere schwierige Schüler brauchen da starke und konstante Beziehungsarbeit. Und dafür eignen sich unter anderem die sozialen Netzwerke, da sie die Kommunikation einfach machen.
Dass sich Schüler gezwungen fühlen könnten, diese Medien zu nutzen, da sie sonst von bestimmten Informationsgruppen ausgeschlossen werden könnten, habe ich bisher auch nicht erlebt. Ab und zu gibt es Schüler, die sich dem aus Datenschutzgründen verweigern. Die rufe ich dann eben persönlich an oder spreche mit ihnen in der Schule. Ein Verbot finde ich unsinnig, da man mit seinen Schülern ohnehin in irgendeiner Form in Kontakt steht. Und wer eben gerne kumpelhaft ist, soll das doch bitte sein dürfen. Was ist so schlimm daran, wenn man ein entspanntes Verhältnis zu den Menschen hat, die einen täglich umgeben.

Man nennt das 'Neue Autorität'

Deswegen muss nicht gleich der Respekt verloren gehen. Autorität geht nicht durch Nähe oder eine größere Vertrautheit verloren, im Gegenteil. Je mehr Beziehungsarbeit ich leiste, umso respektvoller und verständnisvoller sind die Schüler mir gegenüber. Das nennt man die „Neue Autorität“, und die wird als Schulkonzept mittlerweile an vielen Schulen umgesetzt. Es geht hierbei nicht um „Pseudofreundschaften“, sondern um eine neue Umgangsweise miteinander.
Wen das stört, der kann sich dem ja entziehen, indem er oder sie sich distanziert – und auch den sozialen Netzwerken eine Absage erteilt.

Dorothea Kleffner (43) arbeitet an der Kurt-Tucholsky-Stadtteilschule in Hamburg Altona. Sie unterrichtet Deutsch, Gesellschaftskunde und Kunst. Zusätzlich bietet sie Film-, Mal- und Theaterprojekte an und organisiert eine Schülerfirma. Sie ist Mutter von drei Kindern und hat eine Vollzeitstelle.

www.fluter.de/like-den-lehrer. 01.05.2020

Anhang 6.1 Transkript zu Vignette A[73]

Was heißt *viel Aufhebens um etwas machen*? (Etymologischer Hintergrund)

Nr.	Gesprochener Text /r/-Lautung apikal	Art der Visualisierung: Tafelanschrift[74]
01	wir HAM grad=n TEXT geLEsen	
02	und da war	
03	in den ERSten ZWEI ZEIlen war so=n AUSdruck	
04	drin,	
05	von dem ich AUSgehe,	
06	dass der vielLEI::CHT zum verSTÄNDnis	
07	SCHWIErig is;	
08	ähm (.) viel !AUF!hebens um eine sache	L zeigt mit dem Finger auf
09	machen.	unter-schiedliche S, die alle den
10	habt ihr das geSEHen? (--)	Kopf schütteln
11	has=du ne AHnung,	
12	was des HEISST? (--)	
13		
14	KEIner? (-)	
15	also GUT,	
16	dann (.)	
17	WARtet mal (-)	
18	ähm (---)	L schreibt an die Tafel *um eine*
19	GUT (.)	*Sache viel Aufhebens machen*
20	also DES ist der AUSdruck,	
21	der da beNUTZT wird	
22	und wenn ich=s GANZ diREKT versuche,	
23	euch zu SAgen,	
24	dann heißt des EIgentlich;	
25	ich MESSE einer SAche GRO:ße beDEUtung bei	
26	oder ich NEHme sie sehr WICHtich.	
27	oKAY?	
28	also (-) ähm (---)	L schreibt an die Tafel
29	ETwas WICHtich NEHmen. (-)	→ *etwas wichtig nehmen*
30	DES is=n ganz MERKwürdicher AUSdruck,	
31	wir KÖNnen ihn eigentlich HEUte nicht mehr	
32	WIRKlich erSCHLIE:ßen,	
33	wenn wir ihn nicht !KEN!nen,	
34	dafür muss=ma in (.) die ver!GAN!genheit	
35	zuRÜCKgehen	
36	und ZWAR (.) in DIE gang …/ verGANgenheit,	
37	wo (.) MENschen noch (.) mit (.) WAFfen	
38	aufeiNANder LO:Sgegangen sin,	
39	ins MITtelalter. (-)	
40	ähm (.) s gab mal ne ZEIT, (.)	
41	da hat ma=EINfach (-) ja (.) kon…/ äh	
42	konFLIKte mit WAFfen ausgeTRAgen,	

[73] Die Transkripte aller Videos wurden von einer sprachwissenschaftlichen Arbeitsgruppe an der Universität Regensburg unter der Leitung von Christiane Thim-Mabrey erstellt (vgl. zum Vorgehen beim Transkribieren Thim-Mabrey und Lindl, 2021, in Vorb.).

[74] Objektsprache sowie das, was die Lehrkraft konkret an die Tafel schreibt, wird kursiv gedruckt.

43	mit GRO:ßen SÄ:beln zum BEIspiel oder DE:gen	
44	und (.) ähm, beVO:R dieser kampf LOSgegangen	
45	is,	
46	hat ma diese WAFfen vor sich auf den BO:den	
47	geLEGT.	
48	und (.) !WÄH!rend man dann diese waffen	
49	!AUF!gehoben hat,	L ahmt den Akt des Aufhebens nach
50	dann hat ma die nicht !EI:N!fach nur (.)	
51	AUFgehoben,	
52	sondern schon WÄHrenddessen hat ma	
53	angefangen, (.)	
54	zu PROTZen	
55	und zu PRAHlen,	
56	den GEGner eigentlich (.) mit WORten KLEIN	
57	zu machen	
58	und zu ZEIgen,	
59	wie TOLL man is.	
60	also,	
61	man hat den !VO:R!gang des auf!HE:!bens (--)	
62	GRÖßer gemacht,	
63	WICHTich gemacht,	
64	da ne RICHtich GRO:ße SAche draus gemacht	
65	und geNAU das is=es;	
66	also,	
67	den !VO:R!gang des AUFhebens (.) sehr (.)	
68	WICHtich (.) NEH:men,	
69	sehr wichtich MACHen	
70	und !EI!gentlich muss ma SAGen;	
71	ETwas (.) !ZU! (.) WICHtich nehmen.	
72	also wurde eine !RIE:!sige aktION draus	L fügt an die Tafel ein *zu* ein:
73	gemacht;	→ *etwas **zu** wichtig nehmen*
74	nicht EINfach nur still AUFheben	
75	und dann LO:Slegen,	
76	sondern !AUF!heben	
77	und (.) UH (.), LOSlegen	
78	und dann (.) KÄMpfen.	
79	jetzt KLAR?	L gestikuliert wild bei dem Wort *uh*
80	DANN schau=ma JETZT nochmal in den TEXT,	
81	wo des STEH:T (.)	
82	und können uns des BESser erSCHLIEßen.	
83		

Markierungen: Bedeutung der Markierungsmittel[75] (für alle Transkripte)	
Zeilenumbruch	Am Ende einer syntaktischen Einheit
Leerzeile	Nach einer inhaltlichen Zäsur
akZENT	Betonung (Primär- und Hauptakzent)
ak!ZENT!	besonders starke Betonung
(.)	Mikropause

[75] Nach Selting et al. (1998); aber mit eigenen Erweiterungen.

(-), (--), (---)	kurze, mittlere, längere Pause
:, ::, :::	Dehnung eines Vokals (je nach Länge)
ne=das könn=wir	Verschleifungen innerhalb von Einheiten
…// …/ […]	vollständiger Satzabbruch Neuansatz für einen Reformulierungsversuch unverständliche Passage

Anhang 6.2 Transkript zu Vignette D

Was heißt „Viel Aufhebens um etwas machen“? (Synonyme Ausdrücke)

Nr.	Gesprochener Text /r/-Lautung apikal	Art der Visualisierung: Tafelanschrift[76]
01	wir ham in dem TEXT jetzt eine !RE!dewendung	
02	kennengelernt, (--) in (.) ZEIle ZWEI. (.)	
03	habt ihr geSEHEN? (.)	
04	ich GLAU:be,	
05	die ist NICHT ganz verSTÄNDlich.	
06	ich SCHREIB sie euch nochmal EINzeln heRAUS.	L schreibt an die Tafel
07	(---)	*Viel Aufhebens um eine Sache machen*
08	gut. (-)	
09	also, (-)	
10	DIEse REdewendung …/ (-)	
11	ich verSUCH euch das jetz mal (.) MIT (.)	
12	verSCHIEdenen (.) !AN!deren REdewendungen,	
13	von denen ich GLAUB,	
14	sie sind EINfacher zu erklären.	
15	und ZWAR gibt=s DIEses (.) eine (.) MÜcke …/	
16	von einer MÜcke zu einem eleFANten …/	
17	also AUS einer MÜcke …// (---)	L schreibt an die Tafel
18	scho mal geHÖRT? (-)	→ *aus einer Mücke einen Elefanten machen*
19	JA?	
20	ah::m (-)	
21	zu …/	
22	O::der (-)	
23	ähm(---)	L schreibt an die Tafel
24	viel !WIR:!bel um etwas machen.	→ *viel Wirbel um etwas machen*
25	auch DES is etwas,	
26	was ihr !SCHON!,	
27	glaub ich,	
28	be!NUTZT!,	
29	wenn ihr (.) EINE …/ (.)	
30	verSUCHT zu erKLÄren,	
31	dass eine SACHe,	
32	die EIgentlich eher ganz !EIN!fach is,	
33	aber (.) von einem MITschüler zum BEIspiel	
34	!GA::NZ! gʁoß AUFgebauscht wird	
35	oder er MACHT eben !GANZ! viel WIRbel	
36	dadʁum.	Bei dem Wort *ganz* macht die L einen großen Armkreis über ihren Kopf
37	das ist nicht was EINfaches,	
38	sondern eben eine !GRO::!ße NUMmer,	
39	könnte man sagen	
40	oder ich mache sehr viel !WIND!	
41	oder ich mache sehr viel tam!TAM! um etwas.	
42	wenn ich=s !GANZ! (.) EINfach versuche zu	
43	sagen,	
44	kommt !EI!gentlich ʁaus; (-)	
45	ich NEHme eine SAche zu WICHtich,	

[76] Objektsprache sowie das, was die Lehrkraft konkret an die Tafel schreibt, wird kursiv gedruckt.

46	ich !MES!se ihr (.) ZU viel beDEUtung bei.	
47	und die be!TO!nung liegt auf dem !ZU!, (.)	
48	also auf dieser ver!STÄR!kung. (---)	L schreibt an die Tafel
49	oKAY, (-)	→ *einer Sache <u>zu</u> viel Bedeutung beimessen;*
50	also HIER nochmal …/	L kreist das (bereits unterstrichene) *zu* ein (mit einer nassen Kreide);
51	na, (.)	
52	die is NASS,	
53	die SIEHT ma nich gut.	
54	einer SAche beDEUtung beimessen	
55	mit dem FOkus auf !ZU! (.) große beDEUtung.	
56	JETZT verSTANden? (.)	
57	(und) !JETZ! schau=ma nochmal in den TEXT.	L zieht bei dem Wort *zu* beide Augenbrauen in die Höhe

Anhang 6.3 Transkript zu Vignette B

Warum ist ein Argument nicht gelungen? (Funktional-inhaltliche Begründung)

Nr.	Gesprochener Text /r/-Lautung uvular	Art der Visualisierung: Overheadprojektor (OHP)[77]
01	guten MORgen liebe SCHÜlerinnen und	
02	SCHÜler,	
03	wir beSCHÄFtigen uns ja agu …/ ä=aktuELL	
04	damit,	
05	WIE=man geLUNgen argumenTIERT.	
06	a:m (.) und HEUte möchten (.) ich …/	
07	möchte ich mir mit EUCH EINmal EIN (.)	
08	!NICHT! so gelungenes BEIspiel ANschauen,	
09	euch !ZEI!gen,	
10	waRUM es nicht überZEUGT,	
11	um DA:raus dann ABzuleiten,	
12	was eine geLUNgene argumentaTION ausmacht.	
13	als ERStes müssen wir uns das BEIspiel mal	
14	ANSchauen;	
15		
16	hier argumenTIERT ein koLLEG …/ ein koLLEGE	
17	oder eine koLLEgin von mir im INternet	
18	DAfür,	
19	dass ein verBOT,	
20	dass SCHÜler mit f …/ ähm=ähm, mit LEHrern	
21	auf FACEbook oder ÄHNlichen sozialen	
22	NETZwerken beFREUNdet sind,	
23	ähm SINNlos ist.	
24	dazu SCHREI:BT derjenige oder DIEjenige;	
25	*ein verBOT finde ich UNsinnik,*	L liest das Argument auf dem OHP vor und verfolgt mit seinem Folienstift die entsprechenden Textstellen auf der Folie
26	*da man mit seinen SCHÜlern ohnehin in*	
27	*IRgendeiner FORM in konTAKT steht.*	
28	*und WER eben gerne KUMPelhaft ist,*	
29	*SOLL das doch bitte SEIN dürfen.*	
30	*!WAS! ist so SCHlIMM daran,*	
31	*wenn man ein entSPANNtes verHÄLTnis zu den*	
32	*MENschen hat,*	
33	*die einen TÄGlich umgeben?*	
34	SCHAU=ma uns die argumentaTIONSstruktur	
35	einmal geNAUer an,	
36	dann stellen wir FEST;	
37	hier wird beHAUPtet,	
38	dass das verbot *!UN!sinnig* ist,	L umkreist das Wort *unsinnig*; L unterstreich *ohnehin in irgendeiner Form* und umkreist das Wort *Kontakt*
39	!WEIL! man mit den *SCHÜlern ohnehin in*	
40	*!IR!gendeiner form in konTAKT steht.*	
41	is das eine geLUNgene beGRÜNdung DIEses (.)	
42	ERSten ALLgemeinen (.) ähm arguMENTS?	
43	ich würde sagen;	
44	auf gar keinen FALL,	
45	denn na!TÜR!lich stimmt es,	

[77] Objektsprache sowie das, was die Lehrkraft konkret auf die Folie schreibt, wird kursiv gedruckt.

46	dass ich in IRgendeiner form in konTAKT mit	
47	meinen SCHÜlern steh,	
48	klar.	
49	s=is eine (.) SEHR=SEHR allgemeine AUSsage,	
50	die erstmals GAR nichts GROSSartiges	
51	AUSsagt.	
52	aber DIEses konTAKTverhältnis zu meinen	
53	SCHÜlern is ja ein ganz beSONderes;	
54	s=is ein hier!AR!chisches verHÄLTnis,	
55	das sich über unser SCHULsystem und das	
56	geSELLschaftliche sysTEM erGIBT (.)	
57	und ich STEHE=ja also=in einem ganz	
58	be!SON!deren verHÄLTnis zu meinen SCHÜlern.	
59	ich soll meinen SCHÜlern etwas	
60	!BEI!bringen,	
61	hab aber auch die AUFgabe,	
62	für disziPLIN zu SORgen	
63	und sie zu beWERten	
64	und zu beNOten.	
65	!DAS! wird also hier verGESsen	
66	und damit FEHLT hier ein GANZ wichtiger	
67	beSTANDteil dieses sysTEMS.	
68		
69	WEIterhin wird dann beHAUPtet;	
70	*wer eben gerne KUMPelhaft ist,*	L umkreist die Wörter *kumpelhaft* und *sein*
71	*soll das doch bitte !SEIN! dürfen.*	
72	ja, naTÜRlich,	
73	wenn jemand gerne ein KUMpelhafter MENSCH	
74	is,	
75	dann soll er das auch bitte (.) ähm	
76	AUSleben dürfen,	
77	aber funktioNIERT das in einem	
78	hierARchischen sysTEM?	
79	im ENDeffekt eben !NICHT! so gut.	
80	wenn ich mir einen !KUM!pel von mir	
81	ANschaue,	
82	mit dem ich mich REgelmäßik TREFfe,	
83	(.) dann kommuniZIERe ich mit dem auf	
84	AUgenhöhe.	
85	!KLAR! kritiSIER ich ihn auch mal,	
86	wenn ich etwas NICHT so gut finde,	
87	A:ber ich werde eben !NIE!mals HERgehen	
88	und ANfangen zu SAgen;	
89	auf DIEsen AUSdruck oder auf DIEsen	
90	SACHverhalt hast du jetzt DIE und die NO:te	
91	bekommen.	
92	ich WERd=ihn nicht über die MA:ßen hinaus	
93	MA:SSregeln für IRgendETwas,	
94	was mir NICHT so gut geFÄLLT,	
95	ich WERde mich VIEL stärker zuRÜCKhalten.	
96	ge!NAU! das muss ich aber tun SCHÜlern	
97	gegenüber,	
98	von DAher is diese AUSsage auf …/ !NICHT!	
99	auf das SCHULsysTEM überTRAGbar,	

100	!VIEL! zu stark verallgeMEInert	
101	und NICHT konKRET, (-)	
102	soDASS man es AUCH an dieser STELle so	
103	nicht STEHenlassen kann.	
104		
105	und WENN=ma uns dann das !FA!zit anschauen,	L markiert das Fazit durch einen vertikalen Strich am linken Rand des Satzes
106	dann SEHen wir,	
107	dass dort STEHT;	
108	*WAS ist so SCHlIMM daran,*	
109	*wenn man ein entSPANNtes verHÄLTnis hat?*	
110	da is na!TÜR!lich nichts SCHLIMM daran,	
111	aber es is eigentlich ZIEMlich eGA:L,	
112	ob ich mit SCHÜlern (.) auf FACEbook oder	
113	whatsAPP eine FREUNDschaft habe (.)	
114	oder NICHT, (.)	
115	ähm, ich kann TROTZdem sehr entSPANNT	
116	meinen SCHÜlern gegenüber sein	
117	und ich kann sehr !AN!gespannt sein.	
118	diese AUSsage is im ENDeffekt auch wieder	
119	(.) SEHR stark verallgeMEInernd	
120	und TUT zu der argumentaTION selber nichts	
121	zur SACHE,	
122	soDASS wir INSgesamt SAgen können,	L deutet auf den gesamten Text auf dem OHP
123	dass ALle diese AUSsagen kaum auf das SCHUL	
124	…/ sysTEM SCHUle (.) passen,	
125	!SEHR! allgeMEIN sind,	
126	die beSONderheit des SCHULsystems nicht	
127	beINhalten	
128	und von DAher auf gar keinen fall	
129	IRgendeiner ART und WEIse überZEUgen	
	können.	

Anhang 6.4 Transkript zu Vignette E

Warum ist ein Argument nicht gelungen? (Strukturelle Begründung)

Nr.	Gesprochener Text /r/-Lautung uvular	Art der Visualisierung: Overheadprojektor (OHP)[78]
01	guten MORgen liebe SCHÜlerinnen und	
02	SCHÜler.	
03	wir beSCHÄFtigen uns NA:CH wie vor mit dem	
04	argumenTIEren.	
05	wir HAben den SINNvollen aufbau einer	
06	geLUNgenen argumentaTION kennengelernt	
07	und ich möchte jetz einfach EUCH mal an	
08	einem !SCHLECH!ten beispiel zeigen (.)	
09	äh (.) wie=s NICHT funktionieren kann	
10	um dann nochmal daRAUSgehend AUFzubauen,	
11	wie man=s SINNvollerweise macht.	
12		
13	DIEses BEIspiel (-) geht zu dem THEma;	
14	is=es SINNvoll,	
15	mit (-) SCHÜlern auf facebook beFREUNdet zu	
16	sein als LEHrer? (-)	
17	ahm (.) wir wiederHOlen ganz kurz;	
18	eine geLUNgene argumentaTION beSTEHT (.)	L schreibt die Bestandteile einer
19	norMAlerweise in EInem argument aus einer	Argumentation auf eine OHP-
20	beHAUPtung	Folie (*Behaupt. / Begr. /*
21	– KÜRze ich kurz AB –	*Veranschaulichung*)
22	aus einer beGRÜNdung und dann aus einer	
23	verANschaulichung,	
24	wir nennen das oft !BEI!spiel,	
25	ich KÖNNTE dann noch eine FOLge daraus	
26	ableiten.	
27	LIEGT das jetzt in DIEsem fall VOR?	
28	DER autor oder die auTOrin beginnt mit der	
29	AUSsage;	
30	*ein verBO:T finde ich U:Nsinnik.*	
31	DAran SCHLIE:SST sich ein kauSAler	L rumkreist die Behauptung;
32	NEbensatz mit der konjunktiON eingeleitet	L umkreist die Subjunktion *da*
33	*!DA!* an.	
34	ab !HIER! muss im ENDeffekt der …/ die	
35	beGRÜNdung vorliegen,	
36	sodass die beHAUPtung EINfach lautet;	
37	ein verBO:T finde ich UNsinnik.	
38	ich würde jetz beHAUPten;	
39	ds=is eine VIEL zu ALLgemeine beHAUPtung,	
40	die über !ALLEM! steht (-)	
41	und entspricht !NICHT! dem,	
42	was wir in der SCHUle ge!LERNT! haben,	
43	und wir geSAGT haben;	
44	die EINzelbehauptungen, die müssen im	
45	ENDeffekt ko…/konkREte asPEKte anführen.	

[78] Objektsprache sowie das, was die Lehrkraft konkret auf die Folie schreibt, wird kursiv gedruckt.

46	NE=	
47	=man HÄTte hier zum beispiel BESser	
48	schreiben können;	
49	äh:m (.) es ist EINfacher,	
50	mit SCHÜlern in konTAKT zu treten,	
51	sodass wir die beHAUPtung an und für SICH	
52	bereits ERSTmal verWERfen können.	
53		
54	JETZT stellt sich aber die FRAGE;	
55	bezieht sich die beGRÜNdung auf diese	L zeigt mit dem Stift erst auf die Begründung und dann auf die Behauptung
56	beHAUPtung?	
57	denn DAS hab ich euch ja AUCH beigeBRACHT,	
58	dass die beGRÜNdung sich immer exAKT darauf	
59	beZIEHEN muss.	
60	die LAUTet hier;	
61	*DA man mit seinen SCHÜlern OHNEhin in*	L unterstreicht den vorgelesenen Nebensatz
62	*irgeneiner form in konTAKT steht.*	
63	ist DAS jetz ein GRUND dafür,	
64	dass ein verbot UNsinnig ist? (-)	
65	ich würde sagen;	
66	NEIN.	
67	es beGRÜNdet das nicht, (.)	
68	!DENN! ich stehe naTÜRlich mit meinen	
69	schülern in konTAKT,	
70	aber geNAUso könnte ich jetzt AUCH sagen;	
71	WEIL ich mit ihnen in konTAKT stehe,	
72	wäre es SINNvoll,	
73	wenn ich mich EINmal in der WOche abends	
74	mit ihnen TREFfe	
75	und=n BIER trinken gehe.	
76	von DAher SIEHT man also;	
77	es bezieht sich jetzt NICHT auf dieses	
78	verBOT,	
79	es is VIEL zu ALLgemein gehalten	
80	und ist !NIE!mals !NIE!mals eine geLUNgene	
81	beGRÜNdung zu diesem TEIL.	
82		
83	DIEse beGRÜNdung wird jetz in dem WEIteren	
84	verLAUF genauer verANSchaulicht,	
85	und zwar ab dem *!UND!.* ((schluckt))	L umkreist das *und*
86	(jet)=stellt sich hier AUCH wieder die	
87	FRAge;	
88	is das jetzt ein geLUNgenes BEIspiel?	L zeigt mit dem Stift erst auf Beispiel, dann auf die Begründung, dann auf die Behauptung.
89	bezieht sich das auf die beGRÜNdung und auf	
90	die beHAUPtung?	
91	ich les=es euch einmal !VO:R!;	
92	*und wer eben gerne !KUM!pelhaft ist*	
93	– das is der ERSte wichtige Teil –	
94	*soll das doch bitte !SEIN! dürfen.*	
95	und DANN kommt eine rheTOrische !FRA!ge;	
96	*was ist so !SCHLIMM! daran,*	
97	*wenn man ein ent!SPANN!tes ver!HÄLT!nis zu*	
98	*den !MEN!schen hat,*	
99	*die einen TÄGlich umgeben?*	

100	ist DAS jetzt ein BEIspiel,	
101	das dem entSPRICHT,	
102	was !WIR! (.) kennengelernt haben?	
103	ein BEIspiel, (.)	
104	das EInen (.) konKRE:ten FALL betrifft	
105	UND aber DAdurch auf alle verallgeMEInerbar	
106	is,	
107	WEIL=s eine verANschaulichung DARstellt?	
108	ich sage;	
109	das ist !NICHT! der fall.	
110	wir FINden überhaupt kein BEIspiel hier,	
111	es is !KEIN! konkreter EINzelfall,	
112	den man verallgeMEInern kann,	
113	sondern es is eine ALLgemeine !MEI!nung,	
114	(.)	
115	die man NICHT verallgeMEInern kann	
116	und damit is=es ein GANZ schlechtes	
117	BEIspiel, (.)	
118	das wir SO nicht verWENden können.	

Anhang 6.5 Transkript zu Vignette C

Wie widerlegt man ein Argument? (Methode lautes Denken)

Nr.	Gesprochener Text (/r/-Lautung apikal)	Art der Visualisierung: Tafelanschrift und Beamer, durch den die behandelte Argumentation auf weißer Projektionsfläche eingeblendet ist[79]
01	ich möcht=euch gern heut ZEIgen,	
02	wie man ein arguMENT der GEgenseite,	
03	des einen NICHT besonders überZEUGT,	
04	widerLE:gen kann.	
05	DA:zu nehmen wir den text,	
06	den ihr schon KENNT, (.)	
07	WO:=ne autTOrin im FLU:ter schreibt,	
08	sie FINdet die sozialen netzwerke !GUT! für	
09	die beZIEhungspflege zwischen LEHrern und	
10	SCHÜlern. (.)	
11	um (-) zu ner widerLEgung zu kommen,	
12	muss ich mir des arguMENT nochmal genau	
13	ANschauen (.)	
14	und darüber NACHdenken, (.)	
15	was ich daran GUT finde	
16	und wo ich mit meim WIderspruch anknüpfen	
17	KANN.	
18	ne?	
19	und !DAS! mach ich euch jetzt VOR,	
20	indem ich !LAUT! denke,	
21	sodass=ihr geNAU mitkriegt,	
22	WIE meine geDANken ABlaufen. (-)	
23	ja,	
24	ich LES=mir die ersten ZEIlen nochmal durch	
25	und bin (.) DA:mit eigentlich ganz	
26	zuFRIEden.	
27	hier wird geSAGT,	
28	dass (.) äh VIEle kollegen heutzutage	
29	nachmittags noch konTAKT pflegen mit den	
30	SCHÜlern über FACEbook (.)	
31	und das STIMMT auch mit meiner erFAHrung	
32	zuSAMmen (.)	
33	und !LIEGT! wahrscheinlich WIRKlich daran,	
34	dass die kommunikaTION relativ EINfach	
35	funktioNIERT über die sozialen NETZwerke.(.)	
36	da HAB ich eigentlich nix daGE:gen.	
37	des (.)	
38	find=ich,	
39	ka=man so STEHen lassen.	
40	also noTIER=ich mir mal (-) in meinem PLA:N	
41	für die wiger …/ widerLE:gung den geDANken;	

[79] Objektsprache sowie das, was die Lehrkraft konkret an die Tafel schreibt, wird kursiv gedruckt.

42	verEINfachung der kommunikaTION. (--)	L schreibt an die linke Tafelseite
43	und wei=ich des (.) oKAY finde,	*Vereinfachung der*
44	setz=ich daHINter n gelben HAken (.)	*Kommunikation*
45	u:nd bring für MICH so die erKLÄrung,	
46	dass man (.) über FACEbook WIRKLich (.)	
47	öh:m	
48	informatIONEN (.) EINfach AUStauschen kann.	
49	(---)	
50	wenn i=mir jetzt !AN!schau,	
51	was die (.) äh:: (.) auTOrin des textes für	L schreibt an die rechte
52	eine FOLgerung draus ZIEHT,	Tafelseite
53	dann SEH=ich,	*Einfacher Austausch von*
54	sie SAGT;	*Informationen*
55	DAraus folgt ne STARke und konsTANte	
56	beZIEHungsarbeit (-) mit den SCHÜ:lern (.)	
57	und die is vor allem WICHtich (.) FÜ:R die	
58	SCHWÄcheren und SCHWIErigeren SCHÜler. (-)	
59	A::LSO untersuch=ich DIEsen punkt AUCH noch,	
60	(.)	
61	o:b mich der überZEUGT (.)	
62	oder o:b ich eher FINde;	
63	da kann ich meinen WI:derspruch ANknüpfen	
64	=ne.	
65	naja,	L schreibt an die linke Tafelseite
66	des konsTANT find=ich ganz GU:T,	↓
67	weil WIRKlich,	*Starke und konstante*
68	ö	*Beziehungsarbeit*
69	wenn ich NACHmittags kommuniZIEre über	
70	FACEbook	
71	ja (.)	
72	!DAUER!haftere beZIEHungspflege da is,	
73	als wenn ich nur VORmittags die schüler im	
74	KLASsenzimmer seh. (.)	
75	(xxx) muss ich nur über des !STAR!ke	
76	nachdenken	
77	und da hab ich jetz meine ZWEIfel.	
78	ich setz=a=mal a FRAgezeichen drüber, (.)	
79	WEI::L ich mir EIgentlich unter ner STARken	
80	beziehungsarbeit was !AN!deres vorstell.	
81	ich GLAUB,	
82	dass DA:zu (.) der (.) DIrekte perSÖNliche	
83	konTAKT nötig is	
84	und DEN haben wir schon E:her im	
85	!KLAS!senzimmer. (---)	L ergänzt die Tafelanschrift um
86	da ham=wir mehr ZEIT,	ein Fragezeichen neben dem
87	mitnander zu REden	Pfeil
88	und ich denk,	
89	über FACEbook tauscht ma oft WIRKlich nur	
90	kurze INfos au::s;	
91	was MACHST du grad?	
92	was machs ICH grad?	
93	oder was (.) is jetzt DAmit los?	L schreibt an die rechte
94	und kommt WE:niger in=an (.) TIEfergehenden	Tafelseite
95	konTAKT	

	und für ne !STAR!ke beziehung is der SCHON	*Direkter persönlicher Kontakt* […]
96	nötich, denk ich	
97	=ne?	
98	da reicht das NICHT,	
99	was man über FACEbook so hin und HER	
100	schickt. (.)	
101	also, (-)	
102	ich denke,	
103	dass der DIrekte konTAKT da wirklich NÖ:tig	
104	wäre (-)	
105	und der is nich geGEben,	
106	also FIND=ich; (-)	
107	!DA! (.) kann ich widerSPREchen. (.)	L ergänzt die letzte Tafelnotiz mit […] *nötig*
108	und SOweit HÄTT=ich eigentlich jetz=n PLA:N,	
109	(-)	
110	wie ich dieses argument widerLE:gen kann (-)	
111	und (-) erGÄNz (.) des mal HIER (.) als	L malt einen vertikalen Blitz zwischen die beiden Tafelhälften
112	ÜBERschrift. (---)	
113		
114		
115		
		L ergänzt die Tafelüberschrift *WIDERLEGUNG EINES ARGUMENTS*

Anhang 6.6 Transkript zu Vignette F

Wie widerlegt man ein Argument? (Methode Modelltext)

.

Nr.	Gesprochener Text (/r/-Lautung apikal)	Art der Visualisierung: Beamer, durch den die behandelte Argumentation auf weißer Projektionsfläche eingeblendet ist[80]
01	ich möcht=euch gern EINmal ein SCHRIFTliches	
02	BEIspiel zeigen,	
03	wie so=ne widerLEgung eines GEgenarguments	
04	AUSschauen kann (-)	
05	und hab deshalb SELber (.) einen (.)	
06	BEIspieltext verFASST. (.)	
07	der bezieht sich WIEder auf den text im	
08	FLUter,	
09	wo die auTOrin beHAUPtet,	
10	dass man (.) über die sozialen NETZwerke die	
11	beZIEHung zwischen LEHrern und SCHÜlern ganz	
12	gut PFLE:gen kann. (.)	
13	un=WENN ich so=ne widerLEgung !MACH!,	
14	dann schau=ich IMMER drauf,	
15	dass der TEXT,	
16	den ich widerLEgen MÖCHte,	
17	DA auch (.) öfters mal drin VORkommt,	
18	damit ich den !WÜR!dige (.)	
19	und (.) dem (.) verTREter der GEgenseite	
20	ZEIge;	
21	ich hab des schon verSTANden,	
22	was du sagst,	
23	ich NEHM=des auch !AUF!,	
24	UND (-) MANches davon find=ich auch GANZ	
25	überZEUgend.	
26	DESwegen nehm=ich jetz !HIER! mal den PUNKT	L umkreist in dem eingeblendeten Text das Wort *vereinfachen* im ersten Satz (mit lila Stift)
27	mit REIN, (.)	
28	DASS (.) die sozialen NETZwerke (.) die	
29	kommunikaTION verEINfachen (-)	
30	und DASS (.) DAduɐx (.) n SCHNELler (.)	
31	informaTIONSausTAUSCH (.) GANZ gut	
32	funktioNIEren kann. (--)	L umkreist in dem eingeblendeten Text *schneller Austausch von Informationen* (mit lila Stift)
33	marKIERN wir DES mal BEIdes mit !LI!la, (-)	
34	DES is also die FARbe für die (-) eleMENte,	
35	DIE (.) aus dem (.) GEgenargument	
36	!RAUS!genommen sind,	
37	ne. (--)	
38	!HIER! is auch noch ein ziTAT; (-)	
39	die STARke und konsTANTe beZIEHungsarBEI:T,	
40	(-)	
41	des is ja genau !DAS!, (-)	
42		

[80] Objektsprache sowie das, was die Lehrkraft konkret anschreibt, wird kursiv gedruckt.

43	was nach beHAUPtung (.) der auTO:rin (.)	L umkreist *starke und konstante Beziehungsarbeit* (mit lila Stift)
44	MÖ:glich=is über FACEbook und CO. (-)	
45		
46	JA, (.)	
47	aber DA muss ich jetz ANsetzen mit meiner	
48	EIgentlichen widerLE:gung, (-)	
49	weil ich getzt glaube,	
50	dass ge!NAU! !DAS!	
51	– diese !STAR!ke beZIEhungsarBEIT –	L nimmt einen grünen Stift zur Hand (bei den Worten *genau das*)
52	EI:gentlich nicht NÖT…/ nicht MÖ:Glich is,	
53	WENN des nur über FACEbook läuft und OH:ne	
54	perSÖNlichen konTAKT.	
55	ALso (.) muss ich jetz hier beGRÜNden,	
56	waRUM ich glaub,	
57	des is nicht MÖ:Glich, (.)	
58	so wie die auTOrin eigentlich beHAUPtet hat.	L zieht mit dem grünen Stift einen vertikalen Strich rechts neben den relevanten Textabschnitt
59	und HIER sind jetzt die KERNpunkte !MEI!ner	
60	begründung drin, (.)	
61	des marKIER=ich in !GRÜN!.	
62	!ICH! glaub nämlich;	
63	um ne GUte be!ZIEH!ung aufzubauen,	
64	!DA! müssen SCHÜler und LEHrer tatsächlich	L umkreist einzelne Wörter aus dem zuvor als relevant markierten Textabschnitt (*Beziehung, Zeit miteinander verbringen, Klassenzimmer, persönlichen Kontakt*)
65	ZEIT mitnander verBRINgen (-)	
66	und DES passiert am BESten (-) NICHT (.)	
67	vor=m BILDschirm (-) ONLine,	
68	SONdern im KLASsenzimmer (.) im	
69	per!SÖN!lichen kontakt.	
70	des sind also !MEI!ne (.) wesentlichen	
71	PUNKte.	
72	und (.) ich widerSPRECHE dem,	
73	dass des über FACEbook geht,	L zeichnet einen Blitz an das Ende des Textes, in dem es um schnelle Informationen in Facebook geht
74	weil da MEIStens nur KURze infos (.) in (.)	
75	SCHNELlen sätzen AUSgetauscht werden.	
76	des is was ANders als (.) die PFLE:ge einer	
77	starken beZIEHung,	
78	DA:für brauch=ich scho=n TIEfergehenden	
79	AUStausch. (---)	
80	HIER ham wir also nun meine beGRÜNdung (.)	L zeigt mit dem Stift auf den zuvor als relevant markieren Textteil
81	mit der ich reaGIEre auf die punkte,	
82	die ich aus de:m GEgenargument RAUS!genommen	
83	hab (.)	
84	und !SO!mit (.) des GEgenargument widerLEge.	
85	am SCHLUSS rund=ich des ganze AB,	L deutet mit dem Stift auf den letzten Textabschnitt
86	indem ich (.) NOCHmal mich zuRÜCKbezieh zur	
87	THE:se: (.)	
88	die sozialen NETZwerke (.) helfen !NICHT!	
89	dazu,	
90	n gutes verHÄLTnis zwischen SCHÜlern und	
91	LEHrern (.) AUFzubauen.	

Anhang 7 Modelltext für Erklärung in Vignette F

Wie widerlege ich ein Argument?

Sicherlich vereinfachen soziale Netzwerke die Kommunikation dadurch, dass kurze Nachrichten schnell geschrieben und meist sofort vom Adressaten gelesen werden. Für einen schnellen Austausch von Informationen sind sie also gut geeignet. So ein schnelles Informieren ist aber etwas grundsätzlich Anderes als „eine starke und konstante Beziehungsarbeit", wie sie die Autorin besonders für schwierige Schüler fordert. Um eine gute Beziehung zueinander aufzubauen, müssen Lehrer und Schüler tatsächlich Zeit miteinander verbringen, wie man es im Klassenzimmer ja auch tut. Das Entscheidende ist hier der persönliche Kontakt, ein schneller Satz bei Facebook reicht dazu nicht.

Die sozialen Netzwerke helfen deswegen sicherlich nicht dabei, ein gutes Verhältnis zwischen Schülern und Lehrern aufzubauen.

Anhang 8.1 Erhebungsmanual Deutsch MZP1 SuS

Guten Morgen, liebe Schülerinnen und Schüler, ich freue mich, dass ich heute hier bei euch sein kann.

Ich möchte mich zunächst vorstellen: das ist *(begleitende SHK)* und **mein Name** ist Lisa Gaier. Wie ihr ja wisst, kommen wir von der **Universität in Regensburg**. Wir arbeiten dort in einem **großen Projekt** mit, das FALKE heißt. In diesem Projekt wollen wir herausfinden, wie Lehrerinnen und Lehrer im **Deutschunterricht gut erklären**.

Ihr seid Schüler und **hört** jeden Tag viele Erklärungen in unterschiedlichen Fächern.

Es geht heute darum, was aus **eurer Sicht** eine gute Erklärung ist. **Euer Urtei**l ist dabei für uns **sehr wichtig,** weil bisher noch nie jemand Schüler gefragt hat. Deswegen haben wir einen **Fragebogen** entwickelt, in dem ihr jetzt gleich **Videos mit Erklärungen aus dem Deutschunterricht** seht. Nach jedem Video sollt ihr dazu Fragen beantworten.

Was bei der **Beantwortung** der Fragen wichtig ist:

Es gibt **keine richtige und keine falsche** Antwort.

- Bitte seid bei der Beantwortung **ehrlich und ernsthaft**. Die Befragung ist **anonym,** d.h. niemand weiß später, welche Antwort von dir oder von dir stammt. Ihr könnt also ganz unbesorgt eure eigene Meinung sagen.
- **Lest** dazu bitte **alles** gut durch.
- Wenn ihr eine Frage oder ein Wort nicht versteht, **meldet** euch bitte.
- Überlegt euch bitte nicht, wie die Erklärung auf die ganze Klasse wirkt. Überlegt nur: „**Wie finde ich selbst** die Erklärung?"

Habt ihr bis hierher eine Frage? Dann geht es jetzt gleich los.

- Ihr müsst den Fragebogen **selbst** immer weiterklicken: Wenn ihr mit dem Video fertig seid oder mit der Beantwortung der Frage, klickt ihr **rechts unten** auf „**weiter**".
- Relativ zu Beginn werdet ihr aufgefordert, die **Lautstärke** einzustellen; das macht ihr ganz unten rechts, indem ihr auf das **Lautsprechersymbol** klickt.
- Ihr wisst ja, dass die Befragung aus **zwei Teilen** besteht. Nächste Woche kommen wir noch einmal. Ihr bekommt jetzt ein **Codewort**. Das braucht ihr nächste Woche noch einmal, damit wir die einzelnen Teile zusammenführen können. Ihr müsst es also **aufheben**! Wenn das die ganze Klasse schafft, bekommt ihr einen kleinen Preis.

--Codewörter ausgeben--

Schreibt jetzt euer **eigenes** Codewort in das Hausaufgabenheft; lernt es **auswendig** und legt es in euer **Federmäppchen**. Darin bleibt es bis zur nächsten Woche.

--Warten, bis alle fertig sind---

Ihr bekommt jetzt die Links; die tippt ihr in die Browserleiste ein und los geht es. Ihr habt die ganze Stunde Zeit. Wenn ihr fertig seid, meldet euch bitte.

Los geht's und **viel Spaß**!

Anhang 8.2 Erhebungsmanual Deutsch MZP2 SuS

Guten Morgen, liebe Schülerinnen und Schüler, ich freue mich, dass ich heute noch einmal hier bei euch sein kann.

Ihr kennt mich ja schon. Wie ihr ja wisst, kommen wir von der **Universität in Regensburg** und erforschen **gute Lehrererklärungen.**

Heute brauchen wir noch einmal **eure ehrliche Meinung**! Bitte seid wieder **ganz ernsthaft und aufrichtig** bei der Beantwortung der Fragen, - genau wie letzte Woche. Das hat prima geklappt!

Der Fragebogen funktioniert genauso wie der letzte Woche. Als erstes müsst ihr euer Codewort eingeben, dann die Lautstärke einstellen.

Habt ihr eine Frage?

Dann los! **Viel Spaß und vielen Dank**!

Anhang 9.1 Paraphrasierungskatalog MZP1 SuS

Sprechwissenschaft	
Items	**Paraphrasierung**
1. Der Lehrer / Die Lehrerin hatte eine angenehm klingende Stimme.	*Die Stimme des Lehrers / der Lehrerin (klang so, dass du gerne zuhörst (z. B. weil die Stimme klar, locker und voll klang und nicht kratzig oder angestrengt.)*
2. Der Lehrer / Die Lehrerin hatte eine deutliche Aussprache.	*Der Lehrer /Die Lehrerin hat deutlich gesprochen, sodass man die Wörter und Sätze gut verstehen konnte. Der Lehrer / Die Lehrerin hat NICHT genuschelt und die Laute in den Wörtern NICHT „verschluckt".*
3. Der Lehrer / Die Lehrerin hatte eine angemessene Sprechgeschwindigkeit.	*Der Lehrer / Die Lehrerin hat nicht zu schnell und nicht zu langsam gesprochen.*
4. Der Lehrer / Die Lehrerin hat die Sprechpausen so gesetzt, dass ich gut folgen konnte.	*Der Lehrer /Die Lehrerin hat beim Sprechen die Pausen an den richtigen Stellen gemacht.* *Der Lehrer / Die Lehrerin hat beim Sprechen nicht zu wenige und nicht zu viele, nicht zu kurze und nicht zu lange Pausen gemacht.*
5. Der Lehrer / Die Lehrerin hatte eine abwechslungsreiche Sprechweise.	*Der Lehrer / Die Lehrerin hat NICHT eintönig und NICHT langweilig oder leiernd gesprochen. Der Lehrer / Die Lehrerin hat spannend gesprochen und Wichtiges betont.*
6. Der Lehrer / Die Lehrerin hatte einen Körperausdruck, der die Erklärung unterstützt hat.	*Wie der Lehrer / die Lehrerin beim Erklären sich und seine Hände und Arme bewegt hat, aber auch wie er mit dem Gesicht „gesprochen" hat – das hat gut gepasst.*
7. Der Lehrer / Die Lehrerin hat sich den Schülerinnen und Schülern zugewendet.	*Der Lehrer / Die Lehrerin hat sich beim Erklären mit dem Körper zu den Schüler/innen gedreht und sie angeschaut.*
8. Die Art und Weise, wie der Lehrer / die Lehrerin spricht, gefällt mir.	*Alles zusammen – der Klang der Stimme, die Sprechweise und der Körperausdruck des Lehrers / der Lehrerin – gefällt dir. (Insgesamt gefällt es dir, wie der Lehrer gesprochen hat.)*

Persönlichkeitswirkung	
Items	**Erläuterungen**
1. Der Lehrer / Die Lehrerin wirkte in diesem Video auf mich… **…sympathisch.** **…begeistert.** **…natürlich.**	*Wenn man jemanden nett oder lieb oder freundlich findet – dann wirkt die Person auf einen sympathisch.* *Du hast den Eindruck, dass das Thema dem Lehrer / der Lehrerin wirklich wichtig ist und sie sich freut, es dir zu erklären.* *Der Lehrer / Die Lehrerin wirkt wie ein Lehrer / eine Lehrerin, die auch an deiner Schule unterrichten könnte. Er / Sie verhält sich nicht übertrieben oder gestellt. (Dir kommt nichts komisch / geschauspielert vor.)*

Anhang 9.2 Paraphrasierungskatalog MZP2 SuS

Sprachwissenschaft	
Items	**Paraphrasierung**
1. Der Lehrer / Die Lehrerin hat sich gut verständlich ausgedrückt.	*Wo du zugehört hast, hast du gedacht, das habe ich gut verstanden.*
2. Manchmal hat der Lehrer / die Lehrerin extra gesagt, dass etwas sehr wichtig ist.	*Das bedeutet: Der Lehrer / Die Lehrerin hat etwas erklärt und dann noch dazugesagt, dass das sehr wichtig ist.*
3. Bei manchen Wörtern habe ich nicht gewusst, was sie bedeuten.	*Der Lehrer / Die Lehrerin hat manchmal ein Wort gesagt, wo (!) du nicht genau wusstest, was das ist. – Eventuell noch ergänzen: Oder ein Wort, das du nicht gekannt hast.*
4. Manche Sätze hat der Lehrer / die Lehrerin zu lang gemacht.	*Das bedeutet, der Lehrer / die Lehrerin hat zu viel reingepackt in den Satz.*

Fachspezifisch	
Items	**Paraphrasierung**
1. Die Erklärung hat geholfen, das Erklärte genau zu verstehen.	*Durch die Erklärung hast du das Gefühl, dass du ganz genau weißt, was erklärt wurde, nicht nur ein bisschen.*
2. Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.	*Die Erklärung war so, dass du zwar ungefähr verstehst, worum es ging, aber nicht ganz genau.*
3. Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.	*So eine Erklärung (wie im Video) kennst du aus deinem Deutschunterricht, weil deine Lehrerin oder dein Lehrer manchmal auch so erklärt.*
4. Die Erklärung hat mich zum Mitdenken angeregt.	*Während du die Erklärung angehört hast, warst du mit deinen Gedanken beim Inhalt der Erklärung; du hast nicht an etwas anderes gedacht.*

Anhang 9.2 Paraphrasierungskatalog MZP2

Themenspezifisch Video a	
Items	**Paraphrasierung**
1. Es hat beim Verstehen der Redewendung geholfen, dass erklärt wurde, wie sie entstanden ist.	*Die Lehrerin hat ja erklärt, woher die Redewendung kommt. Du hast das Gefühl, dass dir das dabei geholfen hat, diese Redewendung zu verstehen / zu wissen, was diese Redewendung heißt. Verstehen heißt, dass du keine Fragen mehr dazu hast!*
2. Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass erklärt wurde, wie sie entstanden ist.	*Die Lehrerin hat ja erklärt, woher die Redewendung kommt. Du hast das Gefühl, dass du dir dadurch am besten merken kannst, was diese Redewendung heißt. Merken heißt, dass du es auch morgen, übermorgen usw. noch weißt!*
3. Bei Redewendungen sollte im Deutschunterricht regelmäßig erklärt werden, wie sie entstanden sind.	*Du findest es am besten, wenn im Deutschunterricht beim Thema Redewendungen oft erklärt wird, woher diese Redewendungen kommen.*

Themenspezifisch Video b	
Items	**Paraphrasierung**
1. Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Inhalt des Arguments erklärt hat.	*Der Lehrer hat ja erklärt, dass ihn das Argument nicht überzeugt, weil Dinge behauptet werden, die so nicht richtig sind. Du hast das Gefühl, dass dir das dabei geholfen hat, zu verstehen, warum das Argument nicht gut ist. Verstehen heißt, dass du keine Fragen mehr dazu hast!*
2. Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Inhalt des Arguments erklärt hat.	*Der Lehrer hat ja erklärt, dass ihn das Argument nicht überzeugt, weil Dinge behauptet werden, die so nicht richtig sind. Du hast das Gefühl, dass du dir dadurch am besten merken kannst, warum das Argument nicht gut ist. Merken heißt, dass du es auch morgen, übermorgen usw. noch weißt!*
3. Es sollte regelmäßig erklärt werden, warum ein Argument inhaltlich nicht überzeugt.	*Du findest es am besten, wenn beim Thema Argumentieren oft erklärt wird, welche Dinge/Inhalte an einem Argument nicht richtig sind.*

Anhang 9.2 Paraphrasierungskatalog MZP2 SuS

Themenspezifisch Video c	
Items	**Paraphrasierung**
1. Das „laute Denken“, das die Lehrerin vorgemacht hat, hat mir geholfen zu verstehen, wie man die Aussage widerlegt.	*Die Lehrerin hat ja an der Tafel ihre Gedanken laut ausgesprochen. So hat sie gezeigt, wie man auf die Aussage reagiert. Du hast das Gefühl, dass dir das dabei geholfen hat, zu verstehen, wie man auf diese Aussage reagieren kann.* *Widerlegen heißt, auf ein Argument reagieren, mit dem man nicht einverstanden ist.* *Verstehen heißt, dass du keine Fragen mehr dazu hast!*
2. Das „laute Denken“, das die Lehrerin vorgemacht hat, hat mir am meisten geholfen, mir gut merken zu können, wie man die Aussage widerlegt.	*Die Lehrerin hat ja an der Tafel ihre Gedanken laut ausgesprochen. So hat sie gezeigt, wie man auf die Aussage reagiert. Du hast das Gefühl, dass du dir dadurch am besten merken kannst, wie man auf diese Aussage reagieren kann.* *Widerlegen heißt, auf ein Argument reagieren, mit dem man nicht einverstanden ist.* *Merken heißt, dass du es auch morgen, übermorgen usw. noch weißt!*
3. Das „laute Denken“, so wie es die Lehrerin vorgemacht hat, sollte im Schreibunterricht regelmäßig verwendet werden.	*Du findest es am besten, wenn im Deutschunterricht beim Thema Argumentieren oft vorgemacht wird, wie man auf eine Aussage reagiert. Dazu soll deine Lehrerin/dein Lehrer seine/ihre Gedanken laut aussprechen.*

Anhang 9.2 Paraphrasierungskatalog MZP2 SuS

Themenspezifisch Video d	
Items	**Paraphrasierung**
1. Es hat beim Verstehen der Redewendung geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden.	*Die Lehrerin hat ja ganz viele andere Ausdrücke/Redewendungen genannt, die alle das gleiche meinen. Du hast das Gefühl, dass dir das dabei geholfen hat, diese Redewendung zu verstehen / zu wissen, was diese Redewendung heißt.* *Verstehen heißt, dass du keine Fragen mehr dazu hast!*
2. Um sich die Redewendung gut merken zu können, hat es am meisten geholfen, dass andere Ausdrücke mit der gleichen Bedeutung genannt wurden.	*Die Lehrerin hat ja ganz viele andere Ausdrücke/Redewendungen genannt, die alle das gleiche meinen. Du hast das Gefühl, dass du dir dadurch am besten merken kannst, was diese Redewendung heißt.* *Merken heißt, dass du es auch morgen, übermorgen usw. noch weißt!*
3. Bei Redewendungen sollten im Deutschunterricht regelmäßig andere Ausdrücke mit der gleichen Bedeutung genannt werden.	*Du findest es am besten, wenn im Deutschunterricht beim Thema Redewendungen oft viele andere Ausdrücke/Redewendungen genannt werden, die alle das gleiche meinen.*

Themenspezifisch Video e	
Items	**Paraphrasierung**
1. Ich verstehe, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat.	*Der Lehrer hat ja erklärt, dass die Begründung und die Beispiele nicht zur Behauptung passen. Du hast das Gefühl, dass dir das dabei geholfen hat, zu verstehen, warum das Argument nicht gut ist.* *Verstehen heißt, dass du keine Fragen mehr dazu hast!*
2. Ich kann mir gut merken, warum das Argument nicht gelungen ist, weil der Lehrer den falschen Aufbau des Arguments erklärt hat.	*Der Lehrer hat ja erklärt, dass die Begründung und die Beispiele nicht zur Behauptung passen. Du hast das Gefühl, dass du dir dadurch am besten merken kannst, warum das Argument nicht gut ist. Merken heißt, dass du es auch morgen, übermorgen usw. noch weißt!*
3. Es sollte regelmäßig erklärt werden, warum die einzelnen	*Du findest es am besten, wenn beim Thema Argumentieren oft erklärt wird, dass die*

	Bestandteile eines Arguments sich nicht aufeinander beziehen.	*Begründung und die Beispiele nicht zur Behauptung passen.*

Anhang 9.2 Paraphrasierungskatalog MZP2 SuS

Themenspezifisch Video f	
Items	**Paraphrasierung**
1. Dass die Lehrerin ihr vorformuliertes Gegenargument erklärt, hat mir geholfen zu verstehen, wie man die Aussage widerlegt.	*Die Lehrerin hat ja ein fertiges Gegenargument mitgebracht und erklärt. So hat sie gezeigt, wie man auf die Aussage aus dem Text reagiert. Du hast das Gefühl, dass dir das dabei geholfen hat, zu verstehen, wie man auf diese Aussage reagieren kann.* *Widerlegen heißt, auf ein Argument reagieren, mit dem man nicht einverstanden ist.* *Verstehen heißt, dass du keine Fragen mehr dazu hast!*
2. Dass die Lehrerin ihr vorformuliertes Gegenargument erklärt, hat mir am meisten geholfen, mir gut merken zu können, wie man die Aussage widerlegt.	*Die Lehrerin hat ja ein fertiges Gegenargument mitgebracht und erklärt. So hat sie gezeigt, wie man auf die Aussage aus dem Text reagiert. Du hast das Gefühl, dass du dir dadurch am besten merken kannst, wie man auf diese Aussage reagieren kann.* *Widerlegen heißt, auf ein Argument reagieren, mit dem man nicht einverstanden ist.* *Merken heißt, dass du es auch morgen, übermorgen usw. noch weißt!*
3. Vorformulierte Gegenargumente sollten im Schreibunterricht regelmäßig verwendet werden.	*Du findest es am besten, wenn im Deutschunterricht beim Thema Argumentieren oft fertige Gegenargumente mitgebracht werden und erklärt werden.*

Anhang 9.2 Paraphrasierungskatalog MZP2 SuS

Strukturiertheit	Erläuterungen
1. Es ist klar, was genau der Lehrer / die Lehrerin erklären wollte.	*Du könntest also mit wenigen Sätzen sagen, um was es in der Erklärung ging.*
2. Die Erklärung hatte einen roten Faden.	*Etwas hat einen roten Faden. Das bedeutet, es geht in der Erklärung von Anfang bis Ende hauptsächlich um eine Sache. Die Lehrerin / Der Lehrer ist nicht abgeschweift oder es ging nicht plötzlich um ganz andere Dinge. (Das kennst du bestimmt vom Aufsatzunterricht / Geschichten schreiben)*
3. Der Lehrer / Die Lehrerin hat die Erklärung Schritt für Schritt aufgebaut.	*Schritt für Schritt aufbauen bedeutet, eins nach dem anderen zu erklären, in einer guten Reihenfolge. (Vielleicht auch mit dem Leichten zu beginnen und immer schwieriger zu werden, sodass du alles gut verstehen kannst)*
4. Der Lehrer / Die Lehrerin hat zu viele Einzelheiten erklärt.	*Der Lehrer / Die Lehrerin hat zu viele Dinge genannt, die du zum Verstehen nicht gebraucht hast.*
5. Die Erklärung war gut strukturiert.	*Strukturiert, das heißt, dass etwas eine Gliederung hat. Das kennst du vielleicht von einem Aufsatz, wenn man sich überlegen muss, was erzähle ich zuerst und was kommt später dran. Wenn man das gut macht, ist der Aufsatz strukturiert.*

Adressatenorientierung	Erläuterungen
1. Die Erklärung war für mich zu schwierig.	*Du konntest die Erklärung teilweise nicht verstehen.*
2. Die Inhalte der Erklärung hatten etwas mit meinem Leben zu tun.	*Dinge, von denen der Lehrer / die Lehrerin gesprochen hat, kennst du auch aus deinem Alltag.*
3. Die Erklärung passte für mich.	*Die Erklärung war für dich als 9./10.-Klässler genau richtig.*
4. Ich konnte der Erklärung mit meinem Wissen gut folgen.	*Du hast gemerkt, dass du mit dem, was du schon weißt, die Erklärung gut verstehen kannst.*

Anhang 10 Deskriptive und psychometrische Kennwerte auf Itemebene
Übersicht für vertiefte Vignetten

Schülerinnen und Schüler (SuS): n_{max} = 113 Studierende (Stud): n_{max} = 78 Lehrkräfte (L): n_{max} = 35 Didaktiker/-innen (D): n_{max} = 31 **Skala/Itemtext**	**Video A**				**Video B**				**Video C**			
	SuS *M* *(SD)*	**Stud** *M* *(SD)*	**L** *M* *(SD)*	**D** *M* *(SD)*	**SuS** *M* *(SD)*	**Stud** *M* *(SD)*	**L** *M* *(SD)*	**D** *M* *(SD)*	**SuS** *M* *(SD)*	**Stud** *M* *(SD)*	**L** *M* *(SD)*	**D** *M* *(SD)*
Strukturiertheit *Cronbachs Alpha:*	*0,66*	*0,83*	*0,81*	*0,85*	*0,73*	*0,84*	*0,90*	*0,90*	*0,70*	*0,84*	*0,82*	*0,94*
Es ist klar, was genau die Lehrkraft erklären wollte.	1,94 (1,00)	1,88 (0,92)	1,52 (0,63)	1,67 (0,91)	2,23 (1,29)	2,59 (1,01)	2,84 (1,21)	2,52 (1,21)	2,45 (1,24)	2,00 (1,14)	1,65 (0,80)	1,43 (0,60)
Die Erklärung hatte einen roten Faden.	2,48 (1,14)	2,10 (0,98)	1,77 (0,85)	1,62 (0,87)	2,58 (1,18)	2,51 (1,00)	2,90 (1,22)	2,71 (1,12)	2,71 (1,28)	1,98 (1,15)	1,55 (0,72)	1,52 (0,68)
Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut.	2,35 (1,09)	2,18 (1,04)	1,97 (0,98)	1,71 (1,01)	2,52 (1,32)	2,55 (1,04)	3,00 (1,34)	2,62 (1,20)	2,45 (1,13)	1,92 (1,15)	1,55 (0,72)	1,43 (0,68)
Die Lehrkraft hat zu viele Einzelheiten erklärt.*	3,88 (1,43)	3,22 (1,34)	3,29 (1,49)	3,00 (1,45)	3,37 (1,47)	3,31 (1,29)	4,01 (1,57)	3,90 (1,30)	3,37 (1,39)	2,80 (1,31)	2,52 (1,36)	1,76 (0,77)
Die Erklärung war gut strukturiert.**	2,42 (0,98)	2,44 (0,93)	2,32 (1,14)	2,10 (1,22)	2,71 (1,19)	2,86 (1,00)	3,39 (1,26)	3,24 (1,45)	2,71 (1,34)	2,02 (1,11)	2,00 (1,12)	1,57 (0,75)
Adressatenorientierung *Cronbachs Alpha:*	*0,60*	*0,80*	*0,62*	*0,79*	*0,78*	*0,78*	*0,88*	*0,95*	*0,57*	*0,53*	*0,85*	*0,85*
Für Schülerinnen und Schüler war die Erklärung zu schwierig.*	2,47 (1,29)	2,44 (1,09)	1,97 (1,05)	1,86 (0,79)	2,69 (1,42)	3,02 (0,92)	3,65 (1,36)	3,57 (1,47)	2,57 (1,27)	2,37 (1,05)	2,06 (0,77)	1,86 (0,91)
Die Erklärung berücksichtigte die Lebenswelt der Schülerinnen und Schüler.	4,61 (1,27)	3,42 (1,13)	3,77 (1,45)	3,76 (1,05)	4,05 (1,51)	2,96 (1,17)	3,48 (1,29)	3,48 (1,63)	4,26 (1,40)	3,00 (1,00)	2,52 (0,96)	2,71 (1,19)
Die Erklärung passte für Schülerinnen und Schüler dieser Jahrgangsstufe.	2,63 (1,21)	2,36 (0,94)	2,87 (1,12)	2,52 (1,17)	2,82 (1,51)	2,59 (1,00)	3,26 (1,18)	3,19 (1,50)	2,94 (1,45)	2,18 (0,76)	2,03 (0,84)	1,67 (0,80)
Die Erklärung war angemessen didaktisch reduziert.	---	2,48 (0,93)	2,77 (1,12)	2,48 (1,12)	---	2,86 (0,89)	3,55 (1,34)	3,33 (1,62)	---	2,20 (0,76)	2,10 (0,94)	1,67 (0,73)
Schülerinnen und Schüler konnten der Erklärung mit ihrem Wissen gut folgen.	2,13 (1,12)	2,10 (0,84)	2,06 (1,03)	1,95 (0,92)	2,55 (1,36)	2,78 (1,01)	3,32 (1,19)	3,24 (1,67)	2,50 (1,24)	2,00 (0,68)	1,94 (0,77)	1,76 (0,94)
Die Erklärung war adressatengerecht.**	---	2,48 (0,97)	2,58 (1,05)	2,38 (1,24)	---	2,69 (0,96)	3,42 (1,26)	3,24 (1,64)	---	2,20 (0,71)	2,10 (0,98)	1,67 (0,91)
Sprech- und Körperausdruck *Cronbachs Alpha:*	*0,79*	*0,76*	*0,75*	*0,76*	*0,84*	*0,82*	*0,67*	*0,95*	*0,86*	*0,84*	*0,82*	*0,82*
Die Lehrkraft hatte eine angenehm klingende Stimme.	2,22 (0,96)	1,68 (0,78)	2,06 (1,18)	1,55 (0,69)	1,87 (1,06)	1,97 (0,83)	2,31 (0,79)	2,45 (1,04)	2,20 (1,22)	1,89 (0,91)	1,38 (0,72)	1,64 (0,92)
Die Lehrkraft hatte eine deutliche Aussprache.	1,87 (0,88)	1,46 (0,61)	1,69 (0,60)	1,55 (0,69)	2,04 (1,05)	2,14 (0,82)	2,50 (0,97)	2,55 (1,04)	2,22 (1,10)	2,03 (0,83)	1,56 (0,63)	2,73 (1,27)
Die Lehrkraft hatte eine angemessene Sprechgeschwindigkeit.	2,05 (1,04)	1,68 (0,75)	1,75 (0,86)	1,55 (0,69)	2,53 (1,09)	2,81 (1,27)	4,06 (0,93)	4,00 (1,67)	2,07 (0,90)	2,11 (0,97)	1,44 (0,51)	1,82 (0,98)
Die Lehrkraft hat die Sprechpausen so gesetzt, dass ich gut folgen konnte.	2,11 (1,03)	1,92 (0,89)	1,87 (0,96)	1,45 (0,52)	2,65 (1,35)	3,03 (1,36)	4,00 (0,82)	3,82 (1,40)	2,31 (1,17)	2,43 (0,90)	1,75 (0,68)	1,73 (0,65)
Die Lehrkraft hatte eine abwechslungsreiche Sprechweise.	2,87 (1,07)	2,51 (1,35)	2,19 (1,17)	2,09 (0,83)	2,65 (1,32)	2,78 (1,11)	3,75 (1,00)	4,18 (1,33)	2,84 (1,30)	2,78 (1,23)	1,94 (0,85)	2,45 (1,04)

Die Lehrkraft hatte einen Körperausdruck, der die Erklärung unterstützt hat.	1,42 (0,98)	1,35 (0,72)	1,38 (0,81)	1,18 (0,41)	2,87 (1,28)	2,97 (1,40)	3,25 (1,00)	3,45 (1,70)	3,60 (1,20)	3,03 (1,14)	2,44 (0,89)	2,36 (1,29)
Die Lehrkraft hat sich den Schülerinnen und Schülern zugewendet.	2,18 (1,11)	2,27 (1,26)	2,37 (1,41)	2,00 (1,27)	2,02 (1,30)	1,59 (0,76)	2,44 (1,26)	2,55 (1,76)	2,35 (1,24)	2,19 (1,15)	1,81 (0,91)	1,82 (1,40)
Die Art und Weise, wie die Lehrkraft spricht, gefällt mir.**	2,27 (1,01)	2,08 (1,12)	2,06 (1,12)	1,82 (0,60)	2,38 (1,10)	2,51 (1,02)	3,69 (0,79)	3,73 (1,35)	2,65 (1,16)	2,49 (1,10)	1,81 (0,75)	2,00 (1,18)
Sprachliche Verständlichkeit *Cronbachs Alpha:*	*0,54*	*0,55*	*0,32*	*0,31*	*0,61*	*0,42*	*0,50*	*0,57*	*0,44*	*0,41*	*0,14*	*0,75*
Die Lehrkraft hat sich gut verständlich ausgedrückt.	1,95 (0,79)	1,80 (0,88)	1,71 (0,74)	1,62 (0,59)	2,32 (1,22)	2,51 (0,98)	3,13 (1,09)	3,19 (1,47)	2,59 (1,34)	1,78 (0,85)	1,65 (0,71)	1,33 (0,73)
Manchmal hat die Lehrkraft extra gesagt, dass etwas sehr wichtig ist.	2,56 (1,15)	2,66 (1,21)	3,29 (1,66)	3,24 (1,30)	3,12 (1,33)	3,61 (1,10)	3,81 (1,14)	4,19 (1,08)	3,44 (1,21)	2,94 (1,18)	2,84 (1,32)	2,81 (1,45)
Bei manchen Wörtern wissen die Schülerinnen und Schüler eventuell nicht, was sie bedeuten.*	2,98 (1,58)	2,76 (1,27)	2,68 (1,54)	2,76 (1,34)	2,88 (1,55)	2,82 (1,15)	3,51 (1,57)	3,53 (1,50)	2,39 (1,40)	2,27 (0,84)	2,03 (1,22)	2,00 (1,14)
Manche Sätze hat die Lehrkraft zu lang gemacht.*	3,60 (1,30)	3,10 (1,16)	3,16 (1,32)	2,48 (1,08)	3,50 (1,41)	3,73 (1,20)	4,65 (1,17)	3,86 (1,39)	3,32 (1,11)	2,71 (1,17)	2,58 (1,23)	2,38 (1,20)
Persönlichkeitswirkung *Cronbachs Alpha:*	*0,68*	*0,81*	*0,83*	*0,09*	*0,80*	*0,76*	*0,79*	*0,76*	*0,71*	*0,81*	*0,66*	*0,61*
Die Lehrkraft wirkte in diesem Video auf mich sympathisch.	2,40 (1,20)	1,97 (0,96)	2,19 (1,05)	1,64 (0,67)	2,20 (1,10)	2,24 (1,01)	3,25 (0,93)	3,36 (1,29)	2,18 (1,12)	2,14 (1,03)	1,62 (0,50)	1,45 (0,93)
Die Lehrkraft wirkte in diesem Video auf mich begeistert.	2,60 (1,15)	2,49 (1,24)	2,44 (1,09)	2,55 (0,93)	2,71 (1,15)	3,14 (0,98)	3,63 (1,03)	3,73 (1,10)	3,00 (1,15)	2,95 (1,25)	2,31 (0,87)	2,18 (0,75)
Die Lehrkraft wirkte in diesem Video auf mich natürlich.	2,38 (1,10)	2,41 (1,30)	2,38 (1,36)	2,18 (1,08)	2,05 (1,31)	2,46 (1,07)	3,13 (0,89)	3,27 (1,19)	2,42 (1,27)	2,11 (0,97)	1,50 (0,63)	1,82 (1,25)
Erklärtiefe *Cronbachs Alpha:*	*0,74*	*0,84*	*0,86*	*0,86*	*0,81*	*0,87*	*0,88*	*0,91*	*0,80*	*0,87*	*0,86*	*0,90*
Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen.	2,07 (1,06)	2,00 (0,90)	1,77 (0,85)	1,86 (0,85)	2,49 (1,23)	2,94 (1,18)	3,32 (1,28)	3,29 (1,55)	2,68 (1,25)	2,06 (0,92)	2,13 (0,85)	1,71 (0,96)
Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.*	3,31 (1,37)	2,60 (1,18)	2,26 (1,12)	2,57 (1,57)	3,41 (1,43)	3,33 (1,09)	3,00 (1,13)	2,76 (1,14)	3,58 (1,14)	2,71 (1,12)	2,52 (1,15)	1,86 (1,01)
Die Erklärung war kognitiv aktivierend.	3,05 (1,43)	2,52 (1,04)	2,35 (1,14)	2,67 (1,46)	3,26 (1,37)	3,39 (1,19)	3,29 (1,27)	3,43 (1,33)	3,46 (1,47)	2,47 (0,98)	2,32 (1,08)	2,19 (1,21)
[…] hat dabei geholfen zu verstehen, dass […].	2,19 (1,09)	1,80 (0,97)	1,81 (1,11)	1,86 (1,24)	2,62 (1,36)	2,78 (0,92)	3,26 (1,41)	3,10 (1,48)	2,67 (1,30)	1,71 (0,82)	1,68 (0,83)	1,57 (0,98)
[…] hat dabei geholfen, sich […] gut merken zu können.	2,46 (1,33)	2,04 (1,16)	2,06 (1,32)	2,00 (1,38)	2,73 (1,30)	3,20 (1,08)	3,61 (1,45)	3,48 (1,37)	3,04 (1,37)	2,16 (1,16)	2,19 (1,25)	2,24 (1,26)
Fachliche Relevanz *Cronbachs Alpha:*	*0,32*	*0,37*	*0,69*	*0,16*	*0,17*	*0,44*	*0,41*	*0,20*	*0,31*	*0,12*	*0,71*	*0,03*
Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt.	---	2,84 (0,98)	2,52 (1,03)	2,48 (1,17)	---	3,08 (1,15)	3,42 (1,23)	3,24 (1,58)	---	2,06 (0,80)	1,94 (0,81)	1,57 (0,87)
Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.	3,64 (1,30)	3,00 (1,01)	2,23 (1,12)	2,86 (1,20)	2,93 (1,20)	2,43 (1,06)	2,58 (1,12)	2,29 (0,85)	3,46 (1,37)	2,47 (0,98)	1,97 (0,75)	3,14 (1,49)
Bei […] sollte regelmäßig […] erklärt werden.	2,50 (1,43)	2,38 (1,12)	2,26 (1,21)	2,19 (1,17)	2,53 (1,43)	2,20 (0,98)	1,68 (0,95)	1,95 (0,87)	3,10 (1,47)	1,86 (0,98)	1,87 (1,23)	1,57 (0,68)

Bem.: Einheitliche Skalierung aller Items: 1 Stimme voll zu; 2 Stimme zu; 3 Stimme eher zu; 4 Stimme eher nicht zu; 5 Stimme nicht zu; 6 Stimme gar nicht zu; interne Konsistenzen (Cronbachs Alpha) pro Skala, Video (A, B, C, D, E, F) und Statusgruppe (Schülerinnen und Schüler vs. Studierende, Lehrkräfte, Didaktiker/-innen) und arithmetische Mittelwerte

(M) sowie Standardabweichungen (SD) auf Itemebene. Formulierungen der Übersichtlichkeit halber nur SLD-Gruppen. Formulierungen der Schüleritems s. Anhang 3.1 und 3.2 […]: Platzhalter für video- bzw. themenspezifische Formulierungen; *: Item wird in der zugehörigen Skala *rekodiert* berücksichtigt; **: Item wird in der zugehörigen Skala *nicht* berücksichtigt; ---: Item ist in der Skala der jeweiligen Statusgruppe *nicht* enthalten. (Tab. erweitert aus Gaier & Schilcher, 2021, in Vorb.)

Übersicht für einfache Vignetten

Schülerinnen und Schüler (SuS): n_{max} = 113 Studierende (Stud): n_{max} = 78 Lehrkräfte (L): n_{max} = 35 Didaktiker/-innen (D): n_{max} = 31 **Skala/Itemtext**	**Video D**				**Video E**				**Video F**			
	SuS *M* *(SD)*	**Stud** *M* *(SD)*	**L** *M* *(SD)*	**D** *M* *(SD)*	**SuS** *M* *(SD)*	**Stud** *M* *(SD)*	**L** *M* *(SD)*	**D** *M* *(SD)*	**SuS** *M* *(SD)*	**Stud** *M* *(SD)*	**L** *M* *(SD)*	**D** *M* *(SD)*
Strukturiertheit *Cronbachs Alpha:*	*0,66*	*0,84*	*0,86*	*0,87*	*0,57*	*0,87*	*0,74*	*0,83*	*0,55*	*0,80*	*0,90*	*0,95*
Es ist klar, was genau die Lehrkraft erklären wollte.	2,01 (0,99)	2,18 (0,97)	2,52 (1,26)	2,59 (1,37)	2,17 (1,21)	2,32 (1,12)	2,55 (0,89)	2,36 (0,90)	2,76 (1,34)	2,18 (0,94)	1,65 (0,76)	2,10 (1,22)
Die Erklärung hatte einen roten Faden.	2,65 (1,05)	2,63 (1,13)	2,87 (1,23)	2,77 (1,27)	2,59 (1,23)	2,30 (0,86)	2,55 (0,77)	2,41 (0,85)	2,90 (1,20)	2,16 (0,74)	1,68 (0,75)	2,29 (1,19)
Die Lehrkraft hat die Erklärung Schritt für Schritt aufgebaut.	2,63 (1,20)	2,88 (1,07)	3,10 (1,04)	2,82 (1,10)	2,30 (1,06)	2,12 (0,85)	2,39 (0,67)	2,36 (0,85)	2,80 (1,17)	2,10 (0,71)	1,65 (0,80)	2,14 (0,96)
Die Lehrkraft hat zu viele Einzelheiten erklärt.*	3,76 (1,33)	3,45 (1,30)	4,06 (1,21)	3,41 (1,14)	3,42 (1,39)	3,22 (1,27)	3,81 (1,42)	4,00 (1,69)	3,47 (1,25)	2,74 (1,12)	2,68 (1,22)	2,52 (1,17)
Die Erklärung war gut strukturiert.**	2,84 (1,15)	2,94 (1,10)	3,29 (1,01)	2,95 (1,13)	2,48 (1,15)	2,36 (0,92)	3,00 (0,89)	2,86 (1,21)	2,89 (1,25)	2,34 (0,80)	1,87 (0,85)	2,43 (1,17)
Adressatenorientierung *Cronbachs Alpha:*	*0,58*	*0,77*	*0,70*	*0,77*	*0,60*	*0,90*	*0,76*	*0,88*	*0,47*	*0,78*	*0,79*	*0,84*
Für Schülerinnen und Schüler war die Erklärung zu schwierig.*	2,41 (1,43)	2,71 (1,12)	2,52 (0,89)	2,59 (1,10)	2,52 (1,29)	3,12 (1,17)	3,61 (1,28)	3,52 (1,57)	2,67 (1,36)	2,48 (0,86)	2,26 (0,82)	2,43 (1,29)
Die Erklärung berücksichtigte die Lebenswelt der Schülerinnen und Schüler.	4,26 (1,38)	3,14 (1,17)	3,61 (1,17)	3,32 (1,17)	4,21 (1,32)	3,34 (1,14)	3,61 (1,15)	3,67 (1,24)	4,24 (1,44)	3,12 (1,08)	3,00 (1,07)	3,38 (0,87)
Die Erklärung passte für Schülerinnen und Schüler dieser Jahrgangsstufe.	2,90 (1,22)	2,69 (1,07)	3,23 (1,28)	3,23 (1,02)	2,74 (1,26)	2,56 (0,93)	3,10 (0,98)	3,52 (1,17)	3,25 (1,40)	2,52 (0,71)	2,19 (0,65)	2,67 (1,07)
Die Erklärung war angemessen didaktisch reduziert.	---	2,86 (1,04)	3,39 (1,15)	3,23 (1,11)	---	2,82 (1,06)	3,35 (1,05)	3,62 (1,32)	---	2,56 (0,61)	2,16 (0,74)	2,71 (1,38)
Schülerinnen und Schüler konnten der Erklärung mit ihrem Wissen gut folgen.	2,23 (1,17)	2,61 (0,87)	2,55 (1,15)	2,68 (0,10)	2,36 (1,14)	2,58 (0,99)	3,10 (1,08)	3,33 (1,32)	2,65 (1,32)	2,38 (0,60)	2,06 (0,77)	2,38 (1,16)
Die Erklärung war adressatengerecht.**	---	2,88 (1,09)	3,16 (1,21)	3,05 (1,00)	---	2,78 (1,09)	3,16 (1,16)	3,38 (1,32)	---	2,60 (0,81)	2,06 (0,81)	2,48 (1,29)
Sprech- und Körperausdruck *Cronbachs Alpha:*	*0,72*	*0,80*	*0,82*	*0,77*	*0,73*	*0,87*	*0,87*	*0,61*	*0,85*	*0,80*	*0,88*	*0,75*
Die Lehrkraft hatte eine angenehm klingende Stimme.	2,29 (0,94)	1,73 (0,83)	1,76 (0,83)	1,80 (0,94)	1,78 (0,79)	1,97 (0,77)	2,35 (1,06)	1,73 (0,60)	2,51 (1,18)	2,17 (1,02)	1,53 (0,51)	1,33 (0,49)
Die Lehrkraft hatte eine deutliche Aussprache.	2,13 (0,90)	1,73 (0,74)	1,65 (0,93)	1,47 (0,52)	1,78 (0,74)	1,97 (0,85)	2,35 (0,93)	2,00 (0,93)	2,47 (1,20)	2,43 (1,07)	1,88 (0,99)	1,73 (0,70)
Die Lehrkraft hatte eine angemessene Sprechgeschwindigkeit.	2,07 (1,02)	1,63 (0,72)	1,71 (0,77)	1,20 (0,41)	2,58 (1,12)	3,33 (1,47)	3,76 (1,25)	3,40 (1,50)	1,95 (1,01)	2,17 (0,99)	1,82 (0,95)	1,27 (0,46)
Die Lehrkraft hat die Sprechpausen so gesetzt, dass ich gut folgen konnte.	2,42 (1,10)	1,87 (0,68)	1,76 (0,83)	1,60 (0,74)	2,36 (1,22)	3,10 (1,32)	3,82 (0,95)	3,47 (1,46)	2,29 (1,10)	2,10 (0,85)	2,12 (1,17)	1,40 (0,63)
Die Lehrkraft hatte eine abwechslungsreiche Sprechweise.	3,00 (1,15)	2,43 (1,36)	2,35 (0,93)	1,73 (0,70)	2,42 (1,20)	2,90 (1,16)	3,06 (1,03)	3,00 (0,85)	3,04 (1,30)	2,90 (1,00)	2,82 (1,24)	2,67 (0,82)

Die Lehrkraft hatte einen Körperausdruck, der die Erklärung unterstützt hat.	1,85 (1,04)	1,50 (0,86)	1,59 (0,80)	1,53 (1,13)	2,55 (1,30)	3,17 (1,23)	2,82 (1,18)	3,00 (0,93)	3,64 (1,25)	3,67 (0,99)	3,24 (1,30)	2,87 (1,06)
Die Lehrkraft hat sich den Schülerinnen und Schülern zugewendet.	2,07 (1,02)	2,33 (1,16)	1,94 (0,66)	2,33 (1,45)	2,29 (1,12)	2,30 (1,15)	2,71 (1,21)	2,73 (1,16)	3,04 (1,36)	3,00 (1,49)	2,76 (1,09)	2,93 (1,10)
Die Art und Weise, wie die Lehrkraft spricht, gefällt mir.**	2,51 (0,98)	2,10 (1,16)	2,18 (0,88)	1,87 (0,92)	2,09 (0,89)	2,80 (1,06)	3,24 (1,03)	3,00 (9,93)	2,85 (1,22)	2,53 (1,01)	2,00 (1,06)	2,07 (0,59)
Sprachliche Verständlichkeit *Cronbachs Alpha:*	*0,22*	*0,68*	*0,21*	*0,66*	*0,37*	*0,73*	*0,71*	*0,02*	*-0,06*	*0,63*	*0,49*	*0,24*
Die Lehrkraft hat sich gut verständlich ausgedrückt.	2,28 (1,01)	2,14 (0,85)	2,26 (0,77)	2,27 (1,16)	2,13 (1,23)	2,36 (1,10)	3,00 (0,86)	2,86 (1,17)	2,46 (1,14)	1,88 (0,77)	1,61 (0,67)	1,67 (0,86)
Manchmal hat die Lehrkraft extra gesagt, dass etwas sehr wichtig ist.	2,72 (1,22)	2,55 (1,40)	2,65 (1,25)	3,00 (1,35)	2,98 (1,34)	3,02 (1,22)	3,45 (1,18)	3,59 (1,26)	3,31 (1,31)	2,86 (1,18)	2,90 (1,33)	3,05 (1,24)
Bei manchen Wörtern wissen die Schülerinnen und Schüler eventuell nicht, was sie bedeuten.*	2,63 (1,59)	3,08 (1,35)	3,00 (1,46)	2,95 (1,43)	2,57 (1,49)	3,16 (1,28)	3,13 (1,41)	3,41 (1,43)	2,50 (1,30)	2,54 (0,93)	2,19 (1,05)	2,05 (0,86)
Manche Sätze hat die Lehrkraft zu lang gemacht.*	3,63 (1,13)	3,24 (1,26)	3,03 (1,28)	2,86 (1,25)	3,21 (1,45)	3,68 (1,36)	4,65 (1,40)	4,05 (1,17)	3,50 (1,17)	2,98 (1,12)	2,65 (0,95)	2,24 (1,04)
Persönlichkeitswirkung *Cronbachs Alpha:*	*0,53*	*0,80*	*0,54*	*0,69*	*0,75*	*0,80*	*0,78*	*0,82*	*0,85*	*0,71*	*0,80*	*0,62*
Die Lehrkraft wirkte in diesem Video auf mich sympathisch.	2,42 (0,88)	1,93 (1,08)	2,18 (0,88)	1,87 (0,83)	2,04 (0,88)	2,30 (1,06)	2,71 (1,31)	2,73 (1,03)	2,49 (1,39)	2,10 (0,85)	1,53 (0,62)	1,60 (0,63)
Die Lehrkraft wirkte in diesem Video auf mich begeistert.	2,78 (1,05)	2,47 (1,01)	2,59 (1,06)	2,87 (1,36)	2,60 (1,06)	2,83 (1,09)	3,18 (1,13)	3,00 (1,07)	3,09 (1,31)	3,10 (0,76)	3,00 (1,17)	2,87 (1,06)
Die Lehrkraft wirkte in diesem Video auf mich natürlich.	2,38 (0,97)	1,97 (0,81)	2,35 (1,06)	2,07 (0,88)	1,96 (0,84)	2,37 (1,07)	2,82 (1,33)	2,80 (1,01)	2,55 (1,34)	2,03 (0,85)	1,71 (0,92)	1,73 (0,70)
Erklärtiefe *Cronbachs Alpha:*	*0,71*	*0,85*	*0,87*	*0,83*	*0,72*	*0,88*	*0,88*	*0,90*	*0,74*	*0,83*	*0,85*	*0,83*
Die Erklärung hat geholfen, das Erklärte vertieft zu verstehen.	2,34 (1,13)	2,92 (1,11)	3,29 (1,44)	3,36 (1,29)	2,37 (1,13)	2,58 (1,07)	3,00 (1,00)	3,36 (1,47)	2,76 (1,24)	2,38 (0,92)	2,03 (0,71)	2,43 (1,17)
Die Erklärung hat nur oberflächliche Zusammenhänge erläutert.*	3,69 (1,20)	3,67 (1,14)	3,62 (1,33)	3,82 (1,30)	3,32 (1,31)	3,04 (0,95)	2,90 (1,27)	3,14 (1,21)	3,64 (1,21)	2,90 (0,95)	2,48 (1,06)	2,00 (0,95)
Die Erklärung war kognitiv aktivierend.	3,54 (1,49)	3,35 (1,20)	3,52 (1,09)	4,00 (1,23)	3,25 (1,36)	2,98 (1,20)	3,03 (1,05)	3,68 (1,25)	3,61 (1,41)	3,08 (0,99)	2,32 (0,87)	2,81 (1,29)
[…] hat dabei geholfen zu verstehen, dass […].	2,27 (1,19)	2,43 (1,10)	2,77 (1,28)	2,77 (1,15)	2,52 (1,35)	2,40 (1,07)	2,87 (1,02)	3,00 (1,27)	2,63 (1,22)	2,18 (0,87)	1,87 (0,81)	2,43 (1,08)
[…] hat dabei geholfen, sich […] gut merken zu können.	2,56 (1,16)	2,78 (1,29)	3,19 (1,30)	3,59 (1,22)	2,76 (1,23)	2,70 (1,28)	3,13 (1,12)	3,50 (1,41)	2,98 (1,20)	2,66 (1,13)	2,42 (1,18)	2,90 (1,14)
Fachliche Relevanz *Cronbachs Alpha:*	*0,46*	*0,51*	*0,66*	*0,38*	*0,48*	*0,60*	*0,49*	*0,36*	*0,52*	*0,53*	*0,52*	*0,19*
Die Erklärung hat den Aufbau einer spezifischen Kompetenz im Fach Deutsch unterstützt.	---	2,98 (1,07)	3,23 (1,12)	3,41 (1,30)	---	2,42 (0,95)	2,87 (0,99)	3,05 (1,46)	---	2,26 (0,69)	2,10 (0,75)	2,05 (0,97)
Erklärungen wie die eben gesehene kommen im Deutschunterricht vor.	3,12 (1,36)	2,53 (0,97)	2,84 (1,32)	2,32 (1,36)	2,85 (1,31)	2,14 (0,88)	2,55 (1,06)	2,18 (0,91)	3,11 (1,28)	2,44 (0,81)	1,97 (0,55)	2,57 (1,12)
Bei […] sollte regelmäßig […] erklärt werden.	2,50 (1,33)	2,53 (1,22)	2,71 (1,35)	2,68 (1,17)	2,47 (1,25)	2,18 (0,94)	1,87 (0,96)	2,09 (1,02)	3,05 (1,40)	2,68 (1,12)	2,16 (0,93)	2,57 (1,12)

Bem.: Einheitliche Skalierung aller Items: 1 Stimme voll zu; 2 Stimme zu; 3 Stimme eher zu; 4 Stimme eher nicht zu; 5 Stimme nicht zu; 6 Stimme gar nicht zu; interne Konsistenzen (Cronbachs Alpha) pro Skala, Video (A, B, C, D, E, F) und

Statusgruppe (Schülerinnen und Schüler vs. Studierende, Lehrkräfte, Didaktiker/-innen) und arithmetische Mittelwerte (M) sowie Standardabweichungen (SD) auf Itemebene. Formulierungen der Übersichtlichkeit halber nur SLD-Gruppen. Formulierungen der Schüleritems s. Anhang 3.1 und 3.2 […]: Platzhalter für video- bzw. themenspezifische Formulierungen; *: Item wird in der zugehörigen Skala *rekodiert* berücksichtigt; **: Item wird in der zugehörigen Skala *nicht* berücksichtigt; ---: Item ist in der Skala der jeweiligen Statusgruppe *nicht* enthalten. (Tab. erweitert aus Gaier & Schilcher, 2021, in Vorb.)

Anhang 11 Übersicht über sämtliche FALKE-D-Kodes mit Ankerbeispielen

Kode/*Subkode*	**Definition und Ankerbeispiele**
Global	**Definition:** Urteil, das sich auf Erklärqualität insgesamt bezieht, z.B. klar; verständlich; gut **Ankerbeispiel typisch**: *Erklärung überzeugt insgesamt; einleuchtende, da logische Erklärung; ordentlich erklärt; plausibel* **Randständiges Ankerbeispiel:** *Erklärung erfüllt Zweck; kompetenzorientiert*
Fachlich korrekt	**Definition:** inhaltlich/fachlich korrekt/gut **Ankerbeispiel typisch**: *fachlich gut; fachlich korrekt; klare Definition* **Randständiges Ankerbeispiel:** *nicht zu Ende erklärt; hat nichts mit... zu tun*
Direkter Bezug zur Methode	**Definition:** Methode wird direkt genannt (und gewertet) **Ankerbeispiel typisch:** *Strategie mit lautem Denken ist überzeugend; Herkunft des Ausdruckes erklärt* **Randständiges Ankerbeispiel**: *Interessante Hintergrundinformationen*
Direkter Vergleich Variante 1 und 2	**Definition:** Vergleich zwischen Erklärungen, z. B. besser; nun; jetzt; mehr als; ...er; noch… **Ankerbeispiel typisch**: *jetzt besser; erneut; immer noch* **Randständiges Ankerbeispiel:** *nicht zu viele verwandte Sprichwörter, sondern ein prägnantes*
Vertieft vs. oberflächlich	**Definition:** Alles, was sich auf Verarbeitungstiefe und Merken bezieht; z.B. (keine) Weiterverarbeitung; besser behalten; Gebrauchskontext; in Bezug setzen; Anwendung usw. **Ankerbeispiel typisch:** *bleibt im Gedächtnis; wird gut nachvollziehbar; (nicht so) gut merken* **Randständiges Ankerbeispiel:** *kein Feedback, wie man es besser machen könnte*
Fehlende Abstraktion bez. auf Designvorgaben	**Definition:** Kommentare bezogen auf Designvorgaben, z.B. Schüler nicht einbezogen; für Klausurtraining zu...; unterrichtlicher Zusammenhang; Ziel des Unterrichts unklar; Beschreibung von Schülerverhalten **Ankerbeispiel typisch:** *sehr frontal; Schüler nicht wirklich einbezogen* **Randständiges Ankerbeispiel:** *jedoch passen einige Schüler nicht auf*
Sprechen	**Definition:** Alles, was sich auf Prozess des Sprechens bezieht, z.B. Sprechgeschwindigkeit, sowie auf Körperausdruck, Mimik, Gestik (nur genannt nicht zur Veranschaulichung/Unterstützung) **Ankerbeispiel typisch**: *schnelles Tempo; leicht stockend* **Randständiges Ankerbeispiel:** *zu geplant; auswendig gelernt*
Sprache	**Definition:** Bezieht sich auf verwendete Sprache **Ankerbeispiel typisch:** *verständliche Sprache angewandt; gute Sprache; schülergemäße Sprache* **Randständiges Ankerbeispiel:** *griffige Sprache; fachliche Ausdrücke verwendet*

Sonstiges	**Definition:** Alle Beschreibungen (ohne zu werten): z.B. Aufstellen von Fragen und Beantworten dieser Verbesserungsvorschläge: z.B. ... wäre hilfreich gewesen/es fehlt... Kritik am Format: z.B. Inhalt nicht für monologische Erklärung geeignet Nicht Einzuordnendes: z.B. Thematik Waffen; stimme zu; differenziert; Erklärung plastisch
Global Veranschaulichung	**Definition:** Veranschaulichung insgesamt, ohne weitere Konkretisierung **Ankerbeispiel typisch:** *sehr anschaulich; schlecht veranschaulicht* **Randständiges Ankerbeispiel:** *Schritte visualisiert*
Veranschaulichung Beispiel	**Definition**: Beispiele werden konkret genannt; z.B. pos. Gegenbeispiel; Beispiele werden erläutert **Ankerbeispiel typisch:** *an vielen Beispielen erklärt; gute Beispiele* **Randständiges Ankerbeispiel**: *am Beispiel entlang gezogen*
Veranschaulichung Medial	**Definition:** Alle Medien, die erwähnt werden, z.B. Tafelanschrift; Visualisierung; hält es fest; an Tafel fixiert; Markierungen, Farben **Ankerbeispiel typisch*:*** *sehr gut die Tafel* **Randständiges Ankerbeispiel:** *und hält es fest*
Veranschaulichung Gestik	**Definition:** Veranschaulichende Wirkung von Gestik, Mimik, Körpersprache (wichtig: unterstützen; verdeutlichen usw.) **Ankerbeispiele typisch:** *gestisch verdeutlicht; unterstützt Aussagen durch Gestik* **Randständiges Ankerbeispiel:** *eine ausladende Gestik*
Veranschaulichung Inhalt	**Definition:** Alles, was sich konkret auf Inhalt / Methode bezieht, z.B. Veranschaulichung durch geschichtl. Hintergrund **Ankerbeispiel typisch:** *inhaltlich ist seine Erklärung wesentlich anschaulicher* **Randständiges Ankerbeispiel:** *anschaulich dank Positivlösung*
Global Struktur	**Definition:** Auf Struktur insgesamt bezogen, z.B. strukturiert; klar strukturiert **Ankerbeispiel typisch:** *gut strukturiert; etwas unstrukturiert* **Randständiges Ankerbeispiel:** *strukturierend; Logik*
Struktur Gegenstandsbestimmung	**Definition:** Alles, was sich auf Ziel und Gegenstand bezieht, z.B. von Anfang an klar, worum es geht **Ankerbeispiele typisch:** *klar, worum es geht; Ziel klar; nennt Zielsetzung zu Beginn* **Randständiges Ankerbeispiel**: *Übertriebenheit wird nachgeschoben; eigentliches Sprichwort nicht im Fokus*
Struktur Schritt für Schritt	**Definition:** Bezieht sich auf Aufbau und inhaltliche Struktur; z.B. roter Faden; gelungene Entwicklung; stringent **Ankerbeispiele typisch:** *sinnvoll aufgebaut; Schritt für Schritt aufgebaut* **Randständiges Ankerbeispiel:** *zum Schluss zusammengefasst; abschließende Zusammenfassung der angeführten Argumente*

Struktur Fokussierung	**Definition:** Inhaltlich weder zu viel noch zu wenig; beim Wesentlichen bleiben **Ankerbeispiele typisch:** *präzise; zu viele Einzelheiten; zu lang umschrieben; kurz genug* **Randständiges Ankerbeispiel:** *unübersichtlich; unstrukturiert; knapp*
Global Adressatenorientierung	**Definition:** Allgemeines zur Adressatenorientierung, konkrete Erwähnung der Schüler, z.B. zu schwierig; zu leicht; schülernah; schülergerecht **Ankerbeispiele typisch:** *schülerangepasst; wird von Schülern nachvollzogen* **Randständiges Ankerbeispiel:** *gymnasiales Niveau?; bleibt unterkomplex*
Adressatenorientierung Vorwissen	**Definition:** Bezieht sich konkret auf Vorwissen der Schüler, z.B. setzt... bei Schülern voraus; Vorwissen; ... ist evtl. auch nicht bekannt **Ankerbeispiel typisch:** *an Vorwissen der Schüler anschließend; gut an Vorwissen angeknüpft* **Randständiges Ankerbeispiel:** *Vor der Erklärung fand eine kurze Wiederholung statt*
Adressatenorientierung Motivation	**Definition:** Alles, was sich auf Emotionen bezieht, z.B. hält Interesse aufrecht; zu langatmig; spannend **Ankerbeispiel typisch:** *langweilig; sehr uninteressant erklärt* **Randständiges Ankerbeispiel:** *Zuschauer bei sich halten können*
Adressatenorientierung Lebensweltbezug	**Definition:** konkreter Lebensweltbezug erwähnt **Ankerbeispiel typisch:** *Erklärung weicht sehr von Lebenswelt der Schüler ab; für Schüler aus deren Sicht eventuell schlecht nachvollziehbar* **Randständiges Ankerbeispiel:** *Erklärung nimmt Bezug auf Sprachwirklichkeit der Schülerinnen und Schüler*

Erläuterungen: Linke Spalte Kodes und *Subkodes;*
Rechte Spalte zunächst **Definition** und abstrahierte Beispiele zur Orientierung; ***Ankerbeispiel typisch*** und ***randständiges Ankerbeispiel*** wörtlich aus Material zitiert; **typisch** entspricht der Definition und/oder wird häufig genannt; **randständig** bedeutet kein prototypisches Beispiel, fällt aber noch in diese Kategorisierung.