

**Das Interventionsverhalten von Studierenden mit divergierender
prozessdiagnostischer Fähigkeit "Deuten"**

von

Moritz Walz
aus Karlsruhe

Angenommene Dissertation zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

Fachbereich 7: Mathematik / Naturwissenschaften
Universität Koblenz-Landau

Berichterstatter:

Prof. Dr. Jürgen Roth, Universität Koblenz-Landau

Prof. Dr. Katja Lengnink, Justus-Liebig-Universität Gießen

Tag der Disputation: 13.08.2020

Danksagung

Das Schreiben einer Doktorarbeit verlangt nicht nur dem Verfasser einiges an Kraft und Zeit ab. Viele Wegbegleiterinnen und Wegbegleiter opfern ebenfalls ihre Zeit, um angeregt über die Inhalte zu diskutieren und die Arbeit dadurch voranzubringen, oder um einen von Problemen abzulenken und in schwierigen Phasen bei Laune zu halten. Den tief empfundenen Dank, den ich diesen Personen gegenüber empfinde, werde ich vermutlich nicht adäquat in Worte fassen können. Trotzdem würde ich es hiermit gerne versuchen.

An erster Stelle will ich meinem Doktorvater Prof. Dr. Jürgen Roth danken, der mich von der ersten bis zur letzten Minute mit voller Tatkraft unterstützt hat. Er hatte stets ein offenes Ohr für meine Anliegen und half mir mit seinem fundierten Wissen in den verschiedensten Bereichen, aber auch mit seiner respektvollen, lösungsorientierten und vertrauensvollen Art. Er brachte immer wieder neue Gedankengänge und Perspektiven mit ein, um die Arbeit voranzubringen. Seine konstruktiven Beiträge trugen maßgeblich zum Gelingen dieser Arbeit bei. Die umfassende Betreuung übertraf jegliche Erwartungen und ließ keinerlei Wünsche offen.

Mein Dank gilt außerdem Frau Prof. Dr. Katja Lengnink für die Vielzahl ihrer hilfreichen und kritischen Anmerkungen, die teilweise neue Aspekte und Facetten der Arbeit bzw. der Interpretation der Ergebnisse beleuchteten. Darüber hinaus schaffte sie durch ihre offene Art eine sehr angenehme Atmosphäre in den gemeinsamen Treffen und im persönlichen Austausch.

Herzlich bedanken möchte ich mich darüber hinaus beim Institut für Mathematik, in dem ich mich schnell heimisch gefühlt habe, und insbesondere bei der Arbeitsgruppe „Didaktik der Mathematik (Sekundarstufen)“, welche im Rahmen von diversen Forschungsworkshops ihre Gedanken und Anmerkungen zu meiner Arbeit eingebracht und mir dadurch sehr geholfen hat. Darüber hinaus fungierten einige Personen der Arbeitsgruppe als Expertinnen und Experten zur Bewertung der Videovignetten sowie beim Raten der Interventionsvideos. Hervorzuheben sind hierbei Tobias Rolfes, Christian Fahse und Patrizia Enenkiel, die zusätzlich auch Anlaufstelle bei Fragen rund um die statistische Auswertung waren, sowie Michaela Lichti. Außerdem möchte ich mich bei den Doktorandinnen bedanken, die ebenfalls mit ViviAn arbeiten und mit denen ich mich insbesondere über das Konstrukt „Diagnosekompetenz“, aber auch über vieles mehr, austauschen konnte. Vielen Dank Marie-Elene Bartel, Rita Hofmann und nochmals Patrizia Enenkiel. Ebenfalls bedanken will ich mich bei den Mitarbeiterinnen und Mitarbeitern der Didaktik der Mathematik (Primarstufe), die im Rahmen der Forschungsworkshops mitdiskutiert und als Teil der großen Expertengruppe die Interventionsvideos bewertet haben.

Außerdem gilt mein Dank dem Projekt „MoSAiK“ der Universität Koblenz-Landau, welches mir im Rahmen der ersten Förderphase die finanziellen Mittel zur Verfügung stellte, um meine Forschung durchführen zu können und es mir obendrein ermöglichte, an vielen nationalen und internationalen Tagungen teilzunehmen. Durch Personen unterschiedlicher Fachgebiete, die an dem Projekt beteiligt waren, ergaben sich gänzlich neue Perspektiven und darüber hinaus auch bis heute bestehende Freundschaften.

Über die gesamte Zeit gesehen hatte ich mehrere Hilfskräfte beschäftigt, die mir mit teils kleineren Tätigkeiten, wie beispielsweise der Transkription der Interventionsvideos und der Reflexionsvideos, aber auch mit längerfristig angelegten Tätigkeiten, insbesondere der Kodierung, halfen. Zwei davon möchte ich gerne hervorheben, da sie über mehrere Jahre mit mir an der qualitativen Auswertung gearbeitet und durch ihre genaue und umsichtige Kodierung einen Teil zur Arbeit beigetragen haben: Annika Haß und Lena Müller.

Nicht minder möchte ich Barbara Hoch danken, die so freundlich war, meine Arbeit Korrektur zu lesen und mir noch einige nützliche Rückmeldungen gab.

Ein besonderer Dank gilt meiner Lebensgefährtin, Patrizia Enenkiel, die am meisten mit mir über die Arbeit diskutiert und philosophiert hat, die Verständnis dafür zeigte, dass ich die letzten Jahre den Großteil des Urlaubs am Schreibtisch verbracht habe, und die mir in der kompletten Zeit den Rücken stärkte.

Nicht zuletzt möchte ich mich bei meinen Freunden und meiner Familie bedanken, die durch meine Arbeit an der Dissertation häufig auf meine Anwesenheit verzichten mussten. Meinen Eltern, Heide und Jürgen Walz, danke ich, dass sie mir mein Studium ermöglicht und mir dabei in allen Bereichen unter die Arme gegriffen haben, in denen es ihnen möglich war. Sie standen auch während der Arbeit an der Dissertation immer hinter mir und haben nie an mir gezweifelt. Meinem Opa, Herbert Bevier, der mir in vielen Bereichen als Vorbild dient, möchte ich diese Arbeit widmen.

Landau, 04.05.2020

Zusammenfassung

Eine zutreffende Diagnose über den aktuellen Kenntnisstand der jeweiligen Schülerinnen und Schüler ist notwendig, um adäquat in Gruppenarbeitsprozesse intervenieren zu können. Von diesem Zusammenhang wird in der Literatur weitgehend ausgegangen, jedoch gibt es bisher kaum empirische Studien, die diesen belegen. Die vorliegende Arbeit widmet sich schwerpunktmäßig dem Interventionsverhalten von Studierenden. Dabei wird die prozessdiagnostische Fähigkeit „Deuten“ zugrundegelegt, um unterschiedliches Interventionsverhalten auf diese Fähigkeit zurückführen zu können. Sowohl beim Aufbau diagnostischer Fähigkeiten als auch bei der (Weiter-)Entwicklung des eigenen Lehrerhandelns gilt Reflexion als hilfreich. Entsprechend wird auch das Zusammenspiel von Prozessdiagnose und Reflexionsverhalten sowie von Interventionsverhalten und Reflexionsverhalten untersucht.

Für die Erhebung der prozessdiagnostischen Fähigkeit „Deuten“ wurden drei Videovignetten erstellt und in das Videodiagnosetool ViviAn eingebunden. Die Videovignetten zeigen jeweils vier Schülerinnen, die sich mit dem Thema „Terme“ beschäftigen. Im Rahmen eines Lehr-Lern-Labores wurden über vier Semester hinweg alle teilnehmenden Studierenden dazu angehalten, die Videovignetten zu bearbeiten. Ebenso konzipierten sie jeweils zu dritt eine Laborstation im Mathematik-Labor „Mathe ist mehr“ und erprobten diese mit einer Schulklasse. Dabei wurden die Interventionen der Studierenden in die Gruppenarbeitsprozesse der Schülerinnen und Schüler videographiert. Anschließend reflektierten die Studierenden in Kleingruppen über die Erprobungen und über die getätigten Interventionen. Die Reflexionsgespräche wurden ebenfalls videographiert.

Es zeigt sich, dass die Studierenden, die sich zum Zeitpunkt der Erhebung im Masterstudium befanden, noch Entwicklungsspielraum in Bezug auf ihre prozessdiagnostische Fähigkeit „Deuten“ besitzen. Im Hinblick auf die Interventionen waren responsive Interventionen häufiger angemessen als invasive Interventionen, wobei responsive Interventionen auch vergleichsweise häufiger dazu führten, dass mehr Schülerinnen und Schüler nach der Intervention aktiv waren. Studierende mit höherer prozessdiagnostischer Fähigkeit „Deuten“ intervenierten jedoch häufiger invasiv und tätigten dabei trotzdem angemessenere und aktivierendere Interventionen als ihre Kommilitoninnen und Kommilitonen. Entsprechend scheint sich die prozessdiagnostische Fähigkeit „Deuten“ positiv auf die Interventionen der Studierenden auszuwirken und sollte daher bereits im Rahmen des (Lehramts-)Studiums verstärkt geschult werden.

Inhaltsverzeichnis

Tabellenverzeichnis	IV
Abbildungsverzeichnis	VII
1 Einleitung	1
2 Diagnostische Kompetenz von Lehrkräften	5
2.1 Diagnostische Fähigkeiten als Teil der diagnostischen Kompetenz.....	5
2.2 Pädagogische Diagnostik.....	9
2.3 Zeitpunkt, Art und Funktion der Diagnose.....	10
2.4 Komponenten der Diagnose.....	13
2.5 Bisherige Forschung und empirische Ergebnisse.....	16
2.6 Zusammenfassung und Verortung der eigenen Studie.....	17
3 Interventionen in Gruppenarbeitsprozesse	21
3.1 Lehrerimpulse.....	21
3.2 Scaffolding.....	22
3.3 Lehrerinterventionen.....	26
3.4 Modelle zur Erfassung von Interventionen.....	29
3.5 Bisherige Forschung und empirische Befunde.....	33
3.6 Zusammenfassung und Übersicht.....	35
4 Reflexion	36
4.1 Begriffsklärung.....	36
4.2 Modelle zur Erfassung von Reflexion.....	39
4.3 Empirische Ergebnisse.....	42
4.4 Zusammenfassung.....	43
5 Fachdidaktische Ausbildung im Fach Mathematik am Campus	
Landau	45
5.1 Gesamtkonzept.....	45
5.1.1 Vernetzung der mathematikdidaktischen Ausbildungsinhalte.....	45
5.1.2 Videodiagnosetool ViviAn.....	46
5.1.3 Mathematik-Labor „Mathe ist mehr“.....	48
5.2 Konzept des Didaktischen Seminars.....	51
5.2.1 Veranstaltungskonzept.....	51
5.2.2 Ziele der Veranstaltung.....	53
5.3 Zusammenfassung.....	53
6 Grundvorstellungen zu Termen	55
6.1 Variablen.....	56
6.2 Terme.....	66

7	Forschungsfragen.....	71
8	Methodisches Vorgehen	73
8.1	Qualitative Inhaltsanalyse.....	73
8.1.1	Beschreibung	73
8.1.2	Zentrale Begriffe	74
8.1.3	Techniken	75
8.2	Auswertung im Bereich Diagnose	77
8.2.1	Auswahl der Videosequenzen	78
8.2.2	Fachdidaktische Aspekte	78
8.2.3	Diagnoseaufträge	79
8.2.4	Expertenrating	80
8.2.5	Einbettung in ViviAn	80
8.2.6	Ermittlung der prozessdiagnostischen Fähigkeit.....	82
8.2.7	Quantitative Validierung des Testinstrumentes	83
8.3	Auswertung im Bereich Interventionen.....	88
8.3.1	Einbettung in das Didaktische Seminar.....	89
8.3.2	Datengenerierung	89
8.3.3	Analyse der Hilfen.....	89
8.3.4	Analyse der Angemessenheit	92
8.3.5	Weitere Analysen	95
8.4	Auswertung im Bereich Reflexion	96
8.4.1	Einbettung in das Didaktische Seminar.....	96
8.4.2	Datengenerierung und Analyse der Reflexionsvideos	97
8.5	Beschreibung der Population	100
8.5.1	Auswahl der Expertinnen und Experten.....	100
8.5.2	Auswahl der Kodiererrinnen und Kodierer	101
8.5.3	Zusammensetzung der Studierenden.....	101
8.5.4	Zusammensetzung der Lehrkräfte	101
8.5.5	Verschränkung mit anderen Erhebungen	102
8.6	Überblick	102
9	Ergebnisse.....	104
9.1	Diagnostische Fähigkeiten.....	104
9.2	Interventionen in Gruppenarbeitsprozesse	111
9.2.1	Hilfen innerhalb der Interventionen	112
9.2.2	Reihenfolge der Hilfen	118
9.2.3	Angemessenheit.....	120
9.2.4	Anzahl der aktiven Schülerinnen und Schüler	127
9.2.5	Interventionsauslöser.....	129
9.2.6	Zusammenfassung	131
9.3	Reflexion.....	133
9.4	Zusammenhänge zwischen Diagnose und Intervention	140

9.4.1	Hilfen innerhalb der Intervention	140
9.4.2	Reihenfolge der Hilfen	149
9.4.3	Angemessenheit.....	152
9.4.4	Anzahl der aktiven Schülerinnen und Schüler	158
9.4.5	Interventionsauslöser	166
9.4.6	Zusammenfassung	172
9.5	Zusammenhänge zwischen Diagnose und Reflexion	174
9.6	Zusammenhänge zwischen Intervention und Reflexion.....	179
9.6.1	Gruppeneinteilung nach Angemessenheit der Interventionen	180
9.6.2	Gruppeneinteilung nach Anzahl der Interventionen	185
9.6.3	Gruppeneinteilung nach Art der getätigten Interventionen	196
9.6.4	Zusammenfassung	206
10	Diskussion	209
10.1	Diskussion der Ergebnisse	209
10.2	Limitationen.....	214
10.3	Forschungsdesiderate.....	215
11	Literatur.....	216
12	Anhang.....	228

Tabellenverzeichnis

Tabelle 1: Werte und Funktionswerte für die Variable x im Beispiel für die Variable als Veränderliche.....	65
Tabelle 2: Eigenwerte der Faktoren des Screeplots	84
Tabelle 3: Ladungen der Items auf das latente Konstrukt.....	85
Tabelle 4: Übersicht über die punktbiseriale Korrelation aller Items.....	86
Tabelle 5: Übersicht über die endgültige punktbiseriale Korrelation der Items.....	86
Tabelle 6: Absolute und relative Lösungshäufigkeiten sowie Itemschwierigkeiten aller verwendeten Items im Rasch-Modell.....	87
Tabelle 7: Absolute und relative Häufigkeit korrekt beantworteter Diagnoseaufträge der Studierenden	106
Tabelle 8: Absolute und relative Häufigkeit korrekt beantworteter Diagnoseaufträge der Lehrkräfte	109
Tabelle 9: Zeitdauer der verschiedenen Hilfen mit Angabe der jeweilig längsten und der kürzesten Hilfe	113
Tabelle 10: Vergleich der Zeitdauer zwischen dominanten und untergeordneten Hilfen hinsichtlich aller Hilfskategorien.....	114
Tabelle 11: Auftretende Kombinationen von Hilfe-Kategorien innerhalb der Interventionen.....	116
Tabelle 12: Reihenfolge der auftretenden Hilfen	118
Tabelle 13: Anzahl der zuerst genutzten Hilfe innerhalb einer Intervention.....	119
Tabelle 14: Anzahl der Kriterien, die exklusiv zur Stütze der Bewertung der Angemessenheit angegeben wurden	122
Tabelle 15: Auftretende Kombinationen mehrerer Kriterien der Angemessenheit.....	123
Tabelle 16: Die häufigsten beanstandeten Kriterien nach Interventionsart.....	124
Tabelle 17: Anzahl der Kriterien, die exklusiv zur Stütze bei den als unangemessen bewerteten Interventionen angegeben wurden	125
Tabelle 18: Auftretende Kombinationen mehrerer Kriterien der Angemessenheit bei als unangemessen bewerteten Interventionen.....	126
Tabelle 19: Veränderung des Arbeitsverhalten der Schülerinnen und Schüler	128
Tabelle 20: Veränderung des Arbeitsverhalten der Schülerinnen und Schüler getrennt nach angemessenen und unangemessenen Interventionen	128
Tabelle 21: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern	130
Tabelle 22: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“	134
Tabelle 23: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“.....	135
Tabelle 24: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“	137
Tabelle 25: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien hinsichtlich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“	138
Tabelle 26: Übersicht über die Reflexionsmomente aufgeteilt auf die Gruppen von Studierenden und die Hauptkategorien „Rückblick“, „Wesentliche Aspekte“ und „Alternativen“	139
Tabelle 27: Zeitdauer der verschiedenen Hilfen mit Angabe der jeweilig längsten und der kürzesten Hilfe der jeweiligen Gruppen.....	145
Tabelle 28: Durchschnittliche Zeitdauer der verschiedenen Hilfen sortiert nach dominanten und untergeordneten Hilfen der jeweiligen Gruppen.....	147
Tabelle 29: Reihenfolge der auftretenden Hilfen sortiert nach Gruppen	150
Tabelle 30: Anzahl der zuerst genutzten Hilfe innerhalb einer Intervention sortiert nach Gruppen.....	151
Tabelle 31: Die häufigsten beanstandeten Kriterien hinsichtlich der verschiedenen Interventionsarten sortiert nach Gruppen.....	156
Tabelle 32: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe H	160
Tabelle 33: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe M.....	161

Tabelle 34: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe N	162
Tabelle 35: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe H getrennt nach angemessenen und unangemessenen Interventionen	163
Tabelle 36: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe M getrennt nach angemessenen und unangemessenen Interventionen	163
Tabelle 37: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe N getrennt nach angemessenen und unangemessenen Interventionen	164
Tabelle 38: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern bei den Interventionen der Gruppe H.....	170
Tabelle 39: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern bei den Interventionen der Gruppe M	171
Tabelle 40: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern bei den Interventionen der Gruppe N.....	171
Tabelle 41: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“	176
Tabelle 42: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“	177
Tabelle 43: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“	177
Tabelle 44: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“	178
Tabelle 45: Normierte Übersicht über die Reflexionsmomente der Studierenden, eingeteilt nach der Angemessenheit der Interventionen.....	181
Tabelle 46: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen	182
Tabelle 47: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen.....	182
Tabelle 48: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen.....	183
Tabelle 49: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen	184
Tabelle 50: Normierte Übersicht über die Reflexionsmomente der Studierenden, eingeteilt nach der Anzahl der Interventionen	186
Tabelle 51: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen	187
Tabelle 52: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen.....	187
Tabelle 53: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen.....	188
Tabelle 54: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Anzahl der getätigten Interventionen	189
Tabelle 55: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Häufigkeit und der Angemessenheit der getätigten Interventionen	193
Tabelle 56: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Häufigkeit und der Angemessenheit der getätigten Interventionen.....	193
Tabelle 57: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen.....	194

Tabelle 58: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Anzahl und der Angemessenheit der getätigten Interventionen	195
Tabelle 59: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Art der getätigten Interventionen	198
Tabelle 60: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Art der getätigten Interventionen.....	198
Tabelle 61: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Art der getätigten Interventionen.....	199
Tabelle 62: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Art der getätigten Interventionen.....	200
Tabelle 63: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Art der getätigten Interventionen	202
Tabelle 64: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Art der getätigten Interventionen.....	203
Tabelle 65: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Art der getätigten Interventionen.....	204
Tabelle 66: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Art und der Angemessenheit der getätigten Interventionen	205

Abbildungsverzeichnis

Abbildung 1: Übersicht über die drei zu unterscheidenden Ebenen hinsichtlich der Forschungsvorhaben im Mathematik-Labor "Mathe ist mehr"	2
Abbildung 2: Diagnostische Kompetenz (in einem weiteren Verständnis) als Kontinuum nach Leuders et al. (2018).....	6
Abbildung 3: Einbettung diagnostischer Fähigkeiten im COACTIV-Modell zur professionellen Kompetenz von Lehrkräften nach Brunner et al. (2011, 217).....	8
Abbildung 4: Modell der diagnostischen Kompetenz unter Berücksichtigung des Zeitpunkts nach Klug et al. 2013, 39	11
Abbildung 5: Diagnosemodell unter der Berücksichtigung des Zeitpunktes und Ausdifferenzierung der aktionalen Phase nach Reinhold (2018, 133).....	13
Abbildung 6: Charakterisierende Komponenten der Diagnose mit fachdidaktischen Überlegungen als Bezugspunkt nach Münster und von Aufschnaiter (2019) sowie Beretz, Lengnick und von Aufschnaiter (2017)	15
Abbildung 7: Verortung der eigenen Studie in die diagnostische Kompetenz.....	19
Abbildung 8: Konzeptmodell des Scaffoldings von van de Pol et al. (2010).....	24
Abbildung 9: Kategoriensystem lösungsprozessbegleitender Interventionen nach Leiss (2007, 109).....	29
Abbildung 10: Ablaufmodell einer Lehrerintervention bei einer selbstdifferenzierenden Aufgabenstellung von Scherres (2013, 42).....	30
Abbildung 11: Reflexionsprozess nach Boud et al. (2015).....	40
Abbildung 12: Zyklischer Prozess der Reflexion nach Korthagen (1985).....	41
Abbildung 13: Oberfläche der Lernumgebung ViviAn.....	46
Abbildung 14: Schülergruppe beim Arbeiten mit gegenständlichen Materialien	49
Abbildung 15: Schülergruppe, die mit einer Computersimulation arbeitet.....	49
Abbildung 16: Simulation zur Erarbeitung der Volumeninhaltsformel von Quadern.....	50
Abbildung 17: Übersicht über die Aktivitäten im Didaktischen Seminar nach Roth (2020)	51
Abbildung 18: Einteilung der verschiedenen Variablenaspekte.....	63
Abbildung 19: Regelmäßig wachsendes Punktmuster	68
Abbildung 20: Ablaufmodell strukturierender Inhaltsanalyse (allgemein) nach Mayring (2015)	76
Abbildung 21: Ablaufschema einer inhaltlich strukturierenden Inhaltsanalyse nach Kuckartz (2016)	77
Abbildung 22: Screenshot einer in ViviAn eingebundenen Videovignette mit zusätzlich eingeblendetem Diagnoseauftrag unterhalb der Vignette	81
Abbildung 23: Screenshot einer in ViviAn eingebundenen Videovignette.....	81
Abbildung 24: Screeplot im Rahmen der Hauptkomponentenanalyse.....	84
Abbildung 25: Item Characteristic Curves (ICCs)	88
Abbildung 26: Schaubild des Prozesses zur Bewertung der Angemessenheit der Interventionen der Studierenden durch verschiedene Expertengruppen	94
Abbildung 27: Erweiterte Übersicht über die drei zu unterscheidenden Ebenen hinsichtlich der Forschungsvorhaben im Mathematik-Labor "Mathe ist mehr"	103
Abbildung 28: Übersicht über die Testergebnisse der Studierenden hinsichtlich der prozessdiagnostischen Fähigkeit „Deuten“	105
Abbildung 29: Übersicht über die Testergebnisse der Lehrkräfte hinsichtlich der prozessdiagnostischen Fähigkeit „Deuten“	111
Abbildung 30: Verteilung der dominanten und untergeordneten Hilfen innerhalb der getätigten Interventionen der Studierenden	112
Abbildung 31: Auftretende Kombinationen von Hilfe-Kategorien innerhalb der Intervention in gestapelten Diagrammen.....	116
Abbildung 32: Verteilung der Angemessenheit auf die dominanten Hilfe-Kategorien	120

Abbildung 33: Prozentualer Anteil der angemessenen Interventionen bezogen auf die jeweilige Interventionsart	121
Abbildung 34: Durch die Expertinnen und Experten beanstandete Kriterien innerhalb der Interventionen	121
Abbildung 35: Durch die Expertinnen und Experten beanstandete Kriterien innerhalb der als unangemessen bewerteten Interventionen	125
Abbildung 36: Arbeitsverhalten der Schülerinnen und Schüler vor und nach der Intervention der angehenden Lehrkraft	127
Abbildung 37: Verteilung der responsiven und invasiven Interventionen	129
Abbildung 38: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers	131
Abbildung 39: Übersicht über die Reflexionsmomente aller Studierenden hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen"	133
Abbildung 40: Verteilung der dominanten Hilfen innerhalb der Interventionen unterteilt nach Gruppen	142
Abbildung 41: Verteilung der untergeordneten Hilfen innerhalb der Interventionen unterteilt nach Gruppen	143
Abbildung 42: Verteilung aller vorkommenden Hilfen innerhalb der Interventionen unterteilt nach Gruppen	144
Abbildung 43: Angemessenheit der Interventionen sortiert nach Gruppen	152
Abbildung 44: Relative Häufigkeit der angemessenen Interventionen sortiert nach Gruppen	153
Abbildung 45: Angemessenheit der einzelnen Interventionsarten sortiert nach Gruppen	154
Abbildung 46: Durch die Expertinnen und Experten beanstandete Kriterien innerhalb der Interventionen nach Gruppen	155
Abbildung 47: Von den Expertinnen und Experten beanstandete Kriterien bei den als unangemessen bewerteten Interventionen sortiert nach Gruppen	158
Abbildung 48: Arbeitsverhalten der Schülerinnen und Schüler vor und nach den Interventionen der Gruppe H159	
Abbildung 49: Arbeitsverhalten der Schülerinnen und Schüler vor und nach den Interventionen der Gruppe M	160
Abbildung 50: Arbeitsverhalten der Schülerinnen und Schüler vor und nach der Interventionen der Gruppe N 161	
Abbildung 51: Absolute Häufigkeiten der Auslöser der Interventionen der Studierenden mit hoher Ausprägung der diagnostischen Fähigkeit "Deuten"	166
Abbildung 52: Absolute Häufigkeiten der Auslöser der Interventionen der Studierenden mit mittlerer Ausprägung der diagnostischen Fähigkeit "Deuten"	166
Abbildung 53: Absolute Häufigkeiten der Auslöser der Interventionen der Studierenden mit niedriger Ausprägung der diagnostischen Fähigkeit "Deuten"	167
Abbildung 54: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers bei den Interventionen von Gruppe H	168
Abbildung 55: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers bei den Interventionen von Gruppe M	169
Abbildung 56: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers bei den Interventionen von Gruppe N	169
Abbildung 57: Übersicht über die absolute Häufigkeit der Reflexionsmomente aller Studierenden hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen", eingeteilt hinsichtlich der diagnostischen Fähigkeit "Deuten"	175
Abbildung 58: Übersicht über die relative Häufigkeit der Reflexionsmomente aller Studierenden hinsichtlich der drei Hauptkategorien „Rückblick“, „Wesentliche Aspekte“ und „Alternativen“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“	175
Abbildung 59: Übersicht über die absolute Häufigkeit der Reflexionsmomente der zwei Gruppen von Studierenden auf Grundlage der Einteilung nach der Angemessenheit der getätigten Interventionen	180
Abbildung 60: Übersicht über die absolute Häufigkeit der Reflexionsmomente der zwei Gruppen von Studierenden auf Grundlage der Einteilung nach der Anzahl der getätigten Interventionen	186

Abbildung 61: Übersicht über die absolute Häufigkeit der Reflexionsmomente der vier Gruppen von Studierenden auf Grundlage der Einteilung nach der Anzahl sowie der Angemessenheit der getätigten Interventionen.....	190
Abbildung 62: Übersicht über die normierten absoluten Häufigkeiten der Reflexionsmomente der vier Gruppen von Studierenden auf Grundlage der Einteilung nach der Anzahl sowie der Angemessenheit der getätigten Interventionen.....	190
Abbildung 63: Übersicht über die relative Häufigkeit der Reflexionsmomente der vier Gruppen hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen" eingeteilt nach Anzahl und Angemessenheit der getätigten Interventionen.....	192
Abbildung 64: Übersicht über die absolute Häufigkeit der Reflexionsmomente der drei Gruppen von Studierenden auf Grundlage der Einteilung nach der Art der getätigten Interventionen	196
Abbildung 65: Übersicht über die relative Häufigkeit der Reflexionsmomente der drei Gruppen hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen" eingeteilt nach Anzahl und Angemessenheit der getätigten Interventionen.....	197
Abbildung 66: Übersicht über die absolute Häufigkeit der Reflexionsmomente der sechs Gruppen von Studierenden auf Grundlage der Einteilung nach der Art sowie der Angemessenheit der getätigten Interventionen.....	201
Abbildung 67: Übersicht über die relative Häufigkeit der Reflexionsmomente der sechs Gruppen hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen" eingeteilt nach Art und Angemessenheit der getätigten Interventionen	201
Abbildung 68: Übersicht über die drei zu unterscheidenden Ebenen hinsichtlich der Forschungsvorhaben im Mathematik-Labor "Mathe ist mehr" ergänzt durch auf den Forschungsergebnissen aufbauende Optimierungen	213

1 Einleitung

Was macht einen „guten“ Lehrer bzw. eine „gute“ Lehrerin aus? Über welche Kompetenzen sollte er bzw. sie verfügen und an welche Kompetenzen können Fortbildungen anknüpfen? (Reinold, 2016, S. 37ff.)

Diese zentrale Frage der Forschung zum Lehrerberuf erscheint stets aktuell zu sein, da sich die Anforderungen an Lehrkräfte im Laufe der Zeit wandeln und damit auch die Kompetenzen, über die Lehrkräfte verfügen sollten. In Zeiten der COVID-19-Pandemie, in der das Schulkonzept neu gedacht und unter anderem die Digitalisierung des Unterrichts mit Hochdruck vorangetrieben wird, kann man sich die Frage stellen, wie die Beschulung von Kindern und Jugendlichen in der Zeit danach aussehen soll. Aktuell befinden sich Lehrerinnen und Lehrer sowie ihre Schülerinnen und Schüler nicht im Klassenzimmer. Entsprechend ist es den Lehrenden auch nicht möglich, die Arbeitsprozesse der Lernenden zu beobachten und diese womöglich sogar zu unterbrechen, um nützliche Hinweise oder Feedback zum Lernprozess zu geben. Die Frage, welche Fähigkeiten und Fertigkeiten von Lehrkräften relevant sind, um Lernprozesse bei den Schülerinnen und Schülern anzustoßen und diese bestmöglich zu begleiten, rückt dabei immer mehr ins Zentrum.

In der Literatur werden diagnostische Fähigkeiten als wesentlicher Bestandteil der professionellen Kompetenz von Lehrkräften angesehen. Dies wird auch damit begründet, dass adaptive Lehrerinterventionen auf vorausgegangenen Diagnosen basieren (Leiss 2007, 66ff.). Um den Studierenden den Aufbau dieser zentralen Fähigkeit zu ermöglichen, haben Lehramtsstudierende im Fach Mathematik am Campus Landau der Universität Koblenz-Landau die Gelegenheit, im Rahmen eines Lehr-Lern-Labor-Seminars Bezüge zwischen theoretisch vermitteltem Wissen und praktischen Anforderungssituationen herzustellen. In der Veranstaltung entwickeln die Studierenden theoriegeleitet Lernumgebungen für das Mathematik-Labor „Mathe ist mehr“, begleiten Schülerinnen und Schüler bei deren Bearbeitung dieser Lernumgebungen und reflektieren anschließend darüber. Dabei werden insbesondere (1) die prozessdiagnostischen Fähigkeiten der Studierenden, (2) deren Fähigkeit, in Gruppenarbeitsprozesse von Schülerinnen und Schülern adäquat zu intervenieren sowie (3) die Reflexionsfähigkeit der Studierenden mit Blick auf die Entwicklung der Lernumgebungen und das eigene Lehrerhandeln gefordert und gefördert. Damit leistet die Veranstaltung insofern einen Beitrag zur Lehrerprofessionalisierung, als sie Diagnose-, Interventions- und Reflexionsfähigkeiten angehender Mathematiklehrkräfte schult.

Abbildung 1 visualisiert die Ausgangslage unter Berücksichtigung der drei involvierten Akteursgruppen. Die Dozierenden des Seminars gestalten Lernprozesse für die Studierenden, damit diese in die Lage versetzt werden, Lernumgebungen für Schülerinnen und Schüler zu konzipieren. Die Studierenden setzen dies um und erproben ihre Lernumgebung im Mathematik-Labor „Mathe ist mehr“ am Campus Landau mit einer Schulklasse. Dabei beobachten sie die Schülerinnen und Schüler, diagnostizieren deren Lernprozess und reflektieren über die Gestaltung der Lernumgebung. Diese Informationen können anschließend genutzt werden, um das eigene Lehrerhandeln sowie die Gestaltung der Lernumgebung künftig zu adaptieren. Gleiches gilt für die Dozierenden. Diese begleiten die Studierenden bei ihrem Lernprozess, analysieren und evaluieren diesen und können dadurch Rückschlüsse zur Optimierung des Seminarkonzeptes extrahieren.

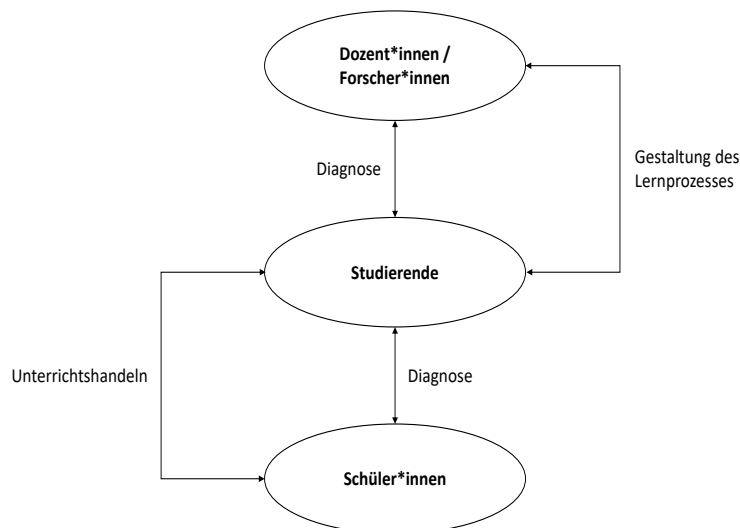


Abbildung 1: Übersicht über die drei zu unterscheidenden Ebenen hinsichtlich der Forschungsvorhaben im Mathematik-Labor "Mathe ist mehr"

Um einerseits den Theorie-Praxis-Bezug herzustellen und dadurch die Ansammlung trägen Wissens zu vermeiden, andererseits die Studierenden auf die Anforderungen im Unterricht möglichst gut vorzubereiten, müsste die erste Phase der Lehramtsausbildung deutlich praxisorientierter gestaltet sein, als sie es aktuell in Rheinland-Pfalz (und vielen anderen Bundesländern) ist. Das ist auch der Tatsache geschuldet, dass es ein logistisches Problem ist, tausende Studierende regelmäßig Erfahrungen im Klassenzimmer sammeln zu lassen. Entsprechend sollte es den Studierenden im Rahmen ihres Studiums an der Hochschule ermöglicht werden, sich über das rein fachliche Wissen hinaus Fähigkeiten anzueignen, die sie später im Klassenzimmer benötigen, um die Schülerinnen und Schüler bestmöglich in ihrem Kompetenzaufbau zu unterstützen.

An diesem Punkt stellt sich die eingangs aufgeworfene Frage nach den relevanten Fähigkeiten und Fertigkeiten, die Studierende in ihrem späteren Berufsleben benötigen, und daran anschließend danach, inwieweit diese möglicherweise zusammenhängen. Welche Kompetenzen sollten bereits im Studium geschult werden? Und falls eine Kompetenz als notwendig angesehen wird, inwieweit kann diese angebahnt werden? Wie viel Zeit sollte dafür zur Verfügung stehen und von welchen anderen Veranstaltungen bzw. Themen, wird diese Zeit, falls nötig, abgezogen? Die Entscheidung für oder gegen die Lehre bestimmter Themen bzw. die Anbahnung bestimmter Fähigkeiten sollte also sorgfältig durchdacht und erforscht werden, damit eine fundierte Entscheidung getroffen werden kann. Diese Arbeit leistet einen kleinen Teilbeitrag dazu.

Kapitel 2 diskutiert diagnostische Fähigkeiten von Lehrkräften. Dabei werden diese zunächst als Teil der diagnostischen Kompetenz dargestellt und darin verortet. Anschließend werden die pädagogische Diagnostik sowie Zeitpunkt, Art und Funktion der Diagnose in den Blick genommen. Außerdem werden verschiedene Komponenten der Diagnose erörtert. An die Vorstellung bisheriger Forschungsergebnisse schließt sich eine Konkretisierung des in der vorliegenden Studie genutzten Diagnosebegriffs.

Kapitel 3 beschäftigt sich mit Interventionen in Gruppenarbeitsprozessen. Dabei werden zunächst verschiedene Begriffe, die den Prozess der Intervention beschreiben, vorgestellt und gegeneinander abgegrenzt. Es werden Schnittmengen zwischen Lehrerimpulsen, Scaffolding und Lehrerinterventionen benannt sowie Unterschiede herausgearbeitet. Anschließend werden verschiedene Modelle zur Erfassung von Interventionen dargelegt. Das Kapitel schließt mit einem zusammenfassenden Überblick.

Kapitel 4 beschäftigt sich mit dem Begriff der Reflexion. Dabei wird zunächst eine Begriffsklärung vorgenommen. Anschließend werden wie in den beiden vorherigen Kapiteln verschiedene Modelle vorgestellt und diskutiert. Diese Modelle, anhand derer die Reflexionstätigkeit geordnet und analysiert werden kann, dienen teilweise auch als Grundlage für die empirischen Ergebnisse, die erläutert werden.

Nach der Behandlung der genannten theoretischen Konstrukte, wird in Kapitel 5 die fachdidaktische Ausbildung im Fach Mathematik an der Universität in Landau vorgestellt. Dazu wird zunächst das Gesamtkonzept dargelegt, wobei aufgezeigt wird, wie die mathematikdidaktischen Ausbildungsinhalte miteinander vernetzt sind. Darüber hinaus werden das Videodiagnosetool ViviAn sowie das Mathematik-Labor „Mathe ist mehr“ vorgestellt.

Kapitel 6 diskutiert die beiden Grundvorstellungen „Term als Rechenschema“ und „Term als Bauplan“. Dazu wird zuerst der Variablenbegriff genauer beleuchtet. Variablen können sowohl statisch als auch dynamisch interpretiert werden. Das Teilkapitel mündet in einer Übersicht und der Vorstellung einer Ordnung für die Sicht auf Variablen. Diese werden schließlich benötigt, um die beiden Grundvorstellungen im Bereich Terme besser nachvollziehen zu können. Das Kapitel endet mit prägnanten Verschlankungen und Beispielen für die Grundvorstellungen.

Aufbauend auf die theoretischen Grundlagen der Arbeit, wird in Kapitel 7 die übergeordnete Forschungsfrage aufgestellt. Diese wird anschließend in mehrere Teilfragen gegliedert.

Kapitel 8 erläutert das methodische Vorgehen der Studie. Dabei wird zunächst auf die qualitative Inhaltsanalyse eingegangen, da diese in allen drei Bereichen (Diagnose, Intervention und Reflexion) als Auswertungsverfahren genutzt wurde. Dabei werden zentrale Begriffe und Techniken vorgestellt. Danach wird genauer auf die Auswertung im Bereich Diagnose geblickt. Hierbei wird die Auswahl der Videosequenzen dargelegt, die darin enthaltenen fachdidaktischen Aspekte erläutert und die Diagnoseaufträge vorgestellt. Diese wurden von Expertinnen und Experten beantwortet und anschließend in das Diagnosetool ViviAn eingebettet. Es wird gezeigt, inwieweit die prozessdiagnostische Fähigkeit „Deuten“ durch in ViviAn eingebundene Videovignetten ermittelt werden kann. Das Teilkapitel endet mit einem Exkurs, in dem die quantitative Validierung des Testinstrumentes vorgenommen wird.

Das nächste Teilkapitel, das sich der Auswertung der Interventionen widmet, beschreibt, wie die Interventionen in das Didaktische Seminar eingebettet sind und wie die Daten generiert wurden. Anschließend wird dargelegt, wie die Hilfen analysiert und wie die Bewertung der Angemessenheit der Interventionen vorgenommen wurden. Außerdem werden weitere Analysen, wie etwa die Analyse der Interventionsauslöser oder die Anzahl der aktiven Schülerinnen und Schüler, beschrieben.

Hinsichtlich der Reflexionsgespräche wird ebenfalls dargelegt, wie diese in das Didaktische Seminar eingebettet wurden und wie die Daten generiert wurden. Außerdem wird auf die Analyse der Reflexionsmomente eingegangen, wobei diese zunächst in drei Hauptkategorien und anschließend noch in weitere Unterkategorien kodiert wurden. Das Kapitel endet mit der Beschreibung der jeweils zugrundeliegenden Populationen. Hierbei werden die Expertinnen und Experten, die Kodiererinnen und Kodierer und die Zusammensetzung der Studierenden sowie

der Lehrkräfte diskutiert. Außerdem wird auf die Verschränkung zu anderen Studien eingegangen.

Kapitel 9 stellt die Ergebnisse der Studie vor. Dabei behandeln die Teilkapitel jeweils einen Teil der Forschungsfrage. Zunächst wird darauf geblickt, inwieweit die Studierenden über die prozessdiagnostische Fähigkeit „Deuten“ verfügen. Dabei wird auch auf die Beantwortung der Diagnoseaufträge durch Lehrkräfte eingegangen und deren Antworten mit denen der Studierenden und den zuvor befragten Expertinnen und Experten verglichen. Auch anhand von Beispielen wird verdeutlicht, inwieweit ein Unterschied zwischen den Studierenden und den Lehrkräften besteht.

In Hinsicht auf die Interventionen in Gruppenarbeitsprozesse werden Ergebnisse zu den Hilfen, die innerhalb der Interventionen genutzt wurden, dargelegt. Dabei wird zwischen dominanten und untergeordneten Hilfen unterschieden und jeweils auch die Zeitdauer berücksichtigt. Anschließend werden die Ergebnisse mit Blick auf die Reihenfolge der Hilfe sowie die Angemessenheit der Hilfen präsentiert. Darüber hinaus werden auch die Anzahl der aktiven Schülerinnen und Schüler sowie die jeweiligen Interventionsauslöser diskutiert.

In Bezug auf die Reflexionsgespräche wird zunächst ein Überblick über die Reflexionsmomente gegeben. Anschließend werden die drei Hauptkategorien genauer beleuchtet. Neben den auftretenden Häufigkeiten werden einige Aussagen der Studierenden genannt und diskutiert. Nach der Analyse der drei Bereiche werden anschließend mögliche Zusammenhänge zwischen den Bereichen vorgestellt.

Als erstes werden dabei Zusammenhänge zwischen Diagnose und Intervention erläutert. Dazu werden die Studierenden in drei Gruppen eingeteilt. Die verwendeten Hilfen innerhalb der Interventionen werden für die drei Gruppen separat betrachtet und miteinander verglichen. Ebenso werden die Reihenfolge der Hilfen, die Angemessenheit der Hilfen, die Anzahl der aktiven Schülerinnen und Schüler und die jeweiligen Interventionsauslöser in den Blick genommen.

Danach werden mögliche Zusammenhänge zwischen der prozessdiagnostischen Fähigkeit „Deuten“ und dem Reflexionsverhalten der Studierenden beleuchtet. Hierzu wird dieselbe Einteilung der Studierenden vorgenommen wie zuvor. Neben der Betrachtung der Hauptkategorien werden auch die Häufigkeiten der Reflexionsmomente in den Unterkategorien miteinander verglichen.

Als letztes werden die Reflexionsgespräche der Studierenden untersucht, indem die Studierenden anhand der getätigten Interventionen neu geordnet werden. Dabei werden verschiedene Einteilungen vorgenommen und jeweils untersucht, inwieweit die Gruppen sich hinsichtlich der Reflexionsmomente unterscheiden.

In Kapitel 10 werden die zuvor genannten Ergebnisse diskutiert. Zunächst werden diese mit bisherigen Forschungsergebnissen verglichen. Dabei können einige Ergebnisse anderer Forscher bestätigt werden, andere wiederum passen weniger zu den vorliegenden Ergebnissen. Abschließend werden Limitationen genannt und bestehende Forschungsdesiderate aufgezeigt.

2 Diagnostische Kompetenz von Lehrkräften

In der gegenwärtigen Diskussion über Lehrerbildung und -fortbildung erfahren nur wenige Aspekte eine solche Aufmerksamkeit wie die diagnostische Kompetenz (Brunner, Anders, Hachfeld & Krauss 2011, 215). Dies liegt vermutlich einerseits daran, dass es sich um ein Konstrukt handelt, bei dem man fest davon ausgeht, dass es zum Lernerfolg der Schülerinnen und Schüler beiträgt (Leuders, Dörfler, Leuders & Philipp 2018). Andererseits gibt es noch einen hohen Forschungsbedarf, um den Zusammenhang der Diagnosekompetenz der Lehrkräfte mit dem Lernerfolg der Schülerinnen und Schüler unter Berücksichtigung verschiedener Einflussfaktoren nachvollziehen zu können und empirisch abzusichern (Leuders et al. 2018).

Es herrscht weitestgehend Konsens darüber, dass die individuellen Voraussetzungen der Lehrkraft, wie etwa das Fachwissen und das fachdidaktische Wissen, aber auch Dispositionen wie Motivation, Überzeugungen und Selbstregulation einen Einfluss auf deren diagnostische Kompetenz haben (Helmke, Hosenfeld & Schrader 2004; Prediger, Tschierschky, Wessel & Seipp 2012; Heinrichs 2015; Beretz, Lengnink & von Aufschnaiter 2017; Leuders et al. 2018). In vielen Studien und Forschungsansätzen wird die Diagnosekompetenz jedoch im Zuge der Handhabbarkeit und Operationalisierung auf eine oder mehrere (diagnostische) Fähigkeiten reduziert (Dübbelde 2013, 10). Die unterschiedlichen Operationalisierungen, die vorgenommen werden, führen auch zu unterschiedlichen Modellen hinsichtlich der diagnostischen Kompetenz (von Aufschnaiter et al. 2015; Dübbelde 2013, 10). Darüber hinaus erfassen einige Autoren diagnostische Fähigkeiten, ohne diese explizit so zu benennen (Lazarevic 2015; Seidel, Blomberg & Stürmer 2010; Treisch 2018). Dadurch ergibt sich insgesamt eine breite und vielschichtige Befundlage.

Um die im Rahmen dieser Arbeit durchgeführte Studie besser nachvollziehen und einordnen zu können, soll das Konstrukt der Diagnosekompetenz zunächst anhand verschiedener Modelle erläutert werden. Dazu werden unterschiedliche Schwerpunktsetzungen und Facetten der Modelle aufgegriffen, um einen Gesamtüberblick über die diagnostische Kompetenz zu geben. Am Ende des Kapitels werden diese nochmals zusammengefasst und die im Rahmen dieser Arbeit durchgeführte Studie verortet.

2.1 Diagnostische Fähigkeiten als Teil der diagnostischen Kompetenz

In vielen Projekten wird die Diagnosekompetenz als Teilaspekt der professionellen Kompetenz von Lehrkräften angesehen und verortet (Dübbelde 2013; Heinrichs 2015; Hoth 2016; Karing, Pfof & Artelt 2011; Sjuts 2007). Dabei versteht man unter der professionellen Kompetenz von Lehrkräften die Kompetenz, die eine Lehrkraft für die erfolgreiche Durchführung aller im Rahmen des Unterrichts geforderten Aufgabenbereiche benötigt (Heinrichs 2015, 23). Leuders et al. (2018) adaptieren das gängige Modell professioneller Kompetenz von Blömeke, Gustafsson und Shavelson (2015), welches Kompetenz als Kontinuum definiert, das Dispositionen (u.a. Wissen und Überzeugungen), situationsspezifische Fähigkeiten (Wahrnehmung, Interpretation und Entscheidungsfindung) sowie Performanz (das zugehörige beobachtbare Verhalten) umfasst, hinsichtlich der Diagnosekompetenz (Abb. 2).

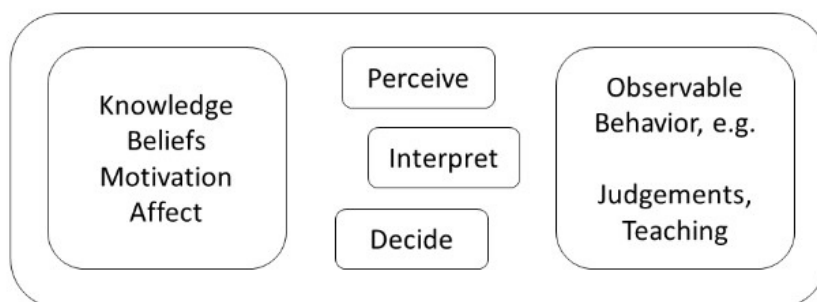


Abbildung 2: Diagnostische Kompetenz (in einem weiteren Verständnis) als Kontinuum nach Leuders et al. (2018)

Die Darstellung bildet links die diagnostischen Dispositionen ab, in der Mitte die diagnostischen Fähigkeiten und rechts die diagnostische Performanz. Die diagnostischen Dispositionen umfassen in dieser Darstellung Wissen und Überzeugungen sowie motivationale und affektive Faktoren, die in einer Person relativ stabil sind. Leuders et al. (2018) verweisen darauf, dass diese auch zur Fähigkeit beitragen, in diagnostischen Situationen erfolgreich zu agieren. Diagnostische Fähigkeiten können als eine Reihe von situationsspezifischen kognitiven Funktionen oder als Prozesse der Wahrnehmung, Interpretation und Entscheidungsfindung betrachtet werden (ebd.). Einerseits stützen sich diese Prozesse auf diagnostische Dispositionen. Andererseits führen sie zu einem beobachtbaren Verhalten in diagnostischen Situationen. Diagnostische Performanz bezieht sich auf das beobachtbare Verhalten in diagnostischen Situationen, wie sie im Berufsleben einer Lehrkraft auftreten. Sie kann als das Produkt der diagnostischen Dispositionen und diagnostischen Fähigkeiten sowie deren Operationalisierungen betrachtet werden (ebd.).

Leuders et al. (2018) weisen darauf hin, dass viele Autoren mit dem Begriff „Kompetenz“ nur den Bereich der Dispositionen meinen und diesen teilweise sogar auf die kognitive Dimension beschränken. Die Performanz wird dabei als die Manifestation dieses latenten Konstrukts angesehen (ebd.). Andere Projekte greifen sich die diagnostischen Fähigkeiten heraus und erforschen diese, vernachlässigen dabei die zugrundeliegenden Dispositionen. An dieser Stelle soll daher eine klare sprachliche Trennung zwischen diagnostischen Fähigkeiten und der Diagnosekompetenz dadurch vorgenommen werden, dass die diagnostischen Fähigkeiten als Teilaspekt der Diagnosekompetenz angesehen werden. Diagnosekompetenz wird mit Leuders et al. (2018) als aus diagnostischen Dispositionen, diagnostischen Fähigkeiten und Performanz bestehend verstanden.

Um nachvollziehen zu können, wie sich die diagnostische Kompetenz in das Konzept der professionellen Kompetenz einbetten lässt, wovon im obigen Modell ausgegangen wird, ohne dies näher zu erläutern, wird nun die Aufteilung des professionellen Wissens von Lehrkräften nach Shulman (1986; 1987) vorgestellt. Shulman (1986; 1987) unterscheidet dabei zwischen *Content Knowledge* (Fachwissen), *Pedagogical Content Knowledge* (fachdidaktisches Wissen) und *Pedagogical Knowledge* (Pädagogisches Wissen). Das Fachwissen wird an der Hochschule erworben und stellt eine Grundvoraussetzung für erfolgreichen Unterricht dar. Das fachdidaktische Wissen der Lehrperson zielt auf die Verbindung zwischen curricularen Fachinhalten und unterrichtsmethodischem Know-how (Shulman 1986). Pädagogisches Wissen umfasst Wissen über allgemeine Prinzipien der Klassenorganisation und des Klassenmanagements (Shulman 1986).

Bromme (1992, 96-98; 1997) erweitert diese Dreiteilung des professionellen Wissens, indem er die beiden Dimensionen „Curriculares Wissen“ und „Philosophie des Schulfaches“ ergänzt. Dadurch ergibt sich aufbauend auf den Überlegungen von Shulman (1986; 1987) folgende Topologie des professionellen Wissens:

- *Fachliches Wissen* (bezieht sich auf Methoden und Erkenntnisse des Faches)
- *Curriculares Wissen* (beinhaltet Vorstellungen über Unterrichtsziele sowie die Auswahl von Lerninhalten)
- *Philosophie des Schulfaches* (zeigt die Relevanz der Fachinhalte und die Beziehung zu anderen Bereichen auf)
- *Pädagogisches Wissen* (betrifft disziplinarische Fertigkeiten, Classroom Management und Bedingungsfaktoren des schulischen Lernerfolgs)
- *Fachspezifisch-pädagogisches Wissen* (bezieht sich auf die angemessene Darstellung und Vermittlung der Inhalte)

Die diagnostische Kompetenz ist sowohl bei Shulman als auch bei Bromme als Subdimension anzusehen (Schwarz, Wissmach & Kaiser 2008). Unklar ist hingegen, inwieweit die diagnostische Kompetenz auf die vorgestellten Dimensionen zurückgreift (Heinrichs 2015, 38). Da sowohl Fachwissen als auch fachdidaktisches Wissen und pädagogisches Wissen beeinflussen können, ob Lernschwierigkeiten hinsichtlich des Materials oder der Aufgabenstellung vorhergesehen werden oder notwendige Voraussetzungen zur Lösung einer Aufgabe erkannt werden können, sieht Bromme (1997) die diagnostische Kompetenz als gutes Beispiel dafür an, wie die unterschiedlichen Bereiche des professionellen Wissens zusammenwirken.

Die Aufteilung des professionellen Wissens nach Shulman (1986; 1987) und die Erweiterung nach Bromme (1992; 1997) sind vielfach zitiert und wurden in verschiedenen Modellen aufgegriffen. Unter anderem baut das Modell professioneller Handlungskompetenz nach Baumert und Kunter (2006), das unter anderem im Rahmen der COACTIV-Studie genutzt wurde (Anders, Kunter, Brunner, Krauss & Baumert 2010; Baumert et al. 2009), auf den Überlegungen der beiden Autoren auf. Baumerts und Kunters (2006) Modell beinhaltet neben Überzeugungen bzw. Werthaltungen, motivationalen Orientierungen und selbstregulativen Fähigkeiten insbesondere auch das Professionswissen (vgl. Abb. 3).

Dieses gliedern Baumert und Kunter (2006) in Kompetenzbereiche auf, welche wiederum in einzelne Kompetenzfacetten aufgeteilt werden können. Die Kompetenzbereiche des Professionswissens umfassen Fachwissen, fachdidaktisches Wissen, pädagogisch-psychologisches Wissen, Organisationswissen und Beratungswissen. Die diagnostischen Fähigkeiten sind einerseits beim fachdidaktischen Wissen und andererseits beim pädagogisch-psychologischen Wissen verortet. Hierbei wird ausdrücklich nur von Fähigkeiten und nicht von Kompetenz gesprochen (Brunner et al. 2011, 215ff.).

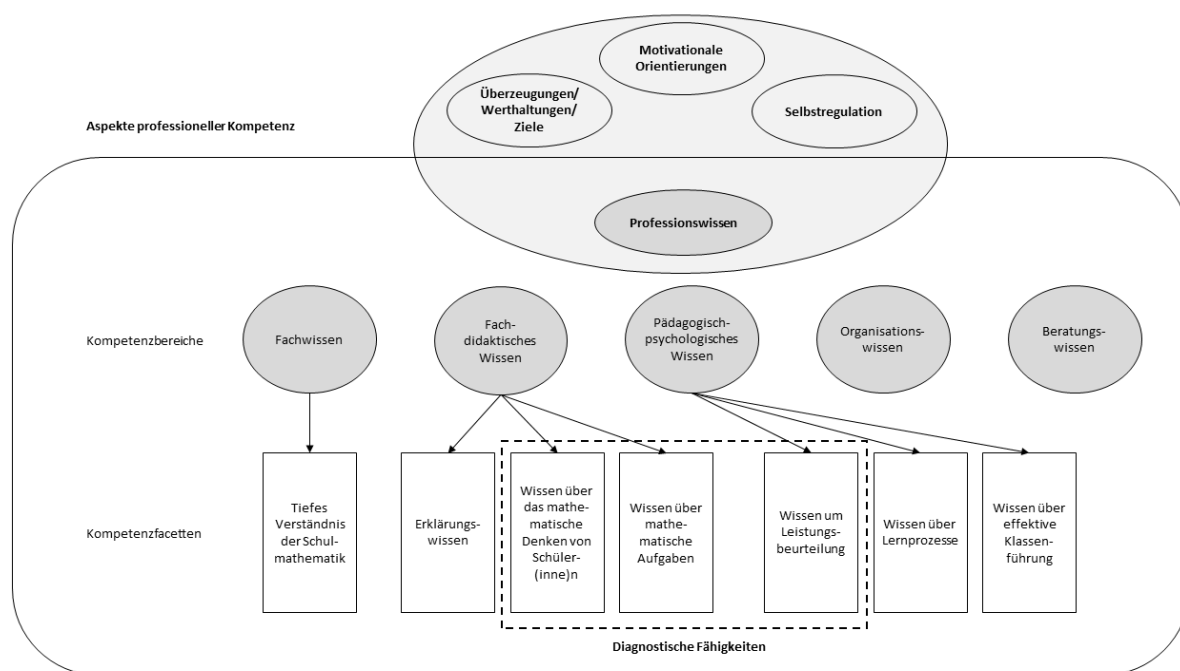


Abbildung 3: Einbettung diagnostischer Fähigkeiten im COACTIV-Modell zur professionellen Kompetenz von Lehrkräften nach Brunner et al. (2011, 217)

Kunter, Kleickmann, Klusmann und Richter (2011, 55) verstehen professionelle Kompetenz als Bündel unterschiedlicher berufsbezogener Voraussetzungen, nämlich des Wissens, der Überzeugungen sowie motivationaler und selbstregulativer Merkmale, die im Wechselspiel miteinander stehen und bestimmen, wie gut eine Lehrkraft die Anforderungen ihres Berufs bewältigen kann. Sie nehmen an, dass professionelle Kompetenz prinzipiell erlern- und veränderbar ist (ebd.). Die professionelle Kompetenz entwickelt sich in der Lehrpersonenbildung (Studium, Referendariat, Fort- und Weiterbildung) und kann durch das Handeln in konkreten beruflichen Situationen mit entsprechender Reflexion angebahnt werden (Kunter et al. 2011, 60).

Der Aspekt der Erlernbarkeit findet sich auch in der Definition von Weinert (2001) wieder, der die diagnostische Kompetenz folgendermaßen definiert:

Dabei versteht man unter Kompetenzen die bei Individuen 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 nutzen zu können (Weinert 2001, 27f.).

In seiner Definition fokussiert Weinert (2001) besonders stark auf Fähigkeiten und Fertigkeiten. Zudem stellt er heraus, dass die zur Problemlösung benötigten Fähigkeiten und Fertigkeiten erlernbar und nicht von Natur aus gegeben sind. Außerdem verweist er auf die motivationale, volitionale und soziale Bereitschaft, die notwendig ist, um eine entsprechende Leistung zu erbringen. Weinert (2001) betont darüber hinaus, dass (fehlende) diagnostische Kompetenz der Lehrperson nachhaltige Auswirkungen auf die notwendige Individualisierung und Differenzierung des Unterrichts hat (ebd.).

Bevor nun auf den Zeitpunkt, die verschiedenen Arten und Aufgaben sowie unterschiedliche Komponenten der Diagnose eingegangen wird, wird die pädagogische Diagnostik vorgestellt. Die pädagogische Diagnostik wurde begrifflich im Rahmen einer neuen Schwerpunktsetzung

weg von der Beurteilung hin zur individuellen Diagnostik und Förderung, von der Diagnosekompetenz abgelöst (Dübbelde 2013, 16). Um die Anfänge der pädagogischen Diagnostik nachvollziehen zu können, wird dabei zunächst ein kurzer Ausflug in die medizinische Diagnostik vorgenommen.

2.2 Pädagogische Diagnostik

Im alltäglichen Verständnis wird der Begriff „Diagnostik“ überwiegend im medizinischen Bereich verortet (Rath 2017, 18). Dabei werden in der Medizin Informationen gesammelt, um auf dieser Grundlage eine geeignete medizinische Diagnose stellen zu können (ebd.). Anschließend wird den Symptomen bzw. Befunden ein Krankheitsbegriff zugeordnet (ebd.). Eine angemessene Diagnose ist hier eine Diagnose, die die Symptome sowie die begleitenden Aspekte in einem Krankheitsbild akkurat erfasst und auf dieser Grundlage eine angemessene Therapie ermöglicht (ebd.). Dieses Verständnis von Diagnose kann auf den Bereich der pädagogischen Diagnostik übertragen werden. Auch die Diagnose im schulischen bzw. pädagogischen Kontext beginnt mit dem Sammeln von Informationen, um schließlich auf dieser Grundlage eine pädagogische Entscheidung zu treffen (ebd.). Jedoch gibt es durchaus Unterschiede hinsichtlich der Informationsbeschaffung in Medizin und Schule.

Der wichtige Unterschied zwischen den beiden diagnostischen Kontexten (Medizin und Schule) besteht in der konkreten Informationssammlung und den dem Urteil folgenden Therapieempfehlungen bzw. pädagogischen Entscheidungen. In der Medizin erleichtern kategoriale Zuordnungsmöglichkeiten in Form von diagnostischen Leitfäden und Manualen das Diagnostizieren sowie die Ableitung der diagnosebasierten Empfehlungen. Solche systematischen, empirisch abgesicherten Zuordnungen als Orientierungsrahmen fehlen in der pädagogischen Diagnostik. Infolgedessen ist der Diagnoseprozess an sich erschwert, da er sich auch stärker als im medizinischen Kontext als ein Prozess unter Unsicherheitsbedingungen und zum Teil sehr komplexen Diagnosesituationen darstellt. (Rath 2017, 18)

Der Begriff „pädagogische Diagnostik“ wurde 1968 von Ingenkamp in Anlehnung an die medizinische und die psychologische Diagnostik geprägt (Jürgens & Lissmann 2015, 58). Pädagogische Diagnostik kann als umfassender Begriff für den Zusammenhang von Schüler- und Unterrichtsbeurteilung angesehen werden (Kutscher 1979, 12). Dabei sind Beurteilungen unter anderem durch die Benotung und Zeugniserteilung im Schulsystem institutionalisiert (Kleber 1976, 11). Bei Beurteilungen kann zwischen interner und externer Beurteilung unterschieden werden, wobei es sich bei der internen Beurteilung um die allgemeine Personenwahrnehmung und bei externer Beurteilung um Notengebung, Zeugnis- und Gutachtenerstellung handelt (Kleber 1976, 37). Jürgens und Lissmann (2015, 58) sehen pädagogische Diagnostik als Methode an, bei der in pädagogischen Kontexten Tätigkeiten ausgeführt werden, um für Lernende begründete Entscheidungen zu treffen. Dabei berufen sie sich auf die Definition von Ingenkamp & Lissmann (2008, 13):

Pädagogische Diagnostik umfasst alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgehalten werden, um individuelles Lernen zu optimieren. Zur pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen sowie die mehr gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen zum Ziel haben.

Somit umfasst die pädagogische Diagnostik ein sehr großes Spektrum an diagnostischen Tätigkeiten. Neben den externen Beurteilungen werden explizit auch die Analyse der Lehr-Lernprozesse sowie die individuelle Förderung einzelner Schülerinnen und Schüler genannt. Kleber (1976, 26) verweist bereits vor über 40 Jahren darauf, dass die Förderung jeder Schülerin und jedes Schülers gemäß deren/dessen Lernvoraussetzungen durch adäquate Modifikation des Unterrichts eine der zentralen Aufgaben von Lehrkräften sei. Diese Sichtweise erfährt heutzutage großen Zuspruch, was an einer Vielzahl unterschiedlicher (Forschungs-)Projekte im Bereich Diagnose festgemacht werden kann (vgl. dazu u.a. Bartel & Roth 2020; Beretz et al. 2017; Brunner et al. 2011; Dübbelde 2013; Enenkiel & Roth 2017; Heinrichs 2015; Hofmann & Roth 2017; Hoth 2016; Karing et al. 2011; Sjuts 2007; Walz & Roth 2017; Weinsheimer 2016; Wildgans, Obersteiner, Fischer & Reiss 2018).

Tendenziell erfolgt hier eine Auseinandersetzung nicht mehr mit pädagogischer Diagnostik, sondern mit diagnostischer Kompetenz. Einige Autoren nehmen keine trennscharfe Unterscheidung der beiden Begriffe vor (vgl. dazu u.a. Büchter und Leuders 2005). Bei anderen Autoren könnte man die diagnostische Kompetenz als notwendige Voraussetzung für pädagogische Diagnostik verstehen (vgl. dazu u.a. Hascher 2011). In den meisten Arbeiten werden die Begriffe nicht voneinander abgegrenzt oder in Bezug zueinander gesetzt (vgl. dazu u.a. Heinrichs 2015). Daher ist es kaum verwunderlich, dass beide Begriffe sich je nach Autor sehr ähneln oder sich deutlich unterscheiden.

Nach Jürgens und Lissmann (2015, 10ff.) liegt der Schwerpunkt der pädagogischen Diagnostik in der Beurteilung von Schülerinnen und Schülern. Die Nutzung der diagnostischen Fähigkeiten zum Erkennen von spezifischen Voraussetzungen und Lernpotentialen, von Aufgabenschwierigkeiten, Lernhindernissen sowie Prozessen und Reaktionen der Lernenden schreiben Jürgens und Lissmann (2015, 11f) der Diagnosekompetenz zu. Mit dem Begriff der „Diagnosekompetenz“ geht also ein Paradigmenwechsel hin zur individuellen Diagnostik und Förderung einher (Dübbelde 2013, 16).

Die nächsten Teilkapitel beleuchten jeweils zentrale Bestandteile der Diagnose. Zunächst wird auf den Zeitpunkt der Diagnose sowie auf verschiedene Arten und Funktionen der Diagnose eingegangen.

2.3 Zeitpunkt, Art und Funktion der Diagnose

Der Zeitpunkt der Diagnose vor, während oder nach einer bestimmten Situation bzw. Handlung, bietet jeweils verschiedene (Fokussierungs-)Möglichkeiten. Vor einer Diagnose können Lehrkräfte sich mit den fachlichen und fachdidaktischen Inhalten, dem Arbeitsauftrag für die Schülerinnen und Schüler und damit verbunden auch mit potenziellen Lösungsstrategien auseinandersetzen. Es kann überlegt werden, wie diagnostiziert werden soll und was überhaupt den Gegenstand der Diagnose darstellen soll. Ebenso sollte auf ein Ziel der Diagnose hingearbeitet werden (Horstkemper 2004). Die Diagnose kann unter anderem genutzt werden, um geeignete Fördermaßnahmen festzulegen, den jeweiligen Schülerinnen und Schülern eine Rückmeldung zu geben oder die Aufgabenschwierigkeit an die Voraussetzungen der Schülerinnen und Schüler anzupassen. Dabei sollten auch mögliche Gütekriterien der jeweiligen Urteilsbildung reflektiert werden (Klug, Bruder, Kelava, Spiel & Schmitz 2013). Während der Diagnose selbst ist diese vorgelagerte Auseinandersetzung hilfreich, um die entsprechenden Informationen durch ein systematisches Vorgehen zu sammeln.

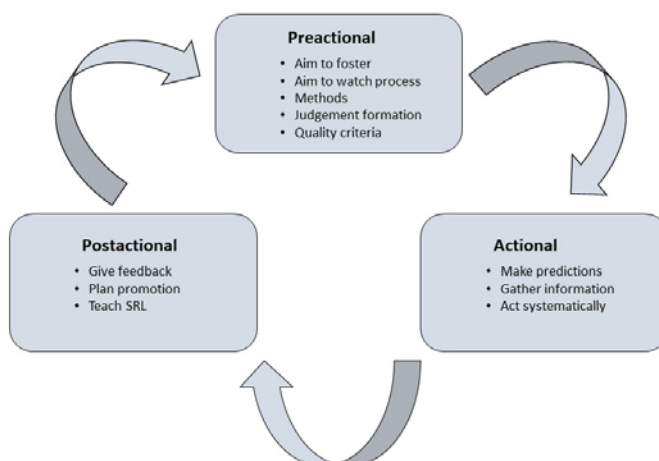


Abbildung 4: Modell der diagnostischen Kompetenz unter Berücksichtigung des Zeitpunkts nach Klug et al. 2013, 39

Klug et al. (2013) zerlegen die Diagnose in drei Dimensionen: präaktional, aktional und postaktional (vgl. Abb. 4). In der präaktionalen Phase ist jede Maßnahme vor der Zusammenfassung der Informationen für die tatsächliche Diagnose von Bedeutung (ebd.). Darüber hinaus werden die diagnostischen Fähigkeiten von Lehrerinnen und Lehrern in dieser Phase aktiviert (ebd.). In der zweiten Phase findet die eigentliche Diagnose statt (ebd.). In dieser Phase ist ein systematisches Vorgehen wichtig, um zuverlässige Ergebnisse zu erzielen (ebd.). Die Autoren sehen es als wichtig an, eine Vorhersage über die Entwicklung sowie zugrundeliegende Lernschwierigkeiten des Lernenden zu treffen (ebd.) Um zu einer Diagnose zu kommen, müssen die dafür relevanten Informationen gefiltert und geeignet interpretiert werden (ebd.). Die postaktionale Phase beginnt unmittelbar nach der eigentlichen Diagnose und beinhaltet die pädagogische Aktion, die sich aus der Diagnose ergibt. Die Autoren erachten dabei auch das Feedback an den Lernenden als wichtigen Faktor für den Lernerfolg (ebd.).

Der Zeitpunkt der Diagnose spielt auch bei der Einteilung verschiedener Diagnosearten eine Rolle. Je nach Diagnoseart wird beispielsweise an einem oder an zwei Zeitpunkten diagnostiziert oder während zwei Zeitpunkten auf den Lernverlauf des jeweiligen Schülers geblickt. Von Aufschnaiter et al. (2015) beschreiben mit der Statusdiagnostik, der Prozessdiagnostik, der Veränderungsdiagnostik und der Verlaufsdagnostik vier Arten der Diagnostik. Bei der *Statusdiagnostik* wird ein Merkmal, eine Fähigkeit oder eine Kompetenz zu einem bestimmten Zeitpunkt erfasst. Findet die Erfassung zu Beginn des Lernprozesses statt, können daraus Fördermaßnahmen resultieren (ebd.). Gegen Ende des Lernprozesses kann aus der Statusdiagnostik beispielsweise die Zuweisung zu weiteren Bildungsabschnitten abgeleitet werden (ebd.). Bei der *Prozessdiagnostik* wird ebenso ein vorliegendes Merkmal oder eine Fähigkeit erfasst. Hierbei wird jedoch weniger auf das Produkt geschaut, sondern vielmehr der Lösungsprozess unter die Lupe genommen (ebd.). Es wird also nachvollzogen, wie die Schülerinnen und Schüler zu ihrer Lösung gekommen sind. Die Statusdiagnostik ist in der Regel in der Prozessdiagnostik enthalten, da das Produkt dort oftmals miterhoben wird. Die Prozessdiagnose bietet sich für die Entwicklung von spezifisch auf die jeweilige Person abgestimmten Fördermaßnahmen an (ebd.). Bei der *Veränderungsdiagnostik* werden mindestens zwei Diagnosen zu unterschiedlichen Zeitpunkten miteinander verglichen (ebd.). Dadurch kann eine Aussage über Zuwachs,

Stagnation oder Abnahme einer Fähigkeit getroffen werden (ebd.). Hierbei wird also erfasst, ob eine Veränderung vorliegt. Eine Aussage darüber, wie eine erfasste Veränderung womöglich zustande kommt, wird schließlich in der *Verlaufsdiagnostik* getroffen. Diese fokussiert die Handlungs- und Denkprozesse zwischen den Messzeitpunkten. Sie umfasst einerseits unterrichtliche Lernangebote, kann andererseits darüber hinaus auch die Hausaufgabenbearbeitung oder Gespräche mit den Eltern oder Mitschülern beinhalten (ebd.).

Nach Hascher (2011) hat die Diagnostik zwei für den Schulalltag wesentliche Zielrichtungen: die Diagnostik der Leistung bzw. des Lernerfolgs und die Diagnostik des Lernprozesses. Bei der Bewertung des Lernerfolgs werden *nach Abschluss eines Lernprozesses* das Ausmaß und die Qualität des erworbenen Wissens und der erworbenen Fähigkeiten im Rahmen des vollzogenen Lernprozesses erhoben (ebd.). Bei der Diagnose des Lernprozesses wird *während des Lernprozesses* diagnostiziert bzw. kontinuierlich analysiert (ebd.). Mit den beiden Formen der pädagogischen Diagnostik werden unterschiedliche Ziele verfolgt. Mithilfe der Leistungsdiagnostik werden der augenblickliche Lernstand der Schülerinnen und Schüler erhoben und auf dieser Grundlage Qualifikationen zugewiesen, was wiederum der Selektion dient (ebd.). Mithilfe der Prozessdiagnostik werden hingegen der Lernverlauf erhellt und daraus zielgerichtete (fach-)didaktische und pädagogisch-psychologische Maßnahmen abgeleitet (ebd.).

Neben diesen Funktionen erfüllen Diagnosen weitere Funktionen. Wesentlich wird hierbei zwischen Selektions- und Modifikationsdiagnostik unterschieden (Heinrichs 2015, 11f). Selektionsentscheidungen dienen dazu, mit Hilfe des diagnostischen Urteils eine Zuordnung von Personen zu unterschiedlichen Maßnahmen oder Bedingungen zu treffen (Heinrichs 2015, 12). Modifikationsentscheidungen dienen der Planung, Unterstützung, Regulierung und Bewertung von Lernprozessen (ebd.). Beide Entscheidungsarten können zu einer Modifikation des Unterrichts führen (ebd.). Sowohl Selektions- als auch Modifikationsdiagnosen können auch trotz bester Absichten unzutreffend oder verzerrt sein.

Bei den diagnostischen Urteilen gibt es nämlich systematische Einflussfaktoren, die eine solche Verzerrung bewirken können. Helmke (2009, 138) nennt die Tendenz zur Mitte (Vermeidung extremer Urteile), die Tendenz zu extremen Urteilen (Vermeidung mittlerer Urteile), den Milde-Effekt (durchweg bessere Beurteilung als sachlich angemessen), den Referenzfehler (fehlerhafte Bezugsnorm), den Halo-Effekt (Wertung aufgrund unzureichender Hinweisreize) und den logischen Fehler (von der Ausprägung eines Schülermerkmals wird auf die Ausprägung eines anderen Schülermerkmals geschlossen) als systematische Einflussfaktoren. Hesse und Latzko (2011, 41ff.) führen darüber hinaus den Pygmalioneffekt (Beeinflussung durch Erwartungen der Lehrkräfte) auf.

Diesen systematischen Einflussfaktoren sollte man sich als Lehrkraft bewusst sein und die eigenen Diagnosen und Bewertungen kritisch reflektieren. Dabei kann es helfen, sich über die der Diagnose zugrundeliegende Bezugsnorm bewusst zu werden. Hierbei kann zwischen individueller, sozialer und objektiver Bezugsnorm unterschieden werden (Heinrichs 2015, 15). Die individuelle Bezugsnorm bezeichnet den Vergleich der Merkmalsinformationen einer Person mit Informationen zur gleichen Person zu einem anderen Zeitpunkt (ebd.). Wird der Vergleich der Merkmalsinformationen nicht mit der gleichen Person (zu einem anderen Zeitpunkt), sondern mit den Informationen über eine andere Person (zum selben Zeitpunkt) herangezogen, spricht man von einer sozialen Bezugsnorm (ebd.). Als objektive Bezugsnorm können Bezugsnormen angesehen werden, bei denen die diagnostischen Informationen mit einem Kriterium verglichen werden. Dabei findet der Vergleich also weder mit dem eigenen Lernzuwachs noch

mit dem Lernstand anderer Schülerinnen und Schüler statt, sondern vielmehr anhand eines (vorher festgelegten) Kriteriums. Jede einzelne Bezugsnorm hat Vor- und Nachteile, weshalb es sinnvoll erscheint, alle Bezugsnormen zu verwenden (Heinrichs 2015, 16).

2.4 Komponenten der Diagnose

Nachdem zuvor die möglichen Zeitpunkte der Diagnose, verschiedene Arten sowie verschiedene Funktionen der Diagnose beleuchtet wurden, sollen nun verschiedene Komponenten der Diagnose vorgestellt werden. Diese bestehen in der Regel unabhängig davon, ob eine einzelne Person, eine Gruppe oder eine ganze Klasse diagnostiziert wird. Auch die zugrundeliegende Bezugsnorm kann frei gewählt werden.

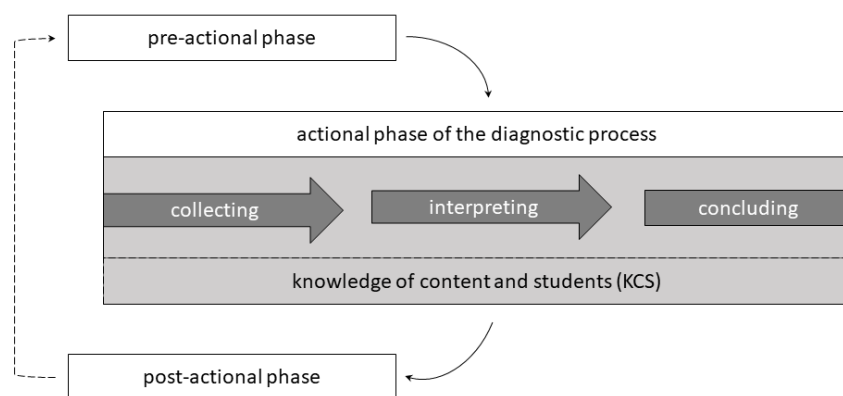


Abbildung 5: Diagnosemodell unter der Berücksichtigung des Zeitpunktes und Ausdifferenzierung der aktionalen Phase nach Reinhold (2018, 133)

Abbildung 5 zeigt das Diagnosemodell von Reinhold (2018), welches auf die Unterscheidung des Diagnosezeitpunktes von Klug et al. (2013) aufbaut und die aktionale Phase um drei Komponenten der Diagnose anreichert. Die drei Komponenten, die Reinhold (2018) in ihrer qualitativen Studie zur Analyse von Interviews nutzt, bestehen in der Erfassung der relevanten Informationen bzw. Daten, einer geeigneten Interpretation dieser sowie darauf aufbauend einer Schlussfolgerung in Form einer didaktischen Handlung. Reinhold (2018) verweist außerdem darauf, dass das Wissen über die (fachlichen) Inhalte sowie über den Kenntnisstand von Schülerinnen und Schülern notwendig ist, um eine Diagnose durchzuführen. Es ist aus ihrer Sicht offensichtlich, dass die verschiedenen Komponenten der Diagnose auf jeweils unterschiedlichen Arten von Wissen beruhen (ebd.). Sie nimmt dabei jedoch keine explizite Zuordnung vor.

Weit verbreitet ist auch der Ansatz von van Es und Sherin (2002), die zunächst zwischen „noticing“ und „interpreting“ unterscheiden. Dabei kann man „noticing“ in etwa mit „collecting (data)“ gleichsetzen. Das Noticing beinhaltet das Wahrnehmen relevanter Informationen während des Unterrichts, die Fähigkeit, Verbindungen zwischen spezifischen Ereignissen und dem größeren Ganzen herzustellen und das was man über den Kontext weiß, zu nutzen, um über eine Situation nachzudenken (ebd.). „Interpreting“ wird in beiden Modellen gleich benannt und umfasst auch inhaltlich dasselbe. Van Es und Sherin (2002) weisen zudem auf die Relevanz dieses Schrittes für Interaktionen im Klassenzimmer hin. In diesem Schritt nimmt man bewusst eine reflektierende Haltung hinsichtlich der Wirkung einzelner Handlungen und Geschehnisse im Unterricht ein. Die Interpretation der Situation soll dann als Grundlage für pädagogische Entscheidungen genutzt werden (ebd.).

Seidel et al. (2010) bauen auf diese Überlegungen auf und nutzen die Begriffe „Noticing“ und „Knowledge-based reasoning“. Unter „Noticing“ verstehen sie die Wahrnehmung von Unterrichtskomponenten. Es bezieht sich auf die wissensgesteuerte Identifikation von Situationen und Ereignissen im Unterricht, die einen entscheidenden Beitrag für den Erfolg von Unterrichtshandlungen darstellen (ebd.). Das Noticing beinhaltet folgende drei Komponenten, die die Lernprozesse auf Seiten der Schülerinnen und Schüler unterstützen:

- *Zielorientierung*: Bereitstellung von Struktur, Zielklarheit und Transparenz
- *Lernbegleitung*: Begleitung des Lernens durch die Lehrpersonen und das Überwachen von Lernprozessen und -entwicklungen
- *Lernatmosphäre*: Schaffung eines unterstützenden und positiven Lernklimas

Die drei Komponenten bilden die Grundlage dafür, welche Situationen und Ereignisse von Lehrpersonen bei der Unterrichtsbeobachtung identifiziert werden. Das Knowledge-based reasoning hingegen zielt auf die wissensgesteuerte Verarbeitung von Unterricht ab. Dabei muss eine systematische Beobachtung unter Einbezug theoretischen Wissens durchgeführt werden. Kennzeichnend dafür sind folgende drei Ebenen:

- Komponenten eines lernwirksamen Unterrichts auf der Basis theoretischen Wissens differenziert *beschreiben*
- Unterrichtssituationen auf der Basis wissenschaftlicher Theorien und Befunde *erklären*
- Wirkungen von Unterrichtssituationen auf weitere Lehr-Lern-Prozesse *vorhersagen*

Anhand dieser Operationalisierung untersuchten Seidel et al. (2010) im Projekt „Observe“ die Struktur und Entwicklung einer professionellen Wahrnehmung von Unterricht bei Lehramtsstudierenden. Dazu werden Videoclips verschiedener Fächer aus Jahrgangsstufe 8 bzw. 9 verwendet und in ein computerbasiertes Instrument integriert. Daneben gibt es weitere Projekte zum noticing und knowledge-based reasoning, wie etwa die Projekte von Seidel und Stürmer (2014) oder Schäfer und Seidel (2015), die sich jedoch auf dasselbe Verständnis der beiden Begriffe stützen und aus diesem Grund an dieser Stelle nicht weiter ausgeführt werden.

Eine differenziertere Operationalisierung diagnostischer Aktivitäten liefern Beretz, Lengnink und von Aufschnaiter (2017). In ihrem Modell legen sie (fach-)didaktische Überlegungen zugrunde (Theorien und Konzepte, etwa zu Grundvorstellungen oder den Stufen des Begriffsverständnisses) und unterscheiden fünf charakterisierende Komponenten des Diagnoseprozesses: Wahrnehmen, Beschreiben, Deuten, Ursachen ergründen und Konsequenzen ableiten (vgl. Abb. 6). Zunächst müssen Lehrkräfte auf geeignete Daten zurückgreifen, die entweder vor dem Hintergrund einer diagnostischen Frage erhoben werden (müssen) oder bereits vorliegen, etwa als Aufgabenlösungen oder Videos mit Transkripten (Beretz et al. 2017). Anschließend erfolgen eine Beschreibung der (förder-)relevanten Beobachtungen, deren differenzierte Deutung unter Einbezug von theoretischen Elementen und Kriterien sowie die Ergründung potenzieller Ursachen und Erklärungen für das gedeutete Verhalten (ebd.). Das Deuten sowie die Ursachenergründung ermöglichen eine Einordnung der wahrgenommenen Situation in Bezug auf die jeweiligen fachdidaktischen Inhalte. Auf dieser Grundlage können schließlich begründete und gezielte Konsequenzen formuliert und ggf. gegeneinander abgewogen werden (ebd.). Bei diesem letzten Schritt handelt es sich um theoretische Überlegungen und nicht um eine konkrete

Umsetzung dieser Überlegungen. Die konkrete Umsetzung von Fördermaßnahmen könnte wiederum Ausgangspunkt für einen erneuten Diagnoseprozess darstellen. Beretz et al. (2017) weisen darauf hin, dass es sich bei ihrem Diagnosemodell nicht zwangsläufig um einen linearen Prozess handeln muss, sondern Diagnostik sich in der Einnahme vielfältiger Perspektiven und unter Berücksichtigung unterschiedlicher Kriterien in einem iterativen Prozess vollzieht. Das bedeutet unter anderem auch, dass einzelne Komponenten ausgelassen oder übersprungen werden können und zwischen verschiedenen Komponenten hin- und hergewechselt werden kann.

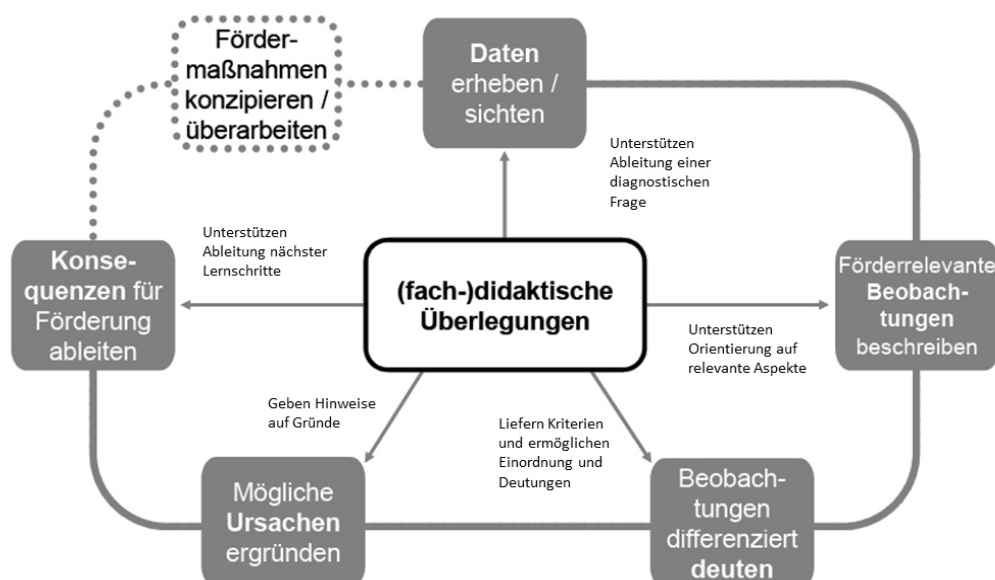


Abbildung 6: Charakterisierende Komponenten der Diagnose mit fachdidaktischen Überlegungen als Bezugspunkt nach Münster und von Aufschnaiter (2019) sowie Beretz, Lengnink und von Aufschnaiter (2017)

Fachdidaktische Überlegungen unterstützen bei der Erhebung bzw. Wahrnehmung der Daten hinsichtlich der Ableitung einer diagnostischen Frage (Münster und von Aufschnaiter 2019). Sie unterstützen außerdem dabei, förderrelevante Beobachtungen zu beschreiben, und helfen bei der Orientierung bzw. Fokussierung auf relevante Aspekte. Dies können beispielsweise Grundvorstellungen in dem entsprechenden mathematischen Themengebiet sein. Auf dieser Grundlage können die wahrgenommenen (und beschriebenen) Aspekte hinsichtlich fachdidaktischer Theorien und Überlegungen gedeutet werden. Hierfür liefern die Theoriebezüge Kriterien und ermöglichen dadurch erst die Einordnung und Deutung (ebd.). Außerdem können Ursachen für die vorliegenden Deutungen aus bestehenden fachdidaktischen Theorien ergründet werden. Zudem geben fachdidaktische Überlegungen Hinweise für die Ableitung von konkreten Fördermaßnahmen und unterstützen bei der Planung nächster Lernschritte (ebd.). Vor diesem Hintergrund wird deutlich, dass die Fachdidaktik im Diagnosemodell tief verankert ist und entsprechend eine große Rolle für die Diagnose spielt.

Das Modell von Beretz et al. (2017) fokussiert hauptsächlich die aktionale Phase des Diagnosemodells von Reinhold (2018), wobei Beretz et al. (2017) (fach-)didaktische Überlegungen als wesentlich erachten, während bei Reinhold (2018) fachliche Kenntnisse sowie Kenntnisse über die jeweiligen Schülerinnen und Schüler hervorgehoben werden. Die Komponente des Beschreibens findet sich (zumindest explizit) nur im Modell von Beretz et al. (2017). Darüber hinaus wird das Interpretieren bei Beretz et al. (2017) in zwei verschiedene Komponenten un-

terteilt: das differenzierte Deuten von Beobachtungen sowie die Ergründung möglicher Ursachen. Hierbei muss jedoch auch darauf hingewiesen werden, dass Reinhold ihre dreigliedrige Aufteilung nochmals in unterschiedliche Schritte unterteilt, sodass sie nur auf den ersten Blick weniger wesentliche Facetten umfasst (Reinhold 2018). Für die Erhebung fachdidaktischer Facetten der Diagnose erscheint das Modell von Beretz et al. (2017) zieführender zu sein, da diese dort integraler Bestandteil sind.

In der bisherigen Forschung zur diagnostischen Kompetenz wurde häufig nicht auf derartige Modelle zurückgegriffen, sondern die Urteilsgenauigkeit der Lehrkräfte in den Blick genommen. Dies lässt sich auch darauf zurückführen, dass die diagnostische Kompetenz teilweise auch entsprechend definiert wurde:

Mit diagnostischer Kompetenz (engl. „diagnostic competence“, „accuracy of judgement“) bezeichnet man die Fähigkeit eines Urteilers, Personen zutreffend zu beurteilen (Schrader 2006, 102).

Die Urteilsgenauigkeit gliedert sich in die Rangordnungskomponente, die Niveauelemente und die Differenzierungskomponente. Die Rangordnungskomponente beschreibt die Fähigkeit, die Schülerinnen und Schüler gemäß ihren Fähigkeiten zutreffend zu ordnen, also in eine Rangordnung einteilen zu können. Die Differenzierungskomponente bezieht sich auf die Fähigkeit, die Merkmalsähnlichkeit in der Klasse möglichst genau einzuschätzen und die Niveauelemente beschreibt die Fähigkeit, die mittlere Merkmalsausprägung in einer Klasse zutreffend einzuschätzen. Eine Ergänzung zu den klassischen Urteilskomponenten stellt der aufgabenspezifische Treffer dar (Karing et al. 2011). Hierbei wird die Lehrereinschätzung mit der Schülerleistung einzelner Aufgaben abgeglichen. Die Besonderheit dabei ist, dass nun auf Ebene einzelner Aufgaben überprüft wird, ob die Lehrkraft die Schülerleistung entsprechend einschätzen kann. Diese Einschätzung, welche die Schwierigkeit der Aufgabe, die Art der Anforderung bzw. den Aufgabentyp berücksichtigt, ist besonders wichtig für die Gestaltung von Interventionen (ebd.). Da diese Operationalisierung der diagnostischen Kompetenz von Studierenden weit verbreitet ist, lassen sich auch diese drei Komponenten als Komponenten der Diagnose bezeichnen.

2.5 Bisherige Forschung und empirische Ergebnisse

In der Forschung zu diagnostischen Fähigkeiten bzw. zur diagnostischen Kompetenz von Lehrkräften wird die Akkuratheit von Lehrerurteilen häufig über einen Vergleich des Lehrerurteils zu bestimmten Schülermerkmalen mit den tatsächlichen Merkmalsausprägungen der Schülerinnen und Schüler bestimmt (Karing et al. 2011). Zur Ermittlung dieser Übereinstimmung werden hauptsächlich die drei Urteilskomponenten Rangordnung, Differenzierung und Niveau verwendet, welche jeweils unterschiedliche Facetten der Urteilsbildung abbilden (Karing et al. 2011).

Empirische Befunde zeigen, dass Lehrkräfte die Rangordnung ihrer Schülerinnen und Schüler insgesamt recht gut einschätzen können (Hoge & Colardaci 1989; Südkamp, Kaiser & Möller 2012). Allerdings gibt es beträchtliche Unterschiede zwischen den diesbezüglichen diagnostischen Fähigkeiten der einzelnen Lehrkräfte (Hoge & Colardaci 1989; Brunner et al. 2011). Nur sehr wenigen Lehrkräften gelingt es, leistungsbezogene und motivationale Merkmale ihrer Klassen korrekt einzuschätzen (Brunner et al. 2011). Die Einschätzungen der verschiedenen Merkmale – Fähigkeitsselbstkonzept, Lernmotivation, Leistungsängstlichkeit, Stresserleben,

soziale Beziehungen – durch die Lehrkraft fallen (noch) schlechter aus als die Leistungsbeurteilung (van Opuysen 2010, Spinath 2005).

Trotz ihrer hohen Popularität ist die Urteilsgenauigkeit mit einigen Problemen behaftet, da bislang noch keine eindeutigen Richtwerte zur Interpretation vorliegen (Praetorius & Südkamp 2017). Außerdem wird kritisiert, dass die Urteilsgenauigkeit sich zu sehr mit der Einschätzung der Gruppe und zu wenig mit dem Individuum beschäftigt (Praetorius & Südkamp 2017). Es ist daher notwendig, die Verfahren zur Erfassung diagnostischer Fähigkeiten zu optimieren und neuartige Zugänge zu wählen (Praetorius & Südkamp 2017). Denn nur bei einer validen und akkuraten Einschätzung der Lernvoraussetzungen kann adaptiver und gewinnbringender Unterricht stattfinden (Rath 2017, 42).

Einer dieser neuartigen Zugänge ist die Erhebung der diagnostischen Fähigkeiten bzw. der professionellen Unterrichtswahrnehmung mit Hilfe von Unterrichtsvideos. In einem solchen Setting stellten Seidel und Prenzel (2007) fest, dass Novizinnen und Novizen im Lehrberuf Unterrichtssituationen hauptsächlich beschreiben. Diese Beschreibungen fallen aufgrund fehlenden theoretischen Wissens oftmals undifferenziert aus. Expertinnen und Experten sind hingegen häufiger in der Lage, auf der Ebene des Erklärens und des Vorhersagens zu operieren (Seidel & Prenzel 2007). Beretz et al. (2017) konnten in ihrer Studie kaum Beschreibungen von Studierenden erkennen, während überwiegend Deutungen vorgenommen und wenige Ursachen ergründet sowie Konsequenzen abgeleitet wurden. Diese teilweise widersprüchlichen Ergebnisse erklären die Autorinnen unter anderem mit dem vorgelagerten Input im Seminar. Es seien jedoch nur selten kontroverse Deutungen vorgenommen worden, was möglicherweise an zu geringem Potenzial für kontroverse Deutungen in den vorliegenden Situationen liegt (Beretz et al. 2017).

Bislang fehlen weitgehend empirische Befunde, die Personenmerkmale in Zusammenhang mit einer guten Diagnose bringen (Rath 2017, 81). Alles in allem konnten bislang kaum bzw. nur geringe direkte Zusammenhänge zwischen Diagnose und Leistungsentwicklung von Schülerinnen und Schülern empirisch nachgewiesen werden (Rath 2017: 22; Karing et al. 2011). Entsprechend besteht an dieser Stelle ein erhöhter Forschungsbedarf.

Im Bereich der Diagnose von Fehlern existieren in der Mathematikdidaktik u.a. Studien von Lindmeier (2011), Heinrichs (2015) und Seifried und Wuttke (2010), welche für diese Arbeit jedoch aufgrund des anders gelagerten thematischen Fokus nicht relevant sind und daher an dieser Stelle nicht weiter ausgeführt werden. Weitere Ergebnisse finden sich auch bei Brunner et al. (2011). Außerdem zählen Hesse und Latzko (2011, 31ff.) einige empirische Befunde auf und erläutern die Untersuchungsdesings einiger oben aufgeführter Forschungsergebnisse. Auch bei Heinrichs (2015, 39ff.) und Leuders et. al. (2018) findet sich jeweils ein Überblick über bisherige Forschungsergebnisse.

2.6 Zusammenfassung und Verortung der eigenen Studie

Zusammenfassend lässt sich festhalten, dass der Begriff der diagnostischen Kompetenz von unterschiedlichen Autoren unterschiedlich definiert und nicht zuletzt auch unterschiedlich operationalisiert wird. Abbildung 7 stellt eine nicht abschließende Übersicht über Aspekte der diagnostischen Kompetenz dar, die im Folgenden einerseits vorgestellt und andererseits zur Verortung der eigenen Studie genutzt wird. Dabei wird mithilfe von Pfeilen verdeutlicht, welche Bereiche einen Einfluss auf einen oder mehrere andere Bereiche haben können. Auf der linken Seite der Abbildung sind persönliche Voraussetzungen aufgeführt, die auf die diagnostische

Kompetenz einwirken. In der Mitte stehen die Diagnoseaufgaben, Art und Zeitpunkt der Diagnose sowie Fokus der Diagnose und Diagnoseauftrag. Auf der rechten Seite stehen die Teilfähigkeiten der Diagnose sowie die Komponenten des Diagnoseprozesses, welche die Performanz sowohl direkt als auch indirekt über die Diagnoseaufgabe beeinflussen. Die Bezugsnorm entscheidet schließlich darüber, womit das beobachtete Handeln verglichen werden soll, und trägt dadurch entscheidend zur Performanz bei. Im Folgenden sollen einzelne Facetten des Schaubilds vertieft und der Teil der empirischen Studie, der sich mit der Diagnose befasst, darin vertort werden.

Fachliches und fachdidaktisches Wissen sowie pädagogisch-psychologisches Wissen werden gemeinhin als Voraussetzungen für Diagnosekompetenz angesehen. Die vorliegende Studie fokussiert insbesondere auf das fachdidaktische Wissen. Fachliches und pädagogisch-psychologisches Wissen werden vergleichsweise nur in niederschwelliger Form angesprochen. In direktem Zusammenhang zu den Voraussetzungen steht der Diagnosefokus, mit welchem das mathematische Thema (bzw. Themengebiet) gemeint ist, das diagnostiziert wird. Da fachliches und fachdidaktisches Wissen die Diagnose beeinflussen, ist es naheliegend, dass auch der Diagnosefokus darüber entscheidet, wie ausgeprägt die diagnostischen Fähigkeiten in diesem Bereich sind, da die meisten Personen in unterschiedlichen Gebieten divergierendes fachliches bzw. fachdidaktisches Wissen aufweisen. Des Weiteren können sich die Dispositionen der Studierenden auf deren Diagnosekompetenz auswirken. Dies wird jedoch im Rahmen der Studie nicht erfasst.

Der Fokus der Diagnose kann als Bindeglied zwischen den persönlichen Voraussetzungen und den in der Mitte des Schaubilds dargestellten Bereichen angesehen werden, da die Wahl des mathematischen Themas einerseits Fachwissen und fachdidaktisches Wissen von dem Diagnostizierenden in diesem Gebiet voraussetzt, andererseits auch auf die Formulierung des Diagnoseauftrags einwirkt. Dieser wiederum muss auch zu der gewünschten Diagnoseart passen. Die Art der Diagnose kann den Zeitpunkt der Diagnose implizit vorgeben, da die Art der Diagnose festlegt, ob diese vor, nach oder während des Lernprozesses stattfindet und ob es sich um eine oder mehrere Diagnosezeitpunkte handelt. Der Zeitpunkt der Diagnose hängt seinerseits direkt mit der Diagnoseaufgabe zusammen. Unter Diagnoseaufgabe wird an dieser Stelle die Unterscheidung zwischen formeller und informeller Diagnose verstanden, während der Diagnoseauftrag sich auf die sprachliche Formulierung des konkreten Auftrages zur Diagnose bezieht. Dieser kann entweder von einer anderen Person vorgegeben werden oder von der/dem Diagnostizierenden selbst stammen. Die Formulierung des Diagnoseauftrages kann sowohl implizit als auch explizit geschehen. Hinsichtlich der Diagnoseaufgabe handelt es sich bei den Diagnoseaufträgen in der vorliegenden Studie um informelle Diagnosen des Unterrichtsgeschehens (vgl. dazu Kap. 8.2). Daraus ergeben sich auch die Diagnoseart (Prozessdiagnostik) und der Diagnosezeitpunkt (während der Situation). Die Antworten der Studierenden auf die Diagnoseaufträge werden mit den Antworten der Expertinnen und Experten verglichen. Durch ein vorgelagertes Expertenrating ist die Bezugsnorm objektiv, da weder auf den Lernzuwachs des einzelnen geschaut wird, noch die Ergebnisse zur Bewertung im Bereich der Diagnose verglichen werden. Vielmehr dienen die Expertenratings als Vergleichsreferenz.

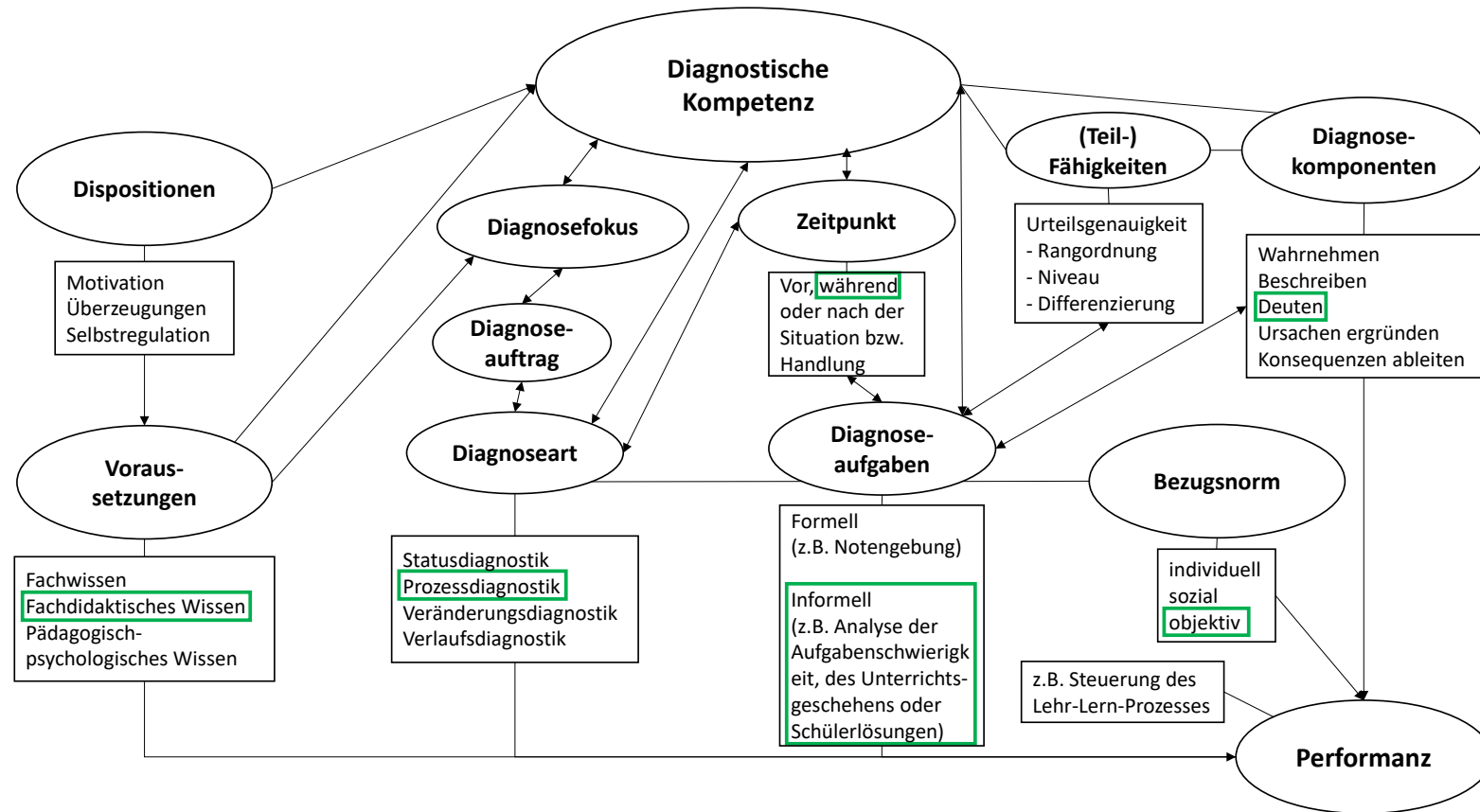


Abbildung 7: Verortung der eigenen Studie in die diagnostische Kompetenz

Diagnostische Fähigkeiten sind ein (wesentlicher) Teilaspekt der Diagnosekompetenz. Es ist aktuell noch unklar, wie viele Teilfähigkeiten die diagnostischen Fähigkeiten umfassen. In bisherigen Forschungsansätzen wurde häufig die Urteilsgenauigkeit von Lehrkräften zugrundegelegt. Eine andere Operationalisierung ergibt sich bei Zugrundelegen der Komponenten des Diagnoseprozesses von Beretz et al. (2017). Beretz et al. (2017) listen fünf Teilfähigkeiten auf, die theoretisch auch alle operationalisiert und erfasst werden können. Aus erhebungstechnischen Gründen beschränkt sich die vorliegende Studie auf die Teilfähigkeit des Deutens (vgl. dazu Kap. 8). Dadurch, dass die Deutung auch voraussetzt, dass die Situation wahrgenommen wurde (und daher theoretisch auch beschrieben werden könnte), ist nicht auszuschließen, dass auch andere Teilfähigkeiten miterhoben werden. Auf die Ursachenergründung wird verzichtet, da die zu beobachtenden Schülerinnen und Schüler den Studierenden nicht bekannt sind, was die Ergründung der vorliegenden Ursachen von Schwierigkeiten und Schülerfehlern immens erschwert. Die Ableitung von Fördermaßnahmen ist in den Interventionen in die Gruppenarbeitsprozesse enthalten, auf die theoretisch im nachfolgenden Kapitel und empirisch in Kapitel 8 und 9 eingegangen wird.

Es herrscht Einigkeit darüber, dass diagnostische Kompetenz kein eindimensionales Konstrukt ist (Helmke und Schrader 1987; Spinath 2005; Brunner et al 2011; Leuders et al. 2018). Weil einige Facetten der Diagnosekompetenz wie beispielsweise die Dispositionen der Studierenden im Rahmen der Studie nicht berücksichtigt werden, wird im Folgenden von diagnostischen Fähigkeiten gesprochen. Da diese sich auf den Lernprozess der Schülerinnen und Schüler beziehen und mittels Prozessdiagnostik begutachtet werden, handelt es sich um prozessdiagnostische Fähigkeiten. Die Diagnoseaufträge beziehen sich explizit nur auf die Fähigkeit „Deuten“ (wenngleich vermutlich auch weitere Teilfähigkeiten, insbesondere das Wahrnehmen, genutzt werden). Entsprechend konzentriert sich die Studie (vorwiegend) auf die prozessdiagnostische Fähigkeit „Deuten“. Die Performanz ist das beobachtbare Verhalten, im Rahmen dieser Studie also die Beantwortung der Diagnoseaufträge zu den Videovignetten (vgl. dazu Kap. 8.2).

Diagnostische Fähigkeiten wie die fortlaufende Registrierung der Lern- und Leistungsfortschritte sowie der Lernschwierigkeiten und Leistungsmängel einzelner Schülerinnen und Schüler werden bisher im Rahmen der Lehrerausbildung und -weiterbildung kaum vermittelt (Weinert 2001). Zwar verweisen van Es und Sherin (2002) darauf, dass diagnostische Fähigkeiten sich im Laufe des Lehrerdaseins entwickeln, jedoch erscheint ein Blick darauf gewinnbringend, inwieweit die prozessdiagnostische Fähigkeit „Deuten“ bei den angehenden Lehrkräften bereits gegen Ende des Studiums ausgeprägt ist und wie diese möglicherweise mit anderen Bereichen (Intervention und Reflexion) zusammenhängt. Insbesondere, weil diagnostische Fähigkeiten eine Lehrkraft dazu befähigen, Schülerinnen und Schüler individuell zu unterstützen und kognitiv zu aktivieren (Grausam 2018, 13f; Brunner et al. 2011, 216). Daher sind diagnostische Fähigkeiten ein wichtiger Bestandteil der Lehrerkompetenz und Lehrerverberuflichkeit (Helmke & Schrader 2006), die es noch weiter zu untersuchen gilt.

3 Interventionen in Gruppenarbeitsprozesse

Die erfolgreiche Berufsausübung ist auch daran zu messen, inwieweit es Lehrkräften gelingt, Lernprozesse von Schülerinnen und Schülern zu initiieren und zu unterstützen, sodass die angestrebten Lernziele erreicht werden (Baumert & Kunter 2011, 30). Daher erscheint es zentral für guten Unterricht, dass Lehrkräfte in Echtzeit auf Schülerhandeln reagieren, also angemessene Mikroadaptationen vornehmen können (Randi & Corno 2005; Schrader & Helmke 2001). Dazu brauchen sie situationsspezifische Fähigkeiten und müssen in der Lage sein, ihre fachlichen und fachdidaktischen Kenntnisse bei ihren Handlungsentscheidungen während der Betreuung von Schülergruppen zu nutzen (Stahnke, Schueler & Roesken-Winter 2016). Es ist eine der zentralen Aufgaben von Lehrkräften, Schülerinnen und Schülern beim Lernen zu helfen (Becker 2008, 231). Herausfordernde Aufgaben alleine reichen nämlich nicht aus, um Schülerinnen und Schüler dauerhaft dazu zu bringen, sich mit den Aufgaben zu beschäftigen und damit verbunden Lernfortschritte zu machen (Baumert, Kunter, Blum, Klusmann, Krauss & Neubrand 2011, 14). Daher müssen Lehrkräfte die Schülerinnen und Schüler mittels verschiedener Hilfen unterstützen. Die Art der Lernhilfe entscheidet dabei schließlich über die Wirksamkeit der Lernprozesse (Becker 2008, 231).

Lehrkräfte benötigen die Fähigkeit, den aktuellen Leistungsstand und damit verbunden auch die Bedürfnisse von Schülerinnen und Schülern zu erfassen, weil sie das so erworbene Wissen einsetzen können, um adäquat mit Schülerinnen und Schülern zu interagieren. Lehrer-Schüler-Interaktion ist ein sehr komplexes und weites Feld, das alle Interaktionsformen zwischen Lehrkraft und Schüler/in meint und letztlich jegliches Handeln im Klassenzimmer beinhaltet. Die vorliegende Arbeit beschränkt sich orientiert an der Forschungsfrage (vgl. dazu Kap. 7) auf Interventionen von Lehrkräften in Arbeitsprozesse von Schülergruppen.

In der Literatur finden sich verschiedene Bezeichnungen für eine solche Unterstützung, beispielsweise „Lehrerimpulse“, „Scaffolding“ oder „Lehrerinterventionen“. Daher soll nun zunächst eine Begriffsklärung stattfinden und die verschiedenen Begriffe dabei in Beziehung zueinander gesetzt werden. Weitere Bezeichnungen im deutschsprachigen und englischsprachigen Raum werden beispielsweise von Krammer (2009, 93ff.) aufgelistet, würden jedoch im Rahmen dieser Arbeit keinen nennenswerten inhaltlichen Mehrwert liefern, weshalb auf die ausführliche Vorstellung zusätzlicher Bezeichnungen verzichtet wird.

3.1 Lehrerimpulse

Unter Lehrerimpulsen werden Denkanstöße verstanden, welche die Schülerinnen und Schüler im Hinblick auf einen bestimmten Inhalt oder ein bestimmtes Ziel aktivieren, motivieren und beim Lernen unterstützen (Wiater 2015, 176). Dabei kann zwischen verbalen und nonverbalen Impulsen unterschieden werden (ebd.). Sie verfolgen das Ziel, Schülerinnen und Schüler zum Weiterdenken, zum Stellungnehmen und zum Werten zu veranlassen, und werden häufig als Hilfsmittel zur Weiterführung des Unterrichtsgesprächs genutzt (Steindorf 2000, 146).

Im Hinblick auf Gruppenarbeitsphasen stellt sich die Frage, ob Lehrkräfte überhaupt in den Arbeitsprozess von Schülerinnen und Schülern eingreifen sollten, und falls ja, wie eine solche Intervention ausgestaltet sein sollte (Link 2011, 66f.). Diegritz et al. (1999) empfehlen, dass Lehrkräfte die Gruppenarbeit möglichst selten, und falls doch, nur kurz, unterbrechen sollten. Die Unterbrechungen sollen nach Diegritz et al. (1999) folgende Richtlinien berücksichtigen:

- Die Lehrkraft sollte sich vor der Intervention gut über das Intragruppengeschehen informieren. Sie sollte das *Geschehen beobachten*, sich in das Gruppengespräch einhören und die Intragruppenkommunikation nicht unvorbereitet und abrupt unterbrechen.
- Um einen hohen Situationsbezug beim Intervenieren zu erreichen, ist ein *hohes Maß an Empathie und Zurückhaltung* erforderlich. Die Lehrkraft sollte versuchen, auf die Gedanken, Anliegen und Probleme der Schülerinnen und Schüler angemessen einzugehen.
- Ein *freundlicher Umgangston* ist im Unterricht allgemein, aber auch bei Lehrerinterventionen günstig. Hohe Umgangsqualität kann erreicht werden, indem man die Schülerinnen und Schüler situationsangemessen lobt, sie ermutigt, tröstet und ihre Meinung anerkennt.
- Lehrkräfte sollten jede Art von Mini-Frontalunterricht vermeiden. Vielmehr sollten sie sich *auf knappe Hinweise beschränken*, die der Gruppe einen Fehler oder eine Sackgasse sichtbar machen und eine Hilfsquelle eröffnen.
- Es ist darauf zu achten, dass sich die Lehrkraft *präzise und verständlich* ausdrückt. Insbesondere sollten die Inhalte des bereits erteilten Arbeitsauftrags nicht relativiert werden.
- Bei Bewertungen von Schülerhandlungen ist auf einen *hohen Situationsbezug* zu achten. Unbegründetes Lob oder ungerechtfertigter Tadel können paradoxe Wirkungen bei den Schülerinnen und Schülern hervorrufen und die Gruppenarbeit zum Erliegen bringen.
- *Interventionen sollten nicht mit Anweisungen oder Fragen beendet werden*, da dies die Schülerinnen und Schüler meist nur verunsichert und der Intragruppenprozess nachhaltig gestört werden kann.
- *Bilanzierende Interventionen*, in denen Schülerinnen und Schüler über den Stand ihrer Arbeit berichten, *sind zu vermeiden*, da diese ihren Auftrag dann meist als erledigt ansehen und in Nebenengagement verfallen.

Die Aufzählung von Diegritz et al. (1999) bezieht sich zwar nicht explizit auf den Begriff der „Lehrerimpulse“, kann jedoch, da es sich um verbale Aufforderungen zum Nachdenken über einen bestimmten Unterrichtsinhalt und um die Aktivierung sowie Motivierung von Schülerinnen und Schülern, also um Lehrerimpulse im Sinne von Wiater (2015, 176f.) handelt, auf diese angewendet werden. Hänze, Schmidt-Weigand und Blum (2007) sind ebenfalls der Auffassung, dass derartige Lernhilfen den produktiven Austausch unterstützen. Dabei sollen die Hilfen nicht nur entweder fachliche oder lernstrategische, sondern jeweils beide Aspekte aufgreifen (ebd.). Die einzelnen Lernhilfen erfüllen schließlich verschiedene Funktionen beim Lösungsprozess (ebd.).

3.2 Scaffolding

Die Qualität des Unterrichtsgesprächs und der Lehrer-Schüler-Interaktionen haben Einfluss auf den Lernerfolg von Schülerinnen und Schülern (Krammer 2009, 36). Im Rahmen des aktuellen Verständnisses von Lernen als konstruktivem, selbstgesteuertem Prozess, in dem die Lernenden in Interaktion mit bedeutungsvollen Inhalten und anderen Personen basierend auf ihrem Vorwissen ihre Kompetenzen aufbauen, weitet sich die Rolle der Lehrperson aus (Krammer 2009,

36). Damit einhergehend richtet sich der Unterricht immer stärker auf die Lernprozesse der einzelnen Schülerinnen und Schüler aus (Krammer 2009, 36f). Die Rolle, die die Lehrkraft in selbstständigkeitsorientierten Lernumgebungen einnehmen soll, ist noch weitgehend ungeklärt bzw. umstritten (Leiss, 2007, 61). Sowohl in der Erziehungswissenschaft als auch in der Fachdidaktik wird kontrovers darüber diskutiert, welche Lehr-Lern-Methoden und welches Lehr- und Lernverhalten adäquat für eine möglichst zielführende Vermittlung der angestrebten Kompetenzen sind (Leiss 2007, 61).

Das auf der Tradition von Piaget und Wygotski aufbauende sozial-konstruktivistische Lehr-Lernverständnis dient der lernpsychologischen Begründung für die individuelle Unterstützung (Krammer 2009, 61). Dabei kommt der sozialen Interaktion aus dieser Perspektive eine hohe Bedeutung für den Aufbau von Wissen zu und die Lehrperson wird zur Begleiterin der Lernenden durch deren Zone der nächsten Entwicklung (Krammer 2009, 61). Diese Zone der nächsten Entwicklung ist nach Wygotski (1987, 83) „das Gebiet der noch nicht ausgereiften, jedoch reifenden Prozesse“. Er schreibt dazu außerdem: „Was das Kind heute in Zusammenarbeit und unter Anleitung vollbringt, wird es morgen selbständig ausführen können. Und das bedeutet: Indem wir die Möglichkeiten eines Kindes in der Zusammenarbeit ermitteln, bestimmen wir das Gebiet der reifenden geistigen Funktionen, die im allernächsten Entwicklungsstadium sicherlich Früchte tragen und folglich zum realen geistigen Entwicklungsniveau des Kindes werden. Wenn wir also untersuchen, wozu das Kind selbständig fähig ist, untersuchen wir den gestrigen Tag. Erkunden wir jedoch, was das Kind in Zusammenarbeit zu leisten vermag, dann ermitteln wir damit seine morgige Entwicklung“ (Wygotski 1987, 83).

Als Antwort auf die Frage nach der Ausgestaltung der Begleitung und den Merkmalen der individuellen Unterstützung wurde Mitte der 70er-Jahre die Metapher des Scaffoldings eingeführt (Krammer 2009, 61). Beim Scaffolding werden Schülerinnen und Schülern Unterstützungsstrukturen zur Verfügung gestellt, welche derart gestaltet sind, dass die Lernenden selbstständig Ideen entwickeln können und gleichzeitig die Sicherheit haben, dass Hilfe verfügbar ist, falls diese benötigt wird (Kobarg 2009, 45f). Der Prozess des Fadings hingegen bezeichnet das langsame Entziehen dieser Unterstützung, wodurch die Schülerinnen und Schüler schließlich selbstständig mit den erworbenen Fähigkeiten umgehen (Kobarg 2009, 46).

Van de Pol, Volman & Beishuizen (2010) verweisen darauf, dass dem Scaffolding-Konzept in den letzten Jahrzehnten eine große Aufmerksamkeit in der Bildungsforschung zuteil wurde. Dabei ist das Scaffolding keine Technik, die in jeder Situation auf die gleiche Weise angewendet werden kann, sondern die Hilfestellung durch die Lehrkraft hängt stark von der Situation und der Aufgabenstellung ab (ebd.). Über die Jahre ist der Begriff des Scaffoldings jedoch auch aus diesem Grund häufig unterschiedlich verwendet worden, sodass die breite Auffassung des Begriffs zu einer gewissen Unschärfe führte (ebd.). Durch die Verwendung unterschiedlicher Definitionen in verschiedenen Studien existiert kein Konsens darüber, was genau Scaffolding ist (ebd.). Jedoch konnten van de Pol et al. (2010) in ihrer Überblicksstudie, in der 66 Artikel untersucht wurden, einige Gemeinsamkeiten herausarbeiten, die sich auch in ihrem Konzeptmodell des Scaffoldings (vgl. Abb. 8) wiederfinden.

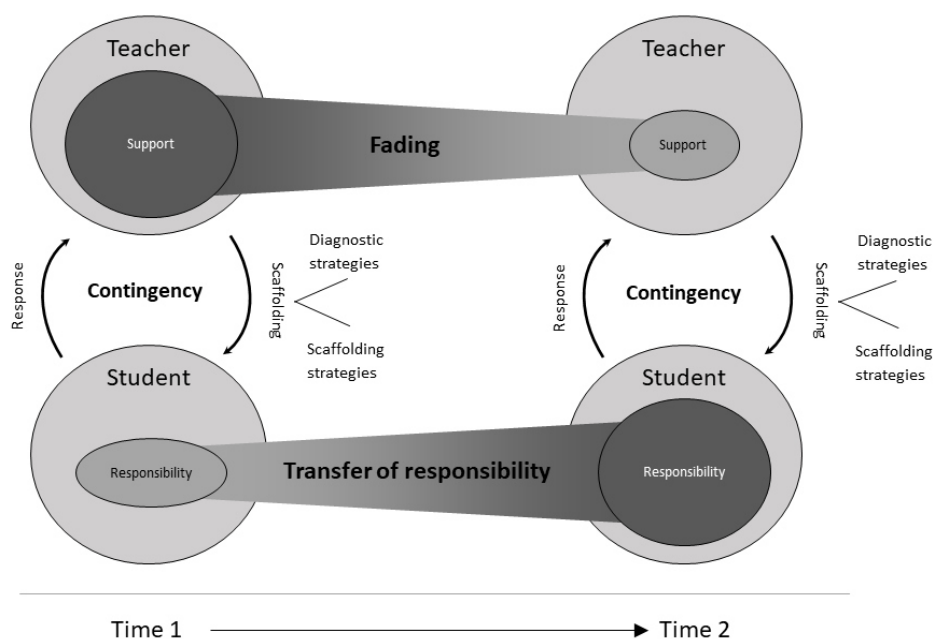


Abbildung 8: Konzeptmodell des Scaffoldings von van de Pol et al. (2010)

Allgemein stellen van de Pol et al. (2010) fest, dass der Kern des Scaffoldings in der Unterstützung eines Schülers bzw. einer Schülerin durch eine Lehrperson besteht, wobei der Lernende an einer Aufgabe arbeitet, die er ohne Unterstützung vermutlich nicht lösen könnte. Dabei ist es wichtig, dass die Lehrkraft den aktuellen Kenntnisstand des Lernenden erfassen kann, um diesen adäquat unterstützen zu können (ebd.). Eine weitere Gemeinsamkeit ist das Fading, also das schrittweise Entziehen von Unterstützung. Dies hängt natürlich mit der jeweiligen Entwicklung der Fähigkeiten der Schülerin bzw. des Schülers zusammen. Darüber hinaus findet beim Scaffolding auch ein Transfer von Verantwortung statt, sprich der Lernende bekommt nach und nach mehr Kontrolle über sein eigenes Lernen (ebd.).

Link (2011, 74) stellt fest, dass es zwischen Scaffolding und Tutoring keine trennscharfe Unterscheidung gibt. Wood, Bruner & Ross (1976) untersuchen das Tutoring als Unterstützung von einzelnen Lernenden beim Lösen eines Problems durch eine Person mit größerer Expertise, indem Kinder von drei bis fünf Jahren Holzpyramiden zusammensetzten und dabei von einer Tutorin unterstützt werden. Dabei geht es um die Frage, ob und unter welchen Umständen es der Tutorin gelingt, die Kinder in die Lage zu versetzen, das Problem anschließend selbstständig lösen zu können. Darüber hinaus wird untersucht, ob sich das Unterstützungsverhalten bei Kindern verschiedenen Alters unterscheidet. Die Videoanalyse ergibt, dass das Unterstützungsverhalten in erfolgreichen Situationen qualitativ und quantitativ dem Alter der Kinder angepasst wurde und entsprechend die Annahmen über den Lernstand eines Kindes und die Schwierigkeit der Aufgabe sowie deren Passung entscheidend sind (Krammer 2009, 75).

Den Begriff des Scaffoldings verwenden Wood et al. (1976) dabei zur Beschreibung der Form der adaptiven Unterstützung, die die Lernenden befähigt, das Problem zukünftig selbstständig lösen zu können (Krammer 2009, 74f.). Somit geht der Begriff weit über das reine Vor-

zeigen und Imitieren von Verhalten hinaus und zielt auf das Verstehen des Problems ab (Krammer 2009, 75). Daher definieren Wood et al. (1976) den Begriff des Scaffoldings folgendermaßen:

...a process, that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts. This scaffolding consists essentially of the adult „controlling“ those elements of the task that are initially beyond the learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence. (Wood et al. 1976, 90)

Wood et al. (1976, 98) nennen und unterscheiden dabei folgende zentralen Merkmale und Unterstützungsfunktionen im Scaffolding:

- *Recruitment*
- *Reduction in degrees of freedom*
- *Direction maintenance*
- *Marking critical features*
- *Frustration control*
- *Demonstration*

Beim Recruitment soll das Interesse des Lernenden für das Problem geweckt werden. Das zweite Merkmal (Reduction in degrees of freedom), das auch als Unterstützungsfunktion angesehen werden kann, meint die Einschränkung der Anzahl der Lösungswege. Eine weitere Unterstützung ist das Aufrechterhalten der Motivation der Schülerinnen und Schüler, sodass diese sich weiterhin mit der Aufgabe beschäftigen (direction maintenance). Marking critical features ist ebenfalls eine Form der Unterstützung und bedeutet, dass während der Aufgabenbearbeitung auf bedeutende Merkmale der Aufgabe aufmerksam gemacht wird. Während des Problemlöseprozesses ist es auch wichtig, die Frustration gering zu halten (frustration control). Es ist außerdem möglich, dass ein Lösungsschritt oder ggf. auch mehrere Lösungsschritte durch die Lehrkraft vorgemacht bzw. gezeigt werden (demonstration).

Krammer (2009, 75f.) fasst diese Merkmale zu drei Hauptfunktionen der Unterstützung bzw. drei Ebenen zusammen, auf denen die Unterstützung wirken soll:

- *Emotionale Ebene:* Motivation wecken und aufrechterhalten sowie Unterstützung im Umgang mit Frustration
- *Prozedurale Ebene:* Strukturieren und Aufrechterhalten des Lösungsprozesses, beispielsweise durch die Einschränkung der möglichen Lösungswege
- *Inhaltliche Ebene:* Hinweise auf für die Aufgabenlösung bedeutsame Aufgabenmerkmale und Vorzeigen von wichtigen Lösungsschritten

Die im Rahmen des Teilkapitels vorgenommene Beschreibung des Scaffoldings ist nicht erschöpfend. Die bisher genannten Autoren und Beschreibungen sollten in ausreichendem Maße dazu beigetragen haben, einen Einblick in gängige Konzepte und Merkmale des Scaffoldings zu gewinnen. Eine weitere Taxonomie des Scaffoldings findet sich beispielsweise bei Dubs (2009, 91ff.). Darüber hinaus stellt Krammer (2009, 77ff.) die Scaffolding-Ansätze weiterer Autoren vor.

3.3 Lehrerinterventionen

Das Scaffolding und Lehrerinterventionen sind zwei Konzepte, die eng miteinander verwoben sind. Um dies nachvollziehen zu können, wird der Begriff „Lehrerinterventionen“ zunächst vorgestellt, um anschließend auf Gemeinsamkeiten und Unterschiede der beiden Begriffe einzugehen.

Für Fürst (1999) sind Lehrerinterventionen während der Gruppenarbeit Unterbrechungen „der Intragruppenkommunikation durch die Lehrkraft“ (Fürst 1999, 122). Durch die Intervention haben Lehrkräfte die Möglichkeit, den Verlauf sowie die Ergebnisse einer Gruppenarbeit zu beeinflussen (Fürst 1999, 107). Ähnlich sind für Kiper und Mischke (2009) Interventionen strategische Handlungen, die auf das Erreichen eines Ziels ausgerichtet sind und ein Bündel von Verhaltensweisen umschließen, das man individuell, situativ und gleichzeitig zielgerichtet nutzt (Kiper & Mischke 2009, 20).

Für Leiss (2007, 65) stellen alle verbalen, paraverbalen und nonverbalen Eingriffe der Lehrkraft in den Lösungsprozess der Schülerinnen und Schüler Lehrerinterventionen dar. Inwieweit eine Lehrkraft es schafft, Schülerinnen und Schülern im Rahmen selbstständigkeitsorientierter Arbeitsphasen angemessene Hilfestellungen zu geben, hängt nach Leiss (2007, 64) weniger von der Anzahl oder der Länge der Hilfen ab, sondern vielmehr davon, inwiefern diese adaptiv sind.

Als adaptive Lehrerinterventionen werden (...) solche Hilfestellungen des Lehrers definiert, die individuell den Lern- und Lösungsprozess der Schüler minimal unterstützen, dass die Schüler maximal selbstständig weiterarbeiten können (Leiss 2007, 65).

Dementsprechend bedeutet Adaptivität die optimale Passung der Lehrerhandlung in Bezug auf die individuellen, kognitiven und sozialen Voraussetzungen der Schülerinnen und Schüler (Leiss 2007, 65). Dies findet sich auch bei einigen Autoren, die den Begriff des Scaffoldings nutzen (vgl. dazu u.a. van de Pol et al. 2010; Wood et al. 1976; Krammer 2009). Jedoch wird der Begriff von anderen Autoren auch etwas starrer interpretiert (van de Pol et al. 2010), sodass die beiden Begriffe „Scaffolding“ und „adaptive Lehrerinterventionen“ nicht gleichgesetzt werden können. Dies lässt sich auch daran erkennen, dass das Scaffolding in der Regel langfristiger angelegt ist als eine Lehrerintervention. Während beim Scaffolding die Unterstützung nach und nach zurückgefahren und die Verantwortung für den Lernprozess ebenfalls schrittweise von der Lehrkraft auf den Lernenden übertragen wird, handelt es sich bei einer Lehrerintervention unter Umständen nur um eine kurze Interaktion zwischen Lehrkraft und Lernendem. Insofern geht Scaffolding in der Regel über (einzelne) Lehrerinterventionen hinaus. Beiden Begriffen ist jedoch gemein, dass der Kenntnisstand und die Lernvoraussetzungen der Schülerinnen und Schüler jeweils individuell erfasst werden sollten, um auf dieser Grundlage eine didaktische Handlung vorzunehmen.

Diese Adaptivität findet sich auch im etwas weiter gefassten Begriff „adaptive Lehrkompetenz“. Nach Brühwiler bezeichnet er die Fähigkeit einer Lehrperson, „ihren Unterricht so auf die individuellen Voraussetzungen der Schülerinnen und Schüler auszurichten und während des Unterrichts laufend anzupassen, dass für möglichst viele Schülerinnen und Schüler günstige Bedingungen für das Erreichen der Lernziele geschaffen werden“ (Brühwiler 2014, 74). Die adaptive Lehrkompetenz kann dabei sowohl in der Unterrichtsplanung als auch während des Unterrichts genutzt werden, um diesen anzupassen (Brühwiler 2014, 74f.). Dabei kann es erforderlich sein, Handlungsanpassungen an situative Gegebenheiten vorzunehmen (Brühwiler

2014, 75). Beispielsweise sollte eine Anpassung des Unterrichts stattfinden, wenn eine Lehrkraft bemerkt, dass die Schülerinnen und Schüler den neuen Lernstoff nicht verstanden haben (ebd.). Umgekehrt kann es auch vorkommen, dass die Schülerinnen und Schüler die Lernziele zügiger als erwartet erreichen und entsprechend auf gleichartige Übungen verzichtet werden sollte, um nicht unnötig Lernzeit zu opfern und ggf. Langeweile und Desinteresse bei den Schülerinnen und Schülern hervorzurufen (ebd.).

Ein wesentliches Ziel der adaptiven Lehrkompetenz ist die optimale Passung zwischen Unterstützungsbedarf und Unterstützungsangebot (ebd.). Beim Lernen im Klassenverband erfahren aufgrund der heterogenen Lernvoraussetzungen in der Regel nur einzelne Schülerinnen und Schüler eine optimale Passung (ebd.). Dadurch sind manche Schülerinnen und Schüler ständig überfordert und andere lernen weniger, als sie aufgrund ihrer individuellen Lernvoraussetzung könnten (ebd.).

Die adaptive Lehrkompetenz beinhaltet die adaptive Planungskompetenz und die adaptive Handlungskompetenz. Mit der adaptiven Planungskompetenz wird die Fähigkeit von Lehrpersonen bezeichnet, bei der Planung von Unterricht die verschiedenen Voraussetzungen für einen gelingenden Unterricht zu berücksichtigen (Brühwiler 2014, 88). Mit adaptiver Handlungskompetenz ist die Fähigkeit von Lehrpersonen gemeint, unterschiedliche Informationen und situative Veränderungen während des Unterrichts zu erfassen, diese in ihre Handlungsentscheidungen miteinzubeziehen und dadurch den Unterricht an die aktuellen Bedürfnisse der Schülerinnen und Schüler auszurichten, damit diese eine bestmögliche individuelle Unterstützung erhalten (ebd.).

Brühwiler (2014, 91) geht davon aus, dass Planungs- und Handlungskompetenz zwar zeitlich und funktional unterschiedlich angelegt, aber nicht unabhängig voneinander sind. Er nimmt an, dass adaptive Handlungskompetenz bis zu einem gewissen Grad durch eine hohe Planungskompetenz kompensierbar sei (Brühwiler 2014, 93). Dies wird damit begründet, dass die Berücksichtigung vieler möglicher Unterrichtssituationen und dazu passender didaktischer Maßnahmen auf unvorhergesehene Situationen vorbereiten. Deswegen müssen derart geplante Unterrichtssequenzen seltener Anpassungen unterzogen werden (ebd.). Jedoch schreibt auch Brühwiler, dass es unmöglich sei, alle Eventualitäten im komplexen Unterrichtsgeschehen vorzusehen (ebd.).

Aus diesem Grund erscheinen Nachjustierungen in Form von Interventionen in die Arbeitsprozesse von Schülerinnen und Schülern plausibel. Fürst und Haag (1998) unterscheiden dabei zwischen invasiven und responsiven Interventionen. Invasive Interventionen gehen von der Lehrkraft aus, wohingegen responsive Interventionen durch die Schülerinnen und Schüler initiiert sind. Seifried und Klüber (2006, 153ff.) bauen auf der Unterscheidung von Fürst und Haag (1998) auf und ändern die beiden Bezeichnungen aufgrund der ggf. negativen Konnotation des Begriffs „invasiv“ in „lehrerinitiiert“ und „lernerinitiiert“ (Seifried & Klüber 2006, 156). Die Qualität der Intervention erfassen Seifried und Klüber (2006, 156f.) mit Hilfe nicht genauer erläuterten Expertenratings auf einer fünfstufigen Likert-Skala unter Berücksichtigung der nachfolgenden vier Kriterien:

- *Orientierung*
- *Aufgabenbezug*
- *Umgangsqualität*
- *Lenkung*

Das Kriterium *Orientierung* bezieht sich darauf, ob die Lehrkraft sich über das Intergruppen-geschehen informiert, etwa durch Beobachtung oder Nachfragen (Seifried & Klüber 2006, 156). Interventionen ohne Einholung von Informationen über den Arbeitsstand der Gruppe zeugen von geringer Orientierung (ebd.). Unter *Aufgabenbezug* verstehen Seifried und Klüber (2006, 156f), ob die Lehrkraft den aktuellen Stand der Bearbeitung der Lerngruppe reflektiert und die Interventionen auf dieser Grundlage ausgestaltet. Situationsgerechtes Handeln ist ein wesentliches Merkmal eines hohen Aufgabenbezuges (ebd.). Mit *Umgangsqualität* ist die Wertschätzung gemeint, die die Lehrkraft Schülerinnen und Schülern entgegenbringt. Entsprechend bedeutet eine hohe Umgangsqualität, dass die Lehrkraft die Schülerinnen und Schüler lobt, ermutigt oder tröstet sowie freundliche und herzlich mit ihnen umgeht (Seifried & Klüber 2006, 157). Anhand des Kriteriums *Lenkung* wird überprüft, ob die Lehrkraft ihre eigenen Gedanken und Wünsche hinsichtlich der Gruppenarbeit in den Vordergrund rückt, was einer hohen Lenkung entsprechen würde, oder ob sie sich als gleichberechtigten Partner versteht, was einer niedrigen Lenkung entsprechen würde (ebd.).

Dass es grundsätzlich auch ohne Intervention funktionieren kann und diese im selbstständigkeitsorientierten Unterricht wirklich nur genutzt werden sollte, wenn die Schülerinnen und Schüler ohne Intervention nicht sinnvoll an der aktuellen Aufgabe weiterarbeiten könnten, bekräftigen Leiss und Tropper (2014, 9). Überfachliche Kompetenzen wie Zeitmanagement oder Konzentrationsfähigkeit sollen die Schülerinnen und Schüler dazu befähigen, Lernprozesse möglichst autonom zu gestalten (ebd.). Auch Fürst (1999) rät dazu, möglichst selten in kooperative Arbeitsphasen zu intervenieren. Die Beobachtung der Lernenden sollte aus der Distanz stattfinden, beispielsweise vom Lehrertisch aus oder mittels Videoaufzeichnungen (Diegritz et al. 1999, 346). Die physische Präsenz der Lehrkraft verleitet die Schülerinnen und Schüler dazu, die Verantwortung für den Lernprozess an den Lehrer zurückzugeben (Leiss & Tropper 2014, 8).

Demgegenüber scheint es jedoch ebenso wenig zielführend zu sein, Lernende komplett ohne Unterstützung sich selbst zu überlassen (Leiss & Tropper 2014, 10). Die Notwendigkeit der Unterstützung leiten Leiss und Tropper (2014, 10) daraus ab, dass das selbstständige Arbeiten von Schülerinnen und Schülern in kooperativen Lernumgebungen hinsichtlich der Konversationsqualität und der Tiefe des Verstehens häufig defizitär ist. Entsprechend ist die Unterstützung der Lerngruppe keinesfalls überflüssig, sondern adaptive Lernhilfen tragen zur individuellen Förderung und Unterstützung der Schülerinnen und Schüler bei. Diese Unterstützungsmaßnahmen, welche im Regelfall auf einer Diagnose aufbauen, können in Makroadaptionen und Mikroadaptionen unterschieden werden (Schrader & Helmke 2001). Makroadaptionen stellen Anpassungen ganzer Lerneinheiten dar, während Mikroadaptionen spontane Abstimmungsmaßnahmen im Lernprozess sind (ebd.).

Zunächst muss auf Grundlage situationsrelevanter Schülervoraussetzungen entschieden werden, ob interveniert werden soll oder ob die Lernenden auch ohne Intervention sinnvoll weiterarbeiten können (Leiss & Tropper 2014, 15). Falls eingegriffen wird, muss abgewogen werden, welche Unterstützungsmaßnahme durchgeführt werden soll (ebd.). Im Anschluss an die Intervention sollte diese bezüglich ihres Erfolgs validiert werden (ebd.).

Die Aufmerksamkeit für beginnende Schwierigkeiten und eine kalibrierte Unterstützung, die gleichwohl die Autonomie des Lernens und die Integrität der Personen respektiert, kann nicht nur die Motivation stabilisieren, sondern stellt wahrscheinlich selbst eine essenzielle Komponente wirksamer Lernumgebungen dar (Baumert et al. 2011, 14).

Insgesamt scheint es also eine Vielzahl an Begründungen für Interventionen in Gruppenarbeitsprozesse zu geben. Entscheidend ist dabei die Art und Weise des Eingriffs. Im nächsten Teilkapitel werden nun anhand verschiedener Modelle unterschiedliche Möglichkeiten zur Erfassung von Interventionen vorgestellt.

3.4 Modelle zur Erfassung von Interventionen

Der individualisierten Unterrichtsgestaltung kommt auch aus empirischer Sicht eine hohe Relevanz zu. Mit der Differenzierung und Adaptivität im Unterricht wird nach der optimalen Förderung der Lernprozesse der einzelnen Lernenden im Unterrichtsverband gesucht und damit auch den Bestrebungen nach einer kognitiv aktivierenden und schülerorientierten Unterrichtsgestaltung entsprochen, welche neben der klaren Strukturierung des Unterrichts und dessen Inhalten und der effizienten Klassenführung als relevante Merkmale eines lernförderlichen Unterrichts gelten. (Krammer 2009, 30)

Leiss (2007, 96ff.) nutzt im Rahmen der Auswertung seiner Studie mittels Qualitativer Inhaltsanalyse die Kategorien Auslöser, Ebene und Absicht (deduktiv erstellt), wobei alle Kategorien nochmals in mehrere Unterkategorien unterteilt werden und diese wiederum (induktiv) in weitere Unterkategorien aufgesplittet werden (vgl. Abb. 9).

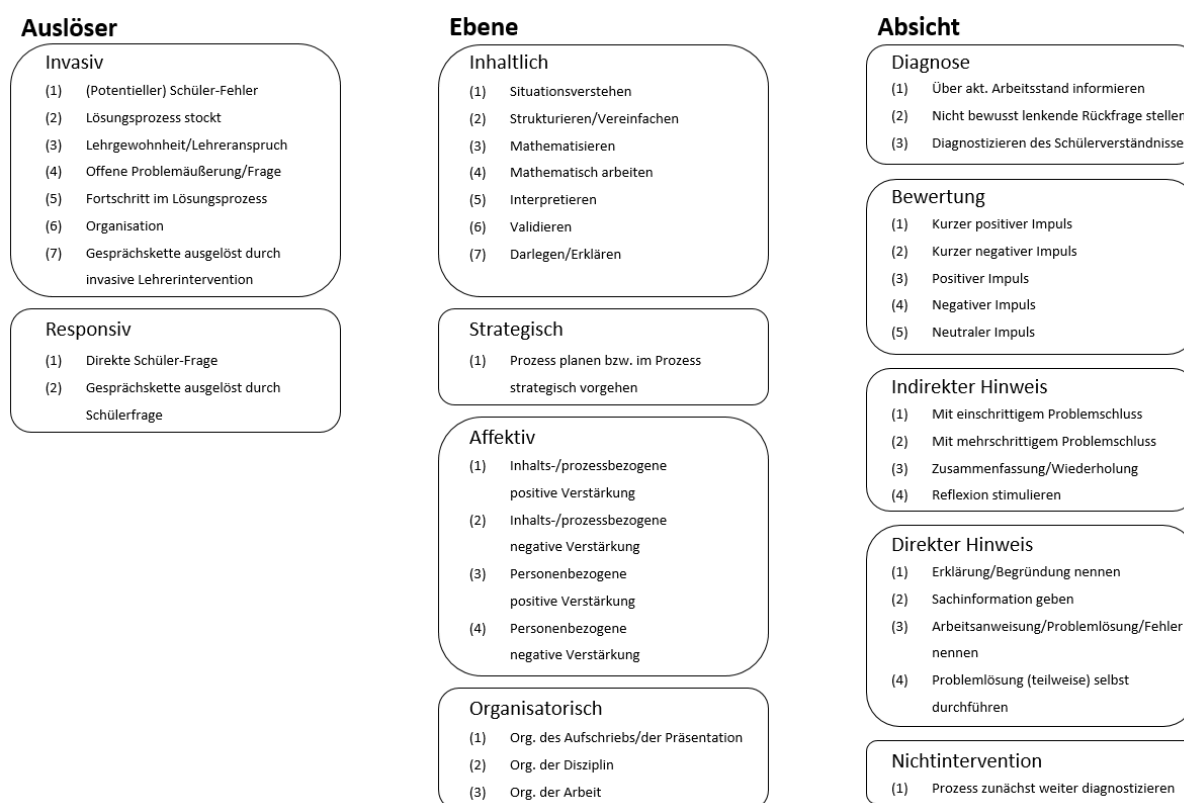


Abbildung 9: Kategoriensystem lösungsprozessbegleitender Interventionen nach Leiss (2007, 109)

Die Kategorie Auslöser wird in invasive und responsive Auslöser unterteilt, wobei Leiss (2007, 105f.) sieben invasive und zwei responsive Auslöser unterscheidet. Die Ebene einer Intervention wird in die Unterkategorien inhaltlich, strategisch, affektiv und organisatorisch untergliedert (Leiss 2007, 106f.). Inhaltliche Interventionen beinhalten konkrete Hinweise in Bezug auf

den Inhalt einer Aufgabe. Strategische Interventionen setzen den Schwerpunkt auf die Vermittlung metakognitiver (Lern-)Strategien, beispielsweise auf die Lenkung der Informationsaufnahme bzw. der Aufmerksamkeit. Interventionen, welche auf die Lernleistungen der Schülerinnen und Schüler Einfluss nehmen, indem positive oder negative Impulse gegeben werden, werden als affektive Interventionen bezeichnet. Organisatorische Interventionen sind solche, die dafür sorgen, dass der Bearbeitungsprozess der Schülerinnen und Schüler organisiert ist und dadurch reibungslos ablaufen kann.

Während inhaltliche Interventionen nochmals in sieben Unterkategorien unterteilt werden, gibt es bei strategischen Interventionen eine, bei affektiven Interventionen zwei und bei organisatorischen Interventionen drei Unterkategorien (Leiss 2007, 106f.). Die Absicht der Interventionen wird in die Unterkategorien Diagnose, Bewertung, indirekter Hinweis, direkter Hinweis und Nichtintervention unterteilt (Leiss 2007, 107f.). Auch hier enthalten die einzelnen Kategorien nochmals bis zu vier Unterkategorien. Insgesamt zeigt sich, dass mithilfe des Kategoriensystems eine Vielzahl unterschiedlicher Gesichtspunkte erfasst wird.

Scherres (2013) baut auf die Überlegungen von Leiss (2007) auf und versteht Lehrerinterventionen in ihrer Arbeit als Prozess, der mit der Diagnose beginnt und optimalerweise nach der getätigten Intervention dazu führt, dass die Schülerinnen und Schüler eigenverantwortlich weiterarbeiten. Den idealtypischen Ablauf einer Lehrerintervention stellt sie folgendermaßen dar:

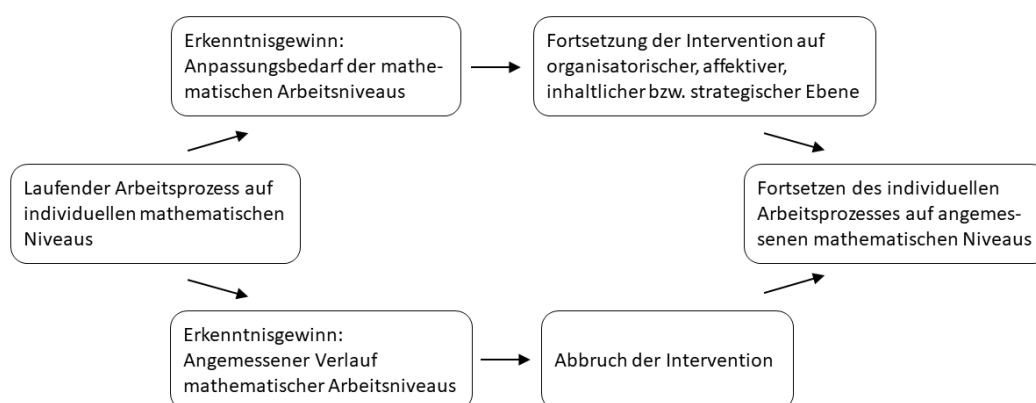


Abbildung 10: Ablaufmodell einer Lehrerintervention bei einer selbstdifferenzierenden Aufgabenstellung von Scherres (2013, 42)

In Scherres (2013) Ablaufmodell (vgl. Abb. 10) kann man erkennen, dass eine Intervention dann abgebrochen werden soll, wenn festgestellt wird, dass die Schülerinnen und Schüler in der Lage sind, sich selbstständig sinnvoll mit den mathematischen Inhalten auseinanderzusetzen. Diagnostiziert die Lehrkraft jedoch einen Anpassungsbedarf des mathematischen Niveaus, sollen organisatorische, affektive, inhaltliche oder strategische Maßnahmen durchgeführt werden. Der entscheidende Punkt ist dabei die Diagnose, bei der die Angemessenheit des mathematischen Arbeitsniveaus für die einzelnen Schülerinnen und Schüler evaluiert werden muss (Scherres 2013, 42f.). Für eine solche individuelle Diagnose müssen sowohl Leistungspotenzial als auch der aktuelle Stand im Bearbeitungsprozess und die jeweiligen mathematischen Inhalte (also fachliche und fachdidaktische Aspekte des Themas) bekannt sein (Scherres 2013, 43).

Interventionen in die Gruppenarbeitsprozesse von Schülerinnen und Schülern können auch als das beobachtbare Verhalten von Lehrkräften (bzw. von Lehramtsstudierenden) im Sinne der diagnostischen Kompetenz angesehen werden (vgl. Abb. 2). Dieses beobachtbare Verhalten

sollte nach Auffassung verschiedener Autoren möglichst abgestimmt auf die Bedürfnisse der Schülerinnen und Schüler und somit möglichst adaptiv sein (Leiss 2007, Stender 2016). Wie bereits beschrieben, sieht Leiss (2007, 82) adaptive Lehrerinterventionen als Hilfestellungen der Lehrperson an, die den individuellen Lern- und Lösungsprozess von Schülerinnen und Schülern minimal unterstützen, sodass sie maximal selbstständig arbeiten können. Eine adaptive Lehrerintervention stellt dabei auf Grundlage von Wissen und/oder einer Diagnose der Lehrperson einen inhaltlich und methodisch angepassten minimalen Eingriff in den individuellen Lösungsprozess des Schülers dar, wodurch dieser in die Lage versetzt wird, eine (potenzielle) Lernbarriere zu überbrücken und selbstständig weiterzuarbeiten (Leiss 2007, 65ff.). Diesen Überlegungen liegt das von Aebli formulierte Prinzip der minimalen Hilfe zugrunde: „Bei alledem gilt das Grundprinzip, dass der Lehrer dem selbstständigen Nachdenken der Schüler solange seinen Lauf lässt, als sie auf dem Wege der Lösung des Problems weiterkommen. Aber auch wenn sie Hilfe brauchen interveniert er nicht auf massive Weise“ (Aebli 1997, 300). Dieses Prinzip der minimalen Hilfe wird von Zech (2002) durch sein fünfstufiges Handlungsmodell für Lehrpersonen präzisiert (Stender 2016, 78). Dabei erläutert Zech (2002, 315ff.) die verwendeten Kategorien von Interventionen folgendermaßen:

- *Motivationshilfen* ermutigen Schülerinnen und Schüler dazu, weiter an der Aufgabe zu arbeiten, ohne dabei einen Bezug zur konkreten Aufgabe zu haben. Diese Form der Hilfe kann also unverändert bei verschiedenen Fragestellungen und auch in anderen Unterrichtsfächern verwendet werden.
- *Rückmeldehilfen* nehmen Bezug auf den konkreten Arbeitsstand und ggf. auch auf den Arbeitsverlauf der Schülerinnen und Schüler. Sie geben diesen Informationen darüber, ob sie auf dem richtigen Weg sind.
- *Allgemein-strategische Hilfen* können fachübergreifend oder fachspezifisch sein. Sie sollen den Schülerinnen und Schülern helfen, ihren eigenen Arbeitsprozess zu steuern, ohne dabei auf die Fachinhalte einzugehen.
- *Inhaltsorientiert-strategische Hilfen* sind Hilfen, die den Lernenden eine Problemlösemethode vermitteln, die überwiegend im Fach Mathematik angewendet wird und auf den Inhalt der zugrundeliegenden Aufgabe abzielt.
- *Inhaltliche Hilfen* sind Hilfen, die den Schülerinnen und Schülern konkrete inhaltliche Informationen geben. Dies kann beispielsweise der nächste Arbeitsschritt sein oder auch das zu verwendende mathematische Verfahren. Darüber hinaus können auch (Teil-)Ergebnisse vorgegeben werden.

Grundsätzlich gilt dabei, dass zunächst möglichst wenig (inhaltliche) Hilfe gegeben werden sollte und den Lernenden somit ermöglicht wird, anhand der Aufgaben und der Lernumgebung selbstständig die mathematischen Inhalte und Beziehungen zu entdecken. Zech (2002) sieht es als wichtige Vorbereitungsaufgabe von Lehrkräften an, sich mit möglichen Hilfen jeder Art auseinanderzusetzen, um sie im Problemlöseunterricht je nach Bedarf variabel anwenden zu können.

Stender (2016, 159f.) weist darauf hin, dass sowohl die Intervention an sich, als auch die Zeit vor bzw. nach einer Intervention zur Einschätzung ihrer Wirkung betrachtet werden muss. Wie lange dieser Zeitraum zu wählen ist, kann nicht eindeutig beantwortet werden, da die Wirkung der Intervention unmittelbar nach der Intervention oder aber erst mit größerem zeitlichem

Abstand auftreten kann (ebd.). Sowohl Stender (2016) als auch Leiss (2007) nutzen die Qualitative Inhaltsanalyse, um die jeweiligen Interventionen zu analysieren. Stender (2016, 179) greift dabei die Hilfen nach Zech (2002) auf und ergänzt diese um „organisatorische Hilfen“ und „sichtbares Zuhören“. Um über den Erfolg einer Intervention zu entscheiden, wird kodiert, ob die Hilfe beachtet wurde, ob die Schülerinnen und Schüler nach der Hilfe weitergearbeitet haben und ob es Nachfragen gab (Stender 2016, 172). Außerdem wird kodiert, ob es sich um eine invasive oder um eine responsive Intervention handelte (Stender 2016, 177).

Eine weitere Möglichkeit der Analyse von Interventionen in Gruppenarbeiten findet sich bei Leiss (2007), der auf die von ihm übersetzte Arbeit von Serrano (1996) hinweist. Dabei beschreibt er folgende Kategorien von Lehrerhandeln:

- *Beobachtung*
 - Reine Beobachtung
 - Beobachtung verbunden mit Kommentaren zum Lernfortschritt
 - Beobachtung verbunden mit aufgabenbezogenen Bemerkungen
- *Eins-zu-Eins Interaktion*
 - Disziplinbezogen
 - Unterstützend
 - Erklärend / Vormachend
 - Hinweise gebend
 - Zum Denken anregend
 - Evaluierend
 - Organisatorisch oder Motivierend
- *Interaktion mit der gesamten Lerngruppe*
 - Mit Bezug zur Aufgabe oder Situation
 - Ohne Bezug zur Aufgabe oder Situation

Einige der Kategorien finden sich auch im Modell von Zech (2002). So würden Kommentare zum Lernfortschritt unter „Rückmeldehilfen“ und motivierende Aussagen unter „Motivationshilfen“ fallen. Andere Kategorien können mehrere Hilfen ansprechen, beispielsweise könnten Interaktionen mit der gesamten Lerngruppe mit Bezug zur Aufgabe oder Situation inhaltsorientiert-strategische Hilfen oder inhaltliche Hilfen beinhalten. Ebenso könnten den Interaktionen mit der gesamten Lerngruppe ohne Bezug zur Aufgabe oder Situation alle anderen Hilfen subsumiert sein. Darüber hinaus findet sich mit den disziplinierenden Hilfen auch mindestens ein neuer Aspekt, welcher bei Zech (2002) nicht zugeordnet werden kann. Während Stender (2016) auch die organisatorischen und evaluierenden Aspekte als neue Aspekte ansieht, könnte man die organisatorischen Aspekte auch in einem weiteren Verständnis den allgemein-strategischen Hilfen und die evaluierenden Aspekte den Rückmeldehilfen zuordnen.

Das vorgestellte Modell wird von Leiss (2007, 79ff.) im Rahmen seiner theoretischen Auseinandersetzung mit Lehrerinterventionen zu vier Ebenen von Interventionen zusammengefasst:

- *Organisatorische Interventionen* zielen auf die Gestaltung der Lernumgebung
- *Affektive Interventionen* sollen die Motivation verbessern
- *Strategische Interventionen* beinhalten Hilfestellungen auf der Metaebene

- *Inhaltliche Informationen* beziehen sich konkret auf die Inhalte der zu bearbeitenden Problemstellung

Darüber hinaus hält Leiss (2007, 81f.) weitere Eigenschaften für relevant, welche nachfolgend vorgestellt werden. Die formale Äußerungsabsicht erfasst, ob es sich um eine Aussage, Frage oder Aufforderung handelt. Bei der prozessbezogenen Äußerungsabsicht hingegen wird erfasst, ob eine Diagnose stattgefunden hat, ob es sich um eine Bewertung handelt oder ob ein Hinweis gegeben wurde. Des Weiteren wird die Länge bzw. die Dauer der Intervention ebenso wie die Bezugsebene (problembezogen, beispielbezogen, allgemein) betrachtet. Die Repräsentationsform (verbal, ikonisch, gestisch, materiell), der Adressat (Einzelperson, Gruppe, Klasse) und die Häufigkeit (eine oder mehrere Interventionsimpulse) stellen ebenfalls Analysemöglichkeiten dar. Außerdem wird erfasst, ob es sich um invasive oder responsive Interventionen handelt (Leiss 2007, 105f.).

Die verschiedenen Interventionsmöglichkeiten finden unter anderem auch im Mathematik-Labor „Mathe ist mehr“ der Universität Koblenz-Landau Anwendung, wenn Studierende in die Gruppenarbeitsprozesse von Schülergruppen intervenieren. Eine detailliertere Darstellung des Mathematik-Labors erfolgt in Kapitel 5. Im nächsten Teilkapitel werden nun noch einige empirische Ergebnisse vorgestellt, bevor das Kapitel überblicksartig zusammengefasst wird.

3.5 Bisherige Forschung und empirische Befunde

Seidel und Shavelson (2007) belegen in ihrer Meta-Analyse, dass aktive Einbindung (der Lernenden in den Unterricht), kognitive Aktivierung, Rückmeldungen (an die Lernenden) und das Überwachen von Lernfortschritt und Verstehensprozessen bedeutende Effekte auf den Lernfortschritt (und darüber hinaus auch auf motivational-affektive Merkmale) der Lernenden haben (Krammer 2009). Darüber hinaus stellt Link (2011, 213) fest, dass strategische Interventionen das selbstständige Arbeiten von Schülerinnen und Schülern in Problemlöseprozessen fördern können.

Kobarg (2009, 20ff.) listet in ihrer Arbeit verschiedene Ergebnisse von Metaanalysen auf und stellt dabei unter anderem fest, dass die Art der Fragen, die den Schülerinnen und Schülern gestellt werden, die Reaktion der Lehrperson auf Schülerantworten und die Rückmeldungen der Lehrperson bedeutsam für die Schülerleistung sind. Darüber hinaus haben auch die positive Verstärkung von Schülerinnen und Schülern, das kooperative Lernen im Unterricht sowie personalisierte und adaptive Instruktionen eine deutliche Wirkung auf das Lernen von Schülerinnen und Schülern (ebd.). Die Bedeutsamkeit adaptiver Instruktionen für das Lernen von Schülerinnen und Schülern wurde dabei gleich in mehreren Metastudien belegt (Kobarg 2009, 20ff.).

Dekker und Elshout-Mohr (2004, 62) untersuchen Lehrerinterventionen in Kleingruppen und stellen fest, dass Schülerinnen und Schüler, deren Lehrperson nicht inhaltlich eingriff, bei einem anschließenden Mathematiktest besser abschneiden und dass gleichzeitig die Streuung der Punktezahl geringer ist. Ähnliche Ergebnisse liefert Fürst (1999, 143ff.), der unter anderem herausfand, dass Lehrerinterventionen, die nur in geringem Maße adaptiv sind, zu einer Verschlechterung der Gruppenarbeit bzw. der inhaltlichen Progression führen. Dies liegt unter anderem an einer fehlenden Gruppenorientierung und zu geringem Aufgabenbezug während der Intervention (Fürst 1999, 145).

Seifried und Klüber (2006, 157ff.) untersuchen die Häufigkeit der Lehrerinterventionen hinsichtlich des Auslösers der Interventionen (lehrerinitiiert vs. lernerinitiiert) sowie die Qualität

der Interventionen. Dazu greifen sie auf Videomaterial zurück, welches mehrere Unterrichtsstunden von ganzen Klassen zeigt (Seifried & Klüber 2006, 155). Aus diesem Filmmaterial greifen sie sich insgesamt fünf Unterrichtseinheiten für ihre Analysen heraus (ebd.). In Bezug auf die Auslöser der Interventionen stellten sie fest, dass etwa 70% der getätigten Interventionen von der Lehrkraft ausgehen und entsprechend etwa 30% von der Lerngruppe. Daraus schlussfolgern sie, dass es Lehrkräften schwerfällt, die Lernprozesse den Betroffenen selbst zu überlassen (Seifried & Klüber 2006, 158). Durch eine Vielzahl von Interventionen werden nach ihrer Ansicht die Förderung der Selbstständigkeit sowie die Eigenverantwortlichkeit der Lernenden gestört (ebd.). Die Lernenden stellen sich auf das Kontrollbedürfnis der Lehrpersonen ein und fragen nach, ob der Arbeitsauftrag richtig verstanden wurde, ob die Vorgehensweise zielführend ist oder ob technische bzw. inhaltliche Hilfen gegeben werden können. Bezüglich der Qualität der Lehrerinterventionen vermuten Seifried und Klüber (2006, 159), dass Lehrpersonen im Rahmen ihrer Interventionen eigene Vorstellungen durchsetzen, ohne sich an den spezifischen Bedingungen der jeweiligen Lerngruppe zu orientieren. Jedoch zeigt sich in ihrer Studie, dass unterschiedliche Lerner unterschiedliche Anforderungen an Lehrerinterventionen stellen – was in einer Gruppe positiv wahrgenommen wird, kann in einer anderen Gruppe kontraproduktiv sein und ggf. sogar den Lernprozess behindern (Seifried & Klüber 2006, 161).

Fürst und Haag (1998) untersuchen die Lehrer-Schüler-Interaktion im Gruppenunterricht. Dazu wurden jeweils 10- bis 15-minütige Gruppenarbeiten von Hauptschülerinnen und -schülern der 5. bzw. 6. Jahrgangsstufe ausgewählt, die möglichst kommunikativ waren und ungefähr im mittleren Leistungsbereich lagen (Fürst & Haag 1998, 18). Insgesamt wurden auf diese Weise 40 Gruppenarbeiten, in denen 111 Lehrerinterventionen ermittelt wurden, analysiert (Fürst & Haag 1998, 66). Etwa 70% der Interventionen waren invasiv, die durchschnittliche Dauer der Interventionen betrug lediglich 23 Sekunden (ebd.). Es gab nur sieben Interventionen, die länger als eine Minute dauerten. Es gab Gruppen, bei denen keine Interventionen getätigt wurde, das Maximum lag bei 10 Interventionen pro Gruppenarbeit (Fürst & Haag 1998, 66f.). Zeitlich längere Interventionen führten zu schlechteren Arbeitsergebnissen bei der jeweiligen Schülergruppe (Fürst & Haag 1998, 61). Die Hypothesen, dass invasive Lehrerinterventionen sich negativ und responsive Lehrerinterventionen sich positiv auf die Leistung der Schülerinnen und Schüler auswirken, konnten nicht bestätigt werden (ebd.).

Hinsichtlich der Interventionsauslöser kommt Stender (2016, 191ff.) zu den gleichen Ergebnissen wie Fürst und Haag (1998), nämlich, dass die invasiven Interventionen überwiegen (Stender 2016, 193). Stender (2016, 191ff.) hält außerdem fest, dass es im Rahmen seiner Studie deutliche Unterschiede zwischen den verschiedenen Tutorentams in Bezug auf die Häufigkeit der Interventionen gibt. Darüber hinaus variiert die zeitliche Länge der einzelnen Interventionen (ebd.). Motivationshilfen kommen kaum vor und auch Rückmeldehilfen wurden nur selten kodiert (Stender 2016, 198). „Sichtbares Zuhören“ tritt nur in einzelnen Tutorentandems nennenswert auf (ebd.). Betrachtet man den Erfolg der Interventionen, stellt man fest, dass deutlich mehr (105) Interventionen als erfolgreich angesehen wurden als als nicht erfolgreich (45) (Stender 2016, 199). Die am häufigsten auftretenden Interventionsarten waren strategische Hilfen und inhaltlich-strategische Hilfen (Stender 2016, 198).

Damit kommt er zu gänzlich anderen Ergebnissen als Leiss (2007, 281), der herausfindet, dass strategische Interventionen im Interventionsrepertoire der Lehrpersonen so gut wie gar nicht vorhanden sind. Ebenso ergab sich aus seiner Analyse, dass (1) Lehrkräfte Lernfortschritte regelmäßig durch ein kurzes positives Feedback kommentieren, (2) versucht wird inhaltliche Schwierigkeiten im Lösungsprozess durch Hilfestellungen auf inhaltlicher Ebene zu lösen

und (3) Lehrpersonen bei inhaltlichen Problemen Hilfen anbieten, die nur noch einen kleinen Schritt der Selbstständigkeit verlangen, um das Problem zu lösen (Leiss 2007, 281).

Weitere Untersuchungen zum Verhalten der Lehrperson in Schülerarbeitsphasen beschreibt Krammer (2009, 56ff.).

3.6 Zusammenfassung und Übersicht

Die Begriffe „Lehrerimpulse“, „Scaffolding“ und „Lehrerinterventionen“ haben unverkennbar eine gewisse Schnittmenge. Lehrerimpulse als Denkanstöße, die die Schülerinnen und Schüler dazu auffordern, über einen bestimmten Unterrichtsinhalt nachzudenken oder sich in etwas hineinzuversetzen, sind recht weit gefasst und entsprechend begrifflich noch zu wenig präzisiert. So können sowohl ein Großteil der didaktischen Handlungen im Rahmen des Scaffoldings als auch die Lehrerinterventionen als Lehrerimpulse angesehen werden.

Scaffolding erscheint zunächst etwas präziser, durch die verschiedenen und vielzähligen Verwendungen des Begriffes jedoch auch unscharf (van de Pol 2010). Grundsätzlich haben sowohl das Scaffolding als auch Lehrerinterventionen in der Regel den Anspruch, möglichst adaptiv zu sein (vgl. dazu u.a. Wood et al. 1976; Krammer 2009; van de Pol 2010; Leiss 2007; Leiss & Tropper 2014; Stender 2016). Entsprechend gibt es auch eine gewisse Übereinstimmung beider Begriffe. Scaffolding ist jedoch breiter und langfristiger angelegt, was sich u.a. am Fading und dem Übertrag der Verantwortung von Lehrkraft auf Lernenden erkennen lässt. Dieser Begriff wäre für die vorliegende Studie zu weit gefasst, weshalb mit dem Begriff Lehrerinterventionen bzw. Interventionen in Gruppenarbeitsprozesse (von Schülerinnen und Schülern) gearbeitet wird.

Das zentrale Element dieser Interventionen (im Sinne von Leiss 2007, Leiss & Tropper 2014) ist die Adaptivität. Die Zielsetzung, Schülerinnen und Schüler minimal zu unterstützen, sodass diese möglichst selbstständig weiterarbeiten können, findet sich bereits in Aebli's Prinzip der minimalen Hilfe (Aebli 1997, 300) und ist weitgehend anerkannt. Das fünfstufige Handlungsmodell für Lehrpersonen von Zech (2002) präzisiert dieses Prinzip, indem konkrete Hilfsarten genannt und beschrieben werden. Mittels dieser Hilfsarten kann der Zeitraum, in dem die Lehrkraft den Lernenden Hilfen gibt, analysiert werden. Manche Autoren nutzen die Hilfsarten auch in abgeänderter Form (Leiss 2007; Leiss & Tropper 2014; Stender 2016).

Die empirischen Ergebnisse zeigen, dass die Art der Hilfe durchaus entscheidend für den Lernerfolg von Schülerinnen und Schülern sein kann (Krammer 2009, Seidel & Shavelson 2007, Kobarg 2009, Dekker & Elshout-Mohr 2004, Fürst 1999, Leiss 2007, Stender 2016). Bezüglich der überwiegend verwendeten Interventionsarten gibt es unterschiedliche Befunde (Leiss 2007, Stender 2016). Dies ist insofern zu erklären, als einerseits unterschiedliche Schülerinnen und Schüler unterschiedliche Voraussetzungen und Bedürfnisse haben und andererseits unterschiedliche Lehrkräfte unterschiedliche Fähigkeiten und Präferenzen besitzen. Außerdem kann das gewählte Setting in der jeweiligen Studie einen (systematischen) Einfluss haben. Bezüglich der Unterscheidung zwischen invasiven und responsiven Interventionen legen einige Studien nahe, dass Lehrkräfte überwiegend invasiv intervenieren (Fürst & Haag 1998; Seifried & Klüber 2006; Leiss 2007; Stender 2016).

Sowohl diagnostische Fähigkeiten als auch die Fähigkeit, adäquat in Gruppenarbeitsprozesse von Schülerinnen und Schülern zu intervenieren, können im Rahmen des Lehrerdaseins stetig (weiter-)entwickelt werden. Dazu ist insbesondere auch eine geeignete Reflexion unabdingbar. Daher werden nun im Folgenden der Begriff der Reflexion, verschiedene Modelle zur Erfassung von Reflexion sowie empirische Befunde im Bereich Reflexion vorgestellt.

4 Reflexion

Nach Clarà (2015) besteht sowohl Übereinstimmung darüber, dass Reflexion einen wichtigen Aspekt der Lehrerbildung darstellt, als auch darüber, dass es keine Klarheit darüber gibt, was Reflexion genau ist. Aufgrund dieser Unklarheit sollen nun zunächst einige unterschiedliche, aber teilweise aufeinander aufbauende Reflexionsverständnisse verschiedener Autoren vorgestellt werden. Anschließend wird auf verschiedene Methoden zur Erfassung von Reflexionsmomenten eingegangen und empirische Ergebnisse vorgestellt.

4.1 Begriffsklärung

Reflexion (lat.) bedeutet soviel wie ‚zurückbeugen‘ und kann daher im übertragenen Sinne als eine Position oder Haltung betrachtet werden, die es ermöglicht, Dinge von einem anderen Standpunkt oder aus einem anderen Winkel zu betrachten (Hilzensauer 2008). Dewey gilt bzgl. der Reflexion als einflussreichste Person der modernen Zeit (Boud, Keogh & Walker 1985, 11). Er geht vom Prinzip des forschenden Lernens aus und bezieht dabei den praktischen Handlungsprozess sowie die soziale Kommunikation mit ein (Hilzensauer 2008). Dewey (1933) stellt fünf Stufen des erziehenden Unterrichts dar:

- (1) Primäre Erfahrungen durch praktische Tätigkeit
- (2) Reflexion der dabei erfahrenen Probleme
- (3) Erkundung der Materialien zur Problemlösung
- (4) Formulierung von Hypothesen zur Problemlösung
- (5) Überprüfung der Hypothesen durch praktisches Handeln

Dabei zielt die Reflexion nicht auf Wissen, sondern auf Annahmen über Dinge oder Zusammenhänge ab (Dewey 1933, 3f.). Der Reflexion muss dabei jeweils eine (Handlungs-)Blockade vorausgehen (Dewey 1933, 3f.). Dewey kontrastiert die Reflexion (als bestimmte Form des Denkens) mit einer Form des Denkens, welche unsystematisch, unkritisch und unbegründet ist und daher möglicherweise falschen Annahmen unterliegt (Dewey 1933, 16). Seine Handlungsanalysen konzentrieren sich hauptsächlich auf die Aspekte Routine und Reflexion (Roters 2012, 115). Routinen sollten hinterfragt werden, da diese hauptsächlich durch Tradition, Autorität und impulsives, unreflektiertes Handeln entstehen (ebd.). Wichtig sind daher die Offenheit und Aufgeschlossenheit, sich in einem aktiven und systematischen Prozess der Problemlösung zu nähern und eine kritisch-distanzierte Haltung gegenüber Autoritäten oder tradiertem Wissen entwickeln zu können (Roters 2012, 116).

Ein weiterer bekannter Vertreter im Bereich der Reflexion ist Donald Schön, der teilweise auf die Überlegungen Deweys aufbaut (Hilzensauer 2008). Er unterscheidet zwischen *reflection-in-action* und *reflection-on-action* (Schön 1983). *Reflection-in-action* bezeichnet die unmittelbare Bewertung der Aktion, während diese getätigt wird. *Reflection-on-action* bezeichnet entsprechend das Reflektieren nach einer Aktion. In beiden Fällen ist es wichtig, dass die eigenen Erfahrungen mit den Gefühlen gekoppelt werden und eine theoriegeleitete Reflexion stattfindet (Schön 1983). In seinen Überlegungen zum „reflective practitioner“ konzentriert Schön sich auf die Fähigkeiten eines Praktikers, sein implizites Wissen über Erfahrungen, Gefühle und Theorien im praktischen Handeln zu nutzen (Schön 1983; Hilzensauer 2008).

Boud, Keogh & Walker (1985) vereinen theoretische Kenntnisse verschiedener Reflexionsmodelle und bieten darüber hinaus eine Erklärung über interne (Gedanken-)Prozesse, die beim

Reflektieren ablaufen. Vorangegangene Erfahrungen stellen dabei, wie auch bei Dewey (1933), die zentrale Voraussetzung für das Lernen dar. Sie unterscheiden den Reflexionsprozess in drei Stufen (Boud et al. 1985, 20ff.). Zunächst muss eine konkrete Handlung durchgeführt werden. Anschließend führt man sich diese und insbesondere auch die Gefühle, die man in der Situation durchlebt hat, vor Augen (ebd.). Abschließend findet eine Neubewertung der Erfahrung statt, bei der die positiven Gefühle in der Situation genutzt werden und die negativen Gefühle ausgeblendet werden sollen (ebd.). Ziel ist die Nutzung positiver Erfahrungen für zukünftige Handlungen. Dazu ist die Trennung von Erfahrungen und Gefühlen während der Erfahrung notwendig (ebd.). Einerseits muss eine möglichst detaillierte und objektive Auseinandersetzung mit der Erfahrung stattfinden, andererseits müssen die positiven Erfahrungen identifiziert und mit dem Lernprozess in Zusammenhang gebracht werden (ebd.).

Gibbs (1988) bezieht sich in seiner Ausführung auf Dewey (1933) und Boud et al. (1985). Er betont, dass Erfahrungen reflektiert werden müssen, da diese ohne die Reflexion darüber schnell vergessen sind und dann keine Schlüsse mehr daraus gezogen werden können. Er beschreibt einen sechsstufigen Reflexionszyklus, welcher folgende Schritte beinhaltet (Gibbs 1988):

- (1) *Description* – Ausführliche Situationsbeschreibung
- (2) *Feelings* – Beschreibung der Gefühle in der Situation
- (3) *Evaluation* – Beurteilung der Situation in positive und negative Aspekte
- (4) *Analysis* – Beobachtung einzelner Komponenten der Situation
- (5) *Conclusion* – Zusammenschau und Handlungsalternativen
- (6) *Action Plan* – Beschreibung künftiger Handlung in ähnlichen Situationen

Im ersten Schritt werden Erfahrungen möglichst detailliert beschrieben. Darauf aufbauend werden die Gefühle in den jeweiligen Situationen dargelegt. Damit trennt Gibbs (1988) die Erfahrungen und die dabei empfundenen Gefühle voneinander, wie es auch bei Boud et al. (1985) der Fall ist. Auf dieser Grundlage wird eine Beurteilung der Situation vorgenommen, wobei positive und negative Aspekte voneinander unterschieden werden. Anschließend werden einzelne Komponenten der Situation betrachtet. In der Zusammenschau werden dann wesentliche Aspekte herausgearbeitet und auf dieser Grundlage künftige Handlungsoptionen abgeleitet. Im letzten Schritt sollen diese ausführlich beschrieben werden und somit eine Optimierung der Handlungen in zukünftigen, ähnlichen Situationen stattfinden. Ziel ist dabei das autonome Denken und Entscheiden.

Beim sechsstufigen Reflexionszyklus von Gibbs (1988) geht es um die Reflexionstätigkeit an sich, wobei nicht darauf eingegangen wird, wie genau sich die Reflexionsfähigkeit entwickelt. Dem widmen sich King und Kitchener (1994, 13ff.), indem sie die Entwicklung von Reflexion in ihrem „reflective judgement model“ beschreiben und darin sieben Stufen identifizieren:

- *Stufe 1*: Wissen ist vermeintlich konkret und wird nicht als abstrakt verstanden. Es kann mit Sicherheit durch direkte Beobachtung erhalten werden. Überzeugungen bedürfen keiner Rechtfertigung, sondern werden als wahr angenommen.
- *Stufe 2*: Wissen wird als sicher vorausgesetzt und kann durch die Sinne oder durch Autoritätspersonen erlangt werden. Überzeugungen sind ungeprüft und höchstens durch

eine Autoritätsperson gerechtfertigt. Es wird angenommen, dass die meisten Fragen genau eine richtige Antwort haben.

- *Stufe 3:* Wissen ist entweder sicher oder nur vorübergehend unsicher. Falls es unsicher ist, können nur eigene Überzeugungen genutzt werden, bis absolutes Wissen (durch Autoritätspersonen) erlangt wird. In Gebieten, in denen Antworten vorhanden sind, werden Überzeugungen durch Autoritäten gerechtfertigt. In Gebieten, in denen es keine Antwort gibt, werden Überzeugungen als persönliche Meinung verteidigt.
- *Stufe 4:* Wissen ist unsicher und Wissensansprüche sind eigenartig, da situationsbedingte Variablen vorgeben, dass Wissen immer mehrdeutig ist. Überzeugungen werden durch die Angabe von Gründen und die Verwendung von Anhaltspunkten gerechtfertigt, aber die Argumente und die Beweisauswahl sind eigenartig.
- *Stufe 5:* Wissen ist kontextuell und subjektiv, da es durch Wahrnehmungen und Beurteilungskriterien einer Person gefiltert wird, weshalb lediglich Interpretationen von Ereignissen oder Problemen bekannt sein können. Überzeugungen werden durch Untersuchungsregeln und kontextspezifische Interpretationen von Anhaltspunkten gerechtfertigt.
- *Stufe 6:* Wissen wird auf der Grundlage von Informationen aus verschiedenen Quellen zu individuellen Schlussfolgerungen über unstrukturierte Probleme aufgebaut. Interpretationen basieren auf Bewertungen von Anhaltspunkten über verschiedene Kontexte hinweg und auf Einschätzungen von angesehenen Personen. Überzeugungen werden gerechtfertigt, indem Anhaltspunkte und Meinungen zu einem Thema aus verschiedenen Perspektiven verglichen und Lösungen erarbeitet werden.
- *Stufe 7:* Wissen ist das Ergebnis eines angemessenen Forschungsprozesses, welcher Lösungen für unstrukturierte Probleme liefert. Die Angemessenheit dieser Lösungen wird evaluiert, indem die wahrscheinlichste und plausibelste Erklärung durch den aktuellen Forschungsstand angenommen wird. Es findet eine erneute Evaluation statt, wenn neue Informationen oder Perspektiven zugänglich sind. Überzeugungen werden nach Wahrscheinlichkeit auf Grundlage einer Vielzahl von interpretatorischen Erwägungen begründet, wie etwa dem Gewicht der Anhaltspunkte, dem Erklärungsgehalt der Interpretationen, dem Risiko fehlerhafter Schlussfolgerungen, den Folgen alternativer Urteile und den Zusammenhängen dieser Faktoren. Schlussfolgerungen werden als das umfassendste, plausibelste oder überzeugendste Verständnis eines Themas auf Grundlage der verfügbaren Informationen verteidigt.

Die einzelnen Stufen werden in drei Bereiche unterteilt. In den frühen Stufen 1,2 und 3 wird Wissen entweder direkt durch persönliche Beobachtung generiert oder durch eine Autoritätsperson erlangt. Personen, die sich in dieser (pre-reflective) Phase befinden, betrachten alle Probleme, als wären sie mit einem hohen Maß an Sicherheit und Vollständigkeit gegeben (King & Kitchener 1994, 16). In den mittleren Stufen 4 und 5 wird realisiert, dass Wissen oftmals auch Elemente des Unwissens beinhaltet. Außerdem wird zwischen leichten und schweren Problemen unterschieden (ebd.). Es wird typischerweise argumentiert, dass Urteile zwar auf Beweisen beruhen sollten, die Bewertung jedoch individualistisch sei. In dieser (quasi-reflective) Phase sind Personen daher mit der Lösung schlecht durchschaubarer Probleme überfordert (ebd.). Die Stufen 6 und 7 repräsentieren die fortgeschrittensten Arten des reflektierten Urteils.

Es wird wahrgenommen, dass das Verstehen der Welt bzw. einzelner Situationen bzw. Probleme nicht gegeben ist, sondern aktiv konstruiert werden muss (King & Kitchener 1994, 17). Wissen muss daher immer in dem Kontext, in dem es entstanden ist, gesehen werden. Mehr Wissen und mehr Urteile führen insgesamt zu einem plausibleren Urteil, welches jedoch auf Basis neuer Informationen stetig neu evaluiert werden muss (ebd.). Diese Phase des Reflektierens nennen King und Kitchener (1994, 17) „truly reflective“.

Neben der Einteilung in verschiedene Entwicklungsstufen, kann Reflexion auch in verschiedene Ebenen unterteilt werden. Nach Siebert (1991), der Lernen vor allem aus der Perspektive der Erwachsenenbildung beschreibt, kann Reflexion in folgende drei Ebenen unterschieden werden: Selbstreflexion, Gruppenreflexion und Problemreflexion. Dabei ist Reflexion für ihn die „Fähigkeit des Menschen, Bedingungen und Wirkungen eigenen Denkens und Handelns zu durchschauen“ (Siebert 1991). Selbstreflexion bezieht sich auf eigenes Tun und Handeln, auf eigene Stärken und Schwächen, auf die Evaluation des eigenen Lernprozesses und beinhaltet auch Selbstkritik (ebd.). Eine Gruppenreflexion ist eine Metakommunikation innerhalb einer Gruppe über einen gemeinsamen Lernprozess, bei dem sowohl Inhalts- als auch Beziehungsaspekte angesprochen und diskutiert werden (ebd.). Die Problemreflexion wird durch einen Impuls initiiert und es werden individuell und/oder gesellschaftlich relevante Themen reflektiert (ebd.).

Das in allen Reflexionsformen ausschlaggebende Kriterium, welches Reflexion von anderem Denken unterscheidet, ist die Umwandlung einer inkohärenten Situation in eine kohärente Situation (Clarà 2015, 262ff.). Hartkens (2016, 17f.) identifiziert vier zentrale Merkmale dieser besonderen Denkform, der Reflexion:

- *Rückbezug auf vorherige Tätigkeiten oder Ereignisse*
- *Gewinnung neuer Erkenntnisse*
- *Distanzierung gegenüber dem Reflexionsgegenstand*
- *Bewusster Verlauf bzw. Prozess*

Reflexionen beziehen sich also jeweils auf vorherige Tätigkeiten oder Ereignisse, wobei der oder die Reflektierende sich gegenüber dem Reflexionsgegenstand distanziert. Das Ziel dieses bewussten Prozesses ist die Gewinnung neuer Erkenntnisse. Dies ist insbesondere dann relevant, wenn ähnliche oder gleichartige Situationen auch zukünftig auf die Person zukommen werden. Die vier zentralen Merkmale nach Hartkens (2016, 17f.) finden sich teilweise auch in verschiedenen Modellen wieder, die versuchen, den Vorgang der Reflexion hinsichtlich bestimmter Inhalte bzw. Kriterien zu erfassen.

4.2 Modelle zur Erfassung von Reflexion

Im Folgenden sollen ausgewählte Modelle zur Erfassung der Reflexion vorgestellt werden. Hatton und Smith (1995) untersuchen den Einfluss von Reflexionsstrategien zur Förderung der Reflexionsfähigkeit von 60 Lehramtsstudierenden in Sydney. Sie analysieren Lerntagebücher und unterschieden dabei vier Formen der Reflexion:

- *descriptive writing,*
- *descriptive reflection,*
- *dialogic reflection und*
- *critical reflection* (Hatton & Smith 1995).

Dabei handelt es sich bei der ersten Form, dem *descriptive writing*, eigentlich nicht um Reflexion, sondern vielmehr um eine reine Beschreibung von Ereignissen (Hatton & Smith 1995). *Descriptive reflection* ist also die erste Form der Reflexion, bei der versucht wird, Dinge zu begründen. Diese Begründungen beruhen dabei auf persönlichen Einschätzungen oder gelesener Literatur (ebd.). Bei der dritten Form, der *dialogic reflection*, führt man einen Diskurs mit sich selbst, um mögliche Gründe genauer zu untersuchen (ebd.). Bei der *critical reflection* werden schließlich die jeweiligen Begründungen in einen historischen, sozialen oder politischen Kontext eingebettet (ebd.).

Ein weiteres Modell zur Erfassung von Reflexion liefern Boud et al. (1985). Sie veranschaulichen Reflexion durch drei Kreise, die für die gesammelten Erfahrungen, den Reflexionsprozess und für Ergebnisse stehen (vgl. Abb. 11). Der Ausgangspunkt der Reflexion sind die Erfahrungen, die im Laufe des Lebens gesammelt wurden, aber auch die Beziehungen, in welchen diese Erfahrungen entstanden sind, sowie die Gedanken und Gefühle, die in den Situationen vorhanden waren (Boud et al. 1985). Im Reflexionsprozess wird auf die Erfahrungen zurückgegriffen, wobei der oder die Reflektierende sich zunächst an diese erinnern soll. Anschließend sollen positive Gefühle bezüglich des Lernens und des jeweiligen Themas genutzt und hemmende Gefühle nicht miteinbezogen werden (ebd.). Dazu ist es notwendig sich solche Erinnerungen ins Bewusstsein zu rufen und das Augenmerk auf die positiven Seiten zu richten. Dabei kann die Erwartung zukünftiger Vorteile, die sich aus der jeweiligen Situation ableiten lassen, für die Verarbeitung der Situationen hilfreich sein (ebd.). Das Entfernen hemmender Gefühle ist eine notwendige Voraussetzung dafür, rational über die Geschehnisse nachdenken zu können (ebd.). Beim Wiedergeben der Situation können auch Gefühle anderer ausgedrückt werden. Der wichtigste Schritt ist die Evaluation, bei der die Erfahrungen analysiert werden und aus den neu gewonnenen Erkenntnissen neue Handlungsoptionen und Verhaltensweisen entstehen (ebd.). Diese können zunächst mental durchdacht werden, bevor sie schließlich erprobt werden. Dadurch hilft die Reflexion dabei, sich auf künftige Situationen vorzubereiten. Die jeweilige Veränderung kann dabei klein oder groß sein, es können sich neue Perspektiven auf Erfahrungen ergeben oder Verhaltensweisen können sich ändern. Das Modell kann von einem Einzelnen oder in einer Gruppe durchlaufen werden sowie angeleitet werden oder ohne Anleitung stattfinden (ebd.).

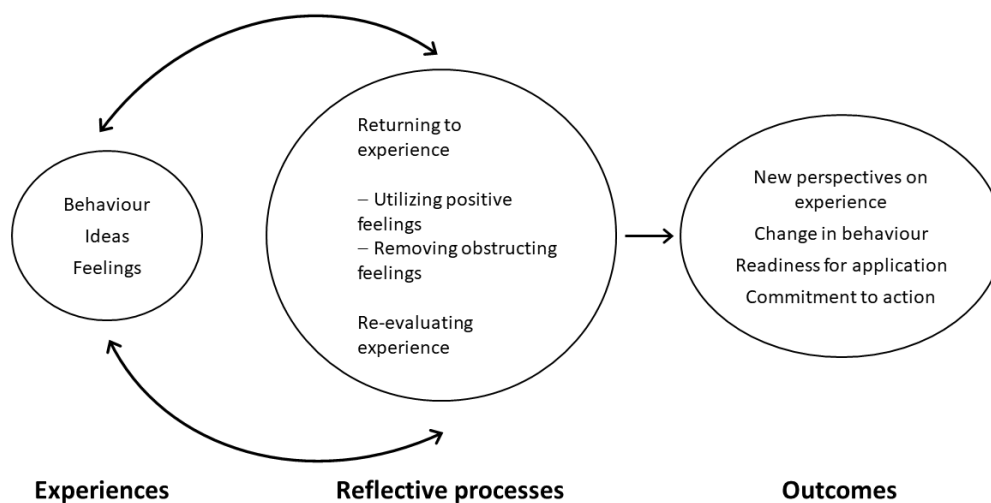


Abbildung 11: Reflexionsprozess nach Boud et al. (2015)

Auch Korthagen (1985) beschäftigt sich mit dem Lernen aus Erfahrung und beschreibt dies als zyklischen Prozess der Reflexion. In seinem Stufenmodell steht vor allem die eigene Wahrnehmung sowie das Aufbauen von bestimmten Denkweisen und Gefühlen bezüglich konkreter Unterrichtssituationen, an denen eine Lehrkraft selbst aktiv beteiligt war, im Vordergrund. Dieser Ansatz knüpft also (ebenfalls) direkt an eigene Erfahrungen und Vorstellungen an. Im Modell werden folgende fünf Phasen unterschieden:

- 1) *Action* (Handeln und Erfahrungen machen)
- 2) *Looking back on the action* (auf diese Erfahrungen zurückblicken)
- 3) *Awareness of essential aspects* (wichtige Aspekte dieser Erfahrung festhalten)
- 4) *Creating alternative methods of action* (nach alternativen Verhaltensweisen suchen)
- 5) *Trial* (ausprobieren der neuen Verhaltensweisen)

In Stufe 1 werden Handlungen getätigt und Erfahrungen gesammelt. Dies kann beispielsweise durch eine Mathematik-Stunde, welche von einem Studierenden gehalten wird, geschehen. Nun erfolgt in Stufe 2 ein Rückblick auf diese Stunde. Einzelne Situationen bzw. Aufgaben(teile) werden diskutiert und beispielsweise Reaktionen der jeweiligen Schülerinnen und Schüler auf die Handlung oder eigene Gefühle in der Situation besprochen. Außerdem kann eine erste grobe Beurteilung erfolgen, inwieweit die Stunde erfolgreich war und wie die Schülerinnen und Schüler sich verhalten haben. In Stufe 3 werden einige Situationen und Aspekte genauer betrachtet und im Zuge dessen die wesentlichen Aspekte herausgearbeitet. Hierbei sollte geklärt werden, inwieweit einzelne Aspekte zum Lernfortschritt der Schülergruppen beitragen konnten oder aus welchen Gründen dies nicht gelungen ist. Anschließend werden auf dieser Grundlage in Stufe 4 alternative Planungen durchleuchtet und alternative Handlungsmöglichkeiten gesucht. Hierbei sollten jeweils Vor- und Nachteile abgewogen werden. Daraus resultieren neue Handlungsansätze, die in Stufe 5 (bei der nächsten Konzeption einer Mathematikstunde) angewendet werden können. Dies bedeutet, dass bei einer erneuten Konfrontation mit einer ähnlichen Situation optimaler Weise die Ergebnisse der eigenen Reflexion genutzt werden, um das eigene Lehrerhandeln anzupassen und sich kontinuierlich zu verbessern (Korthagen 1985, Korthagen & Wubbels 2002). Da es sich um einen zyklischen Prozess handelt, werden die fünf Stufen auch folgendermaßen dargestellt (vgl. Abb. 12):

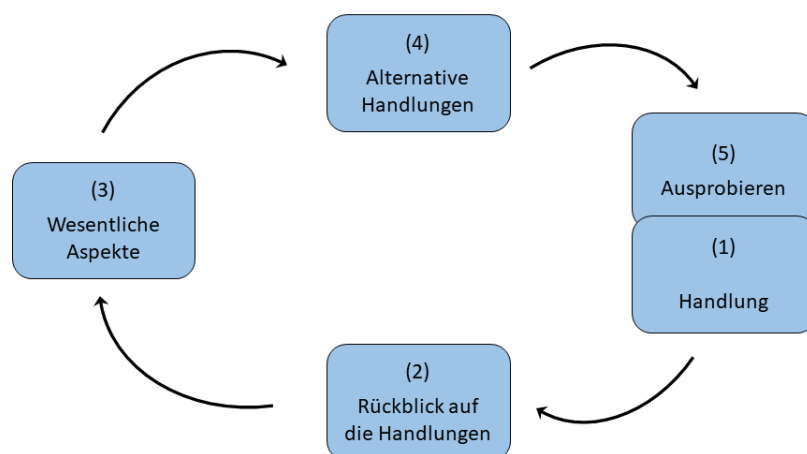


Abbildung 12: Zyklischer Prozess der Reflexion nach Korthagen (1985)

Durch diese Darstellungsweise wird ersichtlich, dass Stufe 1 und Stufe 5 sehr nahe beieinander liegen und zwar insofern, als jeweils eine Handlung durchgeführt wird. Der Unterschied besteht darin, dass das Ausprobieren in Stufe 5 in der Regel eine neue Handlungsoption ist, welche auf Grundlage der ursprünglich getätigten Handlung in Stufe 1 neu gewählt oder modifiziert wurde. Entsprechend dem zyklischen Prozess kann auch diese Handlung wieder als Ausgangspunkt eines neuen Reflexionsprozesses dienen.

Während das Modell von Hatton und Smith (1995) für schriftliche Reflexion ausgelegt ist, sind die Modelle von Boud et al (1985) und Korthagen (1985) tendenziell eher für eine mündliche Reflexion ausgelegt. Es wäre jedoch auch denkbar, das Modell von Hatton und Smith (1995) für eine mündliche Reflexion zu nutzen. Ebenso könnte man die anderen beiden Modelle auf die Reflexion von Lerngebüchern oder anderen schriftlichen Texten übertragen.

Sowohl bei Boud et al. (1985) als auch bei Korthagen (1985) sind (Unterrichts-)Erfahrungen der Ausgangspunkt für die Reflexion. Bei Hatton und Smith (1995) stehen verschiedene Reflexionsstrategien im Fokus. Alle Modelle gehen dabei auf die Gedanken und Gefühle ein. Bei Boud et al. (1985) sollen hemmende Gefühle aktiv ausgeblendet werden, was bei Korthagen (1985) keine Notwendigkeit darstellt, um den Reflexionsprozess zu durchlaufen. Vielmehr können diese zunächst wertfrei geäußert werden und ggf. sogar hilfreich sein, um wesentliche Aspekte der vorliegenden Situation herauszufiltern.

Im nächsten Schritt werden bei Boud et al. (1985) mögliche Handlungsalternativen durchdacht. Korthagen (1985) verweist hierbei auch auf die explizite Abwägung möglicher Vor- und Nachteile. Zwar gibt es auch bei Hatton und Smith (1995) den Auftrag, Dinge gegeneinander abzuwägen. Allerdings ist die Betrachtung in diesem Modell eher retrospektiv und weniger auf künftige Handlungen ausgelegt. Entsprechend endet das Modell auch damit, dass anhand von historischen, sozialen oder politischen Aspekten eine kritische Reflexion stattfindet. Bei den anderen beiden Modellen hingegen geht es darum, in einer künftigen ähnlich gelagerten Situation eine durchdachte Handlung durchzuführen.

Insgesamt sind sich die Modelle von Boud et al. (1985) und Korthagen (1985) trotz gänzlich unterschiedlicher Visualisierung inhaltlich sehr ähnlich. Größter Unterschied ist der Umgang mit negativen bzw. hemmenden Gefühlen, der bei Boud et al. (1985) Voraussetzung für den Reflexionsprozess ist, wohingegen dies bei Korthagen (1985) nicht der Fall ist. Solche und ähnliche Modelle werden genutzt, um Reflexionsmomente zu analysieren. Einige ausgewählte Ergebnisse solcher Studien werden nun vorgestellt.

4.3 Empirische Ergebnisse

Es gibt viele Studien, die das Konstrukt „Reflexion“ empirisch ermitteln wollen, indem Textformen (z.B. Interviews oder verschriftlichte Gruppendiskussionen) anhand verschiedener Kategorien (und ggf. Unterkategorien) analysiert werden (Roters 2012, 123). Entsprechend scheint die Qualitative Inhaltsanalyse grundsätzlich ein gängiges Verfahren zu sein, um das Reflexionsverhalten von (angehenden) Lehrkräften zu durchleuchten.

Gore und Zeichner (1991) stellen in ihrer Analyse von Aktionsforschungsprojekten an der Universität Wisconsin-Michigan fest, dass viele der berücksichtigten angehenden Lehrkräfte kein Verständnis von *reflective practice* zeigen und ihre Reflexion wenig kritisch ist. Dabei wurden 18 angehende Lehrkräfte untersucht.

Beim Vergleich eines reflexiven Praktikums im Vergleich zu einem herkömmlichen Praktikum stellt von Felten (2005, 165) fest, dass das reflektive Praktikum die Reflexion und die Entwicklung des eigenen Handelns in stärkerem Maße fördert als das herkömmliche Praktikum.

Dabei nahmen zunächst 25 und anschließend 55 angehende Lehrkräfte mithilfe eines Fragebogens eine Selbsteinschätzung vor und nach dem Praktikum vor. Sie wurden außerdem nach dem Praktikum von Lehrkräften eingeschätzt (im ersten Schritt von 25 Lehrkräften und anschließend von 32 Lehrkräften). Darüber hinaus wurden qualitative Interviews durchgeführt.

Davis (2006) untersucht in ihrer empirischen Studie schriftliche Reflexionen von angehenden Grundschullehrkräften. Dazu wertet sie Online-Portfolios und Reflexionsaufgaben von 25 Studierenden im 3. Semester aus, wobei diese über ihre eigenen Unterrichtsstunden reflektieren. Davis (2006) unterscheidet dabei zwischen produktiver und unproduktiver Reflexion. Produktiv ist Reflexion aus ihrer Sicht dann, wenn mindestens zwei der von ihr kodierten vier Kategorien („Lerner und Lernen“, „Fachwissen“, „Leistungsbeurteilung“ und „Unterricht“) miteinander in Verbindung gebracht werden. In 26% der Fälle liegt unproduktive Reflexion vor. In 33% der Fälle werden sogar alle vier verschiedenen Kategorien miteinander kombiniert (Davis 2006, 290). Daraus schlussfolgert sie, dass Studierende kontinuierlich dazu angeregt werden müssen, produktiv zu reflektieren.

Weitere Ergebnisse zur Reflexion in der Lehrerbildung, die jedoch keinen Mehrwert im Rahmen dieser Arbeit darstellen, werden von Roters (2012, 123ff.) zusammengetragen.

4.4 Zusammenfassung

Reflexion ist zwar in aller Munde und wird als wichtig angesehen, es besteht jedoch Uneinigkeit darüber, was genau Reflexion ist. Verschiedene Autoren haben über die letzten Jahrzehnte verschiedene Schwerpunkte gesetzt und das Reflexionsverständnis dadurch nach und nach erweitert. Einer der wichtigsten Vertreter ist Dewey (1933), der Reflexion von gesammelten Erfahrungen als Mittel zur Überwindung bestehender Handlungsblockaden ansieht. Schön (1983) unterscheidet insbesondere reflection-in-action und reflection-on-action. Außerdem fokussiert er auf die Fähigkeiten eines Praktikers, implizites Wissen über Erfahrungen, Gefühle und Theorien im praktischen Handeln zu nutzen. Boud et al. (1985) befassen sich unter anderem mit der Trennung von Erfahrungen und Gefühlen. Im Modell von Gibbs (1988) ist es entscheidend, dass während der Reflexion wesentliche Aspekte herausgearbeitet werden und auf dieser Grundlage mögliche Handlungsoptionen für zukünftige ähnlich gelagerte Situationen optimiert werden.

Neben verschiedenen Entwicklungsstufen gibt es auch verschiedene Ebenen der Reflexion. Im Rahmen der vorliegenden Studie sind dabei die Selbst- und die Gruppenreflexion relevant (vgl. dazu Kap. 8.4 und Kap. 9). Zur Erfassung von Reflexionsgesprächen kann unter anderem das Modell von Korthagen (1985) genutzt werden, welches auf konkreten Handlungserfahrungen aufbaut, die zunächst grob ins Gedächtnis gerufen werden, um anschließend wesentliche Elemente zu identifizieren und auf dieser Grundlage Handlungsoptionen für ähnliche Situationen in der Zukunft zu entwickeln.

In den vorgestellten empirischen Studien zeigt sich beispielsweise, dass Studierende wenig kritisch reflektieren (Gore & Zeichner 1991). Des Weiteren wurde gezeigt, dass Reflexion für die Entwicklung des eigenen Handelns förderlich ist (von Felten 2005). Außerdem wurde aufgezeigt, dass das Reflexionsverhalten häufig mit Hilfe der Qualitativen Inhaltsanalyse untersucht wird (Roters 2012; Davis 2006).

Diagnose, Intervention und Reflexion spielen im Rahmen dieser Arbeit insofern eine große Rolle, als sie im weiteren Verlauf den Untersuchungsgegenstand darstellen. Studierende sollen im Rahmen eines Lehr-Lern-Labor-Seminars an der Universität Koblenz-Landau unter anderem ihre diagnostischen Fähigkeiten abrufen, in Gruppenarbeitsprozessen von Schülerinnen

und Schülern intervenieren und darüber reflektieren. Dieses Seminar ist eng an das Mathematik-Labor „Mathe ist mehr“ in Landau angegliedert, welches im nächsten Kapitel vorgestellt wird.

5 Fachdidaktische Ausbildung im Fach Mathematik am Campus Landau

Neben den fachlichen und fachdidaktischen Inhalten, die in den verschiedenen Veranstaltungen im Rahmen des Bachelor- und Masterstudiums der Mathematik am Campus Landau der Universität Koblenz-Landau thematisiert werden, sollen die angehenden Lehrkräfte bereits im Rahmen der ersten Phase der Lehramtsausbildung und somit möglichst frühzeitig hinsichtlich ihrer diagnostischen Fähigkeiten und ihrer Fähigkeit, in Gruppenarbeitsprozesse von Schülerinnen und Schülern zu intervenieren, gefördert werden. Dies geschieht an der Universität Koblenz-Landau am Campus Landau unter anderem im Didaktischen Seminar, welches in Teilkapitel 5.2 vorgestellt wird. Um die Einbettung des Didaktischen Seminars in die mathematikdidaktische Ausbildung verstehen zu können, wird zunächst in Teilkapitel 5.1 das Gesamtkonzept des Mathematiklehramtsstudiums am Campus Landau erläutert.

5.1 Gesamtkonzept

Die Vernetzung der mathematikdidaktischen Ausbildungsinhalte (Kap. 5.1.1) wird unter anderem durch das Videodiagnosetool ViviAn (Kap. 5.1.2) unterstützt. Dabei nimmt auch das Mathematik-Labor „Mathe ist mehr“ eine entscheidende Rolle ein (Kap. 5.1.3). Entsprechend werden die einzelnen Bausteine, die schließlich ineinandergreifen sollen, in den nächsten Abschnitten vorgestellt.

5.1.1 Vernetzung der mathematikdidaktischen Ausbildungsinhalte

Die mathematikdidaktische Lehramtsausbildung für die Sekundarstufen ist an der Universität Koblenz-Landau am Campus Landau über das gesamte Bachelor- und Masterstudium verteilt und aufeinander aufbauend konzipiert (Bartel & Roth, 2017; Roth 2020). Nach einer einführenden Querschnittsvorlesung „Fachdidaktische Grundlagen“ im ersten Semester, sind die Didaktikveranstaltungen im Bachelorstudium (Didaktik der Geometrie, Didaktik der Zahlbereichserweiterungen, Didaktik der Algebra) sowie zu Beginn des Masterstudiums (Didaktik der Stochastik und, für Gymnasialstudierende, Didaktik der Analysis bzw. Didaktik der Analytischen Geometrie und Linearen Algebra) nach Inhalten des Mathematikunterrichts strukturiert. Eine Vernetzung aller Veranstaltungen findet am Ende des Masterstudiums im Rahmen des „Didaktischen Seminars“ statt. Hier bringen die Studierenden ihr gesamtes theoretisches Wissen aus dem bisherigen Studium sowie ihre ersten Praxiserfahrungen aus den Schulpraktika in die Konzeption und Gestaltung von Lernumgebungen für das Mathematik-Labor „Mathe ist mehr“ (Kap. 5.1.3) ein. Das Konzept des Didaktischen Seminars wird weiter unten (vgl. dazu Kap. 5.2.1) ausführlich dargestellt. In einer Wahlpflichtveranstaltung für Studierende des Lehramts an Gymnasien, dem „Fachdidaktischen Forschungsseminar“, werden die während der Labordurchläufe entstandenen Daten (Videos, gescannte Schülerbearbeitungen, Leistungstests) nach verschiedenen Kriterien und Fragestellungen sowie mit den jeweils dazu passenden empirischen Methoden ausgewertet (Roth 2020). Auf diese Weise wird neben der lernprozessbezogenen Diagnose eine ganze Reihe weiterer Diagnoseebenen angesprochen. Sowohl die theorie- und diagnosegestützte Konzeption von Laborstationen, als auch die nachgelagerte empirische Diagnostik, kann, bei entsprechendem Interesse von Studierenden, im Rahmen von Bachelor- und insbesondere Masterarbeiten vertieft werden (Roth 2020).

5.1.2 Videodiagnosetool ViviAn

Damit das theoretische fachdidaktische Wissen, das in den Vorlesungen des Bachelorstudiums vermittelt wird, nicht träge bleibt, und um die Studierenden besser auf die Masterveranstaltungen vorzubereiten, erhalten diese die Möglichkeit, das erworbene Wissen in unterrichtsnahen Lehr-Lern-Situationen anzuwenden. Hier knüpft die Arbeitsgruppe „Didaktik der Mathematik (Sekundarstufen)“ mit dem Einsatz ihres Videodiagnosetools ViviAn (**V**ideovignetten zur **A**nalyse von Unterrichtsprozessen, vgl. <http://vivian.uni-landau.de>) an.



Abbildung 13: Oberfläche der Lernumgebung ViviAn

Abbildung 13 zeigt die standardmäßige Oberfläche der Lernumgebung ViviAn. Im Zentrum ist die Videovignette eingebettet, die in der Regel einen Ausschnitt von etwa zwei bis drei Minuten aus einer videografierten Gruppenarbeitsphase von Schülerinnen und Schülern im Mathematik-Labor „Mathe ist mehr“ zeigt. Bei sogenannten Trainingsvignetten kann das Video jederzeit gestartet und angehalten sowie vor- und zurückgespult werden. Diese Funktionen werden in Testsituationen häufig nicht zur Verfügung gestellt.

Die Gruppenarbeit der Schülerinnen und Schüler wurde von schräg oben gefilmt (Roth 2020). Diese Perspektive ermöglicht es, sowohl die gesamte Lerngruppe, als auch einzelne Lernende, in den Blick zu nehmen. Darüber hinaus ist auch ersichtlich, ob die Schülergruppe das gegenständliche Material nutzt. Um es dem Betrachter zu erleichtern, wahrzunehmen, welcher oder welche der Schülerinnen und Schüler gerade spricht, wird bei allen Verbalisierungen ein gelbes Rechteck neben der entsprechenden Person eingeblendet (vgl. Abb. 13). Die Schülerinnen und Schüler sind im Uhrzeigersinn von S1 bis S4 durchnummeriert, beginnend mit dem Lernenden links unten. Dies ist notwendig, damit die Diagnoseaufträge und -antworten eindeutig zugeordnet werden können.

Rund um das Video sind mehrere Buttons angeordnet, die durch einen Mausklick jeweils ein neues Fenster mit Informationen über die Lernumgebung oder die Schülergruppe öffnen. Links finden sich die Buttons: Arbeitsauftrag, Material und Schülerdokumente. Hinter dem Button „Arbeitsauftrag“ ist der aktuelle Arbeitsauftrag zu finden, der den Schülerinnen und Schüler während der Situation im Video vorliegt. Die gegenständlichen Materialien (und ggf.

auch die Simulation, die in der aktuellen Situation von der Schülergruppe genutzt wird,) können unter „Materialien“ eingesehen werden. Simulationen können an dieser Stelle von den Studierenden nicht nur eingesehen, sondern auch bedient werden. Unter der Schaltfläche Schülerprodukte befinden sich die schriftlichen Aufzeichnungen der Schülerinnen und Schüler. Diese können auch jeweils paarweise gegenübergestellt werden, um einzelne Lösungen bzw. Lösungswege miteinander zu vergleichen.

Das Thema der Lernumgebung sowie die Ziele, die diese anstrebt, sind mit Klick auf den gleichnamigen Button ersichtlich. Darüber hinaus kann man sich unter „Schülerprofil“ mit dem Alter einzelner Schülerinnen und Schüler im Video, der jeweiligen Klassenstufe und der jeweilig besuchten Schulart, vertraut machen. Außerdem ist mit Klick auf den Button „zeitliche Einordnung“ einzusehen, welche Inhalte im Rahmen der Laborstation bereits vor der Situation behandelt wurden und welche Inhalte nach der Situation noch folgen werden.

Rechts unten können die Diagnoseaufträge aufgerufen werden. Diese öffnen sich direkt unter der Videovignette und können vor dem Betrachten des Videos eingesehen werden. Die Diagnoseaufträge fokussieren jeweils auf einen spezifischen inhaltlichen Aspekt des Mathematiklernens, welcher einerseits in der dargestellten Situation in der Videovignette zu erkennen ist und zu welchem andererseits im Rahmen einer Lehrveranstaltung die notwendigen Theoriefacetten und fachdidaktischen Vertiefungen thematisiert wurden (Roth 2020). Dadurch sollen die Studierenden in die Lage versetzt werden, Diagnosen theoriebasiert durchzuführen. In ViviAn können sowohl offene Items als auch Single- und Multiple-Choice Fragen eingebettet werden. Um die Studierenden anzuregen, sich intensiv mit der jeweiligen Situation auseinanderzusetzen, müssen alle geschlossenen Items zusätzlich in einem Freitextfeld begründet werden.

Roth (2020) listet acht Anforderungen auf, die das Diagnosetool ViviAn erfüllt:

- Studierende, die Diagnoseaufträge zu Videovignetten von Schülerarbeitsprozessen bearbeiten, sollten im Wesentlichen über dieselben Informationen zur dargestellten Situation verfügen, wie die betreuende Lehrperson der Schülergruppe.
- Bearbeitungen zu Diagnoseaufträgen werden direkt bei den entsprechenden Aufgaben in ein Textfeld eingetragen und automatisch gespeichert.
- Studierende erhalten direkt nach jeder Bearbeitung eines Diagnoseauftrags eine Expertenbearbeitung als Feedback, die sie mit ihrer eigenen Bearbeitung vergleichen können.
- Studierende können die Diagnoseaufträge zu einem von ihnen selbstgewählten Zeitpunkt innerhalb eines vorgegebenen Zeitfensters an einem frei gewählten Ort durchführen.
- Es gibt ein Menü zur thematischen Übersicht und Auswahl von Videovignetten.
- Es gibt eine Benutzerverwaltung, die Dozierenden individuelle Freigaben von Videovignetten (auch über einstellbare Zeitintervalle) erlaubt und es ermöglicht, diese erst nach Eingang einer unterschriebenen Datenschutzerklärung zu nutzen.
- Es gibt eine Übersicht für Dozierende, welche Videovignetten von den einzelnen Studierenden im Kurs bereits bearbeitet wurden.
- Alle Videodaten der abgebildeten Personen sind geschützt, liegen auf einem Server der Universität Koblenz-Landau unter der vollständigen Kontrolle und ausschließlichen administrativem Zugang von zwei Administratoren der Arbeitsgruppe Didaktik der Mathematik (Sekundarstufen).

Insbesondere die ersten vier Punkte waren ausschlaggebend für die Entwicklung der Lernumgebung (Roth 2020). Durch ViviAn wird den Studierenden die Möglichkeit gegeben, ihr gerade im Rahmen einer Lehrveranstaltung zu einem Inhaltsbereich des Mathematikunterrichts erworbenes theoretisches fachdidaktisches Wissen begleitend zur Lehrveranstaltung einzusetzen, indem sie Diagnoseaufträge zu Videovignetten von Gruppenarbeitsphasen aus dem Mathematik-Labor „Mathe ist mehr“ zu diesem Inhaltsbereich bearbeiten (Roth 2020). Dabei erfahren sie einerseits dessen Bedeutung für lernprozessbezogene Diagnosen und andererseits wird ihnen bewusst, wie ihnen diese Kenntnisse helfen, ihr eigenes Lehrerhandeln zu organisieren. Damit leistet die Arbeitsgruppe einen Beitrag zur Theorie-Praxis-Verknüpfung, vernetzt gleichzeitig die verschiedenen fachdidaktischen Lehrveranstaltungen miteinander und sensibilisiert die Studierenden für die Unterschiedlichkeit der Schülerinnen und Schüler (Roth 2020).

Um den Aufbau und die Förderung der prozessdiagnostischen Fähigkeit möglichst gut unterstützen zu können, wird in verschiedenen Dissertationen untersucht, inwieweit sich die prozessdiagnostische Fähigkeit der Studierenden mit Hilfe von ViviAn entwickeln lässt (Bartel & Roth 2017; Enenkiel & Roth 2017; Hofmann & Roth 2017). Dabei geht es u.a. um die Entscheidung, ob anhand von Videos oder Transkripten gearbeitet werden soll (Bartel & Roth 2017), welche Feedbackform am wirkungsvollsten ist (Enenkiel & Roth 2017) und welchen Einfluss Aufgabenanalysen auf die prozessdiagnostische Fähigkeit von Studierenden haben (Hofmann & Roth 2017). Durch die Implementierung der videogestützten Prozessdiagnosen mit ViviAn in den Vorlesungen „Fachdidaktische Grundlagen“, „Didaktik der Zahlbereichserweiterungen“, „Didaktik der Geometrie“ und „Didaktik der Algebra“ im Bachelorstudium sowie im „Didaktischen Seminar“ im Masterstudium, können die Studierenden ihre prozessdiagnostische Fähigkeit bereits im Studium (weiter-)entwickeln (Roth 2020).

5.1.3 Mathematik-Labor „Mathe ist mehr“

Das Mathematik-Labor „Mathe ist mehr“ ist ein Schülerlabor der Arbeitsgruppe Didaktik der Mathematik (Sekundarstufen) am Campus Landau der Universität Koblenz-Landau (Roth 2020). Wesentliche Grundlage der Arbeit im Mathematik-Labor ist das forschende Lernen, sprich die selbstständige Bearbeitung mathematischer Fragestellungen, die problem- und handlungsorientiert sind (Roth 2013a; Roth 2013b). Schülerinnen und Schüler arbeiten in Kleingruppen an schriftlichen Arbeitsaufträgen, die jeweils in einen Kontext eingebunden sind und in der Schwierigkeit progressiv ansteigen. Dabei können sie bei Bedarf gestufte Hilfen aus einem Hilfeheft abrufen, arbeiten also im besten Fall vollständig autonom und benötigen keine Betreuung durch (angehende) Lehrkräfte. Eine Laborstation besteht dabei immer aus drei Teilen, die jeweils für 90 Minuten konzipiert sind. Die Arbeitshefte beinhalten einen Teil, den alle Schülerinnen und Schüler in der Zeit schaffen sollten und auf den in den nächsten Heften aufgebaut wird, und einen Teil, der bestimmte Inhalte nochmals vertieft und für besonders schnelle Schülergruppen vorgesehen ist. Die Vertiefungsaufgaben sind keine Voraussetzung für die Weiterarbeit im nächsten Heft. Durch den experimentellen Umgang mit gegenständlichen Materialien, Computersimulationen und Videos wird der Aufbau des Verständnisses von Alltagsphänomenen sowie des mathematischen Grundlagenwissens gefördert (Roth 2013b).



Abbildung 14: Schülergruppe beim Arbeiten mit gegenständlichen Materialien



Abbildung 15: Schülergruppe, die mit einer Computersimulation arbeitet

Die Abbildungen 14 und 15 zeigen Schülergruppen im Mathematik-Labor „Mathe ist mehr“, die anhand von gegenständlichen Materialien und Computersimulationen (vgl. Abb. 16) die Oberflächeninhalts- und Rauminhaltsformeln für Quader erarbeiten. Die Simulationen werden mit dynamischen Mathematiksystemen wie GeoGebra erstellt, die auch im Mathematikunterricht verwendet werden. Systematische Variation der Computersimulationen soll den Schülerinnen und Schülern dabei helfen, die Erfahrungen mit realen Modellen zu vertiefen, die Angemessenheit von Mathematisierungen zu überprüfen und die neu gewonnenen Eindrücke und Erkenntnisse zu vernetzen (Roth 2013a). Die Schülerinnen und Schüler kommunizieren über Beobachtungen beim Experimentieren, strukturieren und dokumentieren ihren Arbeitsprozess sowie die daraus resultierenden Ergebnisse und reflektieren anschließend darüber (Roth 2013a). Die Vernetzung mit dem Lernort Schule ist einerseits durch die Bearbeitung von Lehrplanthemen gegeben, andererseits durch die Vor- und Nachbereitung im Unterricht selbst (Roth

2013a). Dazu bekommen die Lehrkräfte Informationen zur Station (benötigte Vorkenntnisse, Lernziele, ...) sowie Vorschläge, wie im Anschluss an die Laborbesuche weitergearbeitet werden könnte.

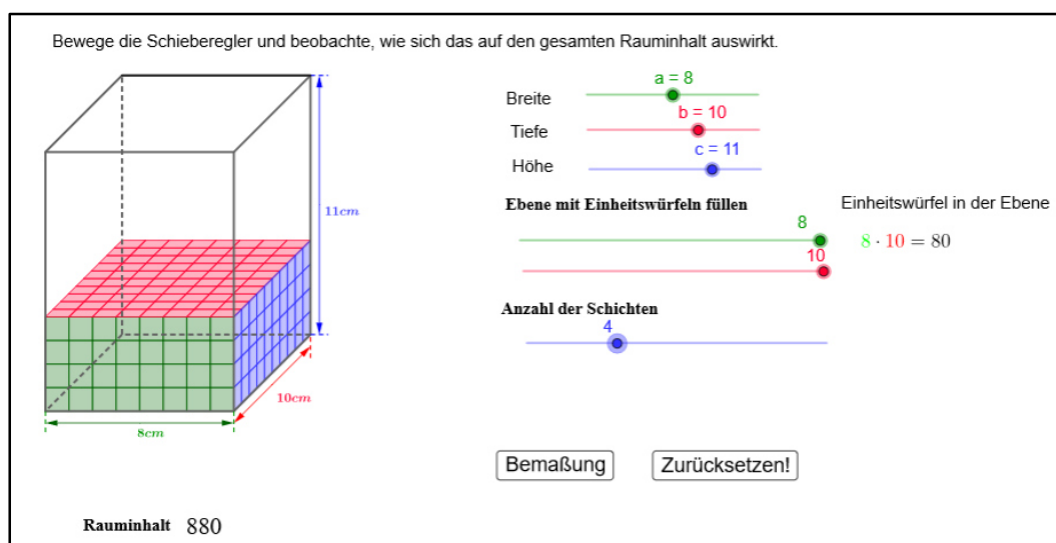


Abbildung 16: Simulation zur Erarbeitung der Volumeninhaltsformel von Quadern

Außerdem ist das Mathematik-Labor „Mathe ist mehr“ ein Forschungslabor, in dessen Zusammenhang ein Großteil der Forschungsaktivitäten der Arbeitsgruppe Didaktik der Mathematik (Sekundarstufen) in Landau stattfinden (Roth 2020). Die verschiedenen Projekte können jeweils zwischen zwei Dimensionen verortet werden, einerseits von der fachdidaktischen Entwicklungsforschung bis hin zur empirischen Grundlagenforschung, andererseits von der Unterrichtsforschung bis hin zur hochschuldidaktischen Forschung (Roth 2020). Für diese Arbeit von besonderer Relevanz ist die Forschung rund um das hochschuldidaktische Forschungsprogramm ViviAn, das bereits in Kapitel 5.1.2 vorgestellt wurde.

Nicht zuletzt ist das Mathematik-Labor „Mathe ist mehr“ auch ein Lehr-Lern-Labor, das zur praxisnahen Ausbildung der Studierenden beitragen soll (Roth 2020). Die Studierenden konzipieren Laborstationen für Schülergruppen und müssen daher ausgiebig über die Kenntnisse und Voraussetzungen nachdenken, die die Schülerinnen und Schüler mitbringen sollen. Diese Laborstationen werden anschließend erprobt. Dabei sind die Studierenden für den kompletten Besuch verantwortlich. Sie begrüßen die Klasse und führen diese in die Arbeit im Mathematik-Labor ein, betreuen die einzelnen Schülergruppen während der drei jeweils neunzigminütigen Laborbesuche und verabschieden diese am Ende der Station. Die Studierenden, die im Laufe ihres Studiums außer den Praktika keine standardisiert eingebundenen Praxisanteile haben, können hier also auch wertvolle Praxiserfahrungen sammeln. Durch eine anschließende Reflexion soll gewährleistet werden, dass die Studierenden ihre Erfahrungen aktiv nutzen, um künftige Unterrichtsentwürfe sowie Interaktionen mit Schülergruppen zu gestalten. Somit stellt die Erprobung der Laborstation einen zusätzlichen Baustein zur Theorie-Praxis-Verknüpfung dar. Ein wesentliches Element hierfür ist die enge Anbindung eines Lehr-Lern-Labor-Seminars an das Mathematik-Labor „Mathe ist mehr“, welches den Titel „Didaktisches Seminar“ trägt und im folgenden Abschnitt vorgestellt wird.

5.2 Konzept des Didaktischen Seminars

Die zentralen Elemente des Lehr-Lern-Labor-Seminars sind die Konzeption einer Lernumgebung für das Mathematik-Labor „Mathe ist mehr“, die Arbeit mit ViviAn, die Betreuung einer Schulklasse sowie eine Reflexion dieser Erprobung. Im Anschluss an die Beschreibung des Veranstaltungskonzepts (Kap. 5.2.1) werden die verschiedenen Ziele des Seminars thematisiert (Kap. 5.2.2).

5.2.1 Veranstaltungskonzept

Das Didaktische Seminar im Masterstudium der Mathematik richtet sich an Studierende mit dem Studienziel Lehramt an Gymnasium bzw. Realschulen Plus. Die Studierenden konzipieren oder überarbeiten in Dreiergruppen Lernumgebungen für das Mathematik-Labor „Mathe ist mehr“, die in ein Lehrplanthema einführen oder dieses vertiefen. Die Konzeption der Lernumgebung wird durch strukturelle Vorgaben angeleitet und von Dozierenden durch regelmäßige Feedbacks unterstützt (vgl. Abb. 17).

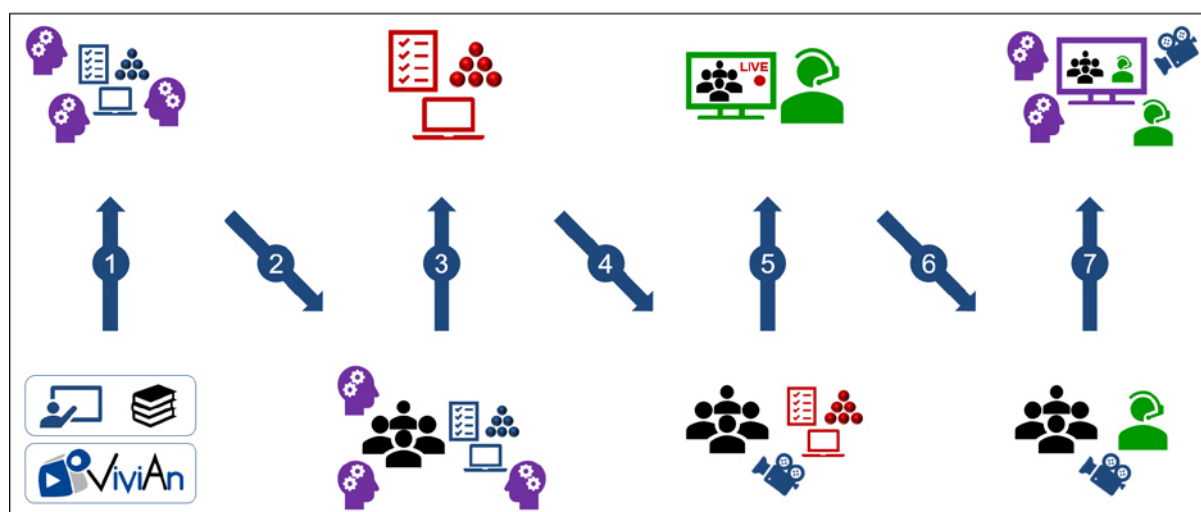


Abbildung 17: Übersicht über die Aktivitäten im Didaktischen Seminar nach Roth (2020)

(1) Vor der Überarbeitung einer Laborstation sollen die Studierenden zunächst drei Videovignetten und dazugehörige Diagnoseaufträge bearbeiten. Anschließend bearbeiten sie die Station zunächst selbst. Durch die Arbeit „als Schülerinnen und Schüler“, das bedeutet, dass sie die Arbeitshefte mit Hilfe des gegenständlichen Materials und der verschiedenen Simulationen bzw. Videos aus Schülerperspektive bearbeiten, durchdringen die Studierenden die Station inhaltlich und haben die Gelegenheit, erste Optimierungsmöglichkeiten festzustellen. Darüber hinaus erleben sie auch, was es bedeutet, selbstständig (unter Anleitung von Arbeitsaufträgen) die Arbeitshefte zu bearbeiten.

(2) Nachdem die Laborstation vollständig bearbeitet wurde, sehen sich die Studierenden eine Durchführung der erprobten Laborstation mit einer Schulklasse entweder live oder per Video an (live setzt voraus, dass in der entsprechenden Phase ein passender Laborbesuch stattfindet). Dadurch können die Studierenden einerseits ein Gefühl für die Art der Arbeit im Mathematik-Labor gewinnen und andererseits weitere Schwächen und Optimierungsbedarfe der Laborstation feststellen.

(3) Danach legen sich die Gruppen literaturgeleitet auf die Lernziele der neuen Laborstation fest. Aufbauend auf den Lernzielen erstellen sie konkrete problembasierte und kontextbezogene Aufgaben, die die Schülerinnen und Schüler mit Hilfe von gegenständlichem Material und Computersimulationen bearbeiten sollen. Auch diese werden von den Studierenden konzipiert und umgesetzt. Wichtig dabei ist jeweils, dass sowohl die Materialien als auch Computersimulationen einen echten Mehrwert gegenüber der reinen Textform bieten. Bei der Konzeption der neuen Lernumgebung sollen die Studierenden die gesammelten Optimierungsmöglichkeiten und Erfahrungen aus der eigenen Bearbeitung und der Beobachtung einer Schülergruppe bei der Bearbeitung einer Laborstation nutzen. Eine Laborstation besteht dabei immer aus drei Stationsteilen, die jeweils für eine Doppelstunde konzipiert sind.

(4) Nach der Fertigstellung wird die Station mit mindestens einer Schulklasse erprobt. Dabei arbeitet eine Schülergruppe in einem eigenen Raum und wird dabei gefilmt. Die restlichen Schülerinnen und Schüler der Klasse arbeiten in Vierergruppen in einem anderen Raum an derselben Station. Die Studierenden, die die Laborstation konzipiert haben, sind für den Laborbesuch verantwortlich und stehen bei Bedarf für Fragen der Schülerinnen und Schüler zur Verfügung.

(5) Während der Erprobung wird die Gruppenarbeit der Filmgruppe in jeder der drei Stationsteile (jeweils ca. 90 Minuten) von einer Studentin bzw. einem Studenten in einer Liveaufzeichnung von einem Nebenraum aus beobachtet. Dabei entscheiden die Studierenden jeweils selbst ob und ggf. wann und wie sie in den Gruppenarbeitsprozess der Schülerinnen und Schüler intervenieren. Darüber hinaus haben sie die Gelegenheit, für sie subjektiv interessante Aspekte digital festzuhalten. Dies findet softwareunterstützt statt, sodass später anhand des gesetzten Markers eindeutig zuordenbar ist, um welche Stelle es sich in der Videoaufzeichnung handelt. Diese Stellen sollen zusätzlich kurz beschrieben werden, damit ersichtlich ist, was für die angehende Lehrkraft an der Stelle interessant war.

(6) Wenn die Studierenden sich für eine Intervention entscheiden, gehen sie in den Videoraum, führen die Intervention durch und werden dabei gefilmt. Die getätigten Interventionen der drei Studierenden werden zu Videovignetten verarbeitet. Diese Interventionsvideos zeigen auch einen kurzen Vorlauf, bevor die eigentliche Intervention beginnt, damit die Intervention von den anderen Studierenden besser eingeordnet und eingeschätzt werden kann.

(7) Die zugeschnittenen Interventionen werden den drei Studierenden, die die Station entwickelt und durchgeführt haben, an einem Folgetermin zur Reflexion vorgelegt. Dabei reflektieren die Studierenden zunächst über die Erprobung der Lernumgebung und anschließend über ihre getätigten Interventionen. Hinsichtlich der Reflexion der Lernumgebung steht es ihnen frei, wie ausführlich sie den Laborbesuch reflektieren und inwieweit sie auf Einzelheiten eingehen. Bei den Interventionen reflektiert jeweils der- oder diejenige zuerst, der oder die die Intervention durchgeführt hat und entsprechend auf dem Video zu sehen ist. Dabei kann er oder sie seinen/ihren Gruppenmitgliedern erklären, weshalb er oder sie interveniert hat, weshalb er oder sie derartig interveniert hat, ob er oder sie wieder intervenieren würde und falls ja, ob er oder sie die Intervention ähnlich oder anders gestalten würde. Anschließend können die Gruppenmitglieder sich dazu äußern und ihre Gedanken teilen. Nachdem die Intervention vollständig reflektiert wurde, wird mit dem nächsten Interventionsvideo analog verfahren. Während ihrer Reflexion werden die Studierenden ebenfalls gefilmt.

Die Veranstaltung wird im Rahmen einer mündlichen Portfolioprüfung, in der die Studierenden ihre Konzeptionen der Lernumgebung theoriegeleitet einzeln vorstellen, begründet darüber reflektieren und etwaige Rückfragen dazu beantworten, abgeschlossen (Roth 2020). Mit der Veranstaltung sind mehrere Ziele verbunden, die im Folgenden vorgestellt werden.

5.2.2 Ziele der Veranstaltung

Die Studierenden sollen auf Basis der fachdidaktischen Theorien aus den Veranstaltungen, die sie bisher besucht haben, sowie ihren Erfahrungen aus den Schulpraktika, eine Lernumgebung konzipieren, adäquat in Gruppenarbeitsprozesse von Schülerinnen und Schülern intervenieren und abschließend darüber reflektieren. Vor der Konzeption der Lernumgebung müssen sie Literaturrecherche in fachlichen und fachdidaktischen Quellen betreiben sowie organisatorische bzw. strukturelle Vorgaben, wie beispielsweise die KMK-Bildungsstandards oder den aktuellen Lehrplan, zu Rate ziehen. Die Studierenden müssen als Gruppe gemeinsam an einem Ziel arbeiten, sich eng abstimmen, eigenständig organisieren und miteinander kommunizieren, um die Lernumgebung in der zur Verfügung stehenden Zeit zu erstellen. Zu Beginn des Seminars bearbeiten die Studierenden verschiedene Diagnoseaufträge zu Videovignetten, die eine Schülergruppe beim Arbeiten im Mathematik-Labor „Mathe ist mehr“ zeigt. Dabei rufen sie ihre prozessdiagnostische Fähigkeit ab, die ggf. (die Arbeit mit ViviAn ist den Studierenden freigestellt) in mehreren Großveranstaltungen im Bachelorstudium geschult wurde (vgl. dazu Bartel & Roth, 2017; Enenkiel & Roth, 2017; Hofmann & Roth, 2017). Bei der Erprobung sollen die Studierenden schließlich eine Schülergruppe mittels Echtzeitvideoübertragung beobachten, Lernschwierigkeiten und aufkommende Probleme diagnostizieren und auf dieser Grundlage ggf. in den Gruppenarbeitsprozess der Schülerinnen und Schüler eingreifen. Diese Erfahrungen und gesammelten Eindrücke werden von den Studierenden reflektiert. Durch die Reflexion über die getätigten Eingriffe sollen die Studierenden sich selbst als Lehrperson wahrnehmen, ihre Handlungen kritisch hinterfragen und diese ggf. beim nächsten Mal anpassen.

5.3 Zusammenfassung

Die Lehramtsausbildung im Fach Mathematik ist am Campus Landau über das gesamte Studium verteilt und aufeinander aufbauend konzipiert. Zu den Großveranstaltungen im Bachelorstudium werden jeweils zum Inhalt der Vorlesung passende Videovignetten zur Verfügung gestellt, die von den Studierenden (auf freiwilliger Basis) bearbeitet werden können. Diese Videovignetten sind in das Videotool ViviAn eingebunden.

Dort können die Studierenden zahlreiche Hintergrundinformationen zu den Schülerinnen und Schülern sowie insbesondere auch zu der oder den Aufgabe(n), die sie in der Videovignette bearbeiten, abrufen. Die Informationen über die Aufgabenstellung, die Mitschriften der Schülerinnen und Schüler sowie das Thema und die Lernziele der Laborstation sollen den Studierenden bei der Bearbeitung der Diagnoseaufträge helfen. Die Studierenden haben dabei in der Regel keine Zeitbegrenzung und können die Videos beliebig oft abspielen und währenddessen pausieren. Somit können sie die Inhalte der fachdidaktischen Großveranstaltungen vertiefen und diagnostische Fähigkeiten (weiter-)entwickeln.

Die Videovignetten stammen von Aufnahmen aus dem Mathematik-Labor „Mathe ist mehr“. Dabei handelt es sich um einen festen Ort am Campus, an dem Schülerinnen und Schüler in ganzen Klassen jeweils ein Lehrplanthema bearbeiten. Üblicherweise kommen die Klassen für dreimal 90 Minuten und arbeiten in Vierergruppen, wobei eine Gruppe in einem separaten Raum arbeitet und videographiert wird. Aus zahlreichen Videos, die über die letzten zehn Jahre entstanden sind, wurden im Rahmen von vier Promotionsprojekten Videovignetten zu verschiedenen Themengebieten und aus unterschiedlichen Klassenstufen erstellt (Bartel & Roth 2020; Enenkiel & Roth 2017; Hofmann & Roth 2017; Walz & Roth 2017).

Im Rahmen des Didaktischen Seminars bearbeiten die Studierenden zunächst ebenfalls Videovignetten, überarbeiten dann in Kleingruppen eine Laborstation des Mathematik-Labors und erproben diese schließlich mit einer Schulklasse. Teilweise werden auch neue Laborstationen konzipiert. Bei der Erprobung müssen die Studierenden jeweils für 90 Minuten eine Gruppe von Schülerinnen und Schülern beobachten und in den Gruppenarbeitsprozess intervenieren, sofern sie dies für notwendig erachten. Im Anschluss an die Erprobung der Laborstation reflektieren die Studierenden in Kleingruppen über diese und über die getätigten Interventionen.

Die Videovignetten, die die Studierenden zu Beginn des Semesters bearbeiten, behandeln das Thema „Terme“. Insbesondere geht es dabei um die Grundvorstellungen zu Termen. Diese werden im nächsten Kapitel vorgestellt.

6 Grundvorstellungen zu Termen

Mathematikunterricht zielt darauf ab, dass Lernende tragfähige Vorstellungen zu den jeweils erarbeiteten Begriffen aufbauen und somit den fachlichen Aspekten der Begriffe inhaltliche Bedeutung geben (Greefrath, Oldenburg, Siller, Ulm & Weigand 2016, 16). Beispielsweise kann eine Pizza in vier gleich große Stücke geteilt werden. Nimmt man sich nun ein Stück davon, so entspricht dies einer Viertelpizza, entsprechend wäre noch eine Dreiviertelpizza vorhanden. Die Bezugsgröße ist dabei in diesem Fall jeweils die komplette Pizza. Diese Grundvorstellung, auch „Teil eines Ganzen“ genannt, hilft Schülerinnen und Schülern bei der Arbeit mit Brüchen (vgl. dazu u.a. Malle 2004).

Vom Hofe definiert den Begriff „Grundvorstellung“ wie folgt:

Der Terminus ‚Grundvorstellung‘ charakterisiert (...) fundamentale mathematische Begriffe oder Verfahren und deren Deutungsmöglichkeiten in realen Situationen. Er beschreibt damit Beziehungen zwischen mathematischen Strukturen, individuell-psychologischen Prozessen und realen Sachzusammenhängen oder kurz: Beziehungen zwischen Mathematik, Individuum und Realität (vom Hofe 1995, 98).

Nach vom Hofe (1995, 97f.) geht es bei Grundvorstellungen insbesondere darum, Beziehungen zwischen mathematischen Inhalten und der individuellen Begriffsbildung herzustellen. Dabei sind folgende Aspekte entscheidend:

- (1) *„Sinnkonstituierung eines Begriffs durch Anknüpfung an bekannte Sach- oder Handlungszusammenhänge bzw. Handlungsvorstellungen“ (ebd., 97),*

Beispiel: Im Pizza-Beispiel wäre das Schneiden der Pizza eine konkrete und bekannte Handlung, die sich die Schülerinnen und Schüler vorstellen können. Die Einteilung eines Kreises in vier gleich große Stücke kann dadurch an eine bekannte Handlungsvorstellung anknüpfen und zur Sinnkonstituierung beitragen.

- (2) *„Aufbau entsprechender (visueller) Repräsentationen bzw. Verinnerlichungen, die operatives Handeln auf der Vorstellungsebene ermöglichen“ (ebd., 98),*

Beispiel: Die Schülerinnen und Schüler können die Pizza als Kreis skizzieren. Dies bietet die Möglichkeit, den Übergang von der konkreten Handlung hin zu einer abstrakteren Darstellung zu ermöglichen. Mit Hilfe der Skizze können nun Stücke (gedanklich) weggenommen bzw. ergänzt werden. Daneben können auch andere Einteilungen vorgenommen werden.

- (3) *„Fähigkeit zur Anwendung eines Begriffs auf die Wirklichkeit durch Erkennen der entsprechenden Struktur in Sachzusammenhängen oder durch Modellieren des Sachproblems mit Hilfe der mathematischen Struktur“ (ebd., 98).*

Beispiel: Anhand der vier gleich großen Pizzastücke erkennen die Schülerinnen und Schüler, dass die Pizza (das Ganze) in vier gleich große Teile zerlegt werden kann. Mit Hilfe der erkannten Struktur in diesem Sachzusammenhang kann die Problemstellung, wer bei gleicher Verteilung der insgesamt vorhandenen Stücke wie viele davon zur Verfügung hat und

welchem Teil des Ganzen dies entspricht, durch mathematische Modellierung des Phänomens gelöst werden.

Im weiteren Verlauf dieser Arbeit werden die Grundvorstellungen zu Termen eine wichtige Rolle spielen. Hintergrund ist, dass zur Einschätzung der prozessdiagnostischen Fähigkeit von Lehramtsstudierenden ein Testinstrument entwickelt und validiert wurde (vgl. dazu Kap. 8.2), bei dem Diagnosen zu drei Videovignetten erstellt werden müssen. Diese Videovignetten zeigen eine Schülergruppe, die gemeinsam an Aufgaben zu Termen arbeitet. Die Diagnoseaufträge zu diesen Videovignetten richten den Blick zu einem großen Teil auf Grundvorstellungen zu Termen (vgl. dazu Kap. 8.2.3).

Da die meisten Terme mindestens eine Variable enthalten, ist es ratsam sich zunächst mit dem Begriff der Variable auseinanderzusetzen, um sich dann den Grundvorstellungen zu Termen zu widmen. Ohne Variablen sind Terme lediglich Zahlen bzw. eine Verknüpfung (+, -, ...) von Zahlen, welche einen Termwert ergeben, der durch Berechnung eindeutig bestimmt werden kann. Wenn ein Term Variablen enthält, kommen mehrere Facetten hinzu, welche im Folgenden zunächst erläutert und schließlich für die Beschreibung und Diskussion von Grundvorstellungen zu Termen genutzt werden sollen. Auf dieser Grundlage soll deutlich werden, was Terme sind und inwieweit Grundvorstellungen bei Termen interpretiert und unterschieden werden können.

6.1 Variablen

Variablen werden in der Regel mithilfe von Buchstaben des lateinischen Alphabets repräsentiert und treten in verschiedenen Kontexten auf. In diesem Teilkapitel werden zunächst die Sichtweisen verschiedener Autoren zu Variablen zusammengestellt. Anschließend werden diese in Bezug zueinander gesetzt und dabei in eine kohärente Darstellung gebracht.

Freudenthal (1983, 473ff.) verweist darauf, dass Variablen keine Erfindung der Mathematik sind, sondern auch in unserem Alltag häufig vorkommen. Als Beispiele nennt er die beiden Begriffe „Löwe“ und „Ding“. Wir geben nicht jedem einzelnen Löwen einen eigenen Namen, sondern bezeichnen einen Löwen, wenn wir ihn sehen, als solchen. Manchmal kommen Adjektive wie *jung* und *alt* oder Angaben wie *der aus dem Amsterdamer Zoo* bzw. *der aus dem Londoner Zoo* hinzu, jedoch wird das Tier jeweils als *Löwe* bezeichnet (Freudenthal 1983, 474). Dies gilt natürlich auch für alle anderen Tiere. Noch einleuchtender wird es bei dem Begriff „Ding“. „Ding“ ist ein Begriff, welcher eine große Vielzahl an Objekten beschreiben kann (Freudenthal 1983, 474). Es ermöglicht es, über etwas zu sprechen, ohne die genaue Bezeichnung zu kennen. „Ding“ steht also als Platzhalter für die eigentlich korrekte Bezeichnung.

Ebenso verhält es sich bei Buchstaben in der Mathematik, die als Symbole etwas repräsentieren (Freudenthal 1983, 477). Wiese man den Buchstaben keine Bedeutung zu, könnte Mathematik als bedeutungsloses Spiel mit Symbolen angesehen werden (ebd.). Variablen können in Ausdrücken entweder durch Zahlen oder durch komplexere Ausdrücke (Terme) ersetzt werden, welche ebenfalls Variablen enthalten können (Freudenthal 1983, 483). Dabei können Variablen zur Generalisierung eines konkreten Sachverhaltes auf ggf. viele weitere, gleichartige Sachverhalte genutzt werden (Freudenthal 1983, 488). Darüber hinaus können Terme durch Variablen strukturell vereinfacht werden, wenn ein gegebener Ausdruck teilweise durch Variablen ersetzt wird, bzw. strukturell erschwert werden, wenn Variablen durch ganze Ausdrücke ersetzt werden (ebd.). Nach Freudenthal (1983, 491f) wurde der Begriff „Variable“ früher dafür genutzt, um etwas Physikalisches, Soziales oder Mathematisches zu beschreiben, das wirklich

variiert, also dynamisch ist, wohingegen mittlerweile eher eine statische Sichtweise eingenommen werde. Sowohl die rein statische Sichtweise, in der Variablen lediglich als Platzhalter interpretiert werden, als auch die Sichtweise, die sich auf die Möglichkeit, eine Vielzahl unterschiedliche Objekte einsetzen zu können, bezieht, sollten im Unterricht nicht strikt getrennt voneinander betrachtet werden (Freudenthal 1983, 491ff.).

Ähnlich ordnen dies Vollrath und Weigand (2007, 79) ein, die darauf verweisen, dass eine Vielzahl unterschiedlicher Begriffe wie „Unbekannte“, „allgemeine Zahl“, „Buchstabenrechnung“ und „Variable“ bis in die 1960er für unterschiedliche Anwendungen genutzt wurden und dies eher verschleierte anstatt die Bedeutung der Buchstaben in der Algebra zu klären. Ab dann beschränkte man sich auf die beiden Begriffe „Variable“ und „Platzhalter“ (Vollrath & Weigand 2007, 81). Dadurch wurden jedoch Unterschiede verwischt, welche zuvor durch die Nutzung verschiedener Begriffe der Variable in verschiedenen Kontexten besser herausgearbeitet wurden (Vollrath & Weigand 2007, 81).

Man merkte, dass es gar nicht so töricht gewesen war, unterschiedliche Kontexte zu betrachten und zu beschreiben, wie man in ihnen mit Variablen arbeitet und welche Vorstellungen bei den Schülern dabei gebildet werden. Als mathematisch bedeutsame Kontexte für unterschiedliche Aspekte beim Arbeiten mit Variablen erwiesen sich: Eigenschaften der Zahlbereiche, Terme und Termumformungen, Funktionen, Gleichungen und Formeln. (Vollrath & Weigand 2007, 81)

Obwohl der Begriff „Variable“ sich in der Mathematik eingebürgert hat, handelt es sich dabei um ein Fremdwort (Vollrath & Weigand 2007, 79). Daher werde oft auch „Platzhalter“ gesagt, um den Schülerinnen und Schülern den Zugang zur Bedeutung des Wortes zu erleichtern (Vollrath & Weigand 2007, 79). In beiden Wörtern schwingen jedoch Vorstellungen mit, welche sich nicht ganz entsprechen (Vollrath & Weigand 2007, 69). Während „Variable“ eher an Veränderungen erinnert, wie man sie beispielsweise im Physikunterricht betrachtet, verbinden Vollrath und Weigand (2007, 79f.) mit „Platzhalter“ eher, dass mögliche Zahlen eingesetzt werden können. Bereits 1989 verweist Vollrath auf diese Unterscheidung der Variable als *Platzhalter* bzw. *veränderliche Größe* (Vollrath 1989, 14ff.). Vollrath und Weigand (2007, 77) halten Variablen für unumgänglich, um allgemeine Aussagen über Zahlen treffen zu können.

Auch für Schoenfeld und Arcavi (1988) ist das Konzept der Variable sowohl für das Unterrichten als auch für das Verstehen von Mathematik zentral. Variablen sind die Grundlage des Übergangs der Arithmetik zur Algebra und notwendig für höhere Mathematik (Schoenfeld & Arcavi 1988). Sie haben mehrere Konnotationen, Bedeutungen und Verwendungsmöglichkeiten (Schoenfeld & Arcavi 1988). Es gibt viele Begriffe, welche jeweils auf einen Teilaspekt von Variablen abzielen, ohne jedoch den Begriff der Variable vollständig abzudecken, beispielsweise nennen Schoenfeld und Arcavi (1988) „Symbol“, „Platzhalter“, „Fürwort“, „Parameter“, „Argument“, „Zeiger“, „Benennung“, „Bezeichner“, „Leerraum“, „Lücke“, „Bezug“, „Fall“ oder „Unbekannte“. Wenngleich die Bezeichnungen aus dem Englischen übersetzt wurden und die Liste nur einem Versuch der Autoren zugrunde liegt, das Wort *Variable* mit einem anderen Wort zu beschreiben, und daher keinen Anspruch auf Vollständigkeit erhebt, zeigt sich jedoch, dass viele unterschiedliche Aspekte bedient werden. In unterschiedlichen Disziplinen werden unterschiedliche Umschreibungen des Begriffs als adäquat erachtet (Schoenfeld & Arcavi 1988). Beispielsweise hielten Informatiker das Wort „Zeiger“ für eine gute Umschreibung, während Sprachwissenschaftler „Fürwort“ für gut befanden (Schoenfeld & Arcavi 1988). Jedoch sind Schoenfeld und Arcavi (1988) davon überzeugt, dass der Variablenbegriff deutlich zu groß sei, als dass er in einem Wort wiedergegeben werden könne. Schoenfeld und Arcavi

(1988) diskutieren zehn Definitionen bzw. Definitionsversuche aus unterschiedlichen, nicht näher genannten Quellen, und fassen zusammen, dass manche darauf abzielen, was eine Variable ist, während andere beschreiben, wie eine Variable genutzt wird.

In sum, the meaning of variable is variable; using the term differently in different contexts can make it hard for students to understand. As teachers, we should be sensitive to this multiple usage. (Schoenfeld & Arcavi 1988)

Schoenfeld und Arcavi (1988) unterscheiden zwischen Variablen als „polyvalent names“ und „variable objects“. Damit verweisen sie einerseits darauf, dass es sich um eine Größe handelt, die eine bestimmte Menge von Werten annehmen und daher durch ein Element dieser Menge ersetzt werden kann. Andererseits kann die Variable kontinuierlich variiert und die Auswirkung betrachtet werden.

Neben diesen beiden Aspekten, die in ähnlicher Weise auch schon bei Freudenthal (1983) angeklungen sind, gibt es noch weitere Aspekte von Variablen. Bereits 1978 untersuchte Küchemann den Variablenbegriff, indem er Schülerinnen und Schüler im Alter zwischen 13 und 15 Jahren dazu befragte. Dabei fasst er aus den zahlreichen Schülerantworten sechs verschiedene Aspekte zusammen, die im Folgenden jeweils an einem Beispiel erläutert werden.

- *Auswertbarer Buchstabe*

Die Variable kann direkt ausgewertet werden. Es sind keine Zwischenschritte mehr nötig, um zum Ergebnis zu gelangen. Im Beispiel $x + 4 = 6$ bedeutet das, dass in einem Rechenschritt bestimmt werden kann, dass die Variable x den Wert 2 hat.

- *Unberücksichtigter Buchstabe*

Zur Erläuterung nutzen wir die Gleichungen $x + y = 6$ und $x + y + 2 = ?$. Die beiden Variablen x und y werden hierbei nicht berücksichtigt. Die zweite Gleichung unterscheidet sich von der ersten insofern, als 2 addiert wird. Daher spielt es keine Rolle, was genau x ist bzw. was genau y ist. Sie können ignoriert werden und trotzdem gelangt man zu einem Ergebnis, in diesem Fall 8.

- *Buchstabe als Objekt*

Beispielsweise kann a als die Seitenlänge eines Quaders angesehen werden. Hierbei wird der Buchstabe lediglich zur Bezeichnung genutzt, beispielsweise für die Länge einer Seite eines geometrischen Objekts. Es handelt sich dabei also nur um einen Namen oder eine Bezeichnung, die in diesem Sinne nicht ausgewertet werden kann.

- *Buchstabe als bestimmte Unbekannte*

Küchemann (1978) gibt an dieser Stelle eine unvollständig gezeichnete Figur an, deren noch unbekannte Seitenanzahl mit n bezeichnet wird. Hat man eine solche Figur vor sich, so kann n zunächst nicht näher bestimmt werden. Es gibt zwar genau einen Wert für n , jedoch ist dieser spezifische Wert unbekannt.

- *Buchstabe als verallgemeinerte Zahl*

Beispielsweise sei $x + y = 6$; $x < y$ und x unbekannt. Die Variable x repräsentiert hierbei mehrere Zahlen. Es sind also mehrere Möglichkeiten gegeben, sodass die Bedingungen erfüllt sind. Beispielsweise könnte $x = 1$ sein, dann wäre $1 + y = 6$ und somit $y = 5$. Analog könnte x auch die Zahlen 2 bis 6 einnehmen und y entsprechend die Zahlen 4 bis 0. Verlässt man die

natürlichen Zahlen und erweitert diese zu den reellen Zahlen oder gar zu den rationalen Zahlen, erhält man eine unendliche Menge an denkbaren Lösungen und somit auch eine unendliche Anzahl an Zahlen, die die Variable x (und somit auch die Variable y) abdeckt.

- Buchstabe als Variable

Geht man der Frage nach, welche Zahl (eigentlich müsste man an dieser Stelle von einem Termwert sprechen) größer ist, $2x$ oder $x + 2$, so ist zunächst unklar, ob $2x$ oder $x + 2$ größer ist. Da keine direkte Aussage darüber getroffen werden kann, muss ein weiterer Zusammenhang gefunden werden. Diesen findet man, indem man die Variable x variiert. Durch systematische Variation der Variablen kommt man darauf, dass $2x$ größer ist, sofern $x > 2$ gilt. Gilt jedoch, dass $x < 2$ ist, dann wäre $x + 2$ größer, und bei $x = 2$ wären beide gleich groß.

Küchemann (1978) verweist in der Diskussion seines Artikels darauf, dass es ihm darum ging, eine Idee zu bekommen, wie Schülerinnen und Schüler mit mathematischen Problemen zurecht kommen, in der Hoffnung, zur Anpassung des Mathematikunterrichts an den Umgang der Schülerinnen und Schüler sowie deren Bedürfnisse, beizutragen (Küchemann 1978).

Franke und Wynands (1991) benennen fünf Aspekte von Variablen, die sich teilweise mit den von Küchemann (1978) gefundenen Aspekten decken, teilweise aber auch neue Sichtweisen liefern. In ihrer Studie wollen sie unter anderem herausfinden, welche Variablenaspekte den Schülerinnen und Schülern besondere Schwierigkeiten bereiten. Ihre fünf Variablenaspekte benennen sie jeweils mit einem Buchstaben: Aspekt E, Aspekt O, Aspekt F, Aspekt G und Aspekt R.

Aspekt E kann man auch als Buchstabe oder Wort bezeichnen, der oder das als Eigenname verwendet wird. Hierbei wird der Buchstabe bzw. das Wort als Eigenname für eine Zahl gebraucht, die dann konstant einen festen Wert hat (Franke und Wynands 1991, 677). Als Beispiele dienen etwa die Kreiszahl π oder die Lichtgeschwindigkeit c .

Bei *Aspekt O* steht der Variablenname für eine Einsetzstelle für ein vorgegebenes Objekt, mit dem operiert wird (Franke und Wynands 1991, 677). Eine Beispielaufgabe lautet: Welchen Wert hat $3a + b$, wenn $b = a - 1$ und $a = 4$?

Aspekt F tritt auf, wenn Variablen als beliebig Veränderliche in Formeln oder funktionalen Zusammenhängen auftreten (Franke und Wynands 1991, 677). Eine Beispielaufgabe dazu lautet: An einer Schule sind L Lehrer und S Schüler. Auf einen Lehrer kommen 10 Schüler. Welche der folgenden Gleichungen stellt diesen Zusammenhang richtig dar?

a) $L = 10 \cdot S$ b) $S = 10 \cdot L$ c) $L : S = 10$ d) $S : L = 10$

Unter *Aspekt G* versteht man Variablen als noch unbekannte, vielleicht existierende Elemente der Grundmenge (Franke und Wynands 1991, 678). Eine Beispielaufgabe lautet: Löse die Gleichung $3 \cdot (2 - x) = 18 - x$.

Mit *Aspekt R* bezeichnen Franke und Wynands (1991, 678) Variablen in allgemeingültigen Aussagen, wie etwa Regeln und Rechengesetzen. Exemplarisch hierfür ist etwa das Kommutativgesetz.

Malle (1993) ist heutzutage der vermutlich meist zitierte Autor, wenn es um Grundvorstellungen zu Termen geht. Seine vor einem Vierteljahrhundert veröffentlichten Sichtweisen zu Termen unterscheiden sich von den Ausführungen von Küchemann (1978) und Franke und Wynands (1991) insofern, als er nicht empirisch Rückschlüsse auf die Variablenaspekte zieht, sondern sich diese anhand theoretischer Überlegungen herleitet. Dies führt dazu, dass die bisher genannten Aspekte teilweise enthalten sind, aber systematischer dargestellt werden, wenngleich

auch Malle (1993) seine Aspekte nicht scharf voneinander trennt und diese teilweise simultan gedacht werden können bzw. müssen.

Er stellt zunächst fest, dass der Variablenbegriff seiner Ansicht nach nicht zufriedenstellend definiert werden könne und daher in der Literatur in der Regel nur verwendet, aber nicht definiert werde (Malle 1993, 44). Falls doch Definitionsversuche getätigt werden, seien diese meist zu einseitig und würden den Variablenbegriff nicht vollständig erfassen (ebd.). Malle (1993, 44) geht sogar so weit zu behaupten, dass niemand genau wisse, was eine Variable sei. Auch er sieht den Variablenbegriff nicht als Erfindung der Mathematik, sondern verweist auf umgangssprachliche Bezeichnungen wie „Ding“, „Sache“, „ein beliebiger“ oder „irgendwelcher“, die den Charakter einer Variablen haben (Malle 1993, 44). Als Vorteil der Buchstabenvariablen gegenüber den Wortvariablen sieht er eine knappere, übersichtlichere, unmissverständlichere und kontextfreie Darstellung, die ein regelhaftes Operieren ermöglicht (Malle 1993, 45). Malle (1993, 46) unterscheidet zunächst drei Aspekte des Variablenbegriffes, nämlich den Gegenstandsaspekt, den Einsetzungsaspekt und den Kalkülaspekt. Diese drei Aspekte sollen nun anhand des Beispiels $2x + 4 = 8$ erläutert werden:

Beim Gegenstandsaspekt wird die Variable als unbekannte oder nicht näher bestimmte Zahl angesehen. Hierbei wird das x zunächst als eine unbekannte Zahl angenommen, mit der einfach umgegangen wird. Für x muss gelten, dass $x + 2 = 4$. Wird 2 zur gesuchten Zahl addiert ergibt sich 4. Also ist 2 die gesuchte Zahl.

Unter dem Einsetzungsaspekt sieht man die Variable als Platzhalter für Zahlen bzw. eine Leerstelle, in die man Zahlen einsetzen darf. Es wird also gefragt, welche Zahl man für x einsetzen kann, damit die Aussage $2x + 4 = 8$ korrekt ist. Wenn man 2 einsetzt, stimmt die Aussage, daher ist 2 die gesuchte Zahl.

Beim Kalkülaspekt wird die Variable als bedeutungsloses Zeichen angesehen, mit dem nach bestimmten Regeln operiert werden darf. Mithilfe einer Äquivalenzumformung wird zunächst auf beiden Seiten 4 subtrahiert, was dazu führt, dass noch $2x = 4$ übrig bleibt. Nun wird auf beiden Seiten durch 2 dividiert, wodurch man $x = 2$ erhält. Also ist 2 die gesuchte Zahl.

Darüber hinaus sieht Malle (1993, 79ff.) die Variable auch als Veränderliche. Zur Klärung dieser Sichtweise unterscheidet er zunächst zwischen Einzelzahlaspekt und Bereichsaspekt. Beim *Einzelzahlaspekt* wird die Variable als beliebige aber feste Zahl aus dem betreffenden Bereich betrachtet, wobei nur genau eine Zahl aus dem Bereich repräsentiert wird. Beim *Bereichsaspekt* wird die Variable als beliebige Zahl aus dem betreffenden Bereich angesehen, wobei jede Zahl des Bereichs repräsentiert wird. Der Bereichsaspekt tritt wiederum in zwei unterschiedlichen Formen auf, dem *Simultanaspekt*, hierbei werden alle Zahlen aus dem entsprechenden Bereich gleichzeitig repräsentiert, und dem *Veränderlichenaspekt*, bei dem alle Zahlen aus dem betreffenden Bereich in zeitlicher Aufeinanderfolge repräsentiert werden. Bei letzterem muss ergänzt werden, dass dies in einer systematischen bzw. gleichmäßigen Form geschieht. Beim Einzelzahlaspekt stellt die Variable also eine Zahl aus dem Bereich dar, die aus diesem Bereich ausgewählt, dann aber festgehalten wird. Beim Simultanaspekt stellt die Variable jede Zahl aus dem Bereich dar, es wurde also noch kein Auswahlvorgang vollzogen. Der Veränderlichenaspekt kommt fast ausschließlich im Zusammenhang mit Funktionen oder funktionalen Betrachtungen von Formeln vor (Malle 1993, 81). Außerdem schreibt Malle (1993, 81) im Hinblick auf mehrere Variablen in einem Kontext:

Kommen in einem Kontext mehrere Variablen vor, von denen einige unter dem Einzelzahlaspekt und einige unter dem Veränderlichenaspekt betrachtet werden, bezeichnet man manchmal die ersteren als Parameter und die letzteren als Veränderliche.

Im Beispiel $f(x) = ax + b$ wären a und b jeweils Variablen, die man unter dem Einzelzahlaspekt auffassen würde, da sie für eine beliebige, aber feste Zahl stehen. Diese werden zwar ausgewählt, dann aber konstant gehalten. Die Variable x wäre in diesem Fall hingegen diejenige, die variiert und daher dem Veränderlichenaspekt zugeordnet werden könnte. Für x werden systematisch verschiedene Zahlen aus dem Definitionsbereich eingesetzt, der dabei gleichmäßig durchlaufen wird. Man achtet dabei auf die sich daraus ergebenden Änderungen des Termwerts.

Drijvers (2003) beschäftigt sich in seiner Arbeit ausführlich mit dem Parameterbegriff. Er unterscheidet Parameter von Variablen und bezeichnet Parameter als Variablen höherer Ebene (Drijvers 2003, 67ff.). Je mehr Variablen ins Spiel kommen, desto komplexer wird es, zumal die Rollen nicht festgelegt sind und während des Lösungsprozesses wechseln können (Drijvers 2003, 68). Da es sich bei Parametern aber letztendlich auch „nur“ um Variablen handelt, wird die Unterscheidung der verschiedenen Aspekte im Folgenden exemplarisch an Variablen festgemacht.

Drijvers (2003, 46) verweist zunächst auf die Unterscheidung von Kindt (1980), wonach Variablen einen Generalisierungsaspekt und einen Dynamikaspekt haben. Dies entspricht einer statischen und einer dynamischen Sicht auf Variablen. Beim *Generalisierungsaspekt* geht es darum, einen Sachverhalt zu begreifen und Muster und Strukturen zu erkennen – also eher statisch-beschreibend das Allgemeine der Situation durch (einen Term mit) Variablen zu erfassen (Barzel & Herget 2006, 6). Beim *Dynamikaspekt* werden Zusammenhänge erkundet, indem für Variablen verschiedene Zahlen eingesetzt und dabei Veränderungen beobachtet und Bewegungen vorgestellt werden (Barzel & Herget 2006, 6).

Drijvers selbst (2003, 66) unterscheidet fünf verschiedene Aspekte von Variablen: Variable als Platzhalter, Variable als veränderliche Größe, Variable als generalisierte Zahl, Variable als Unbekannte und Variable als Symbol.

Die *Variable als Platzhalter* steht für etwas, wofür später numerische Werte eingesetzt werden können. Im Term $2x + 4$ wäre das x also eine Stelle, für die eine Zahl eingesetzt werden kann. Unter dieser Sichtweise ist der Wert bereits festgelegt, unabhängig davon, ob dieser bereits bekannt oder noch unbekannt ist, weshalb eine systematische Variation der Variable nicht vorgesehen ist.

Die *Variable als veränderliche Größe* spiegelt den dynamischen Aspekt wider und gehört zum funktionalen Ansatz der Algebra. Versteht man $f(x) = 2x + 4$ als Funktionsterm, so durchläuft x alle Werte des Definitionsbereiches. Die Variable wird systematisch bzw. gleichmäßig variiert und die jeweilige Veränderung betrachtet. Dadurch besitzt dieses Variablenverständnis einen dynamischen Charakter.

Die *Variable als generalisierte Zahl* bei Mustern und Strukturen kann genutzt werden, um konkrete Situationen bzw. Sachzusammenhänge zu generalisieren und somit Terme bzw. Gleichungen aufzustellen, die die Struktur des vorliegenden Phänomens beschreiben.

Die *Variable als Unbekannte* taucht meistens in Problemlöseaufgaben auf, in denen eine noch unbekannt Zahl gesucht wird, die jedoch prinzipiell bestimmt werden kann. Beispielsweise könnte man sich fragen, für welchen Wert der Variable folgende Gleichung gilt: $2x + 4 = 8$. Der entsprechende Wert ist zunächst unbekannt, kann jedoch mittels gängiger Rechenverfahren bestimmt werden.

Schließlich bleibt noch die *Variable als Symbol*, bei der die Variable lediglich als mathematischer Ausdruck erkannt und hingenommen wird. Diese Rolle hängt mit einem Verständnis von Mathematik (bzw. Algebra) als eigener Sprache zusammen. Im Fall $2x + 4$ würde man den Term durch die Variable x also als etwas Mathematisches erkennen, das als solches hingenommen und nicht weiter hinterfragt würde.

Neben einigen weiteren Autoren, die eigene Sichtweisen zu Termen ausführen, welche jedoch keine Aspekte jenseits der bereits aufgeführten Aspekte beinhalten (vgl. dazu u.a. Warren 1999; Ursini & Trigueros 1997; Usiskin 1999; Griesel 1982), gibt es mittlerweile auch einige Autoren, die eine Zusammenstellung vornehmen und versuchen, einen Überblick über die Thematik zu geben. Bertalan (2007) differenziert zwischen dem Buchstaben als Maßeinheit, dem Buchstaben als Variable und dem Buchstaben als Namen. Siebel (2005) liefert eine tabellarische Einordnung verschiedener Aspekte unterschiedlicher Autoren in Bezug zu Malles Variablenaspekten. Specht (2009) gibt ebenfalls einen Überblick über verschiedene Autoren und ordnet diese den Variablenaspekten von Freudenthal zu. Darüber hinaus unterscheidet sie zwischen Variablenaspekten und Variablenauffassungen und gibt eine Arbeitsdefinition für den Begriff der Variable (Specht 2009, 39). Akinwunmi (2012) gibt ebenfalls einen Überblick über Variablenauffassungen ausgewählter Autorinnen und Autoren. Aufgrund der vorgenommenen Autor/inn/enauswahl, die eine andere Zusammenstellung beinhaltet als die bisher vorhandenen Überblickswerke und entsprechend andere Aspekte beleuchtet bzw. Schwerpunkte setzt, wird im weiteren Verlauf eine eigene Einteilung der Variablenaspekte vorgenommen. Dafür werden nun, nachdem verschiedene Autorinnen und Autoren vorgestellt wurden, zunächst die Inhalte zusammengefasst und aufeinander bezogen.

Die *Unterscheidung zwischen statischer und dynamischer Betrachtung von Variablen* scheint eine zentrale Unterscheidung zu sein, auf die alle genannten Autoren (Küchemann 1978; Kindt 1980; Franke & Wynands 1991; Freudenthal 1983; Schoenfeld & Arcavi 1988; Malle 1993, Drijvers 2003; Vollrath & Weigand 2007) explizit oder zumindest implizit eingehen. Dies scheint daher bei der Betrachtung von Variablen ein erstes, wesentliches Unterscheidungsmerkmal zu sein. Einerseits wird die Variable als etwas Statisches wahrgenommen, für das prinzipiell etwas eingesetzt werden kann. Diese Unbekannte kann durch eine Zahl oder einen Ausdruck, welche/r wiederum eine oder mehrere Variablen beinhalten kann, ersetzt werden. Zur statischen Betrachtungsweise gehören auch die Nutzung von Variablen für allgemeingültige Gleichungen sowie die Wahrnehmung der Variable als mathematisches Symbol. Bei der Nutzung von Variablen für allgemeingültige Gleichungen gelten entsprechende Gleichungen immer, egal welche Zahlen für die Variablen eingesetzt werden. Der dynamische Aspekt von Variablen bezieht sich auf die Veränderung, welche beobachtet werden kann, wenn eine Variable variiert wird. Es wird also beobachtet, wie es sich auf den Funktionswert auswirkt, wenn die Variable systematisch bzw. gleichmäßig verändert wird. Die Variable wird dabei als veränderliche Größe angesehen. In der Regel geschieht dies bei der Betrachtung funktionaler Zusammenhänge.

Einige Autoren gehen über diese dichotome Betrachtungsweise hinaus und untergliedern die Aspekte von Variablen noch genauer. Dabei lassen sich einige Gemeinsamkeiten bei den verschiedenen Autoren feststellen. Die *Variable als Platzhalter* wird von vielen (Freudenthal 1983, Vollrath & Weigand 2007, Drijvers 2003) wörtlich genannt, während Malle (1993) diesen Aspekt als „Einsetzungsaspekt“ bezeichnet. Bei Küchemann (1978) findet sich dieser Aspekt implizit im „ausgewerteten Buchstaben“ und bei Franke und Wynands (1991) im „Aspekt O“. Der Aspekt *Variable als Platzhalter* ist jedoch in allen genannten Veröffentlichungen als solcher zu identifizieren. Ähnliches gilt für den Aspekt der *Variable als Unbekannte*. Drijvers (2003) benennt ihn in dieser Form, während er sich bei Küchemann (1978) im „Buchstaben als Objekt“, bei Malle (1993) im „Gegenstandsaspekt“, bei Franke und Wynands (1991) im „Aspekt G“ sowie bei Vollrath und Weigand (2007) indirekt wiederfindet. Auch der Aspekt *Variable als allgemeine Zahl* ist nicht bei allen Autoren wörtlich zu finden. Während Küchemann (1978) ihn noch recht ähnlich „Buchstabe als verallgemeinerte Zahl“ nennt, bezeichnet Malle (1993)

diesen Aspekt als „Simultanaspekt“ und bei Drijvers (2003) liest man „generalisierte Zahl“. Franke und Wynands (1991) bezeichnen diesen Aspekt als „Aspekt R“. Vollrath und Weigand (2007) nennen ihn zwar in einer Aufzählung verschiedener Begriffe für Variablen, führen ihn jedoch nicht weiter aus, wenngleich er in ihrem Kapitel zu „Schwierigkeiten beim Lernen der Formelsprache“ indirekt angesprochen wird. Die *Variable als Veränderliche* findet man in dieser Bezeichnung bei Malle (1993) und Drijvers (2003). Küchemann (1978) nennt dies „Buchstabe als Variable“, bei Freudenthal (1983) liest man von „Variable in Funktionen“ und Franke und Wynands (1991) bezeichnen dies als „Aspekt F“. Was bei Malle (1993) „Kalkülaspekt“ heißt, könnte man mit der *Variable als Symbol* im Sinne von Drijvers (2003) vergleichen, da die Variable jeweils als bedeutungsloses Zeichen angesehen wird, mit dem gerechnet werden kann. Der Unterschied dieser beiden letztgenannten Begriffe liegt jedoch darin, dass bei Malle (1993) auch die Rechnung an sich impliziert ist und durchaus unter diesem Aspekt durchgeführt werden kann, wohingegen Drijvers (2003) Ausführung zur Variable als Symbol eher dafür spricht, dass man zwar wahrnimmt, dass dort ein mathematisches Zeichen ist und theoretisch etwas gerechnet werden könnte, diese Handlung jedoch nicht durchgeführt wird.

In der Zusammenfassung wurde deutlich, dass viele Aspekte von Variablen sich trotz verschiedener Bezeichnungen recht ähnlich sind. Um das Variablenverständnis, welches dieser Arbeit zugrunde liegt, besser nachvollziehen zu können, werden im Folgenden die Aspekte der verschiedenen Autoren in einer eigenen Systematisierung vereint. Im Wesentlichen baut das Modell (vgl. Abb. 18) auf den Überlegungen von Malle (1993) auf und bezieht einzelne Aspekte anderer Autoren mit ein.

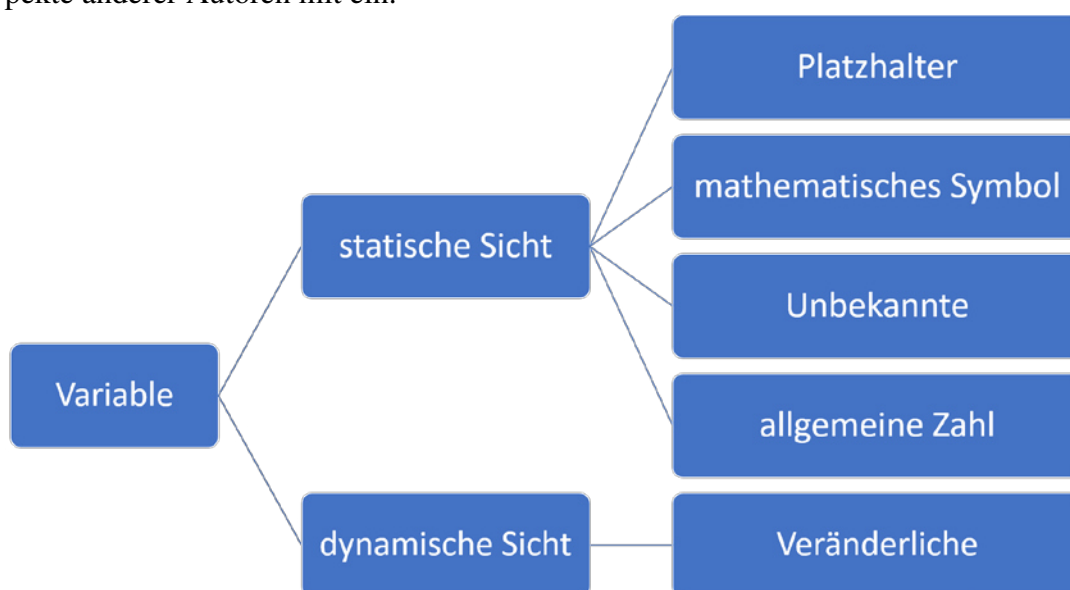


Abbildung 18: Einteilung der verschiedenen Variablenaspekte

Die Einteilung in eine statische und eine dynamische Sicht auf Variablen wurde bereits als ein zentraler Aspekt herausgearbeitet und findet sich daher auch in der Einteilung der verschiedenen Variablenaspekte wieder. Es fällt auf, dass die meisten Aspekte der statischen Sichtweise zugeordnet werden können. Dies bedeutet aber nicht, dass die statische Sichtweise wichtiger wäre als die dynamische Sichtweise. Vielmehr sollten beide Sichtweisen gleichberechtigt und als sich gegenseitig ergänzend betrachtet werden. Im Folgenden werden die einzelnen Variablenaspekte vorgestellt und jeweils anhand eines Beispiels erläutert:

- *Variable als Platzhalter*

Hierbei wird die Variable lediglich als Platzhalter angesehen, welche den „Platz frei hält“, um irgendwann eine bestimmte Zahl oder einen beliebigen Term einzusetzen.

Beispiel: $10\text{€} + 0,09\text{€} \cdot x$

Die Variable x steht an dieser Stelle, um den Platz für eine Zahl (oder eine Verkettung von Zahlen bzw. Variablen) freizuhalten. In diesem einfachen Beispiel könnte es sich um einen Handytarif (Grundgebühr + Kosten für Telefonie) oder eine Abrechnung auf dem Bankkonto (Kontoführungsgebühr + Transaktionskosten) handeln, bei denen jeweils eine Grundgebühr und ein Preis pro verbrauchter Einheit veranschlagt wird. Setzt man den tatsächlichen Verbrauch ein, so erhält man den zu zahlenden Wert. Bei acht Minuten bzw. acht Transaktionen erhalte man entsprechend $10\text{€} + 0,09\text{€} \cdot 8 = 10,72\text{€}$.

- *Variable als mathematisches Symbol*

Hierbei wird die Variable lediglich als mathematisches Symbol wahrgenommen. Befindet sich die Variable in einer Gleichung, so kann die Gleichung unter bekannten Rechengesetzen umgeformt werden, wobei die Variable dabei als bedeutungsloses Zeichen behandelt wird. Dies ist jedoch nur bei einer Variable möglich, daher ist dieser Aspekt aufgrund mangelnder Verallgemeinerung nicht tragfähig.

Beispiel: $7y + 8 = 15$

Einerseits kann die Variable als mathematisches Symbol erkannt werden und man könnte sich infolgedessen von der Gleichung abwenden. Andererseits kann die Gleichung auch mit bekannten Rechenregeln umgeformt werden. In diesem Fall würde auf beiden Seiten der Gleichung 8 subtrahiert werden und anschließend durch 7 geteilt werden. Diese Rechenoperationen lassen sich durchführen, ohne der Variable eine besondere Bedeutung zuzuschreiben.

- *Variable als Unbekannte*

Hierbei steht die Variable für ein Objekt (eine Zahl oder ein Term), welches noch unbekannt ist, prinzipiell jedoch bestimmt werden kann.

Beispiel: $2z + 4 = 8$

Die Variable z steht für eine Zahl, welche zwar noch nicht bekannt ist, jedoch bestimmt werden kann. In diesem Fall muss für z gelten, dass $z + 2 = 4$ ist. Diese Gleichung erhält man, wenn man beide Seiten der ursprünglichen Gleichung durch 2 dividiert. Die gesuchte Zahl addiert mit 2 ergibt also 4. Somit ist 2 die gesuchte Zahl.

- *Variable als allgemeine Zahl*

Hierbei spielt es keine Rolle, welche Zahl für die Variable eingesetzt wird, da die Gleichung allgemeingültig ist. Die Variable steht also für alle Zahlen aus der Grundmenge der Gleichung. Entsprechende Gleichungen gelten immer, unabhängig davon, welche Zahlen eingesetzt werden. Dieser Aspekt kommt daher insbesondere auch in Rechengesetzen vor.

Beispiel: $a \cdot (b + c) = a \cdot b + a \cdot c$

Für die Variablen a , b und c können beliebige Zahlen eingesetzt werden und die Gleichung (in diesem Fall das Distributivgesetz) ergibt jeweils eine wahre Aussage.

- *Variable als Veränderliche*

Hierbei wird die Variable in funktionalen Zusammenhängen verwendet, in denen tatsächlich etwas variiert wird bzw. in denen die Veränderung betrachtet wird. Es wird also betrachtet, was mit dem Funktionswert passiert, wenn die Variable systematisch bzw. gleichmäßig verändert wird.

Beispiel: $x \mapsto 2x + 4$

Es wird betrachtet, was mit dem Funktionswert passiert, wenn die Variable x (systematisch bzw. gleichmäßig) variiert wird. Hierzu könnten z.B. die natürlichen Zahlen von -3 bis $+3$ nacheinander eingesetzt werden:

Tabelle 1: Werte und Funktionswerte für die Variable x im Beispiel für die Variable als Veränderliche

x	$2x + 4$
-3	$2 \cdot (-3) + 4 = -2$
-2	$2 \cdot (-2) + 4 = 0$
-1	$2 \cdot (-1) + 4 = 2$
0	$2 \cdot 0 + 4 = 4$
1	$2 \cdot 1 + 4 = 6$
2	$2 \cdot 2 + 4 = 8$
3	$2 \cdot 3 + 4 = 10$

Man stellt fest, dass der Funktionswert jeweils um genau 2 zunimmt, wenn die eingesetzte Zahl um 1 vergrößert wird (vgl. Tab. 1).

Betrachtet man nun noch einmal das Modell zur Einteilung der verschiedenen Variablenaspekte (vgl. Abb. 18), so kann dieses nun, da die einzelnen Variablenaspekte erläutert wurden, im Hinblick auf die Schule und die Lernentwicklung der Schülerinnen und Schüler interpretiert werden. Bereits in der Grundschule lernen diese implizit die Variable als Platzhalter kennen, indem sie Zahlen in leere Rechtecke, Wolken o.Ä. einsetzen. Anschließend werden sie in der Sekundarstufe 1 zunächst mit der Variable als mathematisches Symbol konfrontiert. Da diese Vorstellung jedoch nicht tragfähig ist, müssen Schülerinnen und Schüler den notwendigen Zwischenschritt durch die Formalisierung nehmen und die Konzepte Variable als Unbekannte und Variable als allgemeine Zahl durchdringen. Auf dieser Grundlage kann die Variable als Veränderliche wahrgenommen und genutzt werden, um Strukturen zu erfassen und mathematisch beschreiben zu können. Das Variablenverständnis wird schließlich auch im nächsten Abschnitt, bei dem die Grundvorstellungen zu Termen vorgestellt werden, teilweise implizit und teilweise explizit, aufgegriffen bzw. genutzt werden.

6.2 Terme

In vielen Lexika wird ein Term als ein mathematischer Ausdruck angesehen, welcher aus Zahlen, Variablen, Symbolen für mathematische Verknüpfungen und Klammern besteht (Barzel & Herget 2006, 4). Auch in Schulbüchern findet man eine solche Definition:

Terme sind Rechenausdrücke, die aus Zahlen, Variablen, Rechenzeichen und Klammern zusammengesetzt sind. (Körner, Lergenmüller, Schmidt & Zacharias 2016, 41)

Darüber hinaus enthalten auch Lehrbücher sehr ähnlich klingende Definitionen, wie beispielsweise:

Unter einem Term versteht man einen Rechenausdruck, der einen Rechenvorgang beschreibt. Er kann mithilfe der bekannten Rechenzeichen und mit Klammern definiert werden. (Rießinger 2016, 3)

Die große Ähnlichkeit der drei Definitionen, welche in verschiedenartigen Veröffentlichungen zu lesen sind, lässt den Schluss zu, dass der Begriff „Term“ einheitlich definiert zu sein scheint. Dies verführt im ersten Moment dazu, eine solche Definition als gegeben und eventuell sogar als vollumfänglich anzusehen, was jedoch mitnichten der Fall ist. Während diese Art der Definition für die Vermittlung im Schulkontext zunächst ausreichend zu sein scheint, genügt sie nicht den Ansprüchen der Fachmathematik. In der mathematischen Fachliteratur werden die Begriffe Variable und Term vorausgesetzt und nicht vorgestellt, stattdessen wird direkt damit umgegangen. Auch für Fachdidaktiker sind damit selbstverständlich bei weitem nicht alle Facetten des Begriffs angesprochen, geschweige denn erläutert. Unter anderem könnte an dieser Stelle die Unterscheidung eines Terms als operationalem Prozess oder als strukturelles Objekt genannt werden (Drijvers 2003, 44). Einige weitere Aspekte des Termbegriffes, welche deutlich über die oben genannten Definitionen hinausgehen, werden nachfolgend vorgestellt. Dabei sollen zunächst die Sichtweisen einiger Autoren auf Terme erläutert und diese anschließend diskutiert werden. Insgesamt fällt auf, dass nur überschaubar viele Autoren sich ernsthaft mit dem Termbegriff auseinandergesetzt haben.

Malle (1993, 56) unterscheidet im Hinblick auf Terme zwei Aspekte des Algebraunterrichts, nämlich Terme aufzustellen und mit Termen zu rechnen. Darüber hinaus verweist er auf die Nützlichkeit von Termen, welche durchaus auch von Schülerinnen und Schülern erkannt werden kann und soll (Malle 1993, 56ff.). Zunächst einmal lassen sich mit Termen bzw. Formeln innermathematische Prozesse und Gesetzmäßigkeiten allgemein beschreiben (Malle 1993, 57). Mit ihnen können außermathematische Sachverhalte verallgemeinert, also modelliert, und Situationen erforscht, also allgemeine Einsichten in besondere Situationen erlangt, werden (Malle 1993, 57ff.). Außerdem können mithilfe von Termen bzw. Formeln abstrakte Problemlösungen geplant und Probleme allgemein gelöst werden (Malle 1993, 59ff.). Darüber hinaus können mit Hilfe von Termen allgemeingültige Argumentationen geführt sowie Wissens Elemente übermittelt werden. Dadurch kann es gelingen, Sachverhalte auf abstrakter Ebene darzustellen (Malle 1993, 64).

Vollrath und Weigand (2007, 82ff.) benennen ebenfalls zwei verschiedene Sichtweisen auf Terme, nämlich *Term als Rechenschema* und *Term als Bauplan*. In Sachsituationen werden Terme mit Variablen meist als Rechenschemata genutzt, was nach Vollrath und Weigand (2007, 83) einer semantischen Sicht entspricht. Die Sichtweise *Term als Rechenschema* fokussiert also eher den Inhalt des Terms (Vollrath & Weigand 2007, 83ff.). Betrachtet man hingegen genauer, welche Größen in welcher Weise in die Berechnung eingehen, entspricht dies der Sichtweise *Term als Bauplan* (Vollrath & Weigand 2007, 84f.). Bei Vollrath und Weigand (2007, 84f.)

sind damit Berechnungen gemeint, welche schließlich Einsichten über die auftretenden Größen und deren Beziehungen liefern. Es handelt sich dabei um eine syntaktische Sicht, sprich um die Annahme eines Regelsystems zur Kombination elementarer Zeichen. Das Erfassen des Bauplans kann dabei durch bildhafte Darstellungen unterstützt werden (Vollrath & Weigand 2007, 84).

Termumformungen werden unter der Sichtweise Term als Rechenschema genutzt, um zu überprüfen, ob unterschiedliche Schemata zu gleichen Ergebnissen führen (Vollrath & Weigand 2007, 85). Mit Gleichheit ist dabei die Wertgleichheit, sprich die Äquivalenz von Termen, gemeint (Vollrath & Weigand 2007, 85). Dies führt schließlich auch zu Vereinfachungen von Termen (Vollrath & Weigand 2007, 85). Betrachtet man Termumformungen als „Baupläne in syntaktischer Sicht“ (Vollrath & Weigand 2007, 85), sieht man die Termumformung als zulässige Veränderung des Plans an, welche zu neuen Erkenntnissen führt. Hierbei wird versucht, in einem Term etwas syntaktisch Sinnvolles zu sehen (Vollrath & Weigand 2007, 85). „Man kann syntaktisch natürlich Termumformungen auch rein schematisch als zulässige Umformungen betrachten“ (Vollrath & Weigand 2007, 85).

Siller und Roth (2016) bauen auf die Überlegungen von Vollrath und Weigand auf und benennen zwei Grundvorstellungen zu Termen analog zu Vollrath und Weigand mit „Term als Bauplan“ und „Term als Rechenschema“. Rechenschema bedeutet hierbei, dass eine Formel genutzt wird, um etwas zu berechnen. Wenn also das entsprechende Schema bereits bekannt ist, müssen nur die entsprechenden Größen eingesetzt werden, um das Ergebnis zu berechnen. Darüber hinaus kann eine Formel auch in eine andere Darstellung überführt werden, wenn dies das Einsetzen der Zahlenwerte und die anschließende Berechnung vereinfacht. Bei der Grundvorstellung „Term als Bauplan“ wird mithilfe des Terms ein Plan erstellt, welcher eine gewünschte Struktur wiedergibt. Es geht darum, wie etwas bestimmt oder strukturiert werden kann. Das Ergebnis durch Einsetzen konkreter Zahlen spielt hier höchstens zur Überprüfung, ob die Struktur bzw. der Bauplan korrekt erstellt wurde (bzw. äquivalent zu anderen Bauplänen ist), eine Rolle.

Termumformungen dienen nach der Grundvorstellung „Term als Rechenschema“ dazu, einen Term derart umzuformen, dass nach dem Einsetzen von Zahlenwerten möglichst wenige Rechenschritte benötigt werden um den Termwert zu bestimmen. Bei der Grundvorstellung „Term als Bauplan“ geht es hingegen darum, ob ein Term von einer Struktur in eine andere umgeformt werden kann. Diese andere Struktur dient, sofern diese adäquat interpretiert wird, dem Erkenntnisgewinn. Die Gleichheit von Termen kann daher einerseits als Wertgleichheit aufgefasst werden (Term als Rechenschema) und andererseits als Strukturgleichheit (Term als Bauplan).

Sinngemäß entspricht die Unterscheidung von „Term aufstellen“ und „mit Termen rechnen“ von Malle (1993) auch den beiden Grundvorstellungen „Term als Bauplan“ und „Term als Rechenschema“, wenngleich sowohl Vollrath und Weigand (2007, 82ff.) als auch Siller und Roth (2016) diese jeweils noch ausführlicher beschreiben und ausdifferenzieren. Vollrath und Weigand (2007) verwenden die Begriffe der „Semantik“ und der „Syntax“, wobei die semantische Sichtweise auf Terme eine eher inhaltsorientierte Sicht ist, welche Vollrath und Weigand der Grundvorstellung „Term als Rechenschema“ zuordnen, und die syntaktische Sichtweise eher auf die Form fokussiert und von Vollrath und Weigand der Grundvorstellung „Term als Bauplan“ zugeordnet wird (Vollrath und Weigand 2007, 78ff.). Siller und Roth (2016) kommen ohne diese beiden Begriffe aus und sprechen im Fall der Grundvorstellung „Term als Bauplan“ davon, dass die jeweilige Struktur eines Phänomens oder eines Problems erkannt und durch-

drungen werden muss. Im Mittelpunkt steht das Erfassen von Strukturen und in diesem Zusammenhang auch das Aufstellen oder Umformen von Termen. Die Grundvorstellung „Term als Rechenschema“ kommt hingegen in Anwendungssituationen zum Tragen und kann dort genutzt werden, um wiederholte, gleichartige Berechnungen für verschiedene Werte möglichst schnell und einfach durchzuführen (Siller & Roth 2016). Termumformungen dienen hierbei der Vereinfachung des Rechenprozesses. Die beiden Grundvorstellungen werden von Vollrath und Weigand (2007) und Siller und Roth (2016) also inhaltlich sehr ähnlich beschrieben, jedoch werden unterschiedliche Begriffe zur Veranschaulichung verwendet und andere Schwerpunkte gesetzt.

Zur Vermittlung der beiden Grundvorstellungen in der universitären Lehre müssen sie möglichst gut handhabbar gemacht werden. Dabei helfen plakative Formulierungen und Vorstellungen, wie beispielsweise „Terme aufstellen“ und „mit Termen rechnen“. Im weiteren Verlauf der Arbeit werden die beiden Grundvorstellungen „Term als Bauplan“ und „Term als Rechenschema“ auf Grundlage der obigen Ausführungen folgendermaßen verwendet:

- **Term als Bauplan**

Bei der Grundvorstellung Term als Bauplan werden Terme entweder erstellt (und somit das vorliegende Phänomen bzw. die vorliegende Situation beschrieben) oder umgeformt. Die Termumformungen entsprechen hierbei Änderungen des Bauplans, welche sich auf die jeweilige Struktur des Phänomens oder des Problems beziehen. Mittels der Grundvorstellung Term als Bauplan können Terme unter anderem auch ineinander überführt werden und deren Gleichheit gezeigt werden. Die Idee ist also das Schaffen oder Durchdringen der jeweiligen Struktur. Terme werden folglich bei der Bearbeitung der Frage nach der mathematischen Struktur eines Phänomens oder eines Problems entwickelt. Die darin enthaltenen Variablen werden von Novizinnen und Novizen häufig unter dem Veränderlichenaspekt betrachtet, da die Variablen zur Durchdringung der Struktur systematisch bzw. gleichmäßig variiert werden. Dies ist anfangs notwendig, um den Aufbau des Terms und somit auch die Bedeutung der einzelnen Variablen für den Term ergründen zu können. Expertinnen und Experten können schließlich aufgrund ihrer Erfahrung oftmals direkt (also ohne den Umweg über die systematische bzw. gleichmäßige Variation) die Struktur des Phänomens erfassen und einen passenden Term dazu aufstellen.

Beispiel: Plättchenmuster

In Abbildung 19 wurde ein regelmäßig wachsendes Punktmuster in Form von Quadraten mit Randpunkten gelegt. Will man nun bestimmen, wie viele Punkte die jeweiligen Muster haben, so kann man dies für einzelne Muster noch durch Zählen herausfinden. Das erste Quadrat besteht aus acht, das zweite aus 12 und das dritte aus 16 Punkten. Die Frage, aus wie vielen



Abbildung 19: Regelmäßig wachsendes Punktmuster

Punkten ein Punktmuster mit beliebiger Kantenlänge besteht, kann jedoch nur beantwortet werden, wenn man die mathematische Struktur der Punktmuster durchdringt. Um die Struktur des Phänomens zu entdecken, kann man die Variable systematisch bzw. gleichmäßig variieren und die Veränderung betrachten. In diesem Beispiel erkennt man unter anderem, dass die Gesamtzahl der Plättchen sich jeweils um 4 erhöht, wenn man die Anzahl der Plättchen an einer Kante um 1 erhöht. Auf dieser Grundlage kann schließlich ein Term erstellt werden und ggf. durch konkrete Zahlen überprüft werden. Man könnte exemplarisch das mittlere Quadrat betrachten,

die Kantenlänge mit der Anzahl der Seiten multiplizieren und dann die Anzahl der Ecken abziehen. Ein möglicher Term dafür wäre $4x - 4$. Für dasselbe Phänomen lassen sich aber häufig auch andere Baupläne aufstellen und so weitere Strukturen entdecken. In dem Plättchenmuster-Beispiel könnte man etwa auch einen Term aufstellen, der die Ecken nicht doppelt miteinbezieht und daher nicht abziehen muss. Dies erhält man beispielsweise, indem man die Kantenlänge um 1 subtrahiert und wieder mit der Anzahl der Seiten multipliziert, in Termschreibweise $4(x - 1)$. Nun kann man sich die Frage stellen, ob beide Baupläne auch wirklich dasselbe Phänomen beschreiben. Inhaltlich sollten beide Terme gleich sein, da sie dieselbe Punktzahl beschreiben. Dies kann auch mathematisch gezeigt werden, indem einer der beiden Terme durch zulässige Termumformungen in den anderen Term überführt wird. In diesem Fall ist dafür nur ein Rechenschritt nötig. Entweder wird im Term $4(x - 1)$ die Klammer ausmultipliziert oder im Term $4x - 4$ die 4 ausgeklammert. Neben den beiden genannten Bauplänen gibt es noch viele weitere Baupläne, die ebenfalls das vorliegende Phänomen beschreiben und entsprechend auch ineinander überführt werden können.

- **Term als Rechenschema**

Bei der Grundvorstellung Term als Rechenschema geht es hauptsächlich darum, etwas auszurechnen. Dabei kann entweder ein bereits aufgestellter bzw. bekannter Term genutzt werden, um Zahlen für die darin enthaltenen Variablen einzusetzen, oder dieser derart umgeformt werden, dass er für den vorliegenden Zweck geeignet ist. Termumformungen dienen hierbei der Vereinfachung der Berechnung. Dies kann beispielsweise dadurch geschehen, dass nach der Termumformung die Anzahl der Variablen reduziert wird oder weniger Einsetzungen und/oder Rechenschritte vorgenommen werden müssen. Terme werden hier also genutzt, um wiederholte, gleichartige Berechnungen für verschiedene Werte schnell und einfach durchführen zu können. Variablen werden hierbei oft als Platzhalter verstanden, da lediglich Zahlen für die Variablen eingesetzt werden und schließlich das Ergebnis betrachtet wird. Somit halten die Variablen den Platz für die konkreten Zahlen frei.

Beispiel: Body-Mass-Index

Will man seinen Body-Mass-Index (BMI) bestimmen, so nutzt man den dafür vorgegebenen Term $\frac{g}{h^2}$, wobei g für das Körpergewicht in Kilogramm und h für die Körpergröße in Metern steht. Wie der Term zustande kommt, ist in diesem Fall nicht relevant. Es geht darum, einen Wert zu bestimmen. Dies geschieht, indem für die beiden Variablen konkrete Zahlenwerte, in diesem Fall die des eigenen Körpergewichts und der eigenen Körpergröße, eingesetzt werden. Rechnet man den Term anhand der konkreten Zahlen aus, so erhält man die gewünschte Kenngröße des BMI.

Beispiel: Berechnung des Durchschnittsteuersatzes

Interessiert man sich dafür, wie hoch der persönliche durchschnittliche Steuersatz ist, so kann man dies mit Hilfe des folgenden Terms herausfinden, indem man die entsprechenden Zahlen für die Variablen einsetzt:

$$\left(\frac{S}{E}\right) \cdot 1,145 \cdot 100\%$$

S steht hierbei für den Steuerbetrag, welcher anhand offizieller Steuertabellen ermittelt werden kann und E steht für das zu versteuernde Einkommen. Setzt man beispielsweise 40.000 € für

das zu versteuernde Einkommen und 8.766 € für den Steuerbetrag ein, so erhält man einen durchschnittlichen Steuersatz von 25,09%. Da lediglich von Interesse ist, wie hoch der durchschnittliche Steuersatz ist, werden die Werte in den Term eingesetzt und dann das Ergebnis betrachtet. Sobald der Term hinterfragt und dessen Struktur reflektiert wird, wird er nicht mehr mit der Grundvorstellung Term als Rechenschema, sondern mit der Grundvorstellung Term als Bauplan angesehen. Man könnte sich über das Zustandekommen des Multiplikators 1,145 wundern und vielleicht zur Lösung kommen, dass zusätzlich zur Einkommenssteuer noch 5,5% Solidaritätszuschlag sowie 9% Kirchensteuer zu erbringen sind, jeweils in Bezug zur Einkommenssteuer und vorausgesetzt, dass überhaupt Kirchensteuer gezahlt werden muss, und unter Vernachlässigung der Tatsache, dass diese in einzelnen Bundesländern unterschiedlich hoch ausfällt. Steht der Bauplan bzw. die Struktur des Phänomens im Fokus, könnte danach gefragt werden, ob der folgende Bauplan äquivalent zu obigem Bauplan ist:

$$\left(\left(\frac{S}{E} \right) + \left(\frac{S}{E} \right) \cdot 0,055 + \left(\frac{S}{E} \right) \cdot 0,09 \right) \cdot 100\%$$

Dazu würde man eine Darstellungsweise in die andere überführen und feststellen, dass beim zweiten Term sichtbar wird, dass die Einkommenssteuer, der Solidaritätszuschlag und die Kirchensteuer jeweils einzeln abgebildet sind. An dieser Stelle ist es natürlich möglich, noch einige weitere Baupläne zu erstellen und zu betrachten.

Mit der Grundvorstellung Term als Rechenschema ließe sich der Term zu dem obigen Ausgangsterm vereinfachen. Intention dabei könnte die Verschlinkung des Terms sein, um beim Einsetzen von Zahlen in den Term möglichst schnell einen konkreten Wert zu erhalten. Termumformungen sind also auch unter der Sichtweise Term als Rechenschema denkbar, jedoch mit dem Ziel, künftig Arbeit und Zeit zu sparen und somit wiederholte, gleichartige Rechnungen für verschiedene Werte schnell und einfach durchführen zu können.

7 Forschungsfragen

Sowohl die Fähigkeit, den Lernstand von Schülerinnen und Schülern möglichst genau einschätzen zu können, als auch die Fähigkeit, ihnen möglichst passgenaue Hilfestellungen geben zu können, sofern sie diese benötigen, sind für eine Lehrkraft wesentlich, um die Lernenden individuell fördern zu können. In Kapitel 2 wurde herausgearbeitet, dass Diagnosen in der Regel auch konkrete Handlungen zur Folge haben sollten, und in Kapitel 3 wurde die Diagnose als notwendige Voraussetzung herausgestellt, um adäquat in Gruppenarbeitsprozesse intervenieren zu können. Entsprechend erscheint es naheliegend, dass es einen Zusammenhang zwischen diesen beiden Fähigkeiten gibt. Dieser Zusammenhang ist bisher jedoch lediglich eine geteilte Annahme und muss noch empirisch untersucht und bestätigt werden.

Ähnliches gilt für die Reflexion, welche in Kapitel 4 beschrieben wurde. Es herrscht weitgehend Einigkeit darüber, dass Reflexion dazu beiträgt, dass Lehrkräfte ihr Unterrichtshandeln verbessern (können). Insbesondere für die Interventionen in die Gruppenarbeitsprozesse ist eine nachfolgende Reflexion wichtig, um künftig in ähnlich gelagerten Situationen verschiedene Handlungsoptionen zu besitzen und sich schließlich nach reiflicher Überlegung und ggf. auch nachdem einige Optionen ausprobiert wurden, bewusst für eine bestimmte Handlungsoption zu entscheiden. Auch hinsichtlich der Diagnostizität erscheint es zielführend, regelmäßig die eigenen Diagnosen zu hinterfragen, um systematisch Einflussfaktoren, die zu einer verzerrten Wahrnehmung führen können, zu vermeiden.

In Kapitel 5 wurde die fachdidaktische Ausbildung im Fach Mathematik am Campus Landau vorgestellt. Dabei wurde aufgezeigt, dass die fachdidaktische Ausbildung sich über das komplette Studium verteilt und systematisch aufeinander aufbaut. Begleitend zu den Großveranstaltungen im Bachelorstudium werden Videovignetten angeboten, die inhaltlich zu den Vorlesungsinhalten passen. Diese Videovignetten sind in ViviAn eingebunden, wodurch den Studierenden auch zahlreiche Hintergrundinformationen zur Verfügung stehen. Dadurch wird versucht, die Situation möglichst nahe an die Realsituation in der Schule anzulehnen, wenngleich dies natürlich nur bedingt möglich ist. Nichtsdestotrotz ist dies eine gute Möglichkeit, um einerseits die fachdidaktischen Inhalte festigen und diese andererseits in einer konkreten Situation anwenden zu können und somit die diagnostische Fähigkeit angehender Lehrkräfte auf- bzw. auszubauen.

Im Didaktischen Seminar, welches im Masterstudium der Mathematik und somit gegen Ende des Studiums angesiedelt ist, werden die Studierenden aufgefordert, Videovignetten in ViviAn zu bearbeiten. Hierbei können sie ihre Erfahrungswerte aus den Bachelorveranstaltungen nutzen, um sicher mit ViviAn umzugehen. Inhaltlich geht es um das Thema „Terme“ (Kap. 6), das im Bachelorstudium noch nicht Gegenstand der Videovignetten war. Darüber hinaus müssen die Studierenden eine Lernumgebung (Laborstation im Mathematik-Labor „Mathe ist mehr“) konzipieren und diese mit einer Schulklasse erproben. Dabei müssen sie auch in Gruppenarbeitsprozesse intervenieren und anschließend in Kleingruppen über die Erprobung sowie über die getätigten Interventionen reflektieren.

Zentrales Forschungsanliegen dieser Arbeit sind die Zusammenhänge zwischen der prozessdiagnostischen Fähigkeit der angehenden Lehrkräfte und deren Interventionen in Gruppenarbeitsprozesse von Schülerinnen und Schülern. Die Interventionen können dabei unter anderem auch als Performanz der Diagnostischen Kompetenz angesehen werden (vgl. dazu Kap. 2 und Kap. 3). Die Vermutung, dass erhöhte diagnostische Fähigkeiten dazu führen, dass Lehrkräfte

adaptivere Hilfen, insbesondere auch bei der Gruppenarbeit, geben können, erscheint naheliegend und wurde in Kapitel 2 und Kapitel 3 ausführlich beschrieben. Da die Reflexion wesentlich zum Aufbau der beiden Fähigkeiten beiträgt, ist auch hier der Zusammenhang zwischen den getätigten Diagnosen und dem Reflexionsverhalten bzw. zwischen den durchgeführten Interventionen und dem Reflexionsverhalten von Interesse. Hypothese dabei ist, dass Studierende mit erhöhter prozessdiagnostischer Fähigkeit auch anders reflektieren als Studierende mit niedriger Ausprägung. Ebenso wird ein Zusammenhang zwischen bestimmtem Interventionsverhalten und dem Reflexionsverhalten vermutet. Hierbei sollen explorativ mögliche Zusammenhänge aufgedeckt werden.

Aufbauend auf die theoretischen Überlegungen und den aktuellen Forschungsstand (vgl. dazu Kap. 2-4), widmet sich diese Arbeit der folgenden Fragestellung:

Worin unterscheiden sich Interventionsverhalten und Reflexionsgespräche von Studierenden mit divergierender prozessdiagnostischer Fähigkeit „Deuten“?

Die übergeordnete Forschungsfrage wird in folgende Teilfragen (TF) untergliedert:

TF 1) Wie ausgeprägt ist die *prozessdiagnostische Fähigkeit* „Deuten“ bei den Studierenden?

TF 2) Wie *intervenieren* Studierende in Gruppenarbeitsprozesse von Schülerinnen und Schülern?

TF 3) Wie *reflektieren* Studierende über die Erprobung einer Lernumgebung und über die von ihnen getätigten Interventionen?

TF 4) Wie unterscheiden sich *Interventionen* von Studierenden mit unterschiedlich stark ausgeprägter *prozessdiagnostischer Fähigkeit* „Deuten“?

TF 5) Wie unterscheiden sich die *Reflexionsgespräche* von Studierenden mit unterschiedlich stark ausgeprägter *prozessdiagnostischer Fähigkeit* „Deuten“?

TF 6) Wie unterscheiden sich die *Reflexionsgespräche* von Studierenden in Abhängigkeit von den getätigten *Interventionen*?

Die ersten drei Teilfragen zielen auf die Bereiche Diagnose, Intervention und Reflexion, wobei diese hierbei zunächst noch separiert voneinander betrachtet werden. Anschließend werden jeweils zwei der drei Bereiche paarweise beleuchtet.

8 Methodisches Vorgehen

Die folgenden Teilkapitel geben einen Überblick über das methodische Vorgehen bei der Erhebung und Analyse der in den drei Bereichen Diagnose, Intervention und Reflexion vorliegenden Daten. Zunächst wird die Qualitative Inhaltsanalyse vorgestellt, da diese bei der Auswertung im Bereich Diagnose, im Bereich Intervention und im Bereich Reflexion genutzt wurde (Kap. 8.1).

Im Anschluss werden die in den drei Bereichen genutzten Auswertungsverfahren jeweils gesondert dargelegt (Kap. 8.2-8.4.). So wird im Bereich der Diagnose zunächst auf die Auswahl der Videovignetten, die Erstellung von dazu passenden Diagnoseaufträgen, die Durchführung eines Expertenratings sowie die Einbettung der Videovignetten in ViviAn eingegangen. Anschließend wird beschrieben, wie die prozessdiagnostische Fähigkeit ermittelt wird und als zusätzliches Gütekriterium eine quantitative Validierung des Testinstrumentes vorgenommen.

Im Bereich Intervention wird die Einbettung in das Didaktische Seminar und damit verbunden auch die Datengenerierung erläutert. Es werden verschiedene Analyseverfahren vorgestellt, insbesondere die Analyse der innerhalb der Intervention getätigten Hilfen und die Analyse der Angemessenheit der Interventionen. Im Bereich der Reflexion wird zunächst ebenfalls auf die Einbettung in das Didaktische Seminar eingegangen. Nach der Beschreibung der Datengenerierung wird das Analyseverfahren erläutert. Abschließend werden in Kapitel 8.5 noch die verschiedenen Populationen der Expertengruppen, Kodiererinnen und Studierenden erläutert. Zudem sollen Verschränkungen mit anderen Erhebungen aufgezeigt werden.

8.1 Qualitative Inhaltsanalyse

Die Qualitative Inhaltsanalyse stellt ein in allen drei Bereichen angewandtes Auswertungsverfahren dar und soll daher vorab dargestellt und erläutert werden. Dazu wird zunächst knapp beschrieben, was eine Qualitative Inhaltsanalyse ist und welches Ziel sie verfolgt. Anschließend werden zentrale Begriffe sowie Techniken, also Vorgehensweisen, der Qualitativen Inhaltsanalyse vorgestellt.

8.1.1 Beschreibung

Das Ziel der Inhaltsanalyse ist die „Analyse von Material, das aus irgendeiner Art von Kommunikation stammt“ (Mayring 2015, 11). Hierbei geht es um die Bedeutung von Texten bzw. Kommunikationsinhalten (Kuckartz 2016, 21). Dabei geschieht die Auswertung qualitativer Daten anhand von Techniken, welche systematisch und intersubjektiv überprüfbar sind und dabei der Komplexität bzw. der Bedeutungsfülle des sprachlichen Materials gerecht werden (Mayring 2015, 10). Das Vorgehen ist regelgeleitet und theoriegeleitet und wird von Mayring auch als „kategoriegeleitete Textanalyse“ bezeichnet (Mayring 2015, 13). Durch das systematische Vorgehen grenzt sich die Inhaltsanalyse gegen einen großen Teil hermeneutischer Verfahren ab (Mayring 2015, 12). Durch die Regelgeleitetheit kann die Analyse von Dritten nachvollzogen und überprüft werden. Die einzelnen Analyseschritte sind theoriegeleitet und auch die Ergebnisse werden anhand des jeweiligen Theoriehintergrunds interpretiert (Mayring 2015, 13). Der qualitative Teil der Analyse besteht in der Anwendung eines Kategoriensystems auf das zu untersuchende Material, welches zunächst erarbeitet und am Material erprobt werden muss (Mayring 2015, 20f.). Auf dieser Basis können dann auch quantitative Analyseschritte

vorgenommen werden (Mayring 2015, 21). Da die zugrundeliegenden Daten dafür bereits qualitativ ausgewertet und somit inhaltlich passend zur jeweiligen Fragestellung sind, ergeben die quantitativen Analysen durchaus Sinn und führen keineswegs zu einem verzerrten, am Gegenstand vorbeilaufenden Ergebnis (Mayring 2015, 21). Vielmehr werden die qualitativen Ergebnisse implizit übernommen und sind somit der Inhaltsvalidität der Ergebnisse dienlich (Mayring 2015, 21). Die Ergebnisse quantitativer Analyseschritte werden dann wieder auf die vorausgehende Fragestellung bezogen und inhaltlich interpretiert (Mayring 2015, 23). Die Qualitative Inhaltsanalyse muss sich, genau wie jede andere wissenschaftliche Methode, anhand von Gütekriterien überprüfen lassen (Mayring 2015, 27).

8.1.2 Zentrale Begriffe

Zwei wichtige Einheiten der Qualitativen Inhaltsanalyse sind die Auswahlinheit und die Analyseeinheit. Kuckartz nennt die beiden Einheiten auch Sampling Unit bzw. Unit of Analysis (Kuckartz 2016, 30). Auswahlinheiten werden nach bestimmten Auswahlverfahren aus der Menge aller Untersuchungsobjekte ausgewählt und stellen die Grundeinheit der Inhaltsanalyse dar (Kuckartz 2016, 30). Es geht also um den prinzipiellen Einschluss bzw. Ausschluss in die Studie (Kuckartz 2016, 30). Analyseeinheiten entscheiden über „die Art des Einbezugs in die inhaltliche Analyse“ (Kuckartz 2016, 30). Eine Auswahlinheit kann also mehrere Analyseeinheiten beinhalten. Als Beispiel nennt Kuckartz (2016) eine Ausgabe der Süddeutschen Zeitung (Auswahlinheit) und darin enthaltene Artikel (Analyseeinheit). Häufig umfassen die Analyseeinheiten genau die Auswahlinheit, beispielsweise im Fall eines Transkripts eines qualitativen Interviews, bei dem in der Regel keine weiteren Untereinheiten gebildet werden (Kuckartz 2016, 31). „Normalerweise entspricht eine Analyseeinheit einem Fall, prototypisch bei einer Interviewstudie, wo jedes geführte Interview eine Analyseeinheit, d.h. einen Fall darstellt“ (Kuckartz 2016, 31). In anderen Fällen, beispielsweise in Gruppendiskussionen, ist zu entscheiden, ob die Gruppe oder die einzelnen Teilnehmenden als Fall definiert werden sollen (ebd.).

Ein weiterer zentraler Begriff ist die „Kategorie“, welche meist im Sinne von „Klasse“ genutzt wird, sprich eine Kategorie ist das Ergebnis der Klassifizierung von Einheiten (Kuckartz 2016, 31). Früh (2017) betont den klassifizierenden Charakter von Kategorien für die Inhaltsanalyse:

Der pragmatische Sinn jeder Inhaltsanalyse besteht letztlich darin, unter einer bestimmten forschungsleitenden Perspektive Komplexität zu reduzieren. Textmengen werden hinsichtlich theoretisch interessierender Merkmale klassifizierend beschrieben. Bei dieser Reduktion von Komplexität geht notwendig Information verloren: Einmal durch die Ausblendung von Mittelungsmerkmalen, die die untersuchten Texte zwar besitzen, im Zusammenhang mit der vorliegenden Forschungsfrage aber nicht interessieren; zum anderen tritt ein Informationsverlust durch die Klassifikation der analysierten Mittelungsmerkmale ein. Nach angegebenen Kriterien werden je einige von ihnen als untereinander ähnlich betrachtet und einer bestimmten Merkmalsklasse bzw. einem Merkmalstypus zugeordnet, den man bei der Inhaltsanalyse 'Kategorie' nennt. Die originären Bedeutungsdifferenzen der einheitlich in einer Kategorie zusammengefassten Mittelungsmerkmale bleiben unberücksichtigt. (Früh 2017, 44)

Das Spektrum dessen, was man unter einer Kategorie versteht bzw. verstehen kann, ist sehr weit (Kuckartz 2016, 33). Eine Kategorie sollte daher möglichst genau definiert werden (Kuckartz 2016, 32). Die Gesamtheit aller Kategorien bilden ein Kategoriensystem (Kuckartz 2016, 38). Kategoriensysteme können linear, hierarchisch oder als Netzwerk aufgebaut sein

(ebd.). Ein lineares Kategoriensystem besteht aus einer Aufreihung von Kategorien, welche sich alle auf der gleichen Ebene befinden, wohingegen ein hierarchisches Kategoriensystem aus verschiedenen über- und untergeordneten Ebenen besteht (ebd.). Bei hierarchischen Kategoriensystemen spricht man von Haupt- bzw. Oberkategorien und Sub- bzw. Unterkategorien (ebd.). Bei der Netzwerkstruktur sind die Elemente auf vielfältige Art, also nicht nur hierarchisch, miteinander verknüpft (ebd.).

Ein weiterer Grundbegriff der Qualitativen Inhaltsanalyse ist der der „Kategoriendefinition“ (Kuckartz 2016, 39). „Die Konstruktion des Kategoriensystems erfordert sorgfältige Arbeit, benötigt eine Menge Zeit und ist folgenreich“ (Kuckartz 2016, 39). Kategoriendefinitionen dokumentieren die grundlegenden Elemente der Inhaltsanalyse für die Rezipienten der Studie und stellen einen Teil des Kodierleitfadens für die Kodierenden dar (ebd.). „Kodierer“ bezeichnet dabei diejenigen Personen, die die Zuordnung von Kategorien zu Teilen des Untersuchungsmaterials, beispielsweise zu Textstellen, vornehmen (Kuckartz 2016, 44).

8.1.3 Techniken

Je nach Autor gibt es viele verschiedene Arten der Qualitativen Inhaltsanalyse. Mayring (2015) unterscheidet beispielsweise die Zusammenfassung, die Explikation und die Strukturierung. Das Ziel der zusammenfassenden Analyse ist es, „das Material so zu reduzieren, dass die wesentlichen Inhalte erhalten bleiben, durch Abstraktion einen überschaubaren Corpus zu schaffen, der immer noch Abbild des Grundmaterials ist“ (Mayring 2015, 67). Bei der Explikation soll zu einzelnen fraglichen Textteilen zusätzliches Material herangetragen werden, welches das Verständnis erweitert, das die Textstelle erklärt bzw. ausdeutet (Mayring 2015, 67). Das Ziel der Strukturierung ist es, „bestimmte Aspekte aus dem Material herauszufiltern, unter vorher festgelegten Ordnungskriterien einen Querschnitt durch das Material zu legen oder das Material aufgrund bestimmter Kriterien einzuschätzen“ (Mayring 2015, 67). Die drei genannten grundlegenden Techniken werden jeweils nochmals in mehrere Techniken untergliedert. Im Folgenden wird lediglich die strukturierende Inhaltsanalyse genauer erläutert, da diese im weiteren Verlauf von Bedeutung sein wird.

Im Fall der strukturierenden Inhaltsanalyse unterscheidet Mayring (2015) zwischen der *formalen Strukturierung*, welche die innere Struktur des Materials nach bestimmten formalen Strukturierungsgesichtspunkten herausfiltert, der *inhaltlichen Strukturierung*, welche das Material zu bestimmten Themen bzw. zu bestimmten Inhaltsbereichen extrahiert und zusammenfasst, der *typisierenden Strukturierung*, welche auf einer Typisierungsdimension einzelne markante Ausprägungen im Material identifiziert und diese genauer beschreibt, und der *skalierenden Strukturierung*, welche zu einzelnen Dimensionen Ausprägungen in Form von Skalenpunkten definiert und das Material daraufhin einschätzt. Alle Formen der strukturierenden Inhaltsanalyse haben folgendes, allgemeines Ablaufschema (vgl. Abb. 20): Nach der Bestimmung der Analyseeinheit und der Festlegung der Strukturierungsdimensionen folgen die theoriegeleitete Bestimmung der Ausprägungen und auf dieser Basis die Zusammenstellung des Kategoriensystems. Darin werden die einzelnen Kategorien definiert sowie Ankerbeispiele und Kodierregeln formuliert. Es folgen zwei Materialdurchläufe, wobei zunächst die Fundstellen bezeichnet werden und anschließend bearbeitet werden. Auf dieser Grundlage können das Kategoriensystem oder einzelne Kategoriendefinitionen revidiert werden. In diesem Fall startet der Durchlauf erneut ab diesem Punkt. Müssen das Kategoriensystem und die einzelnen Kategoriendefinitionen nicht überholt werden, können die Ergebnisse aufbereitet werden.

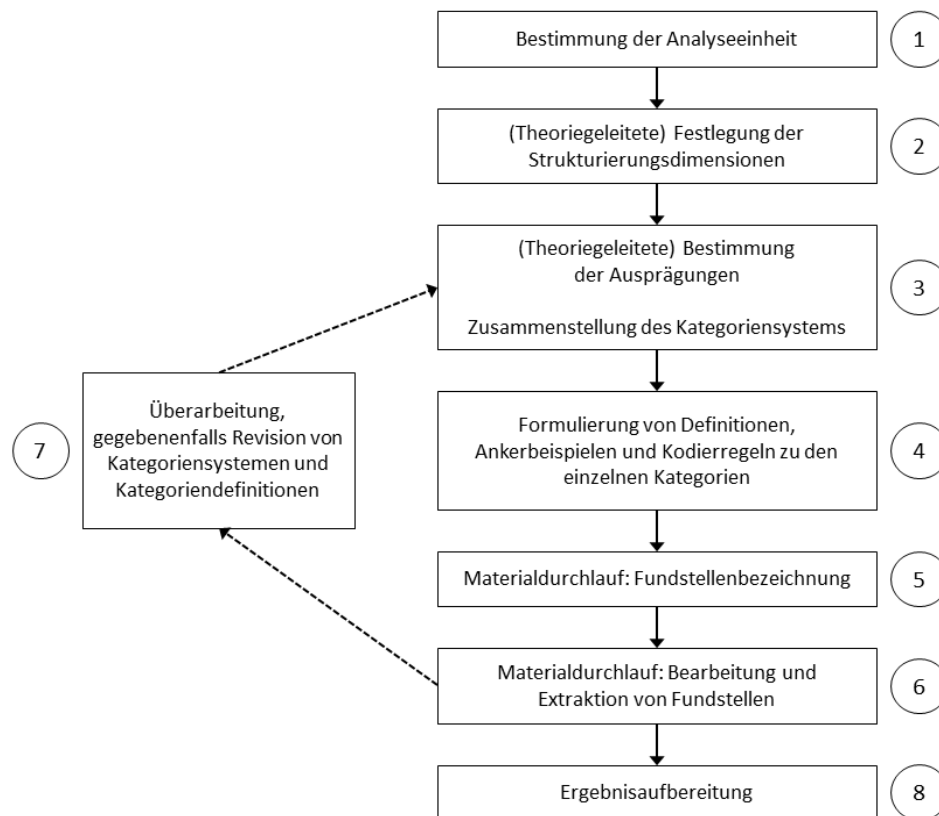


Abbildung 20: Ablaufmodell strukturierender Inhaltsanalyse (allgemein) nach Mayring (2015)

Kuckartz (2016) unterscheidet ebenfalls verschiedene Formen der Qualitativen Inhaltsanalyse. Er nennt diese „inhaltlich strukturierende“, „evaluative“ und „typenbildende“ qualitative Inhaltsanalyse (Kuckartz 2016, 97ff.). Nachfolgend wird nur die „inhaltlich strukturierende qualitative Inhaltsanalyse“ beschrieben, da diese im weiteren Verlauf Anwendung finden wird.

In der inhaltlich strukturierenden Inhaltsanalyse können Kategorien sowohl vollständig induktiv als auch vollständig deduktiv gebildet werden, wenngleich diese beiden Extreme in der Forschung eher selten Anwendung finden (Kuckartz 2016, 97). „In den meisten Fällen kommt ein mehrstufiges Verfahren der Kategorienbildung und Codierung zur Anwendung“ (Kuckartz 2016, 97). Die inhaltlich strukturierende Inhaltsanalyse lässt sich nicht nur auf leidfadenorientierte, problemzentrierte und fokussierte Interviews anwenden, sondern kann beispielsweise auch bei Gruppendiskussionen eingesetzt werden (ebd.). Das Ablaufschema sieht folgendermaßen aus (vgl. Abb. 21): Zunächst werden wichtige Textstellen markiert und thematische Hauptkategorien entwickelt. Anschließend wird das gesamte Material anhand der Hauptkategorien kodiert und alle kodierten Textstellen der Hauptkategorien werden zusammengestellt. Nun werden induktiv anhand des Materials Subkategorien bestimmt und das komplette Material wird erneut kodiert. Im Anschluss folgen weitere Analysen und die Visualisierung der Ergebnisse.

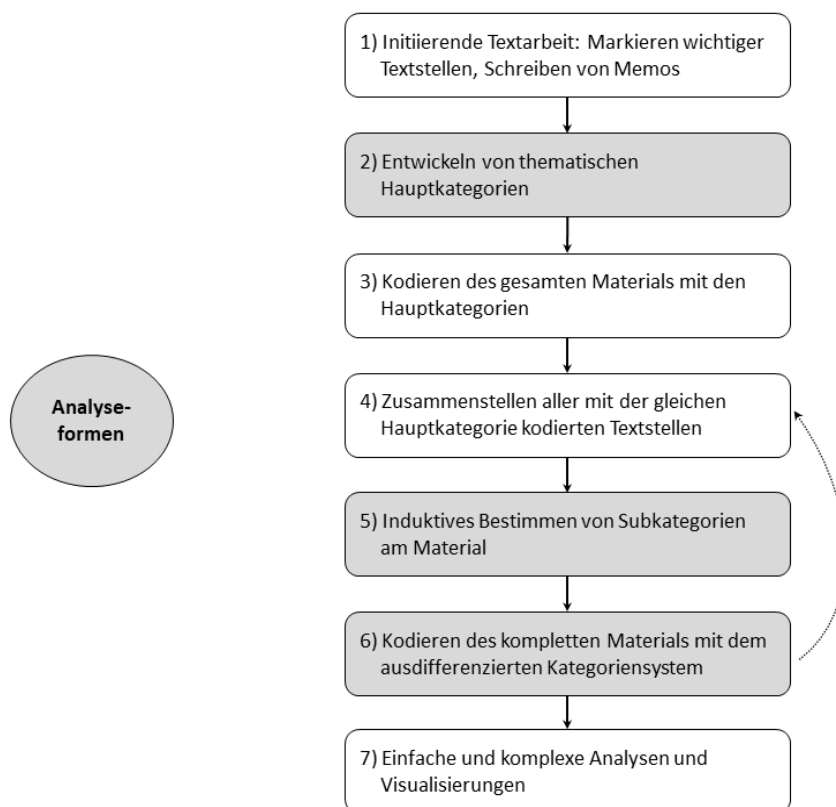


Abbildung 21: Ablaufschema einer inhaltlich strukturierenden Inhaltsanalyse nach Kuckartz (2016)

Im Wesentlichen sind sich die Techniken nach Mayring (2015) und Kuckartz (2016) recht ähnlich. Vergleicht man die strukturierende Inhaltsanalyse nach Mayring (2015) und die inhaltlich strukturierende Inhaltsanalyse nach Kuckartz (2016), stellt man fest, dass es bei beiden Formen zunächst um die Bestimmung relevanter Textstellen bzw. Analyseeinheiten geht. Anschließend werden (theoriegeleitet) Hauptkategorien bzw. Strukturierungsdimensionen gebildet, anhand derer das Material kodiert wird. Beide Autoren weisen darauf hin, dass es ggf. notwendig sein kann, die gebildeten Kategorien nach der Kodierung des kompletten Materials zu revidieren und das Material anschließend erneut zu kodieren. Bei Mayring (2015) wird die Formulierung von Definitionen, Ankerbeispielen und Kodierregeln hervorgehoben. Kuckartz (2016) stellt heraus, dass die Subkategorien anhand des Materials induktiv bestimmt werden.

Für die Auswertung im Bereich Diagnose, welche im folgenden Teilkapitel vorgestellt wird, wurde ein Vorgehen, das an die strukturierende Inhaltsanalyse nach Mayring (2015) angelehnt ist, verwendet. Gleiches gilt auch für Teile der Auswertung im Bereich Interventionen. Alle Verfahren werden jedoch im jeweiligen Teilkapitel ausführlich dargelegt und beschrieben. Die Auswertung im Bereich Reflexion orientierte sich (auch aufgrund der Zuhilfenahme der Software MAXQDA 12) eher am Ablaufschema der inhaltlich strukturierenden Inhaltsanalyse nach Kuckartz (2016). Auch hier wird das Verfahren noch ausführlich und nachvollziehbar beschrieben.

8.2 Auswertung im Bereich Diagnose

In diesem Kapitel wird das methodische Vorgehen im Bereich Diagnose erläutert. Dabei wird zunächst auf die Auswahl der Videosequenzen eingegangen (Kap. 8.2.1). Anschließend werden

die darin enthaltenen fachdidaktischen Aspekte erläutert (Kap.8.2.2). Diese sind auch Inhalt der Diagnoseaufträge, die zu den bei der Erhebung eingesetzten Videovignetten gestellt werden (Kap.8.2.3). Die Diagnoseaufträge wurden vorab durch Expertinnen und Experten validiert. Das Expertenrating wird in dem entsprechenden Teilkapitel vorgestellt (Kap.8.2.4). Anschließend wurden die Videovignetten in ViviAn eingebettet (Kap.8.2.5). Das Expertenrating dient dem Zweck, die prozessdiagnostische Fähigkeit der Studierenden einschätzen zu können (Kap.8.2.6). Obwohl es sich bei der vorliegenden Studie um eine qualitative Studie handelt, wird zum Abschluss dieses Teilkapitels eine quantitative Validierung des Testinstrumentes vorgenommen, um die Aussagekraft der Ergebnisse zu erhöhen (Kap.8.2.7).

8.2.1 Auswahl der Videosequenzen

Aus bereits vorliegenden Videos des Mathematik-Labors „Mathe ist mehr“ am Campus Landau wurden alle Videos im Bereich Algebra gesichtet. Insbesondere das Thema „Terme“ eignet sich für die Erfassung der prozessdiagnostischen Fähigkeit, da davon ausgegangen werden kann, dass dieses den Studierenden inhaltlich bzw. fachlich keine Probleme bereitet. Auf den Videos ist jeweils eine Kleingruppe von Schülerinnen und Schülern zu sehen, welche, in der Regel zu viert, selbstständig an mathematischen Inhalten arbeitet. Ausschnitte, in denen die Schülergruppe möglichst viel über ihre Gedanken zur Lösung der problemhaltigen Aufgaben mitteilt und diese ggf. sogar kontrovers diskutiert, wurden zugeschnitten und thematisch geordnet. Beispielsweise wurden alle Sequenzen, in denen die Gruppe sich implizit mit den Grundvorstellungen „Term als Bauplan“ und „Term als Rechenschema“ beschäftigt, gruppiert. Darüber hinaus wurden auch Sequenzen, in denen die Schülerinnen und Schüler das Material, Simulationen oder Videos nutzen und dabei entweder Erkenntnisse gewinnen oder inhaltliche Fehler begehen, zugeschnitten. Im Anschluss wurden diese Videos gesichtet und aus den jeweiligen Bereichen die ergiebigsten Sequenzen ausgewählt. So entstanden drei Videosequenzen von einer ungefähren Dauer von jeweils zwei Minuten. Alle drei Videos zeigen eine Gruppe aus vier Schülerinnen, die die Laborstation „Figurierte Zahlen“ bearbeiten, in der es um das Thema „Terme“ geht. In Stationsteil eins, aus welchem die drei Videosequenzen entstammen, sollen mit Hilfe von Dreieckszahlen Terme aufgestellt und vereinfacht werden. Hierzu nutzen die Schülerinnen verschiedene Materialien, insbesondere Kugeln, um die Dreieckszahlen selbst zu legen. Sie verwenden verschiedene Simulationen, um zunächst einzelne Dreieckszahlen und später auch einen allgemeinen Term zur Berechnung einer beliebigen Dreieckszahl aufzustellen.

8.2.2 Fachdidaktische Aspekte

Terme bilden einen wesentlichen Bestandteil der Algebra. Aus mathematikdidaktischer Perspektive gibt es dabei die beiden Grundvorstellungen Term als Bauplan und Term als Rechenschema (vgl. dazu Kap. 6). Bei der Grundvorstellung Term als Bauplan werden Terme entweder erstellt oder umgeformt. Die Termumformungen stellen hierbei Änderungen des Bauplans dar, die sich auf die jeweilige Struktur des Phänomens beziehen. Mit Hilfe der Grundvorstellung Term als Bauplan können Terme ineinander überführt werden und (somit) deren Gleichheit gezeigt werden. Grundlegend für diese Sichtweise ist also das Schaffen bzw. Durchdringen der Struktur des Phänomens. Terme werden in diesem Sinne bei der Bearbeitung der Frage nach der mathematischen Struktur eines Phänomens oder eines Problems entwickelt.

Bei der Grundvorstellung Term als Rechenschema geht es vorwiegend darum, etwas auszurechnen. Hierbei wird entweder ein vorliegender Term genutzt, um Zahlen für die darin enthaltenen Variablen einzusetzen, oder ein Term wird derart umgeformt, dass die Berechnung (künftig) mit möglichst wenig Rechenschritten durchgeführt werden kann. Hierzu kann beispielsweise die Anzahl der vorhandenen Variablen reduziert werden. Terme werden bei dieser Grundvorstellung also genutzt, um wiederholte, gleichartige Berechnungen für verschiedene Werte schnell und möglichst einfach durchführen zu können.

Neben den Grundvorstellungen Term als Bauplan und Term als Rechenschema sind auch die binomischen Formeln relevant für die Analyse der Situation, da die Schülerinnen und Schüler sich in den Videosequenzen damit auseinandersetzen (vgl. dazu Kap. 8.2.5). Auf die Herleitung der drei Formeln wird an dieser Stelle verzichtet. Die Kenntnis über die jeweiligen Inhalte wird an dieser Stelle vorausgesetzt.

Da die in der Situation relevanten Inhalte bereits im Bachelorstudium behandelt wurden, sollten auch alle der untersuchten Studierenden über die notwendigen Kenntnisse verfügen. Aus fachlicher Sicht kann davon ausgegangen werden, dass die Studierenden inhaltlich keine Probleme mit den Aufgaben haben, die die Schülerinnen und Schüler bearbeiten. Alle fachdidaktischen Aspekte, die für die Sequenzen und die damit verbundenen Diagnoseaufträge von Relevanz sind (vgl. dazu Kap. 8.2.3), werden den Studierenden vor Bearbeitung der Videovignetten zur Verfügung gestellt.

8.2.3 Diagnoseaufträge

Die ausgewählten Szenen wurden inhaltlich diskutiert und ausgiebig analysiert. Im Anschluss wurden auf Grundlage der fachdidaktischen Überlegungen Diagnoseaufträge zu den einzelnen Videosequenzen formuliert. Diese beziehen sich zu einem großen Teil auf die Grundvorstellungen „Term als Bauplan“ und „Term als Rechenschema“. Hierbei müssen die einzelnen Schülerinnen bzw. deren Äußerungen diagnostiziert werden und mit den verschiedenen Grundvorstellungen in Verbindung gebracht werden.

Bei kommunikativen Schülerinnen mit hohem Redeanteil wird explizit nach der Grundvorstellung gefragt, mit welcher die jeweilige Schülerin argumentiert. Bei den Schülerinnen mit geringerem Redeanteil wird lediglich gefragt, ob diese in der Lage sind, eine geeignete Grundvorstellung zur Bearbeitung der Aufgabe anzuwenden. Hinzu kommen Fragen zur Formulierung des Arbeitsauftrages, den die Schülerinnen bearbeiten, sowie zu fachlichen Inhalten. Hierbei muss beantwortet werden, ob die Schülerinnen die binomischen Formeln in der Sequenz nutzen, ob Bezug auf die geometrische Situation genommen wird und ob die Schülerinnen über die entsprechenden Grundfertigkeiten verfügen, um die Aufgabe lösen zu können.

Des Weiteren gibt es Fragen zur Nutzung der Simulation und zur wahrgenommenen Plausibilität der erhaltenen Ergebnisse, also ob einzelne Schülerinnen die Ergebnisse auf Grundlage ihres Grundwissens kritisch hinterfragen. Abschließend wird gefragt, ob man in der Situation eingreifen würde und falls ja, wie diese Intervention gestaltet werden würde. Zu jeder Sequenz gibt es fünf bis sechs Diagnoseaufträge (vgl. Anhang A). Alle Fragen sind als Single- oder Multiple-Choice Frage formuliert, wobei die gewählten Antwortoptionen jeweils begründet werden müssen.

8.2.4 Expertenrating

Die Videosequenzen wurden 16 Expertinnen und Experten gezeigt, darunter Professoren und Wissenschaftliche Mitarbeiterinnen und Mitarbeiter aus der Mathematikdidaktik mehrerer Universitäten sowie langjährige Mathematiklehrkräfte. Diese schauten sich die entsprechenden Vignetten einmal an und beantworteten im Anschluss die Diagnoseaufträge. Die Videos wurden jeweils nur einmal und ohne Unterbrechung angesehen, da dies der realen Situation im Schulalltag am nächsten kommt, wo es auch nicht möglich ist, sich Szenen mehrmals anzuschauen, anzuhalten oder zurückzuspulen. Die Diagnoseaufträge konnten vorab angesehen werden, um sich damit vertraut zu machen, auf was in der jeweiligen Sequenz geachtet werden sollte. Die von der Mehrheit der Expertinnen und Experten gegebene Antwort wurde als korrekte Antwort angesehen. Diese stimmten mit den im Vorfeld beim Erstellen der Diagnoseaufträge getätigten Überlegungen überein. Die Begründungen der Expertengruppe wurden als Musterlösungen für die jeweils richtigen Antworten verwendet und für die Erstellung des Kodierleitfadens genutzt.

8.2.5 Einbettung in ViviAn

Zur standardisierten Erhebung wurden die Videosequenzen gemeinsam mit den jeweiligen Diagnoseaufträgen in ViviAn eingebettet (vgl. Abb. 22). Zusätzlich zum Video werden in ViviAn mehrere Hintergrundinformationen für den Betrachter hinterlegt. Diese können mittels Mausklick auf den entsprechenden Button aufgerufen werden. Es öffnet sich dann jeweils ein Fenster mit den angeforderten Informationen. So können die jeweiligen Arbeitsaufträge sowie die Materialien, an bzw. mit denen die Schülerinnen in der Sequenz arbeiten, eingesehen werden. In Situationen, in denen die Schülerinnen mit Simulationen arbeiten, sind die entsprechenden Simulationen (und gegenständlichen Materialien) hinter dem Button „Materialien“ hinterlegt. Die Simulationen können vom Betrachter der Videos bedient werden, bevor das Video gestartet wird. Dadurch ist es möglich, sich bereits vorab Gedanken zu Material bzw. Simulationen zu machen und sich beim Betrachten der Videos gänzlich auf die Diagnose zu fokussieren. Ebenso können die Arbeitsergebnisse, die die Schülerinnen in der Sequenz in ihrem Arbeitsheft festhalten, eingesehen und zwischen den Schülerinnen jeweils paarweise verglichen werden. Darüber hinaus gibt es Informationen zu der Station, an der die Schülerinnen arbeiten (Button „Lernumgebung: Thema und Ziele“). Hier wird erklärt, welche Inhalte und Ziele im Rahmen der Laborstation behandelt werden. Zudem erlaubt ViviAn es, sich eine zeitliche Einordnung anzusehen, aus der zu entnehmen ist, an welcher Stelle der Station die Schülerinnen sich aktuell befinden. Unter „Schülerprofile“ können Hintergrundinformationen (Klassenstufe, Schulart, Geschlecht) zu den Schülerinnen eingeholt werden.

Alle genannten Informationen stehen dem Betrachter der Videos zur Verfügung und sollen bereits vor dem Start der Videosequenz studiert werden, damit eine möglichst große Vorentlastung stattfinden kann und die Beantwortung der Fragen auf die prozessdiagnostische Fähigkeit des Betrachters zurückzuführen ist. Zu dieser Vorentlastung gehört auch ein Einführungsvideo, in dem das Videotool ViviAn erläutert wird. Zusätzlich ist es möglich, sich eine Videosequenz der Schülerinnen anzusehen, bei der die Schülerinnen jeweils einen Arbeitsauftrag laut vorlesen. Dadurch kann man sich mit den Stimmen der Schülerinnen vertraut machen. In den Videovignetten selbst wird durch ein gelbes Rechteck gekennzeichnet, welche Schülerin gerade spricht. Dazu erscheint ein solches bei der aktuellen sprechenden Schülerin, solange diese spricht.

Schülerebene

Arbeitsauftrag

Material

Schülerdokumente

Lernumgebung: Thema und Ziele



Metaebene

Schülerprofil

S2 S3

S1 S4

Zeitliche Einordnung

Diagnoseauftrag

Terme Vignette 1

✦ Mit welcher Grundvorstellung argumentiert S1?

Term als Bauplan.

Term als Rechenschema.

Keine Aussage möglich.

✦ Bitte begründen Sie Ihre Antwort.

Abbildung 22: Screenshot einer in ViviAn eingebundenen Videovignette mit zusätzlich eingeblendetem Diagnoseauftrag unterhalb der Vignette

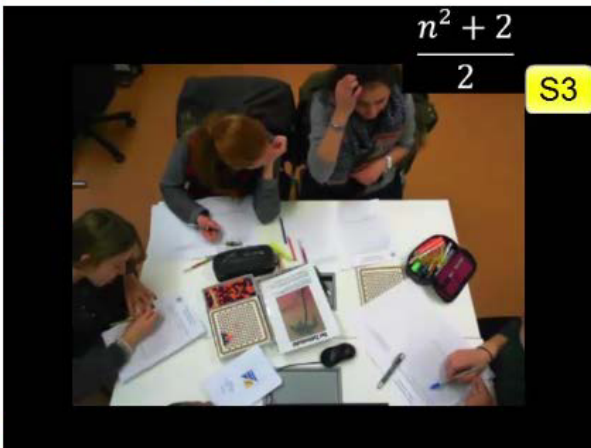
Schülerebene

Arbeitsauftrag

Material

Schülerdokumente

Lernumgebung: Thema und Ziele



Metaebene

Schülerprofil

S2 S3

S1 S4

Zeitliche Einordnung

Diagnoseauftrag

Abbildung 23: Screenshot einer in ViviAn eingebundenen Videovignette

Bei den einzelnen Vignetten liegen jeweils Besonderheiten vor, die das Erfassen der Situation erleichtern sollen. In Vignette 1 und 2 sieht man die Simulation, mit der die Schülerinnen arbeiten. Umgesetzt wurde dies in Vignette 1 derart, dass die Simulation an der Stelle erscheint, an der sich jeweils die Hand der Schülerin befindet, die die Simulation bedient (vgl. Abb. 22). Denn dort befindet sich die Maus, mit der die Simulation bedient wird. Dadurch wird ersichtlich, welche Schülerin die Veränderungen vornimmt und welche der Schülerinnen gerade mitdiskutieren, ohne selbst Veränderungen vorzunehmen. Bei Vignette 2 arbeiten die Schülerinnen innerhalb der Videovignette an verschiedenen Aufgabenteilen. Daher wird jeweils angezeigt, auf welchen Aufgabenteil sich die Äußerungen der Schülerinnen beziehen. Da in Vignette 3 nicht mit einer Simulation gearbeitet wird, ist dort (im Video selbst) lediglich die Schülergruppe zu sehen. Jedoch bekommt der Betrachter die Terme angezeigt, die von den Schülerinnen geäußert werden. Dies geschieht analog zu den gelben Rechtecken immer dann, wenn die Äußerung stattfindet. Aus diesem Grund verfügt Vignette 3 über einen schwarzen Rand, in dem die Terme dann jeweils in weißer Schrift eingeblendet werden (vgl. Abb. 23).

8.2.6 Ermittlung der prozessdiagnostischen Fähigkeit

Zur Ermittlung der prozessdiagnostischen Fähigkeit liegen die Antworten der Studierenden auf Single- bzw. Multiple-Choice-Fragen sowie jeweils eine Begründung der gegebenen Antwort vor. Diese Daten wurden mittels ViviAn erhoben (vgl. dazu Kapitel 8.2.5). Die Vignetten stellen hierbei die Auswahlinheit dar und die einzelnen Beantwortungen der Diagnoseaufträge die Analyseeinheiten. Die Kategorien entsprechen den vorgegebenen Auswahlmöglichkeiten der Diagnoseaufträge. Anhand des Expertenratings wurde ein Kodierleitfaden erstellt und im Laufe der Kodierung induktiv erweitert. Im Kodierleitfaden sind Kategoriendefinitionen und Ankerbeispiele enthalten (vgl. Anhang B). Neben den positiven Beispielen gibt es auch Negativbeispiele, welche dabei helfen sollen, die entsprechenden Kategorien möglichst gut gegeneinander abzugrenzen. Außerdem sind darin verschiedene Hinweise zum grundsätzlichen Vorgehen beim Kodieren enthalten.

Zwei Kodiererinnen erhielten zunächst eine Schulung anhand von Probematerial, welches eigens dafür erstellt wurde. Die Kodierungen wurden besprochen und die Anleitung auf Grundlage der Rückmeldungen nochmals überarbeitet. Mit dieser Kodieranleitung kodierten die beiden Kodiererinnen das gesamte Datenmaterial jeweils zunächst einzeln. Die Übereinstimmung der beiden Kodiererinnen, die mit Hilfe von Cohens Kappa berechnet wurde, liegt im arithmetischen Mittel bei $\kappa = 0.845$, was nach Kuckartz (2016, 210) als sehr gute Übereinstimmung zu werten ist. Die Bandbreite der Übereinstimmung bei den 14 Fragen liegt zwischen $\kappa = 0.639$ und $\kappa = 0.959$. Der Median liegt bei $\kappa = 0.869$. Die Berechnungen wurden jeweils mit der Software R und dem Package „irr“ (Gamer, Lemon, Fellows und Singh 2019) durchgeführt.

Eine solch hohe Übereinstimmung führt in der Praxis häufig dazu, dass ab dann alle restlichen Daten nur noch von einem der beiden Kodiererinnen kodiert werden. Die vorliegenden Daten wurden jedoch vollständig von beiden Kodiererinnen kodiert. Zusätzlich wurden im Anschluss alle Unstimmigkeiten, also alle Antworten, bei denen sich die beiden Kodiererinnen nicht vollends einig waren, weil eine der beiden eine Zuordnung vorgenommen hat, die die andere nicht vorgenommen hat, ausfindig gemacht und den beiden an einem Folgetermin vorgelegt. In einem Konsensgespräch, also einer kommunikativen Validierung, mussten die beiden sich darauf einigen, ob die jeweilige Zuordnung der Textstelle zu der Kategorie weiterhin Bestand haben sollte (Kvale 2007; Guest, MacQueen & Namey 2012). Beide Kodiererinnen hatten

jeweils die Möglichkeit, die andere mit Argumenten davon zu überzeugen, dass die vorgenommene Kodierung die richtige und somit konsensfähig ist. Bei allen aufgetretenen Unstimmigkeiten konnte am Ende ein Konsens gefunden werden, sodass nun eine Konsenskodierung vorliegt, welche als Grundlage für weitere Analysen dient. Diese Konsenskodierung ist aufgrund des aufwendigen und mehrstufigen Prozesses als hochwertiger zu betrachten, als die einzelnen Kodierungen.

Da somit für alle Studierendenantworten auf die Diagnoseaufträge eine Konsenskodierung vorliegt, ist jeweils ersichtlich, ob die Studierenden die Diagnoseaufträge im Sinne der Expertinnen und Experten beantworten konnten. Zur Ergebnisdarstellung werden alle Diagnoseaufträge gleichwertig bepunktet, sprich für jede Antwort im Sinne der Expertinnen und Experten wird ein Punkt vergeben. Dadurch können insgesamt 14 Punkte erreicht werden. Dieses Vorgehen mag zunächst verwundern, da die Auswertung bisher rein qualitativ war und eine Be-punktung eher einem quantitativen Vorgehen zugeordnet wird. Jedoch liefert dies einen guten Überblick über die Verteilung der diagnostischen Fähigkeiten innerhalb der Experimentalgruppe. Dies ist auch deshalb relevant, da die Interventionen und Reflexionsgespräche der Studierenden im weiteren Verlauf auch auf Grundlage einer Einteilung hinsichtlich der diagnostischen Fähigkeiten betrachtet werden. Mayring (2016, 37) schreibt dazu: „Eine wichtige Funktion qualitativen Denkens ist es, sinnvolle Quantifizierungen zu ermöglichen“. Dies ist deshalb wichtig, weil Quantifizierungen zur Absicherung und Verallgemeinerung der Ergebnisse beitragen können (Mayring 2016, 38). Entsprechend erscheint das gewählte Vorgehen inhaltlich sinnvoll und der Datenlage angemessen.

8.2.7 Quantitative Validierung des Testinstrumentes

Zunächst wurde mit Hilfe der Software IBM SPSS Statistics 24 überprüft, ob sich die Annahme bestätigen lässt, dass alle Diagnoseaufträge ein latentes Konstrukt widerspiegeln. Dazu wurde eine Hauptkomponentenanalyse mit allen beantworteten Diagnoseaufträgen von insgesamt 126 Personen durchgeführt (darin enthalten sind neben den Antworten der Experimentalgruppe auch Antworten der Vergleichsgruppe sowie die Beantwortung von Lehrkräften). Die Grundvoraussetzungen wurden anhand des Bartlett-Tests und des Kaiser-Meyer-Olkin-Kriteriums überprüft (Field, Miles & Field 2012, 914ff.). Der Bartlett-Test überprüft die Nullhypothese, dass keine Korrelationen zwischen den Items vorliegen. Aufgrund des signifikanten Ergebnisses von $p < 0,001$ muss die Nullhypothese verworfen werden. Daraus folgt, dass die Items miteinander korrelieren. Das Kaiser-Meyer-Olkin-Kriterium überprüft, ob der Variablenansatz für eine Dimensionsreduktion angemessen ist. KMO-Werte zwischen 0,5 und 0,7 sind als mittelmäßig und Werte zwischen 0,7 und 0,8 als gut anzusehen (Field et al. 2012, 920). Entsprechend ist der vorliegende KMO-Wert von 0,968 als sehr gut zu bewerten.

Bei der Hauptkomponentenanalyse werden die Daten in ein Set linearer Kombinationen zerlegt und sämtliche Varianz analysiert (Döring & Bortz 2016, 481). Der Screeplot (von Faktoren gegen den Eigenwert) weist darauf hin, dass es sich um ein einfaktorielles Konstrukt handelt (vgl. Abb. 24). Es ist klar zu erkennen, dass nach dem ersten Faktor ein Bruch im Graph entsteht. Außerdem ist anhand des Eigenwerts ersichtlich, dass der erste Faktor im Verhältnis zu den anderen Faktoren den Großteil der Varianz aufklärt (Döring & Bortz 2016, 482). Dies kann deshalb entnommen werden, weil die Summe der Eigenwerte die Gesamtvarianz bildet und der Eigenwert des ersten Faktors mit Abstand am größten ist. Ab dem vierten Faktor liegen alle Eigenwerte ungefähr bei 0 und tragen entsprechend nicht mehr wesentlich zur Summe bei.

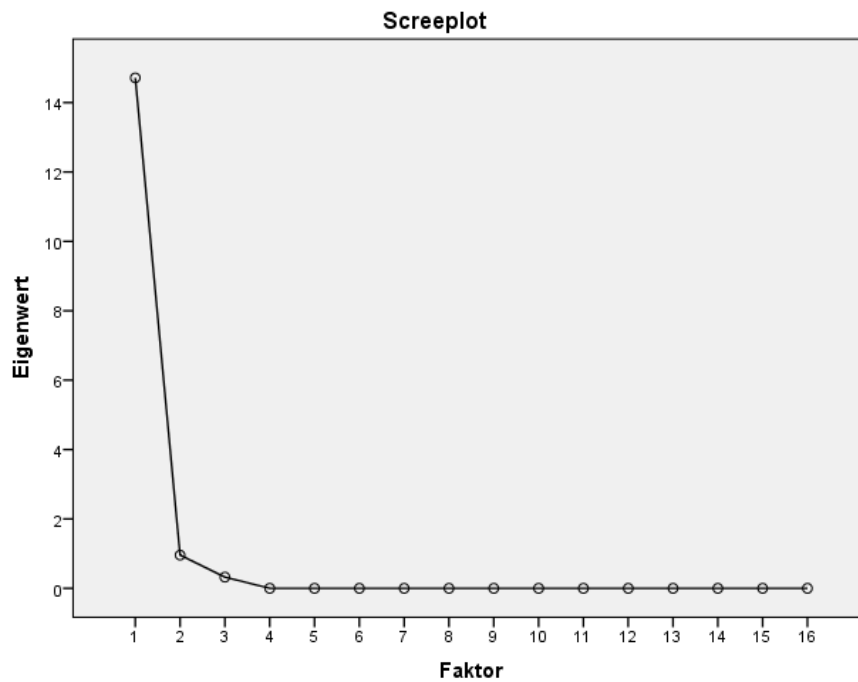


Abbildung 24: Screepplot im Rahmen der Hauptkomponentenanalyse

Auch das Kaiser-Guttman-Kriterium, das besagt, dass nur Faktoren mit Eigenwerten >1 beibehalten werden sollen, führt zu dem Ergebnis, dass die Diagnoseaufträge ein einfaktorielles Konstrukt bilden (Döring & Bortz 2016, 482). Aus Tabelle 2 kann entnommen werden, dass die zweite Komponente, also der zweite Faktor, mit 0,958 aufgrund des Kaiser-Guttman-Kriteriums nicht beibehalten und somit nur der erste Faktor berücksichtigt wird. Dies bestärkt auch die Annahme, dass sich alle Diagnoseaufträge auf die Komponente „Deuten“ des Diagnosemodells von Beretz et al. (2017) beziehen.

Tabelle 2: Eigenwerte der Faktoren des Screeplots

Komponente	Gesamt
1	14,713
2	,958
3	,324
4	,001
5	,001
6	,001
7	,001
8	,000
9	,000
10	,000
11	,000
12	,000
13	,000
14	,000
15	,000
16	,000

Tabelle 3: Ladungen der Items auf das latente Konstrukt

Kommunalitäten		
	Anfänglich	Extraktion
F1_TaR	1,000	,915
F2_nein	1,000	,914
F3_TaB	1,000	,914
F4_kA	1,000	,914
F5_S3	1,000	,914
F6_Ges	1,000	,914
F7_Sim	1,000	,890
F8_kA	1,000	,890
F9_ja	1,000	,890
F10_S3	1,000	,890
F11_niemand	1,000	,890
F12_nein	1,000	,956
F13_nein	1,000	,954
F14_nein	1,000	,956
F15_ja	1,000	,956
F16_ja	1,000	,956

Extraktionsmethode: Hauptkomponentenanalyse.

Tabelle 3 stellt die nicht-rotierten Faktorladungen hinsichtlich einer einfaktoriellen Lösung dar. In der linken Spalte sind die Items aufgeführt. Der mittleren Spalte kann entnommen werden, dass jedes Item zu Beginn einen Eigenwert von 1 hat. Der Wert der Extraktion in der rechten Spalte sagt aus, wie gut die Items den Faktor beschreiben. Da alle Werte $\geq 0,890$ sind, scheinen die Items den Faktor gut zu beschreiben.

Trotz der obigen Hinweise, dass mit den Items ein einfaktorielles Konstrukt abgebildet wird, wurde zur Absicherung auch eine zweifaktorielle Lösung in Betracht gezogen. Hierzu wurde eine Oblimin-Rotation gewählt, da dies zulässt, dass die Faktoren miteinander korrelieren (Jennrich & Sampson 1966). Dadurch kann überprüft werden, ob es einen zweiten Faktor gibt, der eine inhaltliche Interpretation der Items zulässt. Bei der Berechnung laden alle Items mit Werten zwischen -0,280 und 0,313 auf den zweiten Faktor. Hingegen laden alle Items mit Werten zwischen 0,943 und 0,978 auf den ersten Faktor. Die Korrelation zwischen den beiden Faktoren beträgt $r = 0,097$ und ist damit zu vernachlässigen (Cohen 1988, 79). Die Faktorladungen und die sehr geringe Korrelation stärken die Hypothese, dass die Items eine einfaktorielle Struktur aufweisen.

Klassische und propabilistische Testtheorie sind nicht als separate Auswertungsmethoden anzusehen, sondern ergänzen sich vielmehr gegenseitig (Moosbrugger 2008, 216; Rost 2006, 261). Entsprechend können die Items, da diese ein eindimensionales Konstrukt zu bilden scheinen, anhand eines Rasch-Modells genauer untersucht werden.

Nachfolgend werden nun die Analysen beschrieben, die mit Hilfe der Software R und dem Package TAM durchgeführt wurden (Robitzsch, Kiefer & Wu 2019). Zuerst wurde die Trennschärfe der Items betrachtet, um durch Aussortieren der Items mit zu geringer Trennschärfe nur

diejenigen Items beizubehalten, die eine gute Trennschärfe besitzen und somit einen substanziellen Mehrwert für die Bestimmung der Personenfähigkeit liefern. Die Punktbiseriale Korrelation gibt Auskunft über genau diese Trennschärfe. Items, deren punktbiseriale Korrelation < 0 ist, sollten nicht aufgenommen werden (Kelava und Moosbrugger 2012). Items mit Trennschärfen $< 0,2$ werden als schlecht kategorisiert, Items zwischen $0,2$ und $0,3$ als grenzwertig, Items zwischen $0,3$ und $0,4$ als gut und Items mit Trennschärfen $> 0,4$ werden als sehr gut kategorisiert (Ebel & Frisbie 1991).

Tabelle 4: Übersicht über die punktbiseriale Korrelation aller Items

Itemname	Punktbiseriale Korrelation
F1_TaR	0.36
F2_nein	0.26
F3_TaB	0.41
F4_kA	0.40
F5_S3	0.34
F6_Ges	0.40
F7_Sim	0.52
F8_kA	0.25
F9_ja	0.42
F10_S3	0.50
F11_niemand	0.47
F12_nein	0.38
F13_nein	0.50
F14_nein	0.41
F15_ja	0.43
F16_ja	0.37

Tabelle 4 zeigt, dass zwei Werte unter $0,3$ liegen und somit überprüft werden müssen. Es handelt sich dabei um Items, die auch von den Expertinnen und Experten nicht einheitlich bewertet wurden und die daher auch aus inhaltlichen Gründen kritisch zu sehen sind. Entsprechend werden diese beiden Items aussortiert. In der Regel berechnet man die punktbiseriale Korrelation nach Entfernen eines Items erneut. Da dieser Prozess keine neuen Erkenntnisse lieferte, wird auf eine erneute Berichtigung der Kennwerte verzichtet und stattdessen direkt die endgültige punktbiseriale Korrelation der nun insgesamt 14 Test-Items in Tabelle 5 dargestellt.

Tabelle 5: Übersicht über die endgültige punktbiseriale Korrelation der Items

Itemname	Punktbiseriale Korrelation
F1_TaR	0.37
F3_TaB	0.43
F4_kA	0.41
F5_S3	0.33
F6_Ges	0.41
F7_Sim	0.51
F9_ja	0.41
F10_S3	0.53

F11_niemand	0.47
F12_nein	0.39
F13_nein	0.50
F14_nein	0.44
F15_ja	0.43
F16_ja	0.40

Man kann erkennen, dass nun alle Werte größer als 0,3 sind, also keine schlechten oder grenzwertigen Items mehr vorhanden sind. Darüber hinaus gibt es einige Items, deren Werte als sehr gut anzusehen sind (F3_TaB, F4_kA, F6_Ges, F7_Sim, F9_ja, F14_nein, F10_S3, F11_niemand, F13_nein und F15_ja). Somit kann festgehalten werden, dass die Trennschärfe insgesamt zufriedenstellend ist.

Als nächstes wurde ein Modellvergleich zwischen einem 1PL (Rasch-Modell) und einem 2PL-Modell (Birnbaum-Modell) vorgenommen. Hierzu wurden die AIC- und BIC-Werte der beiden Modelle verglichen (Akaike 1987; Schwartz 1978). Sowohl der AIC-Wert (1968,82) als auch der BIC-Wert (2011,37) des Rasch-Modells sind niedriger als der AIC-Wert (1984,56) und der BIC-Wert (2063,97) des Birnbaum-Modells. Darüber hinaus wurde ein likelihood-ratio-Test durchgeführt, um zu überprüfen, ob es einen signifikanten Unterschied zwischen der Beschreibung der Daten durch die beiden Modelle gibt. Der likelihood-ratio-Test ist mit 13 Freiheitsgraden und einem p-Wert von 0,672 nicht signifikant. Daher beschreibt das restriktivere Modell (Rasch-Modell) die Daten genauso gut wie das 2PL-Modell. Auch der absolute Modellfit, welcher mit Hilfe des standardized root mean square residual dargestellt werden kann, ist mit einem Wert von 0,085 relativ gut (Hu & Bentler 1999). Somit kann das Rasch-Modell beibehalten und für weitere Analysen genutzt werden.

Die Infit-Werte aller Items liegen im Bereich zwischen 0,952 und 1,088. Entsprechend sind sie alle im guten Bereich zwischen 0,750 und 1,333 (Wilson 2004, 129). Die Outfit-Werte liegen im Bereich zwischen 0,892 und 1,121. Die verschiedenen Itemschwierigkeiten sind in Tabelle 6 aufgelistet. Man kann erkennen, dass es ein sehr schwieriges Item (F1_TaR) und ein sehr leichtes Item (F9_ja) gibt sowie eine große Bandbreite dazwischen.

Tabelle 6: Absolute und relative Lösungshäufigkeiten sowie Itemschwierigkeiten aller verwendeten Items im Rasch-Modell

Itemname	Anzahl korrekter Antworten	Relative Anzahl richtiger Lösungen (in %)	Itemschwierigkeit (Wert)
F1_TaR	20	0,16	1,918
F3_TaB	97	0,77	-1,400
F4_kA	40	0,32	0,898
F5_S3	88	0,70	-0,978
F6_Ges	58	0,46	0,191
F7_Sim	69	0,55	-0,220
F9_ja	111	0,88	-2,290
F10_S3	79	0,63	-0,605
F11_niemand	32	0,25	1,256
F12_nein	90	0,71	-1,066
F13_nein	88	0,70	-0,978

F14_nein	100	0,79	-1,559
F15_ja	104	0,83	-1,792
F16_ja	63	0,50	0,005

Zu einem ähnlichen Ergebnis kommt man auch, wenn man die Item Characteristic Curves (ICCs) betrachtet (vgl. Abb. 25). Die Kurven zeigen, dass Personen mit unterschiedlicher Fähigkeit adäquat beschrieben werden können. Dies ist daran erkennbar, dass es genug ähnliche Items gibt, um Personen mit ähnlicher, aber unterschiedlicher Fähigkeit zu unterscheiden. Man erkennt außerdem, dass mehr Items im Bereich von -2 bis 0 liegen und daher dieser Bereich auch genauer erfasst wird als der Bereich zwischen 0 und +2. Insgesamt liegt eine zufriedenstellende Verteilung der Kurven vor.

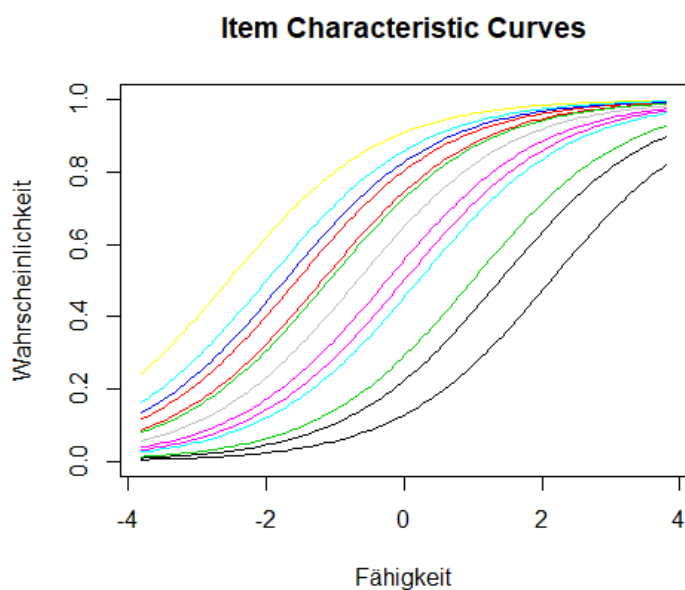


Abbildung 25: Item Characteristic Curves (ICCs)

Nachdem bisher die einzelnen Items betrachtet wurden, wird nun die Gesamtskala betrachtet. Dort ergibt sich sowohl eine zufriedenstellende EAP-Reliabilität von 0,656 als auch eine zufriedenstellende WLE-Reliabilität von 0,624 (Wu, Adams, Wilson & Haldane 2007; Kelava & Moosbrugger 2012; Ebel & Frisbie 1991). Die EAP-Reliabilität sagt aus, inwiefern der Test das misst, was er messen soll (Wu et al. 2007). Die WLE-Reliabilität trifft eine Aussage darüber, inwiefern die Personenfähigkeiten innerhalb des Tests korrekt erfasst werden (Kelava & Moosbrugger 2012; Ebel & Frisbie 1991).

8.3 Auswertung im Bereich Interventionen

Nachdem im vorigen Teilkapitel die Auswertung im Bereich Diagnose vorgestellt wurde, wird in diesem Kapitel die Methodik im Bereich Interventionen erläutert. Zunächst wird dabei beschrieben, wie die Videos, die später analysiert werden, erhoben wurden (Kap. 8.3.1 und Kap. 8.3.2). Anschließend wird dargelegt, wie die Interventionen analysiert werden. Dabei wird auf die genutzten Hilfearten, die die Studierenden im Rahmen ihrer Interventionen verwenden, eingegangen (Kap. 8.3.3). Des Weiteren wird die Angemessenheit der Interventionen thematisiert (Kap. 8.3.4). Darüber hinaus werden die Dauer der Intervention, die Dauer der einzelnen Hilfen innerhalb der Intervention, der Interventionsauslöser, die Anzahl der arbeitenden Schülerinnen

und Schüler sowie die Reihenfolge der Hilfen innerhalb einer Intervention analysiert (Kap. 8.3.5).

8.3.1 Einbettung in das Didaktische Seminar

Im Rahmen des Didaktischen Seminars haben die teilnehmenden Studierenden die Aufgabe, in Kleingruppen eine Laborstation des Mathematik-Labors zu überarbeiten und diese im Anschluss mit einer Schulklasse zu erproben. Bei der Erprobung beobachtet jeweils ein Mitglied der Gruppe die Schülergruppe, welche im Filmraum arbeitet. Dadurch, dass die Stationen von maximal drei Personen überarbeitet wurden und jede Station aus genau drei Stationsteilen besteht, kann jede Studentin bzw. jeder Student die Schülergruppe für einen kompletten Stationsteil durchgehend beobachten. Die anderen beiden Gruppenmitglieder betreuen die restliche Schulklasse im Mathematik-Labor. Jede Studentin bzw. jeder Student hat also ungefähr 90 Minuten Zeit, die Aufgabenbearbeitung und Kommunikation der Schülergruppe im Filmraum zu verfolgen. Die Studierenden haben dabei die Aufgabe, zu intervenieren, falls die Schülerinnen und Schüler nach Ansicht der/des Studierenden ohne einen Eingriff nicht sinnvoll weiterarbeiten kann. Es obliegt der/dem Studierenden, zu entscheiden, ob und wie sie/er interveniert. Das Büro, von dem aus die Studentin oder der Student die Schülergruppe beobachtet, befindet sich auf dem gleichen Flur und ist somit nur einige Meter vom Filmraum entfernt. Die oder der Studierende hat also nur wenige Sekunden Zeit, sich zu überlegen, wie die Interventionen gestaltet und was den Schülerinnen und Schülern mitgeteilt werden soll.

8.3.2 Datengenerierung

Die Interventionen wurden videographiert und zugeschnitten. Dabei ist einerseits die komplette Intervention des Studierenden zu sehen, vom Öffnen bis zum Schließen der Tür des Filmraums, andererseits aber auch immer ein je nach Situation unterschiedlich langer Vorlauf, welcher einen Einblick in das Geschehen vor der Intervention geben soll. Dieser Vorlauf ist bis zu drei Minuten lang. Insgesamt liegen 208 Interventionen aus vier aufeinanderfolgenden Semestern des Didaktischen Seminars vor, beginnend im Sommersemester 2016. Jede bzw. jeder der Studierenden hat dabei mindestens einmal und maximal zehnmal interveniert. Es gab keine explizite Vorgabe, wie oft interveniert werden sollte.

8.3.3 Analyse der Hilfen

Es liegen 208 Interventionen von den 48 Teilnehmerinnen und Teilnehmern des Didaktischen Seminars aus den Semestern vom Sommersemester 2016 bis zum Wintersemester 2017/18 vor. Bei den zugeschnittenen Interventionen sind jeweils ein Vorlauf sowie die komplette Intervention der bzw. des Studierenden zu sehen. Der Vorlauf kann je nach Intervention eher kurz (etwa 30 Sekunden) oder auch länger (zwei bis drei Minuten) ausfallen. Dabei wurde versucht, relevantes Schülerverhalten zum Verständnis der Intervention mit in die Videosequenz aufzunehmen. Auswahlinheit ist hierbei die Bearbeitung der Laborstation durch die Schülergruppe, die Analyseeinheiten sind jeweils die zugeschnittenen Interventionen. Es wurde eine Qualitative Inhaltsanalyse, angelehnt an die strukturierende Inhaltsanalyse nach Mayring (2015), durchgeführt. Die Strukturierungsdimensionen wurden anhand des fünfstufigen Handlungsmodells für Lehrpersonen von Zech (2002) festgelegt und das Kategorienmodell wurde auf dieser Grundlage aufgestellt. Anschließend wurden die Kategorien definiert und Ankerbeispiele sowie Kodierregeln aufgestellt. Die Fundstellenbezeichnung erübrigt sich, da jede Intervention als eine

Analyseeinheit angesehen, sprich nicht weiter untergliedert wurde. Nach dem ersten Materialdurchlauf zeigte sich, dass eine nicht zu vernachlässigende Anzahl an Interventionen nicht zugeordnet werden konnte. Da diese Interventionen größtenteils ähnlicher Natur waren, wurde eine neue Kategorie hinzugefügt, sodass das Material abschließend mit folgendem Kategoriensystem kodiert wurde:

- Disziplinarische Hilfen

Die Schülerinnen und Schüler werden daran erinnert, die Regeln einzuhalten, welche dazu beitragen sollen, dass sie konzentriert an der Aufgabe arbeiten können. Darüber hinaus können sie auch darauf hingewiesen werden, dass sie sich aktuell entweder gar nicht mit dem Inhalt beschäftigen oder, bedingt durch ihre privaten Exkurse, zu langsam arbeiten. Die Schülerinnen und Schüler können ebenfalls dazu aufgefordert werden, dass sie verantwortungsvoll und nur bei den zugehörigen Aufgaben mit dem vorgegeben Material arbeiten sollen.

- Motivationshilfen

Die Schülerinnen und Schüler werden ermutigt, weiter an der Aufgabe zu arbeiten. Dabei wird jedoch kein konkreter Bezug auf die eigentliche Aufgabe genommen. Somit handelt es sich um Formulierungen, welche auch bei anderen Fragestellungen oder auch in anderen Fächern verwendet werden können. Die Intervention erfolgt in der Regel ausschließlich mündlich, vereinzelt auch über Gesten (z.B. Nicken). Die Motivation kann sowohl durch Betonung der guten Arbeit der Schülerinnen und Schüler als auch durch Anreize erfolgen, wie zum Beispiel einen Hinweis auf die nächste Pause oder das Ende der Aufgabe.

- Rückmeldehilfen

Die Schülerinnen und Schüler erhalten Rückmeldungen und Informationen darüber, ob sie bei der Bearbeitung der Aufgabe auf dem richtigen Weg sind. Dabei werden der konkrete Arbeitsstand und der Arbeitsverlauf berücksichtigt. Die Rückmeldung kann sowohl das Aufzeigen eines Fehlers beinhalten als auch die richtige Bearbeitung bestärken. Ebenso können die Schülerinnen und Schüler eine Rückmeldung dazu erhalten, ob sie bei einer vorherigen Aufgabe etwas übersehen oder übersprungen haben. Die Aussagen müssen explizit auf eine Rückmeldehilfe schließen lassen, um dieser Kategorie zugeordnet zu werden. Eine inhaltliche Hilfe, die nur implizit eine Rückmeldung (beispielsweise zu der Richtigkeit einer Aufgabe) beinhaltet, wird nicht als Rückmeldehilfe angesehen.

- Allgemein-strategische Hilfen

Die Schülerinnen und Schüler werden auf Problemlösemethoden aufmerksam gemacht. Es handelt sich hierbei um eine Hilfe, welche auf die Steuerung des Arbeitsprozesses zielt, wobei jedoch nicht auf den fachlichen Inhalt eingegangen wird. Hierzu gehören zum Beispiel allgemeine Anregungen zum Nachdenken über den eigenen Lösungsweg, aber auch Hinweise, welche auf den Umgang mit den Medien (z.B. Computerprogramm) zielen, oder solche, die auf andere Hilfen (z.B. Hilfeheft) verweisen. Der Fokus liegt auf der Vermittlung einer Strategie, die sowohl in anderen Situationen als auch in den meisten Fällen in anderen Fächern verwendet werden könnte.

- **Inhaltsorientiert-strategische Hilfen**

Die Schülerinnen und Schüler erhalten Hilfen, die für das fachspezifische Arbeitsverfahren von Nutzen sind. Jedoch handelt es sich hierbei um keine inhaltliche Hilfe, da den Schülerinnen und Schüler nicht gesagt wird, wie etwas zu tun ist, sondern lediglich, was sie tun sollen. Der Fokus liegt somit auf der Bereitstellung einer Strategie, welche jedoch auf den Inhalt und das Fach bezogen ist und daher im Regelfall nicht direkt auf andere Fächer übertragen werden kann. Die Lehrkraft kann durch gezielte Fragen das Vorwissen der Schülerinnen und Schüler erfassen und sie so leiten, dass sie selbständig zur Lösung bzw. zum Lösungsweg gelangen. Hierzu gehören ebenfalls Definitionen von Begriffen, die den Schülerinnen und Schülern nicht bekannt sind, solange diese nicht auf den direkten Inhalt der Aufgabe schließen lassen.

- **Inhaltliche Hilfen**

Die Schülerinnen und Schüler erhalten inhaltliche Informationen zu ihrer Aufgabe. Ihnen wird mitgeteilt, welchen Arbeitsschritt sie als nächstes durchführen müssen, wobei das Verfahren meistens eindeutig mathematisch benannt wird. In manchen Fällen wird bereits das Teilergebn bzw. das Ergebnis vorgegeben. Die Schülerinnen und Schüler werden also konkret auf die Lösung hingewiesen, sodass sie selbst keine Transferleistung mehr erbringen müssen.

Die Interventionen wurden von einer Kodiererin und einem Kodierer kodiert. Dabei konnten innerhalb einer Intervention auch mehrere verschiedene Hilfe-Kategorien erfasst werden. Beispielsweise wäre dies der Fall, wenn der jeweilige Student bzw. die jeweilige Studentin den Schülerinnen und Schülern zunächst eine Strategie vermittelt, die auch in einer anderen Situation oder einem anderen Fach genutzt werden kann (allgemein-strategische Hilfe) und abschließend noch kurz motivierend auf die Schülerinnen und Schüler einwirkt (Motivationshilfe). In diesem Fall würden also zwei Hilfen innerhalb einer Intervention kodiert. Analog können auch mehrere Hilfen in verschiedenen Konstellationen innerhalb einer Intervention erteilt werden. Ebenso ist es möglich, dass im Rahmen einer Intervention nur genau eine Hilfe gegeben wird.

Aus diesem Grund wurde einerseits jeder Intervention genau eine Kategorie zugeordnet, die innerhalb der Intervention sowohl inhaltlich als auch in ihrem zeitlichen Umfang dominant war. Im obigen Beispiel, bei dem den Schülerinnen und Schülern zunächst eine Strategie vermittelt wird, wobei sie abschließend zudem kurz motiviert werden, wäre die allgemein-strategische Hilfe vermutlich zeitlich und inhaltlich dominanter als die Motivationshilfe. Entsprechend würde diese Hilfe auch als „dominante Hilfe“ kodiert werden. Die Motivationshilfe würde als „untergeordnete Hilfe“ kodiert.

Eine Besonderheit hierbei stellt die inhaltliche Hilfe dar. Tritt eine inhaltliche Hilfe gemeinsam mit einer allgemein-strategischen und/oder inhaltsorientiert-strategischen Hilfe auf, so können letztere in manchen Fällen als Vorbereitung für die inhaltliche Hilfe angesehen werden, wodurch die inhaltliche Hilfe dominant würde. Dies ist der Tatsache geschuldet, dass die Vorgabe einer (Teil-)Lösung den Stellenwert der Hilfe innerhalb der Intervention verändert, da niederschwelligere Hilfen möglicherweise obsolet werden, wenn am Ende eine (Teil-)Lösung vorgegeben wird.

Die Übereinstimmung der beiden Kodierer (Cohens Kappa) liegt im arithmetischen Mittel bei $\kappa = 0.775$. Der geringste Wert liegt bei $\kappa = 0.539$, der höchste Wert bei $\kappa = 0.950$ und der Median bei $\kappa = 0.798$. Somit liegt eine gute Übereinstimmung vor (Kuckartz 2016, 210). Die Berechnungen wurden jeweils mit der Software R und dem Package „irr“ (Gamer et al. 2019) durchgeführt.

Im Anschluss an die Einzelkodierungen fand eine kommunikative Validierung statt (Kvale 2007; Guest, MacQueen & Namey 2012). Im Rahmen dieses Konsensgesprächs wurden alle ungleich kodierten Interventionen erneut von beiden Kodierern betrachtet und beide tauschten sich anschließend argumentativ darüber aus. Durch diesen Vorgang wurde für alle Interventionen eine einheitliche Kodierung vorgenommen, die als Grundlage für weitere Analysen dient. Es kann davon ausgegangen werden, dass die Konsenskodierung aufgrund des mehrstufigen Prozesses und des somit vergleichsweise aufwendigen Vorgehens zu einer guten inhaltlichen Passung beiträgt. Der verwendete Kodierleitfaden befindet sich im Anhang (vgl. Anhang C).

8.3.4 Analyse der Angemessenheit

Neben der Analyse der verwendeten Hilfen innerhalb der Interventionen wurde auch untersucht, ob eine Intervention jeweils angemessen war. Um dies zu bewerten, wurden Kriterien der Angemessenheit bzw. der Unangemessenheit von Interventionen für den Kontext des Mathematik-Labors „Mathe ist mehr“ entwickelt und zusammengestellt. Die Kriterien sind folgende:

- Auswahl der Hilfe-Kategorie

Bei diesem Kriterium geht es um die Wahl der Hilfe. Dabei ist entscheidend, ob die Expertinnen und Experten die getätigte Hilfe-Kategorie als für die Schülerinnen und Schüler in dieser Situation passend einstufen oder ob sie eine andere Hilfe-Kategorie an dieser Stelle als sinnvoller erachten.

- Stufung der Hilfen

Dieses Kriterium kommt dann zum Tragen, wenn mehrere Hilfen innerhalb einer Intervention getätigt wurden. Hierbei wird hinterfragt, ob die Hilfen aufeinander aufbauen bzw. ob die Reihenfolge der Hilfen stimmig ist. Außerdem wird darauf geachtet, ob zunächst Fragen gestellt wurden, um die eigene Beobachtung mit dem realen Stand der Schülerinnen und Schüler zu vergleichen.

- Erkennbare Hypothese zum Schülerproblem

Bei diesem Kriterium geht es darum, ob die Experten eine der Intervention inhärente Hypothese zum Schülerproblem erkennen können oder nicht. Wenn die Hilfe des Studierenden anscheinend unbegründet erfolgt und die Hilfe dadurch wenig zielgerichtet wirkt, könnte dies an dieser Stelle identifiziert werden.

- Inhalt der Intervention

Hierbei geht es um die fachliche Korrektheit der gegebenen Hilfe. Dieses Kriterium kann entsprechend nur bei Hilfen mit inhaltlichem Charakter angewandt werden (inhaltsorientiert-strategische Hilfen oder inhaltliche Hilfen).

- Länge der Intervention

Dieses Kriterium befasst sich mit dem zeitlichen Umfang der Hilfe. Dabei wird bewertet, ob eine Intervention in einer zeitlich angemessenen Form stattfand oder diese deutlich zu lang (länger als 3 Minuten) war.

- Medienunterstützung

Hierbei geht es darum, ob die Hilfe ggf. durch eine geeignete Medienwahl unterstützt wurde. Dieses Kriterium ist nicht für alle Hilfe-Kategorien relevant. Selbst bei inhaltsorientiert-strategischen Hilfen und inhaltlichen Hilfen muss jeweils im Einzelfall entschieden werden, ob zusätzlich zur sprachlichen Hilfe ein geeignetes Medium genutzt werden sollte.

- Intensität der Intervention

Bei diesem Kriterium geht es darum, ob die Intensität der Intervention als angemessen bewertet wird. Beispielsweise können den Schülerinnen und Schülern bei inhaltlichen Hilfen deutlich zu viele Informationen mit auf den Weg gegeben werden.

- Präsenz der Lehrperson

Dieses Kriterium befasst sich mit der physischen Präsenz der angehenden Lehrkraft. Dabei wird insbesondere bei disziplinarischen Hilfen darauf geachtet, ob die Hilfe durch die Lehrerpräsenz angemessen ausgeführt wurde.

- Angemessenheit der Sprache

Bei diesem Kriterium geht es in erster Linie um die adäquate Verwendung von Fachsprache. Darüber hinaus können auch grobe sprachliche Fehler bemängelt werden, die nicht fachsprachlich sind.

Anhand dieser Kategorien rezipierte zunächst eine kleine Expertengruppe, bestehend aus zwei Lehrkräften und zwei wissenschaftlichen Mitarbeiterinnen und Mitarbeitern, die Interventionsvideos und bewertete die Interventionen auf einer Skala von 1 bis 4, wobei 1 eine völlig unangemessene Intervention darstellt und 4 eine völlig angemessene Intervention (vgl. Abb. 26). Eine 2 steht für eine eher unangemessene Intervention, eine 3 für eine eher angemessene Intervention. Bei allen Bewertungen unter einem Skalenwert von 4 wurde mindestens ein Kriterium angegeben, welches bei der Intervention als unangemessen eingestuft wurde, sodass alle festgestellten Mängel begründet wurden. Die Bewertungen der einzelnen Interventionen wurden jeweils summiert und durch die Anzahl der Rater geteilt. Lag der Mittelwert unter 2, so wurde die Intervention als unangemessen gewertet. Bei einem Mittelwert über 3 wurde die Intervention als angemessen bewertet.

Alle Interventionen mit einem Mittelwert zwischen 2 und 3 wurden einer großen Experten-Gruppe, bestehend aus Professorinnen und Professoren, Lehrkräften und wissenschaftlichen Mitarbeiterinnen und Mitarbeitern vorgelegt. Diese hatten ebenfalls die Aufgabe, das Interventionsvideo anzusehen und eine Bewertung zwischen 1 und 4 abzugeben. Auch hier führten Mittelwerte von 3 und höher (angemessen) bzw. 2 und niedriger (unangemessen) zu einer abschließenden Bewertung der Angemessenheit. Mittelwerte zwischen 2 und 3 führten dazu, dass nun zunächst Argumente für die Angemessenheit bzw. Unangemessenheit vorgetragen und in der Gruppe diskutiert wurden. Im Anschluss wurde das Interventionsvideo erneut betrachtet und eine Abstimmung per Mehrheitsentscheid durchgeführt. Bei dieser letzten Schleife bestand lediglich die Möglichkeit, sich zwischen angemessen und unangemessen zu entscheiden.

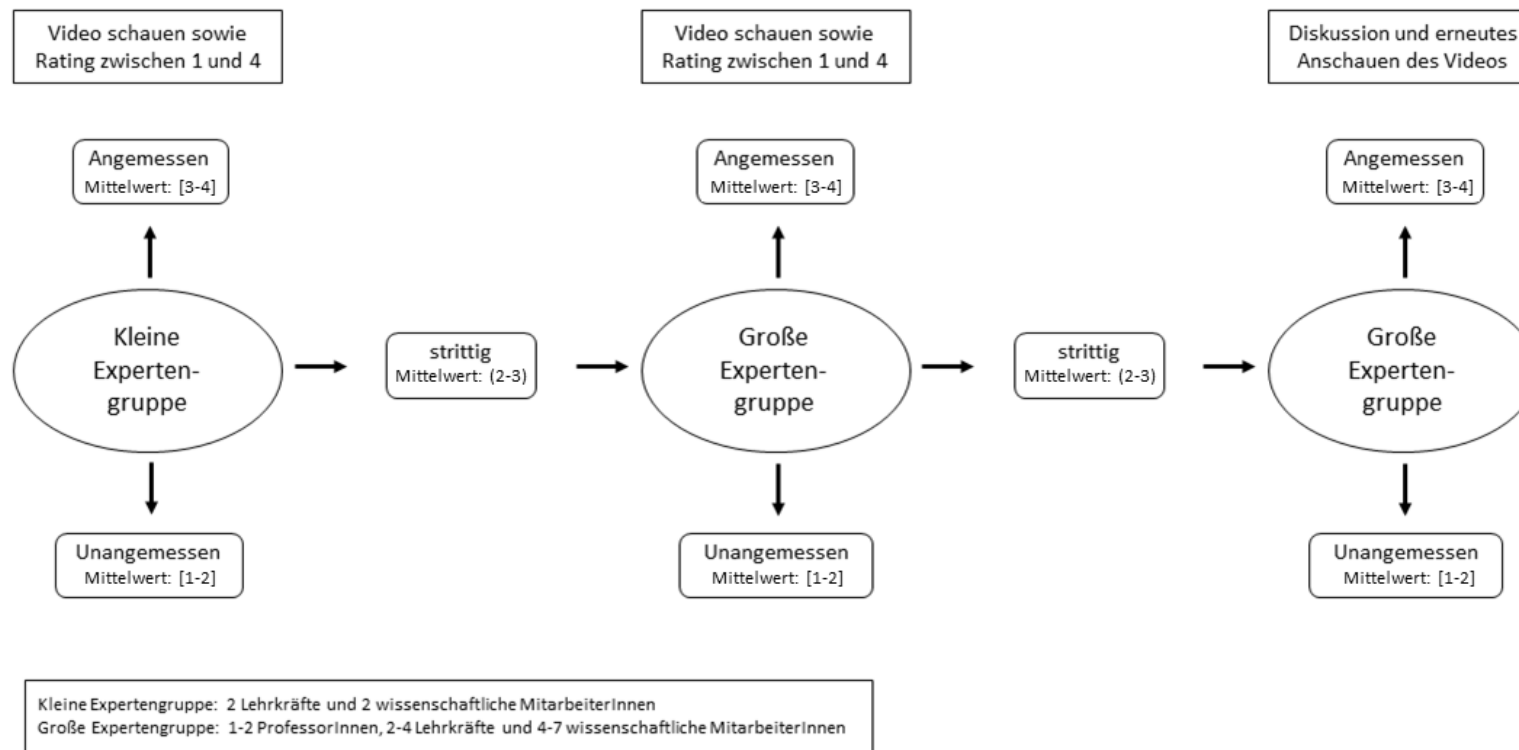


Abbildung 26: Schaubild des Prozesses zur Bewertung der Angemessenheit der Interventionen der Studierenden durch verschiedene Expertengruppen

Dieses aufwendige Verfahren soll gewährleisten, dass die Bewertung der Angemessenheit möglichst fundiert stattfindet. Mit Hilfe der kleinen Expertengruppe wurde zunächst die Masse an Interventionsvideos reduziert, da alle verhältnismäßig eindeutigen Entscheidungen in die eine oder andere Richtung bereits vorgenommen werden konnten. Dadurch konnte die große Expertengruppe entlastet werden. Insgesamt konnten durch die kleine Expertengruppe bereits 161 der 208 Interventionen abschließend auf deren Angemessenheit beurteilt werden. Die verbleibenden 47 Interventionen, die im ersten Durchlauf noch als strittig bewertet wurden (Mittelwert der vier Rater zwischen 2 und 3), wurden von der großen Expertengruppe bewertet. Diese konnte schließlich im ersten Durchgang der großen Expertengruppe einen Großteil der zunächst strittigen Interventionen abschließend bewerten, sodass in weniger als zehn Fällen eine letzte Runde mit dichotomer Entscheidung und einfacher Mehrheit zur abschließenden Bewertung genutzt wurde.

8.3.5 Weitere Analysen

Neben den Auswertungen der dominanten und untergeordneten Hilfen sowie der Angemessenheit werden außerdem die Dauer der Intervention, die Dauer der einzelnen Hilfen innerhalb der Intervention, der Interventionsauslöser (invasiv vs. responsiv), die Anzahl der arbeitenden Schülerinnen und Schüler vor bzw. nach der Intervention sowie die Reihenfolge der Hilfen innerhalb einer Intervention analysiert. Die Reihenfolge der Hilfen wurde im Zuge der Kodierung dominanter und untergeordneter Hilfen mitkodiert.

Als Dauer der Intervention wird bei allen Interventionen die Zeit vom Öffnen der Tür beim Betreten des Filmraumes bis zum Schließen der Tür beim Verlassen des Raumes durch eine Studierende bzw. einen Studierenden angesehen. In wenigen Ausnahmefällen ist dies nicht möglich, beispielsweise, wenn die Intervention direkt nach einer Pause stattfand oder das Ende der Intervention mit dem Ende des Laborbesuches der Schülerinnen und Schüler zusammenfiel. Bei diesen seltenen Ausnahmefällen werden jeweils naheliegende Bezugspunkte gewählt, beispielsweise das gemeinsame Betreten des Raumes der bzw. des Studierenden und der Schülerinnen und Schüler (ggf. ohne das Öffnen der Tür). Die Dauer der einzelnen Hilfen innerhalb der Intervention wurde im Rahmen des Konsensgespräches zur Kodierung der dominanten und untergeordneten Hilfen ermittelt. Dabei wurde neben der Entscheidung über die Hilfeart (dominant bzw. untergeordnet) auch darüber entschieden, wann welche Hilfe beginnt und endet. Insgesamt entspricht die Summe der Zeitdauer aller Hilfen innerhalb einer Intervention der Interventionsdauer.

Zur Ermittlung der Anzahl der arbeitenden Schülerinnen und Schüler werden jeweils 30 Sekunden vor und 30 Sekunden nach der Intervention betrachtet. Alle Schülerinnen und Schüler werden dabei einzeln im Hinblick auf die nachstehenden Kriterien betrachtet. Eine Schülerin bzw. ein Schüler wird als arbeitend eingestuft, wenn sie bzw. er:

- über die Aufgabe bzw. den Arbeitsauftrag kommuniziert
- einem oder mehreren Mitschüler/inne/n aktiv zuhört
- etwas in das Arbeitsheft schreibt
- im Arbeitsheft oder im Hilfeheft liest
- mit dem gegenständlichen Material arbeitet
- mit einer Simulation bzw. einem Video arbeitet

Die beiden zuletzt genannten Punkte sind jeweils in Abgrenzung zu reinem Spielen zu sehen. In Fällen, bei denen keine Beobachtung möglich ist, also das Arbeitsverhalten vor bzw. nach der Intervention nicht beobachtet werden kann, wird dies entsprechend vermerkt. Beispielsweise, wenn die Intervention der oder des Studierenden sich an eine Pause anschließt und die Schülerinnen und Schüler den Filmraum erst mit Beginn der Intervention betreten. In solchen Fällen gibt es de facto keine 30 Sekunden vor der Intervention, in der das Arbeitsverhalten der Schülerinnen und Schüler analysiert werden könnte. Dies kann bei Erprobungen von Laborstationen vorkommen, bei denen die Schulklasse eine weite Anreise hat und daher alle drei Teilstationen an einem Tag durchführt. Ebenso kann es passieren, dass eine Intervention am Ende eines Laborbesuches stattfindet. In solchen Fällen gibt es keine 30 Sekunden nach der Intervention, in der das Arbeitsverhalten der Schülerinnen und Schüler analysiert werden könnte. Dies kann potenziell bei allen Schülergruppen der Fall sein.

Das Arbeitsverhalten vor der Intervention wurde bei allen vorliegenden Videos von zwei Personen kodiert. Bei den verschiedenen Kategorien (Anzahl der arbeitenden Schülerinnen und Schüler) wurde im arithmetischen Mittel eine Übereinstimmung von $\kappa = 0.646$ erzielt. Der geringste Wert liegt bei $\kappa = 0.575$, der höchste Wert bei $\kappa = 0.790$ und der Median bei $\kappa = 0.636$. Insgesamt liegt also eine gute Übereinstimmung vor (Kuckartz 2016, 210). Das Arbeitsverhalten nach der Intervention wurde dann ausschließlich von einer der beiden Personen kodiert.

Der Auslöser der Intervention, sprich ob die Intervention von den Schülerinnen und Schülern (responsiv) oder von der Lehrperson (invasiv) ausging, wurde bei allen 208 Interventionsvideos von zwei Personen kodiert. Bei der Kategorie „responsiv“ liegt die Übereinstimmung der beiden Kodierer bei $\kappa = 0.825$. Bei der Kategorie „invasiv“ liegt die Übereinstimmung bei $\kappa = 0.788$. Im arithmetischen Mittel ergibt sich daher eine Übereinstimmung von $\kappa = 0.807$. Die Übereinstimmung ist daher als sehr gut anzusehen (Kuckartz 2016, 210).

8.4 Auswertung im Bereich Reflexion

Ähnlich wie die Interventionen der Studierenden ist auch die Reflexion über die Erprobung der Laborstation sowie über die getätigten Interventionen fest im Didaktischen Seminar verankert (Kap. 8.4.1). Die videographierten Reflexionsgespräche werden anschließend mit Hilfe der Software MAXQDA 12 analysiert und ausgewertet (Kap. 8.4.2).

8.4.1 Einbettung in das Didaktische Seminar

Nachdem die Station erprobt wurde und alle Gruppenmitglieder die Filmgruppe während eines Stationsteils genauer beobachtet haben, findet ein Reflexionsgespräch mit allen Gruppenmitgliedern statt. Hierbei sollen die Studierenden zunächst über die Erprobung an sich reflektieren. Dabei steht es den Studierenden komplett frei, ob sie der Reihe nach alle Aufgaben von Stationsteil eins besprechen und dann analog für die beiden verbleibenden Teile vorgehen oder ob sie weniger systematisch vorgehen und ihre Erinnerungen und Eindrücke auf andere Art und Weise äußern und reflektieren.

Im Anschluss schauen sich die Studierenden die von ihnen getätigten Interventionen an. Hierbei soll sich zunächst derjenige äußern, der den Eingriff getätigt hat und schildern, weshalb er es getan hat, ob er wieder intervenieren würde und falls ja, wie diese Intervention aussehen würde. Auch sonstige Gedanken und Gefühle in Bezug auf die gesehene Intervention können geäußert werden. Wenn derjenige fertig ist, haben die Gruppenmitglieder die Möglichkeit zu

äußern, ob sie den Eingriff ebenfalls getätigt oder ob sie evtl. anders gehandelt hätten. Die Äußerungen sollten jeweils möglichst fundiert begründet werden.

Nachdem alle Interventionen angesehen und besprochen wurden, haben die Gruppenmitglieder die Möglichkeit, Situationen zu schildern, die ihnen besonders im Gedächtnis geblieben sind. Beispielsweise könnten an dieser Stelle Situationen beschrieben werden, in denen sich Studierende begründet gegen eine Intervention entschieden haben, obwohl die Schülergruppe Probleme hatte oder sogar inhaltliche Fehler begangen hat.

Das Reflexionsgespräch findet ausschließlich unter den Gruppenmitgliedern statt und es befindet sich sonst auch niemand im Raum. Somit soll verhindert werden, dass die Studierenden von einer Interviewerin oder einem Interviewer gelenkt und / oder gehemmt werden. Letzteres könnte beispielsweise der Fall sein, wenn es sich bei dem Interviewer auch um den Seminarleiter handelt.

8.4.2 Datengenerierung und Analyse der Reflexionsvideos

Auswahleinheit und Analyseeinheit sind hierbei jeweils die videographierten Reflexionsgespräche. Die Reflexionsgespräche wurden vollständig videographiert. Alle Studierenden äußerten sich dabei und nahmen aktiv am Reflexionsgespräch teil. Es liegen 17 Reflexionsvideos von Kleingruppen, bestehend aus zwei bzw. drei Teilnehmer/innen, aus den Semestern vom Sommersemester 2016 bis zum Wintersemester 2017/18 vor. Insgesamt liegen Äußerungen und Reflexionsmomente von 46 Teilnehmenden des Didaktischen Seminars vor. Es wurde eine Qualitative Inhaltsanalyse, angelehnt an das Vorgehen der inhaltlich strukturierenden Analyse nach Kuckartz (2016), durchgeführt. Die Hauptkategorien wurden anhand des ALACT-Modells von Korthagen (2002) erstellt. Da die erste und die letzte Stufe dieses Modells jeweils konkrete Handlungen darstellen und entsprechend noch nicht zur Reflexion gehören, wurden drei Kategorien zugrunde gelegt: Looking back on the action (Rückblick), Awareness of essential aspects (Wesentliche Aspekte) und Creating alternative methods of action (Alternativen). Zusätzlich zu diesen deduktiv erstellten Kategorien wurden diese anhand des Videomaterials induktiv in mehrere Unterkategorien aufgegliedert, sodass folgendes Kategoriensystem zugrunde gelegt wurde:

- Looking back on the Action (Rückblick)

Diese Kategorie umfasst vier Unterkategorien, welche jeweils weiter untergliedert sind. Es werden Absichten und Ziele, inhaltliche und organisatorische Aspekte sowie Gedanken und Gefühle aufgegriffen und entweder ein noch recht grober Rückblick gegeben oder Aspekte genannt, auf die nicht weiter eingegangen wird bzw. welche nicht als wesentlich angesehen werden. Im Folgenden werden die einzelnen Unterkategorien mit ihren jeweiligen Untergliederungen genauer betrachtet und erläutert.

- Absichten und Ziele

Die Absichten und Ziele der Studierenden können sich einerseits auf die Lernumgebung und andererseits auf die Intervention(en) beziehen. Somit werden die Überlegungen während der Gestaltung der Lernumgebung sowie die Intentionen der getätigten Interventionen berücksichtigt. Die Studierenden äußern hierbei, was sie sich vorgenommen hatten, was sie ausprobieren wollten und auf welche Dinge sie achten wollten. Zudem

fällt unter diese Unterkategorie der Abgleich der Ziele der Lernumgebung mit der Erprobung und der Abgleich der Ziele der Intervention mit dem getätigten Eingriff. Des Weiteren wird die Auswertung der Intervention(en) berücksichtigt.

- Inhaltliche und organisatorische Aspekte

Eine weitere Unterkategorie beschäftigt sich mit den inhaltlichen und organisatorischen Aspekten, wobei zwischen Aspekten der Intervention(en) und Aspekten der Lernumgebung unterschieden wird. Es handelt sich hierbei um Aspekte, die den Verlauf der Station mitbestimmen bzw. während der Erprobung zu beobachten waren, aber von den Studierenden nicht als wesentlich wahrgenommen wurden.

- Gedanken und Gefühle

Darüber hinaus werden die Gedanken und Gefühle näher betrachtet. Es wird zwischen den Gedanken und Gefühlen der Schülerinnen und Schüler und der Studierenden unterschieden, wobei diese jeweils wieder auf die Lernumgebung oder die Interventionen bezogen werden. Es ist zu beachten, dass diesen Unterkategorien lediglich die Aussagen zugeordnet werden, die sich auf Gedanken und Gefühle während der Erprobung oder Intervention beziehen.

- Grober Rückblick

Die letzte Unterkategorie umfasst einen groben Rückblick, welcher bzgl. der Lernumgebung oder der durchgeführten Intervention(en) aufgestellt wird. Dieser grobe Rückblick enthält Überlegungen zur Station im Allgemeinen oder auch über einen Stations- teil hinweg. Die Betrachtung konkreter Aufgaben gehört nicht mehr zu dieser Unterkategorie.

- Awareness of essential aspects (Wesentliche Aspekte)

Die Studierenden nehmen hierbei bestimmte Aspekte genauer in den Blick und machen sich dadurch nochmals essentielle Aspekte bewusst. Dies können sowohl konkrete inhaltliche Aspekte als auch organisatorische Aspekte sein, welche einen Einfluss auf den Lernzuwachs der Schülerinnen und Schüler haben können. Dabei geht es um die von den Studierenden subjektiv als essentiell wahrgenommenen Aspekte und nicht um die Aspekte, die für den Kodierenden wesentlich sind. Innerhalb dieser Kategorie werden drei Unterkategorien unterschieden, welche zum Teil wiederum untergliedert sind. Im Folgenden werden die Unterkategorien genauer betrachtet, erläutert und voneinander abgegrenzt. Neben einzelnen organisatorischen sowie inhaltlichen Aspekten der Lernumgebung bzw. der Intervention(en) wird das kritische Hinterfragen durch die Studierenden genauer in den Blick genommen.

- Inhaltliche und organisatorische Aspekte

Die Studierenden richten ihren Fokus hierbei auf bestimmte Aspekte und machen sich dadurch nochmals wesentliche inhaltliche oder zentrale organisatorische Aspekte der Lernumgebung bzw. der Intervention bewusst. Es geht dabei um die von den Studierenden subjektiv als wesentlich wahrgenommenen Aspekte und nicht um die Aspekte, die für den Kodierenden wesentlich sind.

- Kritisches Hinterfragen

Darüber hinaus umfasst diese Kategorie die genauere Betrachtung bzw. das kritische Hinterfragen des eigenen Handelns sowie der eigenen Vorbereitung. Um dieser Unterkategorie zu entsprechen, reicht eine Aussage über das eigene Handeln oder die eigene Vorbereitung alleine allerdings nicht aus. Es ist notwendig, dass der Aussage eine kritische Auseinandersetzung in Form einer diesbezüglichen Argumentation oder Hypothesenbildung folgt. Auch die Kritik an anderen sowie an der Vorbereitung der Gruppe oder der einzelnen Person sind von dieser Kategorie auszuschließen.

- Creating alternative methods of action (Alternativen)

Die Studierenden sprechen davon, alternative Handlungsformen zu gestalten. Dies kann sich sowohl auf die Gestaltung der Lernumgebung als auch auf die getätigten Interventionen beziehen. Es wird betrachtet, was beim nächsten Mal anders gemacht bzw. verbessert werden kann. Dabei können auch Vor- und Nachteile einer Alternative abgewogen werden. In dieser Kategorie wird zwischen der Nennung von Alternativen sowie dem Abwägen von Vor- und Nachteilen in Bezug auf die Lernumgebung bzw. die Intervention(en) unterschieden. Insgesamt ergeben sich somit aus den beiden Unterkategorien vier Differenzierungsmöglichkeiten, welche im Folgenden genauer betrachtet werden.

- Nennung von Alternativen

Die Studierenden sprechen davon, alternative Handlungsformen zu gestalten. Dies kann sich sowohl auf die Gestaltung der Lernumgebung als auch auf die getätigten Interventionen beziehen. Es wird betrachtet, was beim nächsten Mal anders gemacht bzw. verbessert werden kann.

- Abwägung von Vor- und Nachteilen

Im Zusammenhang mit der Nennung von Alternativen können auch Vor- und Nachteile bzgl. der Lernumgebung oder von verschiedenen Eingriffen abgewogen werden. Hierbei ist eine kritische Auseinandersetzung mit dem betrachteten Aspekt sowie das Anführen von Argumenten, welche abgewogen werden, notwendig.

Den einzelnen Hauptkategorien wurden jeweils verschiedene Unterkategorien zugeordnet, welche einerseits aus theoretischen Überlegungen entstanden sind, andererseits anhand des Materials (induktives Bestimmen von Subkategorien anhand des Materials). Auf dieser Grundlage wurde das Datenmaterial durch zwei Kodierer, mithilfe der Software MAXQDA 12, kodiert. Zunächst wurden dazu mehrere Reflexionsvideos gemeinsam kodiert. Anschließend wurden abwechselnd Videos von einem Kodierer kodiert und vom anderen Kodierer überprüft. Nachdem beide Kodierer eine sehr hohe Übereinstimmung hatten und kaum noch Unstimmigkeiten auftraten, wurde das restliche Datenmaterial jeweils nur noch von einem Kodierer kodiert. Auftretende Unklarheiten und Unsicherheiten wurden geäußert und besprochen, bis alle Fundstellen eindeutig zugewiesen werden konnten. Bei der Kodierung wurde Wert daraufgelegt, dass die Häufigkeit der getätigten Reflexionsmomente der einzelnen Studierenden möglichst genau erfasst wurde. Gleiche Aussagen des gleichen Studierenden, also reine Wiederholungen der eigenen Worte, wurden nicht nochmals als neues Reflexionsmoment gewertet. Der verwendete Kodierleitfaden befindet sich im Anhang (vgl. Anhang D).

8.5 Beschreibung der Population

Nachdem die Methodik in den verschiedenen Bereichen vorgestellt wurde, wird nun die jeweils zugrundeliegende Population beschrieben. Dazu wird in Teilkapitel 8.5.1 die Auswahl der Expertinnen und Experten erläutert. Anschließend erfolgt die Beschreibung der Auswahl der Kodierinnen und Kodierer (Kap. 8.5.2). Teilkapitel 8.5.3 beschreibt die Zusammensetzung der Studierenden und Teilkapitel 8.5.4 die Zusammensetzung der Lehrkräfte. Abschließend geht Teilkapitel 8.5.5 auf Verschränkungen mit anderen Studien ein.

8.5.1 Auswahl der Expertinnen und Experten

Für die Beantwortung der Diagnoseaufträge bzgl. der prozessdiagnostischen Fähigkeiten wurden Professoren der Mathematikdidaktik, Mathematik-Lehrkräfte und Wissenschaftliche Mitarbeiterinnen und Mitarbeiter im Bereich Mathematik bzw. Mathematikdidaktik herangezogen. Die Professoren qualifizieren sich insbesondere durch ihre Expertise in der Didaktik der Algebra und durch ausgeprägte Kenntnisse im Bereich Diagnostik. Darüber hinaus bringen sie ihre Kompetenzen und Erfahrungen aus ihrem vielfältigen Arbeitsalltag mit in die Beurteilung. Die Mathematik-Lehrkräfte verfügen über einen reichhaltigen Erfahrungsschatz, welchen sie über Jahre im Schulalltag aufgebaut haben. Das Thema Terme ist dabei derart zentral, dass man als Lehrkraft häufig damit in Berührung kommt und somit einerseits fachlich gut aufgestellt sein sollte, andererseits implizit oder sogar explizit auch schon seinen Blick für die Entwicklung der Fähigkeiten und Fertigkeiten der Schüler geschult hat. Unter den Wissenschaftlichen Mitarbeitern befinden sich einige, die sich über Jahre mit Diagnosekompetenz auseinandergesetzt haben und selbst in diesem Bereich forschen. Andere beschäftigen sich mit Videovignetten aus anderen Themengebieten, wobei auch andere Schwerpunkte wie etwa die Gestaltung der Arbeitsaufträge oder das passende Feedback für die Studierenden gesetzt werden. Durch diese Vielfalt soll ein möglichst breites Spektrum an Expertise in die kumulativ angesammelten Musterlösungen, basierend auf den Antworten der Expertinnen und Experten, einfließen.

Zur Bewertung, ob die einzelnen Interventionen in die jeweiligen Gruppenarbeitsprozesse der Schülergruppe angemessen waren oder nicht, wurde zunächst eine Kleingruppe, bestehend aus zwei Lehrkräften und zwei Wissenschaftlichen Mitarbeitern, herangezogen. Die Lehrkräfte verfügen über langjährige Erfahrung im Mathematikunterricht und haben über die Jahre schon verschiedenste Situationen erlebt und bewerten müssen. Auf Grundlage dieses Erfahrungsschatzes qualifizieren sie sich, um auf Grundlage der Kriterien der Angemessenheit bzw. Unangemessenheit von Eingriffen die Interventionen der Studierenden zu beurteilen. Die beiden wissenschaftlichen Mitarbeiter sind beide sehr gut mit dem Mathematik-Labor „Mathe ist mehr“ vertraut und forschen beide im Bereich Diagnosekompetenz. Darüber hinaus arbeiten beide mit Videovignetten und sind mit dem Kodieren vertraut. Die jeweiligen Kenntnisse und Erfahrungen aus den unterschiedlichen Bereichen, über die die vier Rater verfügen, ergänzen sich optimal, sodass die Interventionen sachgemäß beurteilt werden können.

Darüber hinaus gibt es eine größere Gruppe, bestehend aus Professorinnen und Professoren der Mathematikdidaktik, Lehrkräften und Wissenschaftlichen Mitarbeiterinnen und Wissenschaftlichen Mitarbeitern, welche sich die Interventionen anschaut, bei denen die Kleingruppe unterschiedlich geratet hat. Die Professoren, die einerseits über verschiedenste mathematikdidaktischen Kenntnisse verfügen, andererseits aber auch für eine Zeit in Schulen tätig waren und dort Erfahrungen als Lehrkräfte gesammelt haben, eignen sich hervorragend, um die Angemes-

senheit der Interventionen zu beurteilen. Die zusätzlichen Mathematik-Lehrkräfte und Wissenschaftlichen Mitarbeiter der Mathematikdidaktik sind aus den oben genannten Gründen ebenfalls für die Entscheidungsfindung der Angemessenheit von Interventionen geeignet.

8.5.2 Auswahl der Kodiererinnen und Kodierer

Für die Auswertung der Diagnoseaufträge der Studierenden zur Erfassung ihrer prozessdiagnostischen Kompetenz wurden zwei Hilfskräfte akquiriert, die beide Mathematik-Lehramt studieren. Dies sollte dem inhaltlichen Verständnis der Videovignetten, der dazugehörigen Diagnoseaufträge sowie der Expertenantworten dienen. Darüber hinaus wurde mehreren Bewerbern die Aufgabe gegeben, anhand eines Artikels zu Grundvorstellungen im Bereich Terme die beiden Grundvorstellungen in eigenen Worten zu erklären. Da die beiden Grundvorstellungen „Term als Bauplan“ und „Term als Rechenschema“ zentral sind und einen Großteil der Fragen ausmachen, war es wichtig, dass die beiden Kodiererinnen die Begriffe verstehen und eigenständig wiedergeben können. Immerhin müssen sie auch in der Lage sein, die in eigenen Worten formulierten Antworten der Studierenden, welche durchaus von den Musterlösungen der Expertinnen und Experten abweichen können, inhaltlich einschätzen und bewerten zu können.

Jeweils eine der beiden Kodiererinnen wurde anschließend auch für die Auswertung der Interventionen bzw. der Reflexionsvideos eingesetzt. Durch die ausführliche Vorstellung des Mathematik-Labors „Mathe ist mehr“ und des „Didaktischen Seminars“ waren beide mit den erforderlichen Rahmenbedingungen und Hintergrundinformationen vertraut. Beide arbeiteten sich darüber hinaus ausgiebig in die jeweilige Thematik ein. Die zugrundeliegenden Modelle, auf denen das jeweilige Kategoriensystem bei den Interventionen bzw. den Reflexionsvideos beruht, aber auch alternative Modelle in den beiden Bereichen wurden mit dem entsprechenden Kodiererinnen besprochen und diskutiert. Durch diesen erweiterten Blick und die Miteinbeziehung bei der Überarbeitung der Kategoriendefinitionen waren die Kodiererinnen gut auf ihre Arbeit vorbereitet.

8.5.3 Zusammensetzung der Studierenden

Die Studierenden entstammen aus vier „Didaktischen Seminaren“, welche vom Sommersemester 2016 bis zum Wintersemester 2017/2018 an der Universität Koblenz-Landau am Campus Landau stattfanden. Die Teilnehmerzahl der Seminare lag jeweils zwischen 5 und 19 Studierenden, welche jeweils gymnasiales Lehramt oder Realschullehramt im Masterstudium des Faches Mathematik studierten. Insgesamt liegen für 48 Studierende sowohl die Beantwortungen der Diagnoseaufträge als auch die getätigten Interventionen vor. Aufgrund technischer Probleme sind die Daten eines Reflexionsvideos unbrauchbar, sodass in diesem Bereich lediglich die Äußerungen von 46 der Teilnehmerinnen und Teilnehmer ausgewertet werden können.

8.5.4 Zusammensetzung der Lehrkräfte

Zur Validierung des Testinstrumentes im Bereich diagnostischer Fähigkeiten (vgl. dazu Kap. 8.2.7) und um die Freitextantworten der Studierenden auf die Diagnoseaufträge besser einordnen zu können (vgl. dazu Kap. 9.1) bearbeiteten 15 Lehrkräfte die drei Videovignetten (vgl. dazu Kap. 8.2.3 und Kap. 8.2.5). Dabei handelt es sich um sieben Gymnasiallehrkräfte sowie vier Lehrkräfte, die an Realschulen tätig sind und jeweils zwei Lehrkräfte, die an einer Grundschule bzw. einer Förderschule arbeiten. Die durchschnittliche Berufserfahrung beträgt etwas

mehr als 11 Jahre. Die am längsten aktive Lehrkraft blickte zum Zeitpunkt der Erhebung auf 23 Jahre Berufserfahrung zurück. Die Lehrkraft mit der geringsten praktischen Erfahrung befand sich im ersten Jahr nach dem Referendariat. Die Erfahrungswerte der Lehrkräfte streuen über diese beiden Extreme. Die Lehrkraft, die sowohl bei den meisten als auch den geringsten Berufsjahren an Stelle 8 zu finden ist (anders ausgedrückt der Median), verfügt über 10 Jahre Berufserfahrung. Keine der Lehrkräfte hatte zuvor mit ViviAn (vgl. dazu Kap. 5.1.2 und Kap. 8.2.5) gearbeitet. Die Daten wurden von zwei Studierenden im Rahmen ihrer Bachelorabschlussarbeiten erhoben. Hierbei wurde nur auf die Rohdaten zurückgegriffen, die Auswertung wurde von den Kodiererrinnen vorgenommen, die in Kapitel 8.5.2 vorgestellt wurden.

8.5.5 Verschränkung mit anderen Erhebungen

Neun der 48 Studierenden haben bereits im Sommersemester 2015 an einer Studie zum Bruchzahlverständnis und den Grundvorstellungen im Bereich Brüche, insbesondere der Grundvorstellung „Teil eines Ganzen“, teilgenommen. In dieser Studie wurde mit ViviAn gearbeitet und versucht, die diagnostischen Fähigkeiten der Studierenden zu fördern. Außerdem gibt es acht Studierende, die im Sommersemester 2016 an einer ähnlich gelagerten Studie, ebenfalls zum Thema „Brüche“, mittels ViviAn sowie mit dem Ziel der Förderung der diagnostischen Fähigkeiten teilgenommen haben.

Im Wintersemester 2016/2017 und im Wintersemester 2017/2018 fanden außerdem Erhebungen in Veranstaltungen der Didaktik der Geometrie bzw. der fachwissenschaftlichen Grundlagen statt, bei denen die Studierenden ebenfalls mit ViviAn arbeiteten und die diagnostischen Fähigkeiten der Studierenden geschult werden sollten. In diesem Zusammenhang gibt es jedoch keine Überschneidungen, sprich keiner der Teilnehmerinnen und Teilnehmer war im Erhebungszeitraum zwischen Sommersemester 2016 und Wintersemester 2017/2018 Teilnehmerin oder Teilnehmer des Didaktischen Seminars und somit Teilnehmerin oder Teilnehmer der in dieser Arbeit vorgestellten Studie.

8.6 Überblick

Nachdem die verschiedenen Vorgehensweisen in den drei Bereichen vorgestellt wurden, wird an dieser Stelle ein Überblick über die Erhebungen geliefert (vgl. Abb. 27). Zunächst konzipieren die beteiligten Dozierenden das Lehr-Lern-Labor-Seminar, damit die teilnehmenden Studierenden eine eigene Laborstation erstellen und diese mit Schülerinnen und Schülern erproben können. Im Rahmen dieses Seminars bearbeiten die Studierenden Videovignetten in ViviAn (vgl. dazu Kap. 8.2). Damit soll festgestellt werden, inwieweit die Studierenden ihre diagnostische Fähigkeit „Deuten“ bereits aufbauen und ausbilden konnten. Die Untersuchung wird im Rahmen eines Forschungsanliegens durchgeführt, entsprechend ist die oberste Ebene des Schaubilds angesprochen. Da die Studierenden in diesem Fall als Probanden dienen, ist auch die mittlere Ebene des Schaubilds von Relevanz. In den Videovignetten sind jeweils vier Schülerinnen zu sehen, die Terme aufstellen und umformen. Entsprechend ist auch die dritte Ebene bei dieser Erhebung involviert. Dabei interagieren jeweils zwei Ebenen paarweise miteinander. Einerseits wurden aus dem vorhandenen Videomaterial Videovignetten sequenziert; dabei nutzt der Forschende die videografierte Kommunikation der Schülerinnen und Schüler. Andererseits

betrachten die Studierenden die Videovignetten und machen sich Gedanken über fachdidaktische Inhalte im Hinblick auf die gesehenen Schülerhandlungen. Darüber hinaus werden die Antworten der Studierenden durch den Forschenden ausgewertet.

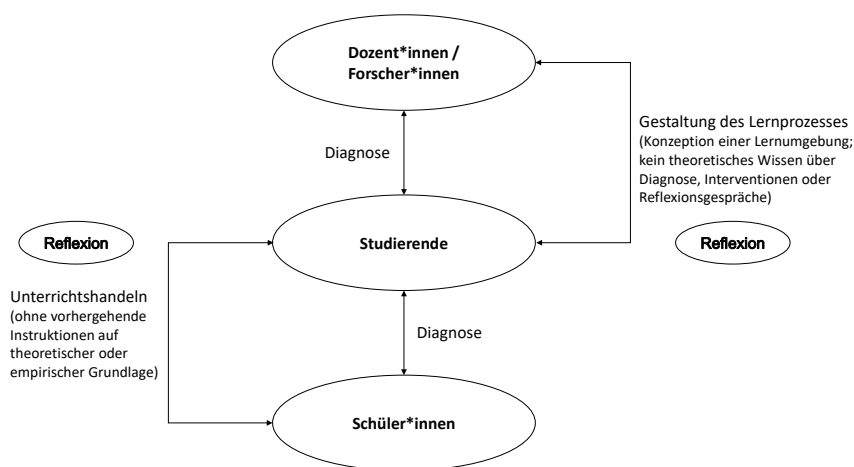


Abbildung 27: Erweiterte Übersicht über die drei zu unterscheidenden Ebenen hinsichtlich der Forschungsvorhaben im Mathematik-Labor "Mathe ist mehr"

Die Studierenden sind bei der Erprobung der Lernumgebung dazu angehalten, die Schülerinnen und Schüler jeweils für eine Teilstation genau zu beobachten und ihnen Hilfestellungen zu geben, sofern sie diese benötigen. Ob und in welcher Form die Studierenden intervenieren, ist ihnen freigestellt. Sie haben im Rahmen des Seminars keinen theoretischen Input bzgl. der Interventionen erhalten. Somit kann untersucht werden, wie die Studierenden von sich aus mit den jeweiligen Problemstellungen umgehen. Die Interventionen in die Gruppenarbeitsprozesse von Schülerinnen und Schüler betreffen ebenfalls alle drei Ebenen. Die Studierenden interagieren mit den Schülerinnen und Schülern. Der Forschende betrachtet sowohl die Interventionen der Studierenden als auch die Arbeitsphasen der Schülerinnen und Schüler vor und nach den Interventionen.

Sowohl die Gestaltung des Lernprozesses als auch das Unterrichtshandeln werden durch die Studierenden reflektiert. Die Studierenden haben zunächst die Möglichkeit, frei über die Erprobung der Laborstation zu reflektieren. Anschließend betrachten sie die getätigten Interventionen ihrer Gruppe. Dabei sind die Lernfortschritte und Lernschwierigkeiten der Schülerinnen und Schüler von Interesse. Gleichmaßen geht es aber auch um die Erkenntnisse der Studierenden und deren Nutzen in künftigen, ähnlich gelagerten Situationen. Die Reflexionsgespräche in Zweier- bzw. Dreiergruppen verliefen ohne Input eines theoretischen Vorgehens hinsichtlich einer gewinnbringenden Reflexion. Somit kann untersucht werden, wie die Studierenden gegen Ende des Mathematiklehramtsstudiums reflektieren. Die Ergebnisse können ggf. genutzt werden, um die Konzeption des Lehr-Lern-Labor-Seminars anzupassen.

Nachdem das methodische Vorgehen für die verschiedenen Bereiche erläutert und die zugrundeliegenden Populationen beschrieben wurden, werden im nächsten Kapitel die Ergebnisse der Studie vorgestellt.

9 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Analysen der im Rahmen der Veranstaltung „Didaktisches Seminar“ über vier Semester hinweg erhobenen Daten in den Bereichen Diagnose, Intervention und Reflexion vorgestellt. Zunächst werden die Ergebnisse bereichsspezifisch dargestellt. Hinsichtlich der diagnostischen Fähigkeit, den in den Videovignetten abgebildeten Lernprozess der Schülerinnen und Schüler zu deuten, werden die Ergebnisse auf Personenebene und auf Diagnoseauftragebene präsentiert. Anschließend werden die Ergebnisse von Lehrkräften vorgestellt und in Bezug zu den Ergebnissen der Studierenden gesetzt (Kap. 9.1).

Hinsichtlich der Interventionen der Studierenden in die Gruppenarbeitsprozesse der Schülerinnen und Schüler werden zunächst die getätigten Hilfen innerhalb der Interventionen, also die dominanten Hilfen sowie die untergeordneten Hilfen präsentiert. Dabei werden auch die zeitliche Dauer der Intervention, die zeitliche Dauer der einzelnen Hilfen innerhalb der Intervention sowie auftretende Kombinationen von Hilfe-Kategorien vorgestellt. Anschließend wird auf die Reihenfolge der gegebenen Hilfen eingegangen. Eine weitere Facette der Interventionen ist die Angemessenheit. Diese wird getrennt nach den verschiedenen Hilfe-Kategorien dargestellt. Darüber hinaus wird auch auf die Kriterien der Angemessenheit eingegangen. Ein zusätzliches Kriterium zur Überprüfung der Angemessenheit ist die Anzahl der Schülerinnen und Schüler, die vor bzw. nach der Intervention aktiv arbeiten. Dies wird zunächst unabhängig und dann in Verbindung mit der Bewertung der Angemessenheit, wie sie durch die Expertengruppen vorgenommen wurde, untersucht. Weiterhin werden die Interventionsauslöser analysiert und schließlich in Verbindung mit der Bewertung der Angemessenheit und der Anzahl der aktiven Schülerinnen und Schüler gesetzt (Kap. 9.2).

Im Bereich Reflexion wird zunächst ein Überblick über das Auftreten von Reflexionsmomenten in den Hauptkategorien „Rückblick“, „Wesentliche Aspekte“ und „Alternativen“ vermittelt. Anschließend wird auf Unterkategorien eingegangen und zwischen Äußerungen, die sich auf die Erprobung der Lernumgebung und Äußerungen, die sich auf die getätigten Interventionen beziehen, unterschieden. Am Ende des Teilkapitels wird eine Übersicht über alle Reflexionsmomente, aufgeteilt nach den einzelnen Gruppen, präsentiert (Kap. 9.3).

Im Anschluss an die separate Vorstellung der einzelnen Bereiche sollen mögliche Zusammenhänge zwischen ihnen unter die Lupe genommen werden. Dazu werden die Studierenden aufgrund ihrer Bearbeitung der Diagnoseaufträge zu den Videovignetten in drei Gruppen eingeteilt. Teilkapitel 9.4 zeigt, inwiefern zwischen den Gruppen Unterschiede bei den getätigten Interventionen zu erkennen sind, und deckt dadurch mögliche Zusammenhänge zwischen der diagnostischen Fähigkeit „Deuten“ und den Interventionen der Studierenden auf. Das gleiche Vorgehen wird auch in Teilkapitel 9.5 angewandt, in dem mögliche Zusammenhänge zwischen der diagnostischen Fähigkeit „Deuten“ und dem Reflexionsverhalten der Studierenden untersucht werden. Teilkapitel 9.6 untersucht, ob Zusammenhänge zwischen den getätigten Interventionen und dem Reflexionsverhalten der Studierenden bestehen.

9.1 Diagnostische Fähigkeiten

Zur Erhebung der diagnostischen Fähigkeiten wurden drei Videovignetten erstellt, in ViviAn eingebunden und passende Diagnoseaufträge formuliert. Die Freitextantworten der Studierenden wurden kodiert und letztendlich auf Grundlage der qualitativen Analyse dichotom bewertet (vgl. dazu Kapitel 8.2).

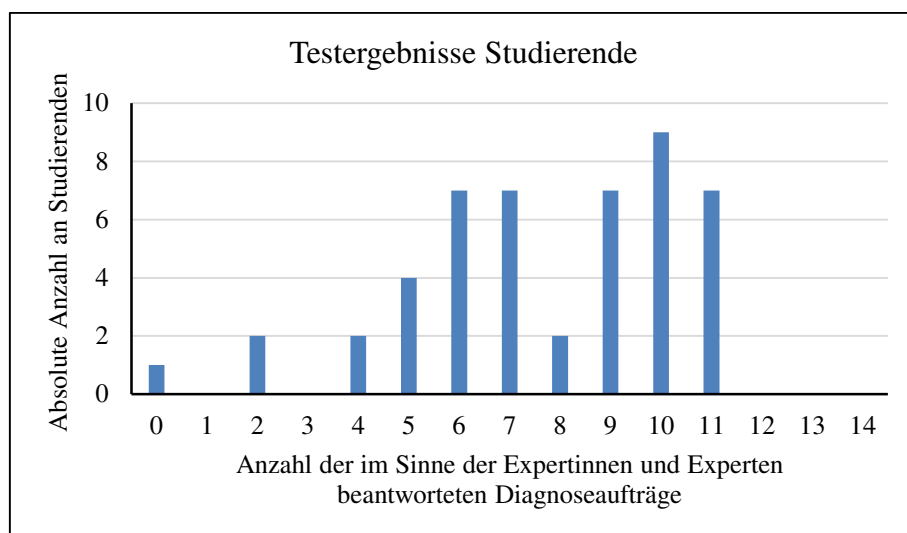


Abbildung 28: Übersicht über die Testergebnisse der Studierenden hinsichtlich der prozessdiagnostischen Fähigkeit „Deuten“

Abbildung 28 gibt einen Überblick über die Verteilung der Anzahl der im Sinne der Expertinnen und Experten beantworteten Diagnoseaufträge. Auf der x-Achse ist die Anzahl der Diagnoseaufträge und auf der y-Achse die Anzahl der Studierenden abgebildet, die gleich viele Diagnoseaufträge beantworten konnten. Man erkennt, dass keiner der Studierenden in der Lage war, zwölf oder mehr Diagnoseaufträge im Sinne der Expertinnen und Experten zu beantworten. Sieben Studierende waren in der Lage, elf Diagnoseaufträge adäquat zu beantworten. Da der Grad der Übereinstimmung (zwischen den Studierendenantworten und den Expertenantworten) als Maß für die diagnostische Fähigkeit der Studierenden angesehen wird, handelt es sich dabei um die Studierenden, die über die verhältnismäßig am besten ausgeprägte diagnostische Fähigkeit verfügen. Natürlich muss dies vorsichtig interpretiert werden, da die Fallzahl für eine quantitative Analyse recht gering ist. Jedoch kann aufgrund der qualitativen Untersuchung in Form der Kombination aus fachdidaktisch formulierten Diagnoseaufträgen (vgl. dazu Kap. 8.2.3), Expertenrating (vgl. dazu Kap. 8.2.4) und Validierung des Testinstrumentes (vgl. dazu Kap. 8.2.7) davon ausgegangen werden, dass das Messinstrument die Fähigkeit, die gemessen werden soll, tatsächlich erfasst. Entsprechend verfügen die neun Studierenden, die zehn Diagnoseaufträge beantworten konnten, verhältnismäßig ebenfalls über hohe diagnostische Fähigkeiten verglichen mit den anderen Studierenden der Experimentalgruppe. Auf der anderen Seite sieht man, dass ein Student nicht in der Lage war, auch überhaupt einen Diagnoseauftrag adäquat zu beantworten. Zwei Studierende konnten jeweils zwei bzw. vier Diagnoseaufträge im Sinne der Expertinnen und Experten beantworten. Mit fünf Diagnoseaufträgen konnten vier Studierende immer noch verhältnismäßig wenig adäquate Antworten geben. Jeweils sieben Studierende beantworteten sechs, sieben bzw. neun Diagnoseaufträge adäquat und bei zwei Studierenden war dies bei acht Diagnoseaufträgen der Fall.

Mit Blick auf das Gesamtergebnis aller Studierenden der Experimentalgruppe wird ersichtlich, dass die Masterstudierenden noch Optimierungsmöglichkeiten im Bereich diagnostischer Fähigkeit „Deuten“ besitzen. Um zu sehen, welche Diagnoseaufträge den Studierenden tendenziell Schwierigkeiten bereiten, bietet sich ein Blick auf die Anzahl korrekter Antworten, sowohl absolut als auch relativ, an (vgl. Tab. 7). Die Diagnoseaufträge, die die Studierenden am seltensten lösen konnten (Diagnoseauftrag 1 und Diagnoseauftrag 11), bezogen sich inhaltlich

beide auf die Grundvorstellung Term als Rechenschema. Die erste Frage zu der ersten Videovignette, die „Mit welcher Grundvorstellung argumentiert Schülerin 1?“ lautet, konnten lediglich drei der 48 Studierenden im Sinne der Expertinnen und Experten beantworten. In Vignette zwei lautet einer der Diagnoseaufträge: „Folgende Schülerinnen verwenden die Grundvorstellung Term als Rechenschema“. Diesen Diagnoseauftrag beantworteten lediglich neun der 48 Studierenden im Sinne der Expertinnen und Experten. Das Diagnostizieren der Grundvorstellung „Term als Rechenschema“ scheint den Studierenden also noch sehr schwer zu fallen.

Tabelle 7: Absolute und relative Häufigkeit korrekt beantworteter Diagnoseaufträge der Studierenden

Item	Anzahl korrekter Antworten	Relative Anzahl richtiger Lösungen (in %)
F1_TaR	3	0,06
F3_TaB	35	0,73
F4_kA	10	0,21
F5_S3	31	0,65
F6_Ges	19	0,40
F7_Sim	28	0,58
F9_ja	41	0,85
F10_S3	29	0,60
F11_niemand	9	0,19
F12_nein	32	0,67
F13_nein	32	0,67
F14_nein	37	0,77
F15_ja	41	0,85
F16_ja	22	0,46

Betrachtet man die Freitextantworten der Studierenden auf Diagnoseauftrag 1, der sich auf Videovignette 1 bezieht und den nur drei der 48 Studierenden im Sinne der Expertinnen und Experten beantworten konnten, genauer, so finden sich darunter häufig Begründungen für die Grundvorstellung Term als Bauplan:

- „Sie argumentiert damit, dass die Simulation ein Rechteck ergibt. Zur Berechnung braucht man die Hälfte des Flächeninhalts.“
- „Sie erkennt, dass die Ausgangsfigur ein Rechteck ist, welches in zwei Dreiecke unterteilt werden kann.“
- „Flächeninhalt eines Rechtecks und dividiert durch Anzahl der Dreiecke“
- „Zusammenbauen zweier Dreiecke zu einem Rechteck und dann wieder zerlegen“.

Die Studierenden beziehen sich hierbei oftmals auf die Simulation in der Videovignette und liefern eine inhaltliche Begründung für den Lösungsweg bzw. das richtige Ergebnis. Dabei scheinen sie teilweise die einzelnen Schülerinnen aus dem Blick zu verlieren und deren Äußerungen entweder nicht richtig wahrzunehmen oder nicht richtig deuten zu können. Dies zeigt sich auch in Antworten, die sehr kurz und plakativ als Begründung für die Antwort „Term als Bauplan“ gegeben werden, wie etwa:

- „Schülerin 1 redet von Rechtecken“
- „Da sie vom Rechteck ausgeht“
- „Erwähnt oft geometrische Figuren zum Erklären“
- „S1 begründet die Antwort mit dem Rechteck“

Die kurzen Antworten, die einen Verweis auf die geometrischen Objekte in der Simulation aufweisen, könnten ein Zeichen dafür sein, dass in Bezug auf Schülerin 1 wenig wahrgenommen wurde. Betrachtet man das Video und die einzelnen Äußerungen der Schülerinnen genauer, so erkennt man, dass die Herleitung der Formel die Leistung von Schülerin 3 ist. Schülerin 1 nutzt diese Formel dann lediglich, um konkrete Werte einzusetzen und anhand dieser Werte zu überprüfen, ob die Formel korrekt ist. Dass Schülerin 1 ihren Fokus ganz klar auf das Rechnen legt, wird in ihren Äußerungen ersichtlich. Sie äußert, dass die Aufgabenstellung sei, wie „sich das direkt berechnen lässt“. Nachdem Schülerin 3 die Formel geometrisch hergeleitet hat und Schülerin 1 daraufhin konkrete Werte eingesetzt hat, ruft sie: „Es stimmt Leute, weil es dann 15 sind und 15 ist auch D5“. D5 steht hierbei für Dreieckszahl 5, also die Summe der Kugeln, aus der ein gleichschenkeliges Dreieck mit der Grundfläche 5 besteht. Da die Herleitung und Verifizierung der Formel für eine Schülerin aus der Gruppe anscheinend zu schnell ging, versuchen die anderen Gruppenmitglieder es ihr zu erklären. Dabei äußert Schülerin 1 „Dreißig geteilt durch zwei ist fünfzehn. Einfach ausrechnen. Es stimmt, jetzt können wir weitermachen“. Hinter diesen Äußerungen ist jeweils die Absicht zu erkennen, den Term zu nutzen, um etwas auszurechnen. Entsprechend ist dabei jeweils die Grundvorstellung Term als Rechenschema die handlungsleitende Grundvorstellung, mit der Schülerin 1 agiert und schließlich auch argumentiert. Die meisten der untersuchten Studierenden waren also nicht in der Lage, die Äußerungen von Schülerin 1 richtig zu deuten.

Auch bei Diagnoseauftrag 11, der sich auf die zweite Videovignette bezieht, sollen die Freitextantworten der Studierenden genau betrachtet werden, da es sich dabei um den Diagnoseauftrag handelt, der am zweitseltensten im Sinne der Expertinnen und Experten beantwortet wurde. Dabei geht es wie bei dem eben diskutierten Diagnoseauftrag auch um die Grundvorstellung Term als Rechenschema, nur wurde hierbei nicht nach einer konkreten Schülerin gefragt, sondern es wurde gefragt, welche der Schülerinnen diese Grundvorstellung in der Videosequenz zur Argumentation nutzt. Die Begründungen sind vielfältig und lassen jeweils Interpretationsspielraum. Einige Begründungen lauteten beispielsweise:

- „S3 verwendet die Grundvorstellung Term als Rechenschema, da sie erkennt, dass man den Term $(n+1)$ subtrahieren muss, um die Lösung zu erhalten“
- „S2 und S3 argumentieren auf der formalen Ebene und geben eine Formel an“
- „S3 arbeitet sofort mit verschiedenen Termen und kann auch erklären warum sie diese aufgestellt hat“

Für diese Antworten scheint die Verwechslung der beiden Grundvorstellungen die plausibelste Erklärung zu sein. Die Schülerinnen haben in der Videosequenz nämlich die Aufgabe, zu graphischen Darstellungen, bei denen jeweils unterschiedliche Kreise bei den Dreieckszahlen farbig markiert sind, mit Hilfe von Simulationen die jeweiligen Formeln zu ihrer Berechnung zu finden. Sie stellen verschiedene Formeln auf und überprüfen, ob sie zu den jeweiligen graphischen Darstellungen aus dem Arbeitsauftrag passen. Zu keinem Zeitpunkt werden vorhandene Formeln genutzt, um Zahlen einzusetzen und etwas zu berechnen.

Bei der Antwort „S3 und S4 argumentieren mit Rechenschema“ erscheint es aufgrund der mangelnden Ausführung der Begründung naheliegend, dass derjenige die relevanten Informationen zur Beantwortung der Frage entweder nicht wahrgenommen hat oder nicht adäquat deuten konnte. Es findet sich keinerlei Hinweis darauf, inwiefern die eigene Aussage durch das Verhalten der Schülerinnen gestützt werden könnte. Dies wäre beispielsweise möglich, indem eine Beschreibung der für die vorgenommene Deutung relevanten Aspekte vorgenommen würde. Bei derart kurzen Antworten kann natürlich auch die mangelnde Motivation des Studierenden, sich mit der Aufgabe auseinanderzusetzen oder diese nachvollziehbar zu begründen, eine Rolle spielen.

Weitere Antworten lauteten „S1 kann mit dem Pfeil beispielsweise nichts anfangen“ und „S1 versteht nicht die Bedeutung des Pfeiles als Hinweis zum späteren Entfernen. Eine Vorstellung als Bauplan hätte ihr dabei geholfen“. Diese beiden Antworten zielen jeweils auf die graphische Darstellung im Arbeitsauftrag ab, den die Schülerinnen in der Situation bearbeiten. Es finden sich jedoch auch hier keinerlei Stützen, die die eigentliche Antwort belegen. Entsprechend muss auch an dieser Stelle geschlussfolgert werden, dass die Studierenden nicht in der Lage waren, die Situation wahrzunehmen bzw. zu deuten. Ein Indiz dafür könnte auch sein, dass einer der beiden Studierenden trotz der kurzen Antwort darauf verweist, dass die entsprechende Grundvorstellung hilfreich gewesen wäre, obwohl dies überhaupt nicht gefragt wurde. Vielleicht handelt es sich dabei an der Stelle um den Versuch, die Frage nicht auszulassen, sondern zumindest etwas Sinnvolles zu antworten.

Manche Studierenden gaben sehr differenzierte Antworten und gingen auf alle Schülerinnen einzeln ein. Exemplarisch dafür ist folgendes Beispiel: „S1 und S4 erstellen eine Formel. S2 findet auch eine Formel aber argumentiert nicht. S3 entwickelt richtige Formeln für jede Teilaufgabe. S4 arbeitet mit der Simulation und erhält eine Formel“. Da in der Situation theoretisch alle vier Schülerinnen mit der Grundvorstellung argumentieren können, erscheint das Vorgehen des Studierenden sinnvoll, auch alle Studierenden einzeln abzuhandeln und die gesetzten Kreuzchen damit auch zu begründen. In diesem Fall hängen seine Ausführungen vermutlich damit zusammen, dass derjenige auch alle Schülerinnen angekreuzt hat, sprich der Meinung ist, alle vier Schülerinnen würden die Grundvorstellung Term als Rechenschema in der Situation für ihre Argumentation nutzen. Zunächst schreibt er, dass zwei Schülerinnen eine Formel aufstellen. An dieser Stelle scheint er die Grundvorstellungen zu verwechseln, da dies eher eine Begründung für die Nutzung der Grundvorstellung Term als Bauplan wäre. Bei Schülerin 2 meint er zu erkennen, dass sie die Grundvorstellung nutzt, schränkt jedoch selbst ein, dass sie nicht damit argumentiert. Auch für Schülerin 3 nimmt er an, dass sie richtige Formeln erstellt hätte und deswegen die Grundvorstellung Term als Rechenschema nutzt. Auch hier scheinen die Grundvorstellungen verwechselt worden zu sein. Im letzten Satz wird noch darauf verwiesen, dass Schülerin 4 mit einer Simulation arbeitet. Dies ergibt jedoch keinen inhaltlichen Mehrwert und zeugt daher an dieser Stelle eher davon, dass derjenige keine geeigneten Stützen für seine vorgenommene Deutung darbieten kann. Insgesamt zeigt sich auch an den Antworten der Studierenden, dass die Aufgabe den Studierenden recht schwerfiel.

Um die Ergebnisse der Studierenden besser einordnen zu können, werden sie nun mit denen der 15 Lehrkräfte verglichen. Dazu können zunächst auf Ebene der einzelnen Diagnoseaufträge die absoluten und relativen Lösungshäufigkeiten verglichen werden. In Tabelle 8 sind die Ergebnisse der Lehrkräfte zu sehen. Es zeigt sich, dass der schwierigste Diagnoseauftrag für die Lehrkräfte ebenfalls einer der beiden Diagnoseaufträge war, bei dem es um die Grundvorstellung „Term als Rechenschema“ ging. Dies stärkt die Vermutung, dass es sich bei diesem Diag-

noseauftrag um ein schwieriges Item handelt. Der Diagnoseauftrag, der den Studierenden jedoch am meisten Probleme bereitete (Diagnoseauftrag 1), wurde von den Lehrkräften zum Großteil korrekt beantwortet. Darüber hinaus ist ersichtlich, dass es zwei Diagnoseaufträge gab, die alle Lehrkräfte adäquat beantworten konnten und darüber hinaus viele Diagnoseaufträge, die ein Großteil der Lehrkräfte beantworten konnte. Die Lehrkräfte scheinen also besser mit den Situationen in den Videovignetten und den Diagnoseaufträgen zurechtzukommen als die Studierenden.

Tabelle 8: Absolute und relative Häufigkeit korrekt beantworteter Diagnoseaufträge der Lehrkräfte

Item	Anzahl korrekter Antworten	Relative Anzahl richtiger Lösungen (in %)
F1_TaR	10	0,67
F3_TaB	15	1,00
F4_kA	5	0,33
F5_S3	13	0,87
F6_Ges	10	0,67
F7_Sim	8	0,53
F9_ja	15	1,00
F10_S3	10	0,67
F11_niemand	3	0,20
F12_nein	13	0,87
F13_nein	14	0,93
F14_nein	12	0,80
F15_ja	12	0,80
F16_ja	8	0,53

Die beiden Aufträge, zu denen die Freitextantworten der Studierenden näher analysiert wurden, sollen nun auch für die Lehrkräfte genauer unter die Lupe genommen werden. Bei Diagnoseauftrag 1, der den Studierenden große Probleme bereitete, aber von zehn der 15 Lehrkräfte im Sinne der Expertinnen und Experten beantwortet werden konnte, lauteten Antworten der Lehrkräfte folgendermaßen:

- „S1 kann mit der von S3 aufgestellten Formel weitere Beispiele von Dreieckszahlen berechnen.“
- „Da sie die aufgestellte Formel nutzt, um andere Bsp. zu berechnen.“
- „S1 vergleicht die berechnete Lösung mit dem Rechteck der Simulation, d.h. kein ´Zusammenbauen´ des Terms, sondern nur Vergleich der Lösung.“
- „Die Schülerin versucht mithilfe eines Beispiels das Ergebnis zu überprüfen und berechnet Ergebnisse zu verschiedenen Beispielen der Simulation und vergleicht diese.“

Bei den Freitextantworten der Lehrerinnen und Lehrer wird recht schnell deutlich, dass sie die relevanten Situationen wahrnehmen konnten und die Handlungen bzw. Äußerungen von Schülerin 1 auch hinsichtlich der Grundvorstellung „Term als Rechenschema“ deuten konnten. Die Begründungen legen jeweils nahe, dass Schülerin 1 stark auf das Berechnen fokussiert ist. Die von Schülerin 3 aufgestellte Formel wird also durch Einsetzen konkreter Zahlen auf Korrektheit

überprüft. Dazu wird das Ergebnis auch anhand der Simulation überprüft. Die Lehrkräfte lassen sich durch die Visualisierung der Rechtecke, anhand derer man theoretisch auch Baupläne entwerfen könnte, jedoch nicht auf eine falsche Fährte locken. Sie erkennen eindeutig, dass es nicht um das Aufstellen eines Terms geht, also nicht mit der Grundvorstellung „Term als Bauplan“ argumentiert wird. Eine Lehrkraft stellt dies in der Freitextantwort explizit heraus: „Die Schülerin berechnet verschiedenste konkrete Varianten aus und vergleicht die Ergebnisse. Dabei geht sie nicht auf den Aufbau des Terms ein oder ähnliches, wobei man auf eine Grundvorstellung Term als Bauplan schließen kann“. Die Antwort begründet zunächst, weshalb es sich um die Grundvorstellung „Term als Rechenschema“ handelt und anschließend, weshalb die Schülerin in der Situation nicht mit der Grundvorstellung „Term als Bauplan“ agiert bzw. argumentiert.

Bei Diagnoseauftrag 11 gaben lediglich drei Lehrkräfte eine Antwort im Sinne der Expertinnen und Experten:

- „Meiner Meinung nach keiner der Schüler, da keiner der Schüler Rechenoperationen anstellt.“
- „Vielleicht verfügen einige der Schülerinnen über diese Grundvorstellung, jedoch wird sie von keiner Schülerin angewendet. Es werden keine Ergebnisse berechnet.“
- „Für mich war nur bei Schülerin 3 die Grundvorstellung ‚Term als Bauplan‘ erkennbar, ansonsten hat keine der Schülerinnen mithilfe einer Rechnung argumentiert.“

In den Antworten wird ersichtlich, dass die Lehrkräfte erkannt haben, dass keine der Schülerinnen eine Rechenoperation durchführt bzw. mithilfe einer Rechnung argumentiert. Bei der letzten Antwort wird darüber hinaus darauf verwiesen, dass Schülerin 3 mit der Grundvorstellung „Term als Bauplan“ argumentiert (was korrekt ist). Neben diesen drei Lehrkräften gab es zwei weitere, die zwar die richtige Antwort ankreuzen, jedoch keine Begründung dafür liefern. Da nur die Freitextantworten ausgewertet wurden, werden die unausgefüllten Freitextfelder entsprechend als unzureichende Antworten gewertet. Weitere unzureichende Antworten waren beispielsweise folgende:

- „Schülerin 3 benutzt die zu Aufgabenteil a) erarbeitete Formel und formt sie um.“
- „Schülerin 3 nutzt die Terme sehr sicher!“

Bei diesen beiden Antworten wurde jeweils begründet, weshalb Schülerin 3 die Grundvorstellung „Term als Rechenschema“ nutzt. Das Ziel der Umformungen ist jedoch eindeutig die Veränderung des Bauplans. Entsprechend wurde zwar erkannt, dass Schülerin 3 Terme umformen kann, jedoch wurde dies der falschen Grundvorstellung zugeschrieben und die Situation entsprechend nicht im Sinne der Expertinnen und Experten gedeutet. Vereinzelt gab es auch Antworten wie „Bin mir aber nicht mehr sicher“, bei denen anscheinend geraten wurde. Derartige Begründungen wurden ebenfalls als unzureichend angesehen.

Nimmt man nicht einzelne Fragen, sondern das Gesamtergebnis in den Blick, so kann die Einschätzung, dass Lehrkräfte die Situationen in den Videovignetten besser diagnostizieren können als die Studierenden, durch die Übersicht der Summe der von den Lehrkräften im Sinne der Expertinnen und Experten beantworteten Diagnoseaufträge zu den Videovignetten belegt werden (vgl. Abb. 29). Es lässt sich erkennen, dass zwei Drittel der Lehrkräfte zehn oder mehr

Diagnoseaufträge adäquat beantworten konnten. Auch die von den Studierenden nicht erreichten Punktzahlen von 12 und 13 wurden von den Lehrkräften erreicht. Jeweils eine Lehrkraft konnte lediglich vier bzw. fünf Diagnoseaufträge beantworten. Dies zeigt, dass der Erwerb der diagnostischen Fähigkeiten auch bei Lehrkräften im Schuldienst noch optimiert werden kann. Eine mittlere Gruppe, bestehend aus insgesamt drei Lehrkräften, konnte jeweils genau sieben, acht bzw. neun Diagnoseaufträge richtig beantworten.

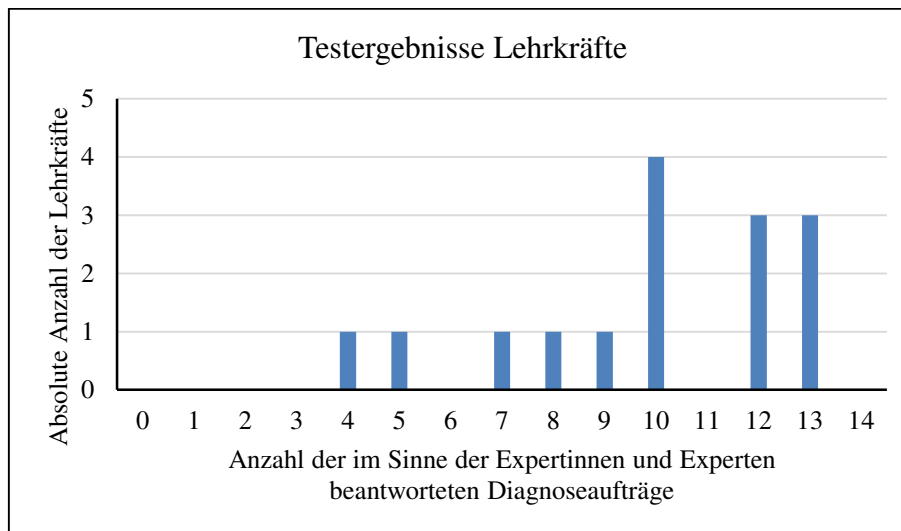


Abbildung 29: Übersicht über die Testergebnisse der Lehrkräfte hinsichtlich der prozessdiagnostischen Fähigkeit „Deuten“

Für weitere Untersuchungen, bei denen die prozessdiagnostische Fähigkeit „Deuten“ jeweils zugrunde gelegt wird, um die Studierenden anhand dieser Fähigkeit zu unterscheiden, müssen diese in Gruppen unterteilt werden. Hierfür wird die Experimentalgruppe in drei gleich große Gruppen unterteilt, sodass jede Subgruppe genau aus 16 Studierenden besteht. Diese Einteilung ergibt folgende drei Gruppen:

- 16 Studierende mit verhältnismäßig hoher diagnostischer Fähigkeit (10 oder mehr adäquat beantwortete Diagnoseaufträge)
- 16 Studierende mit im Vergleich mittlerer diagnostischer Fähigkeit (7 bis 9 adäquat beantwortete Diagnoseaufträge)
- 16 Studierende mit im Vergleich geringer diagnostischer Fähigkeit (6 oder weniger adäquat beantwortete Diagnoseaufträge)

Die Einteilung wird für die Untersuchung möglicher Zusammenhänge zwischen diagnostischer Fähigkeit und Interventionsverhalten (vgl. dazu Kap. 9.4) sowie zwischen diagnostischer Fähigkeit und den Reflexionsgesprächen der Studierenden (vgl. dazu Kap. 9.5) zugrunde gelegt.

9.2 Interventionen in Gruppenarbeitsprozesse

Im Bereich der Interventionen in die Gruppenarbeitsprozesse der Schülerinnen und Schüler wurden mehrere Analysen vorgenommen. Zunächst werden die Ergebnisse hinsichtlich der dominanten und untergeordneten Hilfen innerhalb einer Intervention vorgestellt. Dabei wird auch auf die unterschiedliche Länge und die auftretenden Kombinationen der verschiedenen Hilfen

eingegangen. Anschließend werden die Ergebnisse hinsichtlich der Angemessenheit der Interventionen sowie der Interventionsauslöser präsentiert. Außerdem wird die Reihenfolge der Hilfen genauer in den Blick genommen.

9.2.1 Hilfen innerhalb der Interventionen

Die 208 Interventionsvideos wurden jeweils auf die darin enthaltenen Hilfen hin untersucht, welche dominant oder untergeordnet sein können (vgl. dazu Kap. 8.3). Abbildung 30 zeigt die Verteilung der dominanten und untergeordneten Hilfen innerhalb der durch die Studierenden getätigten Interventionen. Auf der x-Achse sind die absoluten Häufigkeiten aufgetragen und auf der y-Achse die verschiedenen Hilfe-Kategorien.

Betrachtet man die dominanten Hilfen, die in Dunkelblau abgebildet sind, so lässt sich erkennen, dass am häufigsten inhaltsorientiert-strategische Hilfen als dominante Hilfe verwendet wurden. Am zweithäufigsten wurden allgemein-strategische Hilfen als dominante Hilfe genutzt, gefolgt von den inhaltlichen Hilfen. Selten wurden Rückmeldehilfen und disziplinarische Hilfen als dominante Hilfe genutzt. Motivationshilfen kamen als dominante Hilfe fast gar nicht vor.

Mit Blick auf die untergeordneten Hilfen, die in Grau abgebildet sind, ist festzustellen, dass Rückmeldehilfen am häufigsten als untergeordnete Hilfe gegeben wurden. Ebenfalls häufig kamen inhaltsorientiert-strategische und allgemein-strategische Hilfen als untergeordnete Hilfe vor. Motivationshilfen und disziplinarische Hilfen wurden selten als untergeordnete Hilfe erteilt. Inhaltliche Hilfen kamen fast nicht als untergeordnete Hilfe vor.

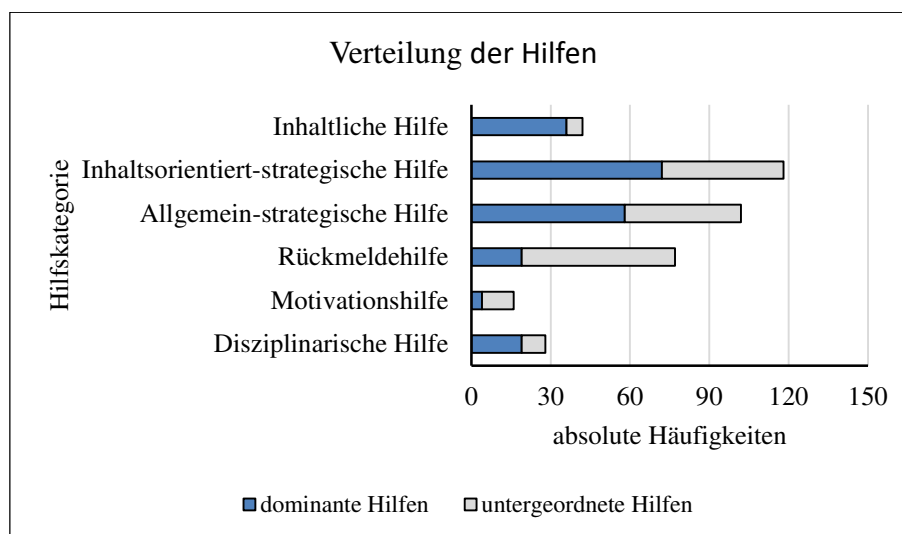


Abbildung 30: Verteilung der dominanten und untergeordneten Hilfen innerhalb der getätigten Interventionen der Studierenden

Betrachtet man die Summe der dominanten und untergeordneten Hilfen, nimmt also alle Hilfen innerhalb einer Hilfskategorie in den Blick, so zeigt sich, dass inhaltsorientiert-strategische und allgemein-strategische Hilfen am häufigsten verwendet wurden, gefolgt von Rückmeldehilfen. Etwas seltener wurden inhaltliche Hilfen genutzt, um die Schülerinnen und Schüler zu unterstützen. Insgesamt kamen nur wenige disziplinarische Hilfen vor und Motivationshilfen wurden

am seltensten gegeben. Darüber hinaus lässt sich festhalten, dass Rückmeldehilfen und Motivationshilfen überwiegend als untergeordnete Hilfen und disziplinarische Hilfen überwiegend als dominante Hilfe erteilt wurden. Inhaltliche Hilfen waren fast immer dominant.

Neben der unterschiedlichen Verteilung der verschiedenen Hilfen kann auch die unterschiedliche Dauer der jeweiligen Hilfe untersucht werden. In Tabelle 9 ist die durchschnittliche Zeitdauer der verschiedenen Hilfskategorien angegeben. Darüber hinaus wurden auch die jeweils kürzeste und die jeweils längste Interventionsdauer angegeben.

Tabelle 9: Zeitdauer der verschiedenen Hilfen mit Angabe der jeweilig längsten und der kürzesten Hilfe

Hilfe-Kategorie	Durchschnittliche Dauer	Kürzeste Dauer	Längste Dauer
Inhaltliche Hilfe	87 Sekunden	5 Sekunden	358 Sekunden
Inhaltsorientiert-strategische Hilfe	55 Sekunden	5 Sekunden	290 Sekunden
Allgemein-strategische Hilfe	35 Sekunden	3 Sekunden	112 Sekunden
Rückmeldehilfe	30 Sekunden	2 Sekunden	295 Sekunden
Motivationshilfe	13 Sekunden	3 Sekunden	34 Sekunden
Disziplinarische Hilfe	19 Sekunden	3 Sekunden	60 Sekunden

Die durchschnittliche Dauer der Hilfen unterscheidet sich deutlich je nach Hilfe-Kategorie. Während Motivationshilfen im Durchschnitt nur 13 Sekunden lang sind, dauern inhaltliche Hilfen durchschnittlich fast eineinhalb Minuten. Insgesamt ist ein Trend dahingehend zu erkennen, dass je inhaltlicher die Hilfen werden, auch die dafür benötigte Zeit zunimmt. Neben den Motivationshilfen sind nämlich auch disziplinarische Hilfen mit 19 Sekunden durchschnittlich recht kurz. Diese beiden Hilfe-Kategorien haben keinen inhaltlichen Bezug. Bei den Rückmeldehilfen, die im Schnitt eine halbe Minute dauern, wurde den Schülerinnen und Schülern eine Rückmeldung darüber gegeben, ob sie sich auf dem richtigen Lösungsweg befinden oder ob eine (Teil-)Aufgabe korrekt gelöst wurde. Auch wenn diese Hilfe-Kategorie sich nicht direkt mit dem Inhalt der Aufgabe beschäftigt, so ist doch ein erster inhaltlicher Bezug deutlich. Allgemein-strategische Hilfen dauern durchschnittlich 35 Sekunden. Inhaltsorientiert-strategischen Hilfen hingegen, die mehr inhaltlichen Bezug haben als die allgemein strategischen Hilfen, dauern 55 Sekunden. Am längsten dauern die inhaltlichen Hilfen, die per Definition am inhaltlichsten sind.

Ein ähnliches Bild ergibt sich in Bezug auf die längsten getätigten Hilfen in der jeweiligen Hilfe-Kategorie. Die längste Motivationshilfe (34 Sekunden) und die längste disziplinarische Hilfe (60 Sekunden) dauern im Vergleich zu den anderen Hilfe-Kategorien am kürzesten. Außerdem ist die Zeitdauer der längsten allgemein-strategischen Hilfe mit 112 Sekunden nicht einmal halb so lang wie die längste inhaltsorientiert-strategische Hilfe, die knapp fünf Minuten dauert. Die längste inhaltliche Hilfe ist mit beinahe sechs Minuten auch die längste Hilfe, die gegeben wird. Einen Ausreißer in der Progression stellt die längste Rückmeldehilfe dar, die mit knapp fünf Minuten ähnlich lange dauert wie die längste inhaltsorientiert-strategische Hilfe. Bezüglich der kürzesten Hilfen innerhalb der Hilfe-Kategorien lassen sich kaum Unterschiede feststellen. Während disziplinarische Hilfen, Motivationshilfen, Rückmeldehilfen und allge-

mein-strategische Hilfen mit zwei bzw. drei Sekunden in einem Satz abgehandelt werden können, dauern die kürzesten inhaltsorientiert-strategischen und inhaltlichen Hilfen immerhin 5 Sekunden. Somit ist auch hier eine Tendenz zu erkennen, die jedoch nicht überinterpretiert werden sollte.

Da es auch einen Unterschied machen kann, ob eine Hilfe innerhalb einer Intervention dominant oder untergeordnet ist, listet Tabelle 10 die durchschnittlichen Zeitwerte für dominante und untergeordnete Hilfen getrennt voneinander auf.

Tabelle 10: Vergleich der Zeitdauer zwischen dominanten und untergeordneten Hilfen hinsichtlich aller Hilfskategorien

Hilfe-Kategorie	Durchschnittliche Dauer (dominant)	Durchschnittliche Dauer (untergeordnet)
Inhaltliche Hilfe	96 Sekunden	31 Sekunden
Inhaltsorientiert-strategische Hilfe	58 Sekunden	51 Sekunden
Allgemein-strategische Hilfe	44 Sekunden	23 Sekunden
Rückmeldehilfe	35 Sekunden	29 Sekunden
Motivationshilfe	16 Sekunden	12 Sekunden
Disziplinarische Hilfe	19 Sekunden	19 Sekunden

Betrachtet man zunächst die durchschnittliche Dauer der dominanten Hilfen, so lässt sich erkennen, dass die oben erwähnte Progression auch hier wieder sichtbar wird. Disziplinarische Hilfen und Motivationshilfen dauern mit 19 bzw. 16 Sekunden durchschnittlich am kürzesten. Dominante Rückmeldehilfen sind im Schnitt 35 Sekunden lang und dominante allgemein-strategische Hilfen dauern durchschnittlich 44 Sekunden. Inhaltsorientiert-strategische Hilfen hingegen dauerten im arithmetischen Mittel beinahe eine Minute und inhaltliche Hilfe nahmen im Durchschnitt mit 96 Sekunden deutlich am meisten Zeit in Anspruch.

In Bezug auf die untergeordneten Hilfen zeigt sich ein ähnliches Bild. Motivationshilfen dauern mit durchschnittlich zwölf Sekunden am kürzesten, gefolgt von disziplinarischen Hilfen mit 19 Sekunden. Es folgen allgemein-strategische Hilfen (23 Sekunden) und Rückmeldehilfen (29 Sekunden), die nun anders als bei den dominanten Hilfen in umgekehrter Reihenfolge auftreten. Ebenso mag es zunächst verwundern, dass die inhaltliche Hilfe, wenn sie als untergeordnete Hilfe erteilt wird, durchschnittlich mit 31 Sekunden kürzer dauert als die durchschnittliche untergeordnete inhaltsorientiert-strategische Hilfe, die mit durchschnittlich 51 Sekunden am längsten dauert. Eine mögliche Erklärung für diesen Befund ist, dass länger andauernde inhaltliche Hilfen vermutlich im Regelfall den Charakter der Intervention derart stark prägen, dass die inhaltliche Hilfe die dominante Hilfe innerhalb der Intervention darstellt. Entsprechend ist es nachvollziehbar, dass inhaltliche Hilfen, sofern sie als untergeordnet kodiert wurden, verhältnismäßig kürzer dauern.

Diese mögliche Begründung wird auch durch den Vergleich der durchschnittlichen Zeitdauer von dominanten und untergeordneten Hilfen gestärkt. Die durchschnittliche Zeitdauer dominanter Hilfen liegt in jeder Kategorie mindestens bei der durchschnittlichen Zeitdauer der entsprechenden untergeordneten Hilfe-Kategorie. Bei der inhaltlichen Hilfe ist diese sogar

mehr als dreimal so lange. Eine größere Kluft gibt es bei keiner anderen Hilfe-Kategorie. Während allgemein-strategische Hilfen immerhin noch fast doppelt so lange dauern, wenn sie die dominante Hilfe innerhalb der Intervention darstellen, ist die Dauer der Hilfen bei den anderen Hilfe-Kategorien nur geringfügig unterschiedlich. Bei der disziplinarischen Hilfe ist die durchschnittliche Zeit der dominanten und der untergeordneten Hilfen sogar gleich lang. Insgesamt stärkt dies obige Interpretation, dass inhaltliche Hilfen einen besonderen Status innehaben, der dazu führt, dass die zeitliche Kluft zwischen dominanten und untergeordneten Hilfen verhältnismäßig groß ist. Bei den meisten anderen Hilfen ist dies nicht zu erkennen.

Neben der Dauer der Interventionen wurden auch die verschiedenen Kombinationen von Hilfen kodiert. Dabei kam es auch vor, dass innerhalb einer Intervention drei oder sogar vier verschiedene Hilfen erteilt wurden. Die am häufigsten auftretende Kombination aus drei Hilfen kommt sechsmal vor und besteht aus einer Rückmeldehilfe zusammen mit einer allgemein-strategischen und einer inhaltsorientiert-strategischen Hilfe. Die häufigste Kombination aus vier Hilfen tritt nur zweimal auf und besteht aus einer Rückmeldehilfe, einer allgemein-strategischen Hilfe, einer inhaltsorientiert-strategischen Hilfe und einer inhaltlichen Hilfe.

Da Kombinationen aus drei oder mehr Hilfen jedoch insgesamt recht selten (in gleicher Weise) auftreten, werden nachfolgend nur die Kombinationen zwischen jeweils zwei Hilfe-Kategorien vorgestellt. Dabei gehen die Kombinationen von drei oder mehreren Hilfen derart mit ein, dass beispielsweise drei Hilfen als drei auftretende Paarungen angesehen werden. Vier Hilfen werden analog als sechs auftretende Paarungen festgehalten. Außerdem gibt es auch Interventionen, bei denen nur eine Hilfe gegeben wurde. Daher entspricht die Anzahl der verschiedenen Kombinationen auch nicht der Gesamtanzahl der erteilten Hilfen. Die Kombinationen geben jedoch einen Einblick, welche Hilfe-Kategorien häufig mit anderen Hilfe-Kategorien zusammen auftreten.

In Tabelle 11 sind die sechs Hilfekategorien jeweils in der ersten Spalte und in der ersten Zeile aufgetragen. Da eine Hilfe-Kategorie innerhalb einer Intervention nicht mehrfach erfasst wurde, ist die Diagonale, in der zu entnehmen ist, wie häufig eine Hilfe-Kategorie mit derselben Hilfe-Kategorie innerhalb der getätigten Interventionen vorkommt, jeweils Null. Zur besseren Lesbarkeit werden die Hilfe-Kategorien in der ersten Zeile horizontal geschrieben. Außerdem werden die Zellen in der Tabelle mit zwei Graustufen markiert: dunkelgraue Zellen beinhalten Kombinationen die häufiger als 40 Mal auftreten, hellgraue markierte Zellen beinhalten Kombinationen, die zwischen 20 und 40 Mal kodiert wurden, und nicht markierte Zellen beinhalten Kombinationen, die weniger als 20 Mal aufgetreten sind.

Man kann erkennen, dass die häufigste Kombination zweier Hilfen innerhalb einer Intervention die Kombination aus inhaltsorientiert-strategischer Hilfe und allgemein-strategischer Hilfe darstellt. Diese Paarung kann insgesamt 51 Mal identifiziert werden. Mit 50 Mal fast genauso häufig tritt die Kombination aus inhaltsorientiert-strategischer Hilfe und Rückmeldehilfe auf. Mit etwas Abstand am dritthäufigsten tritt die Paarung allgemein-strategische Hilfe und Rückmeldehilfe auf. Dass diese drei Paarungen am häufigsten vorkommen, ist insofern nicht verwunderlich, als diese drei Hilfe-Kategorien auch am häufigsten identifiziert werden konnten. Daher ist es naheliegend, dass diese auch am häufigsten in Kombination miteinander auftreten. Insbesondere ist es nicht verwunderlich, dass die Kombination aus inhaltsorientiert-strategischer Hilfe und Rückmeldehilfe häufig vorkommt, da die inhaltsorientiert-strategische Hilfe am häufigsten als dominante Hilfe genutzt wurde und die Rückmeldehilfe am häufigsten als untergeordnete Hilfe.

Tabelle 11: Auftretende Kombinationen von Hilfe-Kategorien innerhalb der Interventionen

Hilfe-Kategorie	Inhaltliche Hilfe	Inhaltsorientiert-strategische Hilfe	Allgemein-strategische Hilfe	Rückmeldehilfe	Motivationshilfe	Disziplinarische Hilfe
Inhaltliche Hilfe	0	28	10	25	2	1
Inhaltsorientiert-strategische Hilfe	28	0	51	50	6	8
Allgemein-strategische Hilfe	10	51	0	29	6	11
Rückmeldehilfe	25	50	29	0	7	5
Motivationshilfe	2	6	6	7	0	2
Disziplinarische Hilfe	1	8	11	5	2	0

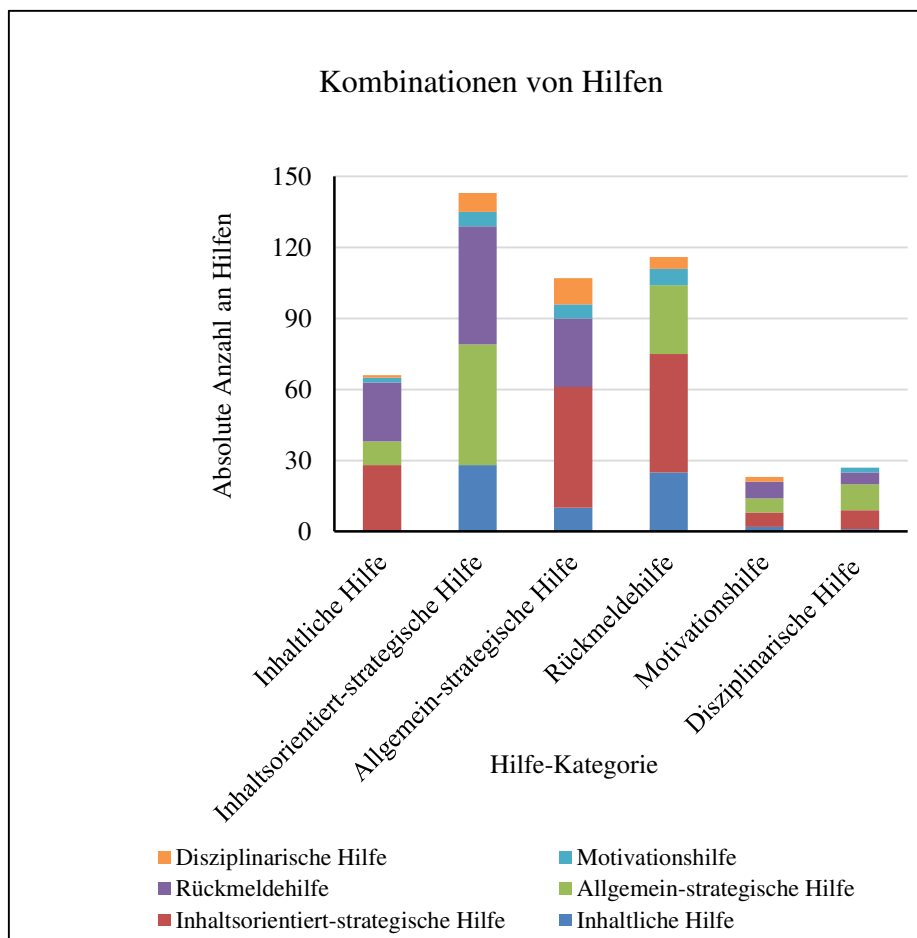


Abbildung 31: Auftretende Kombinationen von Hilfe-Kategorien innerhalb der Intervention in gestapelten Diagrammen

In 28 Interventionen werden inhaltsorientiert-strategische und inhaltliche Hilfe gemeinsam erteilt. Aus der Paarung kann jedoch nicht abgelesen werden, ob zunächst die inhaltsorientiert-strategische Hilfe gegeben wird und anschließend die inhaltliche Hilfe oder umgekehrt. Eine Erklärung hierfür könnte sein, dass die Studierenden nach der inhaltsorientiert-strategischen Hilfe das Gefühl haben, dass diese noch nicht ausreicht und entsprechend noch inhaltlich nachlegen. Umgekehrt könnte es sein, dass die Studierenden das Ergebnis nennen und erst anschließend ein Verfahren vermitteln, das die Schülerinnen und Schüler auch bei anderen, ähnlich gelagerten Aufgaben nutzen können.

Mit 25 Fällen ebenfalls noch verhältnismäßig oft kommt die Paarung Rückmeldehilfe und inhaltliche Hilfe vor. Dies könnten Fälle sein, in denen die Studierenden die Schülergruppe zunächst darauf hinweisen, dass ein Lösungsweg oder eine (Teil-)Aufgabe noch nicht korrekt gelöst wurde und anschließend nicht abwarten, bis die Lernenden ihren Fehler selbstständig korrigieren, sondern in der gleichen Intervention ein Teilergebnis oder gar eine Musterlösung vorstellen.

Selten treten inhaltliche Hilfen zusammen mit Motivationshilfen oder mit disziplinarischen Hilfen auf. Ebenfalls sehr selten ist auch die Kombination aus Motivationshilfe und disziplinarischer Hilfe. Dies liegt vermutlich einerseits daran, dass eine disziplinarische Hilfe den Studierenden eine gewisse Überwindung kostet, da er nicht mit den Lernenden vertraut ist. Andererseits gilt Ähnliches für die Motivationshilfe, die insgesamt nur selten auftritt und entsprechend auch nicht häufig mit anderen Hilfen in Kombination auftreten kann. Dies kann auch Abbildung 31 entnommen werden. Dort sind die aufgetretenen Kombinationen von Hilfen als gestapelte Diagramme visualisiert.

Zusammenfassend lässt sich bezüglich der verwendeten Hilfen innerhalb der Interventionen festhalten, dass die Studierenden überwiegend inhaltsorientiert-strategische und allgemein-strategische Hilfen nutzen. Diese beiden Hilfe-Kategorien sind auch die beiden häufigsten dominanten Hilfen. Inhaltliche Hilfen kommen am dritthäufigsten als dominante Hilfe vor. Rückmeldehilfen und insbesondere Motivationshilfen sind selten dominant. Entsprechend werden weniger inhaltliche Interventionen in geringerem Maße getätigt als Interventionen mit inhaltlichem Bezug. Dies könnte daran liegen, dass die Studierenden die konzipierte Laborstation auch gerne vollständig erproben wollen. Dazu müssen die Schülerinnen und Schüler die Aufgaben in der vorgegebenen Zeit bearbeiten. Hierzu erscheinen Hilfen mit höherem inhaltlichen Gehalt förderlich, denn je mehr Inhalte vorgegeben werden, desto schneller schreiten die Schülerinnen und Schüler in der Bearbeitung der Aufgaben voran.

Im Fall der untergeordneten Hilfen werden Rückmeldehilfen am häufigsten genutzt. Dies spricht dafür, dass die Studierenden durchaus auch darauf bedacht sind, den Schülerinnen und Schülern eine Rückmeldung zu ihrem aktuellen Lernstand zu geben. Andererseits spielen diese Hilfen in den Interventionen eine untergeordnete Rolle, was bedeutet, dass eine andere Hilfe dominanter war. Entsprechend liegt der Fokus der Hilfe nicht darauf, den Schülerinnen und Schülern eine Rückmeldung zu geben, sondern in den meisten Fällen auf einer inhaltsorientiert-strategischen Hilfe, einer allgemein-strategischen Hilfe oder einer inhaltlichen Hilfe. Den Schülerinnen und Schülern wird also in vielen Fällen keine Zeit gegeben, selbstständig darüber nachzudenken, weshalb ein Lösungsweg oder eine Lösung nicht zielführend oder korrekt ist, sondern es wird direkt eine mögliche Lösungsstrategie (mit mehr oder weniger inhaltlichem Bezug) geliefert oder im Extremfall im Rahmen einer inhaltlichen Hilfe die Lösung vorgegeben.

Inhaltliche Hilfen sind ebenso wie disziplinarische Hilfen nur selten untergeordnet. Gerade bei inhaltlichen Hilfen erscheint dies nachvollziehbar, da durch die Vorgabe eines konkreten

Lösungsweges oder gar eines (Teil-)Ergebnisses das Wesen der Intervention dermaßen verändert wird, dass möglicherweise vorher oder im Anschluss gegebene Hilfen vergleichsweise unbedeutend werden. Motivationshilfen werden auch als untergeordnete Hilfe nur selten gegeben. Sie kommen daher auch bei der Betrachtung aller Hilfen (also der Summe der dominanten und untergeordneten Hilfen) kaum vor.

Bezüglich der Zeitdauer der Interventionen lässt sich festhalten, dass die Interventionen länger dauern, je inhaltlicher die Hilfe ist. Entsprechend dauern inhaltliche Hilfe am längsten und inhaltsorientiert-strategische Hilfen am zweitlängsten. Am kürzesten dauern Hilfen ohne inhaltlichen Bezug, also Motivationshilfen und disziplinarische Hilfen. Bei untergeordneten Hilfen ändert sich dies insofern, als inhaltliche Hilfen vergleichsweise kurz sind, wenn sie nicht dominant sind. Dies ergibt sich daraus, dass lange inhaltliche Hilfen dazu führen, dass diese auch als dominant wahrgenommen werden. Von dieser Besonderheit abgesehen, ergibt sich auch bei den untergeordneten Hilfen ein Zusammenhang zwischen Inhalt und Zeitdauer der Intervention.

Die häufigsten Kombinationen von Hilfen, also die Nutzung mehrerer Hilfen innerhalb einer Intervention, sind die Kombination aus inhaltsorientiert-strategischen Hilfen und allgemein-strategischen Hilfen sowie die Kombination aus inhaltsorientiert-strategischen Hilfen und Rückmeldehilfen. Dies ist insofern nicht verwunderlich, als diese drei Hilfe-Kategorien auch insgesamt am häufigsten genutzt werden.

9.2.2 Reihenfolge der Hilfen

Aus den 208 Interventionen wurden genau 208 Reihenfolgen erfasst, alle Reihenfolgen wurden also nur einmal gewertet. Das bedeutet, dass eine Intervention, die aus drei Hilfen besteht, nur einmal und zwar genau in der Reihenfolge der drei Hilfen, erfasst wurde. Beispielsweise wurden in sechs Fällen jeweils zuerst eine Rückmeldehilfe, gefolgt von einer inhaltsorientiert-strategischen und dann von einer allgemein-strategischen Hilfe gegeben. Die Reihenfolge Rückmeldehilfe → Inhaltsorientiert-strategische Hilfe wurde dabei nicht festgehalten, ebenso wenig wie die Reihenfolge Inhaltsorientiert-strategische Hilfe → allgemein-strategische Hilfe.

Insgesamt können 52 verschiedene Reihenfolgen von erteilten Hilfen festgestellt werden. Hieran wird ersichtlich, dass potenziell eine beinahe beliebige Verkettung von verschiedenartigen Hilfen möglich ist und auch eine große Bandbreite innerhalb der Studie abgedeckt wird. Viele davon treten nur selten auf und nur zehn Konstellationen konnten mindestens fünfmal erfasst werden. Diese zehn Reihenfolgen von Hilfen sollen im Folgenden näher betrachtet werden. Dabei werden auch exklusive Hilfen, die ohne weitere Hilfe-Kategorie erteilt wurden, mitaufgelistet (vgl. Tab. 12).

Tabelle 12: Reihenfolge der auftretenden Hilfen

Reihenfolge	Anzahl
Allgemein-strategische Hilfe	35
Inhaltsorientiert-strategische Hilfe	24
Allgemein-strategische Hilfe → Inhaltsorientiert-strategische Hilfe	16
Rückmeldehilfe → Inhaltsorientiert-strategische Hilfe	15
Disziplinarische Hilfe	11
Rückmeldehilfe	10
Inhaltsorientiert-strategische Hilfe → Allgemein-strategische Hilfe	7

Inhaltsorientiert-strategische Hilfe → Inhaltliche Hilfe	7
Rückmeldehilfe → Inhaltsorientiert-strategische Hilfe → Allgemein-strategische Hilfe	6
Rückmeldehilfe → Inhaltliche Hilfe	5

Man kann erkennen, dass die am häufigsten auftretenden Interventionen reine allgemein-strategische Hilfen waren, gefolgt von reinen inhaltsorientiert-strategischen Hilfen. Bei der am dritthäufigsten auftretenden Reihenfolge handelt es sich zum ersten Mal um eine Kombination aus zwei Hilfen. In 16 Fällen wurde nämlich zunächst eine allgemein-strategische Hilfe und anschließend eine inhaltsorientiert-strategische Hilfe gegeben. In 15 Fällen wurde zunächst eine Rückmeldehilfe und anschließend eine inhaltsorientiert-strategische Hilfe erteilt.

Es ist eine Tendenz zu erkennen, dass zunächst weniger inhaltliche Hilfen gegeben werden und erst anschließend eine weitere Hilfe mit höherem inhaltlichen Bezug. Indiz hierfür ist, dass allgemeinstrategische Hilfen in 16 Fällen vor inhaltsorientiert-strategischen Hilfen gegeben wurden, Rückmeldehilfen in 15 Fällen vor inhaltsorientiert-strategischen Hilfen gegeben wurden und inhaltsorientiert-strategische Hilfen in sieben Fällen vor inhaltlichen Hilfen gegeben wurden. Dieser Befund trifft jedoch nicht auf alle Reihenfolgen zu. Beispielsweise wurde in sieben Fällen zuerst eine inhaltsorientiert-strategische Hilfe gegeben, bevor auf eine allgemein-strategische Hilfe zurückgegriffen wurde.

Um zu klären, ob die Studierenden tatsächlich zunächst Hilfen nutzen, die weniger inhaltlichen Bezug haben, wurden alle Interventionen hinsichtlich der ersten erteilten Hilfe untersucht. In Tabelle 13 sind die Ergebnisse aufgelistet. Es wäre zu erwarten, dass die inhaltsorientiert-strategischen Hilfen, die insgesamt am häufigsten genutzt wurden und am häufigsten die dominante Hilfe-Kategorie darstellten, an erster Stelle stehen. Jedoch werden diese von den Studierenden nur am dritthäufigsten als erste Hilfe innerhalb der Intervention genutzt. Am häufigsten werden allgemein-strategische Hilfen als erste Hilfe verwendet. Die zweithäufigste Kategorie sind die Rückmeldehilfen, die in 51 Interventionen und somit in einer Intervention mehr als erste Hilfe verwendet werden als inhaltsorientiert-strategische Hilfen. Dass Motivationshilfen am seltensten als erstes in der Intervention erteilt werden, verwundert nicht, da diese ohnehin nur selten vorkommen. Der geringe Anteil an inhaltlichen Hilfen bestärkt die Vermutung, dass die Studierenden zunächst weniger inhaltliche Hilfen nutzen und sich im Laufe der Intervention eher inhaltlichen Hilfen bedienen.

Tabelle 13: Anzahl der zuerst genutzten Hilfe innerhalb einer Intervention

Erste Hilfe innerhalb der Intervention	Anzahl
Disziplinarische Hilfe	22
Motivationshilfe	6
Rückmeldehilfe	51
Allgemein-strategische Hilfe	71
Inhaltsorientiert-strategische Hilfe	50
Inhaltliche Hilfe	8

Hinsichtlich der Reihenfolge der gegebenen Hilfen innerhalb der Interventionen ist eine Tendenz zu entdecken, dass zunächst weniger inhaltliche Hilfen genutzt werden. Dies erkennt man beispielsweise daran, dass die beiden am häufigsten auftretenden Reihenfolgen aus zwei Hilfen eine allgemein-strategische Hilfe gefolgt von einer inhaltsorientiert-strategischen Hilfe sowie

eine Rückmeldehilfe gefolgt von einer inhaltsorientiert-strategischen Hilfe sind. Diese Einschätzung kann auch durch die zuerst verwendeten Hilfen innerhalb der Interventionen bestätigt werden. Inhaltsorientiert-strategische Hilfen, die insgesamt am häufigsten auftreten, werden nur am dritthäufigsten an erster Stelle verwendet. Es zeigt sich also auch hier die Tendenz, dass zunächst weniger inhaltliche Hilfen genutzt werden.

9.2.3 Angemessenheit

Insgesamt wurden 154 der 208 Interventionen als angemessen bewertet. Entsprechend wurden 54 Interventionen als unangemessen eingestuft. Abbildung 32 zeigt die Verteilung der Angemessenheit auf die dominanten Hilfe-Kategorien. Auf der x-Achse ist die Anzahl der Interventionen und auf der y-Achse die unterschiedlichen (dominanten) Hilfe-Kategorien abgetragen. Angemessene Hilfen sind in Grün und unangemessene Interventionen in Rot dargestellt.

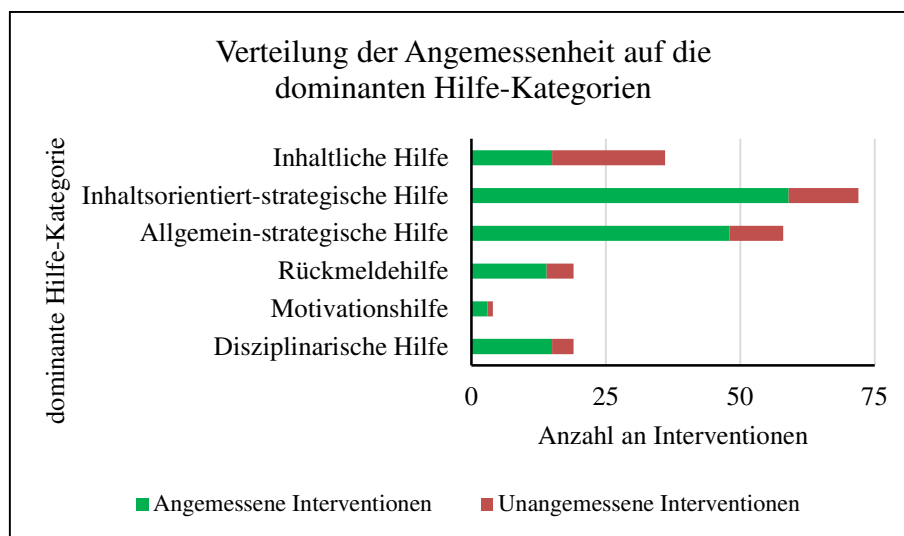


Abbildung 32: Verteilung der Angemessenheit auf die dominanten Hilfe-Kategorien

Man kann feststellen, dass bei allen Hilfe-Kategorien außer der inhaltlichen Hilfe der Anteil an angemessenen Interventionen gegenüber dem Anteil an unangemessenen Interventionen überwiegt. Bei den inhaltlichen Hilfen werden hingegen mehr unangemessene als angemessene Interventionen getätigt. Dies ist auch in Abbildung 33 erkennbar, in der der prozentuale Anteil der angemessenen Interventionen bezogen auf die jeweilige Interventionsart dargestellt wird. Dabei wird zur Ergebnisdarstellung die dominante Hilfe innerhalb der Intervention namensgebend dafür, welche Interventionsart vorliegt. Dies wird deshalb so definiert, weil die dominante Hilfe die Intervention am besten charakterisiert. Während der prozentuale Anteil bei allen anderen Hilfen bei rund 80% liegt, werden inhaltliche Hilfen nur in rund 40% der Fälle angemessen bewertet.

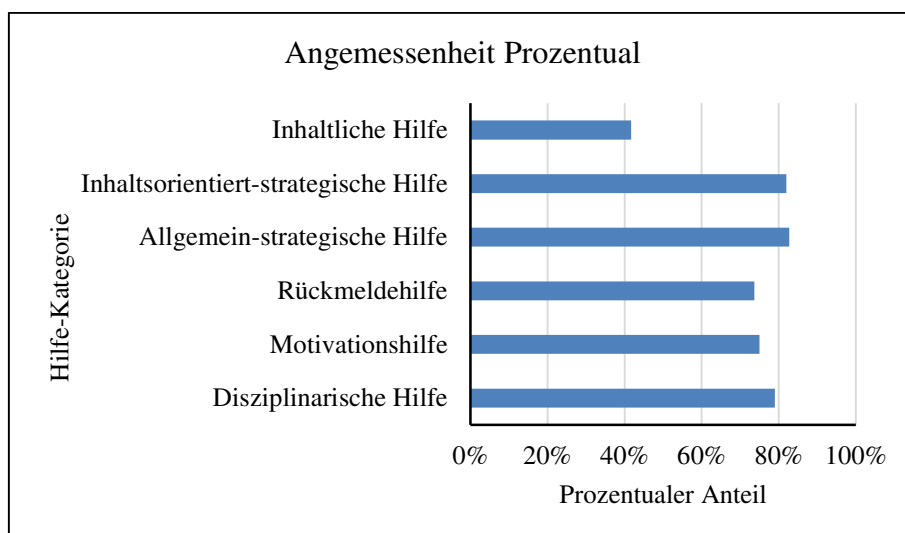


Abbildung 33: Prozentualer Anteil der angemessenen Interventionen bezogen auf die jeweilige Interventionsart

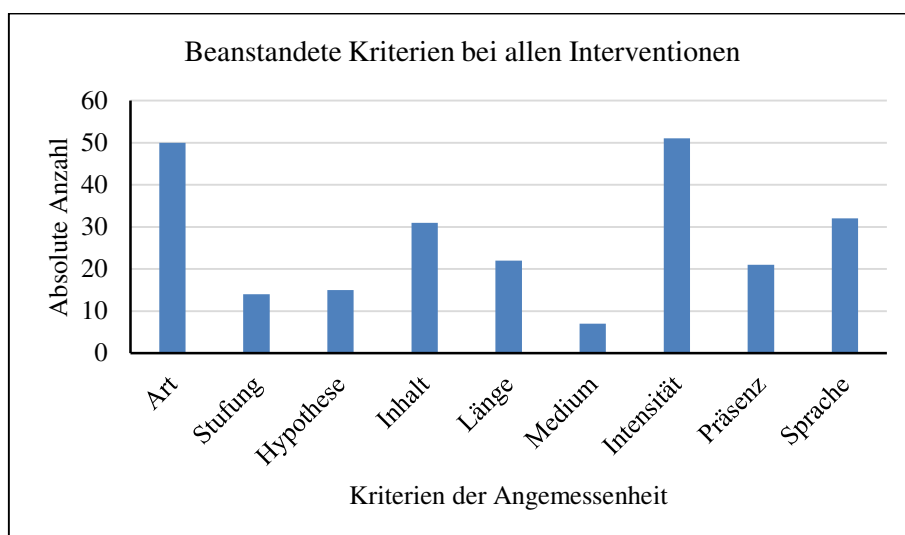


Abbildung 34: Durch die Expertinnen und Experten beanstandete Kriterien innerhalb der Interventionen

Um einen tieferen Einblick zu erhalten, weshalb dies so ist und vor allem auch weshalb Interventionen als unangemessen bewertet wurden, können die Kriterien der (Un-)Angemessenheit genutzt werden (vgl. dazu Kap. 8.3.4). In Abbildung 34 sind die Kriterien abgebildet, die die Expertinnen und Experten bei allen 208 Interventionen beanstandet haben. Man sieht, dass die Art (Auswahl der Hilfe-Kategorie) und Intensität (Intensität der Intervention) am häufigsten (jeweils etwa 50 Mal) beanstandet wurden. Mit jeweils etwa 30 Beanstandungen bilden die Kategorien Inhalt (Inhalt der Intervention) und Sprache (Angemessenheit der Sprache) Kategorien, die ebenfalls häufig kritisiert wurden. Mit jeweils rund 20 Beanstandungen wurden die Kriterien Länge (Länge der Intervention) und Präsenz (Präsenz der Lehrperson) festgehalten. Vergleichsweise selten (jeweils ungefähr zehn Mal) wurden Stufung (Stufung der Hilfen), Hypothese (Erkennbare Hypothese zum Schülerproblem) und Medium (Medienunterstützung) als unangemessen eingestuft.

Hinsichtlich der Art der Hilfe ist relativ klar, dass die Expertinnen und Experten der Meinung waren, eine andere Hilfe-Kategorie wäre zielführender gewesen. Hinsichtlich der Intensität der Intervention ist jedoch nicht klar, ob die gegebene Hilfe zu intensiv war oder ob diese intensiver

hätte sein müssen. Bei einer inhaltlichen Hilfe beispielsweise könnte eine unangemessene Intensität einerseits bedeuten, dass zu viel Inhalt (evtl. sogar in Form des Ergebnisses) vorgegeben wurde, oder aber, dass der inhaltliche Hinweis nicht ausgereicht hat und ausführlicher hätte dargeboten werden sollen. Die Studierenden hatten also einerseits Probleme, die passende Hilfe auszuwählen, andererseits war diese oftmals nicht in dem für die Expertinnen und Experten passenden Maß. Inhaltliche Fehler waren ebenfalls recht häufig, obwohl diese bei einigen Hilfe-Kategorien in der Regel keine Relevanz besitzen (etwa bei disziplinarischen Hilfen oder Motivationshilfen). Wenige Probleme hatten die Studierenden damit, das passende Medium für die Intervention zu wählen. Dieses Kriterium ist jedoch nur in wenigen Interventionen überhaupt relevant, beispielsweise wenn Inhalte erläutert werden und dazu etwas veranschaulicht werden sollte, weil die Schülerinnen und Schüler sonst nicht richtig folgen könnten. In den meisten Interventionen reicht jedoch eine verbale Hilfe ohne zusätzliche Unterstützung durch ein Medium aus, sodass es wenig verwundert, dass dieses Kriterium am seltensten kritisiert wurde.

Da es auch einige Interventionen gab, die die Expertinnen und Experten als angemessen bewertet haben, und es entsprechend auch Interventionen gab, die ohne Nennung eines Makels auskamen, wird ersichtlich, dass bei manchen Interventionen mehrere Kriterien als Stütze angegeben wurden, um die Angemessenheit der Intervention zu bewerten. Daher werden nun zunächst die Kriterien und deren Häufigkeiten aufgelistet, die einzeln genannt wurden. In Tabelle 14 kann man erkennen, dass die Art der Hilfe und die Intensität der Intervention am häufigsten ohne weiteres Kriterium angegeben wurden. Entsprechend zeigt sich auch an dieser Stelle, dass diese beiden Kriterien den Studierenden die größten Probleme bereiteten und daher von den Expertinnen und Experten kritisiert wurden. Das Kriterium „Medium“ wurde nie einzeln angegeben, was sich in die obigen Überlegungen, dass es sich hierbei um ein weniger relevantes Kriterium handelt, fügt.

Die Kriterien „Hypothese“ (1) und „Stufung“ (2) wurden ebenfalls sehr selten als exklusive Stütze für die Bewertung der Angemessenheit angegeben. Alle anderen Kriterien wurden mindestens fünfmal exklusiv angegeben, wobei das Kriterium „Inhalt“ sogar achtmal herangezogen wurde. Auch hier zeigt sich die Tendenz, die bereits bei der Betrachtung aller beanstandeten Kriterien ersichtlich wurde, dass inhaltliche Fehler im Vergleich zu den anderen Kriterien recht häufig auftreten. Dies ergibt sich auch aus der Relevanz der inhaltlichen Fehler, denn ein inhaltlicher Fehler führt vermutlich recht schnell dazu, dass eine Intervention insgesamt als unangemessen angesehen wird. Nach Betrachtung aller Kriterien sowie der exklusiv genutzten Kriterien lässt sich festhalten, dass die drei Kriterien „Art“, „Intensität“ und „Inhalt“ die relevantesten Kriterien zu sein scheinen.

Tabelle 14: Anzahl der Kriterien, die exklusiv zur Stütze der Bewertung der Angemessenheit angegeben wurden

Kriterium	Anzahl
Art	10
Stufung	2
Hypothese	1
Inhalt	8
Länge	5
Medium	0
Intensität	10
Präsenz	5
Sprache	5

An dieser Stelle ist es von Interesse, welche Kriterien häufig gemeinsam mit anderen Kriterien auftreten (vgl. Tab. 15). Es wird ersichtlich, dass fast alle denkbaren Kombinationen tatsächlich auftreten. Während einige nur eher selten auftreten, stechen vor allem die Kombinationen „Sprache“ und „Intensität“ (16 Mal) und „Präsenz“ und „Intensität“ (13 Mal) heraus. Bei diesen beiden Kombinationen wäre vorstellbar, dass die Intensität der Hilfe zu groß war, also zu viel Hilfe gegeben wurde und den Schülerinnen und Schülern dadurch Lerngelegenheiten vewehrt wurden. Es ist auch denkbar, dass die Intensität zu gering war und dadurch weniger Hilfe gegeben wurde, als eigentlich notwendig gewesen wäre. Hinsichtlich der Sprache könnte es sein, dass nicht treffend formuliert wurde und die Interventionen dadurch unnötig in die Länge gezogen wurden. Die Kombination der Kriterien „Präsenz“ und „Intensität“ erscheint insofern plausibel, als beispielsweise bei disziplinarischen Hilfen durchaus auch die physische Präsenz der Lehrkraft gefordert ist, die beispielsweise nicht gegeben ist, wenn eine Ermahnung von der Tür aus erteilt wird.

Mit jeweils zehn Fällen treten die Kombinationen „Art“ und „Inhalt“, „Art“ und „Intensität“ sowie „Länge“ und „Intensität“ noch recht häufig auf. Es fällt auf, dass die Intensität ein Kriterium zu sein scheint, das häufig in Kombination mit anderen Kriterien beanstandet wurde. Dass die Summe der Kombinationen nicht mit der Summe der Beanstandungen übereinstimmt, liegt daran, dass bei drei oder mehr Beanstandungen mehr Kombinationen zustande kommen, die dann entsprechend über die Summe der einzelnen Beanstandungen hinausgehen. Wenn beispielsweise in einer Intervention drei Kriterien beanstandet werden, entstünden daraus drei mögliche Kombinationen von jeweils zwei Kriterien. Für vier beanstandete Kriterien innerhalb einer Intervention wären es bereits sechs und für fünf beanstandete Kriterien sogar zehn mögliche Kombinationen.

Tabelle 15: Auftretende Kombinationen mehrerer Kriterien der Angemessenheit

Kriterium	Art	Stufung	Hypothese	Inhalt	Länge	Medium	Intensität	Präsenz	Sprache
Art	-	5	8	10	3	1	10	3	6
Stufung	5	-	5	3	3	1	9	3	3
Hypothese	8	5	-	5	3	1	5	0	3
Inhalt	10	3	5	-	7	2	9	2	4
Länge	3	3	3	7	-	1	10	3	5
Medium	1	1	1	2	1	-	5	1	2
Intensität	10	9	5	9	10	5	-	13	16
Präsenz	3	3	0	2	3	1	13	-	6
Sprache	6	3	3	4	5	2	16	6	-

Darüber hinaus lohnt ein Blick darauf, welche Kriterien bei welchen Interventionsarten am häufigsten angemerkt wurden. Dazu wurden die zwei häufigsten Kriterien pro Interventionsart in

Tabelle 16 aufgelistet. Man kann erkennen, dass die Intensität bei allen Interventionsarten ein wichtiges Kriterium zu sein scheint. Lediglich bei den Rückmeldehilfen ist dieses Kriterium nicht unter den zwei häufigsten Kriterien zu finden. Das Kriterium „Inhalt“ ist erwartungsgemäß besonders bei inhaltlichen Hilfen entscheidend. Nach der „Intensität“ am häufigsten vertreten ist die „Art“ der Hilfe, die bei vier Interventionsarten das zweithäufigste genannte Kriterium darstellt. Dies ist insofern plausibel, als besonders bei den Hilfen, die aus inhaltlicher Sicht weder zu den inhaltlichsten noch zu den am wenigsten inhaltlichen Hilfen zählen, die Art der genutzten Hilfe-Kategorie eine wichtige Rolle spielt. Die Expertinnen und Experten bewerten sowohl bei inhaltsorientiert-strategischen und allgemein-strategischen als auch bei Rückmeldehilfen häufig die gewählte Interventionsart als unpassend. Die Motivationshilfen können aufgrund der geringen Fallzahl nicht verlässlich interpretiert werden. Bei disziplinarischen Hilfen scheinen insbesondere die Präsenz der Lehrkraft sowie die Intensität der Intervention die ausschlaggebenden Kriterien zu sein. Wie bereits erwähnt, ist diese Kombination gut begründbar, da bei disziplinarischen Hilfen die physische Präsenz der Lehrkraft eine zentralere Rolle spielt als bei anderen Interventionsarten.

Tabelle 16: Die häufigsten beanstandeten Kriterien nach Interventionsart

Interventionsart	Häufigstes Kriterium	Zweithäufigstes Kriterium
Inhaltliche Hilfe	Inhalt, Intensität (je 14)	
Inhaltsorientiert-strategische Hilfe	Intensität (14)	Art (13)
Allgemein-strategische Hilfe	Intensität (11)	Art, Sprache (je 7)
Rückmeldehilfe	Art, Sprache (je 4)	
Motivationshilfe	Präsenz (2)	Art, Länge, Intensität, Hypothese, Sprache (je 1)
Disziplinarische Hilfe	Präsenz (10)	Intensität (9)

Nachdem nun zunächst alle Interventionen und die in allen Interventionen beanstandeten Kriterien untersucht wurden, soll im Folgenden genauer auf die Interventionen eingegangen werden, die als unangemessen eingestuft wurden. Insbesondere dort ist es interessant, welche Kriterien ausschlaggebend dafür sind, Interventionen als unangemessen zu bewerten. In Abbildung 35 sind auf der x-Achse die verschiedenen Kriterien aufgetragen. Auf der y-Achse ist die Anzahl angegeben, wie oft ein Kriterium als Stütze für die Bewertung der Unangemessenheit der Intervention diente. Man kann erkennen, dass die gewählte Art der Hilfe bei den als unangemessen eingestuften Interventionen am häufigsten beanstandet wurde. Häufig hätten die Expertinnen und Experten also eine andere Hilfe-Kategorie als zielführender angesehen. Der Inhalt der Intervention wurde am zweithäufigsten beanstandet. Dies kann auch daran liegen, dass die Expertinnen und Experten fachlich sehr sicher sind und etwaige Fehler entsprechend erkennen. Ebenfalls erwähnenswert sind die Kategorien „Länge“, „Intensität“ und „Sprache“. Die anderen vier Kategorien („Stufung“, „Hypothese“, „Medium“ und „Präsenz“) waren eher selten ausschlaggebend für die Stütze der Bewertung einer Intervention als unangemessen.

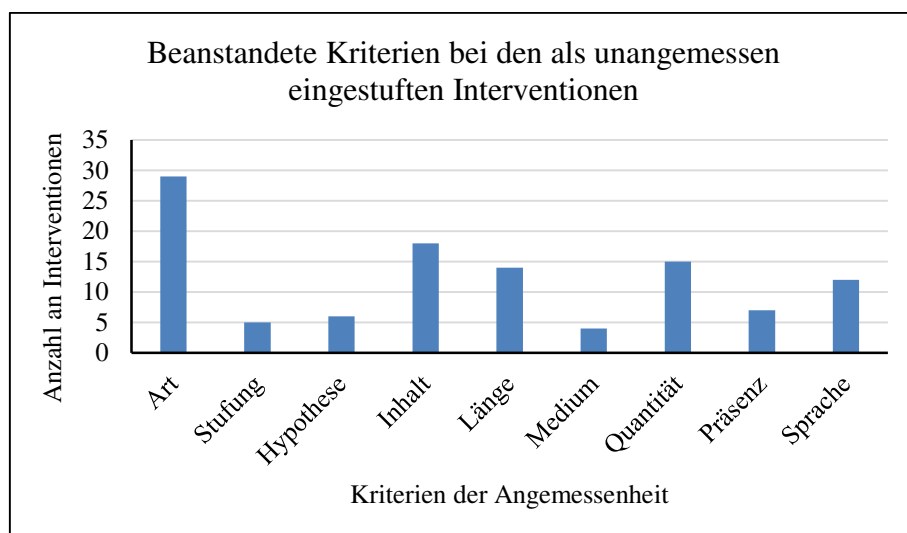


Abbildung 35: Durch die Expertinnen und Experten beanstandete Kriterien innerhalb der als unangemessen bewerteten Interventionen

Nachfolgend wird auf die exklusiv auftretenden Kriterien und auf die auftretenden Kombinationen mehrerer Kriterien der Angemessenheit geblickt. Tabelle 17 verdeutlicht, dass vier der neun Kriterien („Stufung“, „Hypothese“, „Medium“ und „Sprache“) nicht exklusiv als Stütze bei einer als unangemessen bewerteten Intervention herangezogen wurden. Das am häufigsten exklusiv genannte Kriterium ist die Art der Hilfe. Dies passt auch zu dem obigen Befund, dass dies das Kriterium darstellt, welches am häufigsten bei unangemessenen Interventionen identifiziert wurde. Die fachliche Korrektheit („Inhalt“) wurde am zweithäufigsten exklusiv genannt. Auch dies passt zu dem vorigen Befund, da es das zweithäufigste beanstandete Kriterium innerhalb von unangemessen bewerteten Interventionen bildet. Zusätzlich werden die Länge (2), die Intensität (1) und die Präsenz (1) exklusiv angegeben.

Tabelle 17: Anzahl der Kriterien, die exklusiv zur Stütze bei den als unangemessen bewerteten Interventionen angegeben wurden

Kriterium	Anzahl
Art	5
Stufung	0
Hypothese	0
Inhalt	3
Länge	2
Medium	0
Intensität	1
Präsenz	1
Sprache	0

Im Hinblick auf die auftretenden Kombinationen mehrerer Kriterien der Angemessenheit bei als unangemessen bewerteten Interventionen kann man erkennen, dass die Kombinationen „Art“ und „Inhalt“ (8) und „Art“ und „Intensität“ (7) am häufigsten auftreten (vgl. Tab. 18). Eine Reihe von Kombinationen tritt gar nicht auf. Ebenfalls häufig treten „Länge“ und „Intensität“ (6) sowie „Intensität“ und „Sprache“ in Kombination (6) auf.

Tabelle 18: Auftretende Kombinationen mehrerer Kriterien der Angemessenheit bei als unangemessen bewerteten Interventionen

Kriterium	Art	Stufung	Hypothese	Inhalt	Länge	Medium	Intensität	Präsenz	Sprache
Art	-	1	4	8	3	0	7	2	3
Stufung	1	-	1	2	2	0	3	0	2
Hypothese	4	1	-	3	3	0	3	0	2
Inhalt	8	2	3	-	5	2	3	0	3
Länge	3	2	3	5	-	1	6	2	3
Medium	0	0	0	2	1	-	2	1	2
Intensität	7	3	3	3	6	2	-	3	6
Präsenz	2	0	0	0	2	1	3	-	3
Sprache	3	2	2	3	3	2	6	3	-

Der Großteil der 208 Interventionen durch die angehenden Lehrkräfte wurde von den Expertinnen und Experten als angemessen eingestuft. Dies gilt auch für alle Interventionsarten, abgesehen von der inhaltlichen Hilfe. Inhaltlichen Hilfen wurden überwiegend als unangemessen bewertet. Dies kann vielfältige Gründe haben. Eine mögliche Erklärung ist, dass der Großteil der Kriterien der Angemessenheit, die für die Bewertung zugrunde gelegt wurden, bei inhaltlichen Hilfen Anwendung findet, wohingegen nicht alle Kriterien für alle Interventionsarten relevant sind. Eine ebenfalls denkbare Erklärung hierfür ist, dass inhaltliche Fehler vermutlich eher als entscheidendes Kriterium für die Bewertung der Angemessenheit angesehen werden, da eine inhaltliche Hilfe, die fachliche Mängel aufweist, wohl nur in den seltensten Fällen noch als angemessen bewertet wird. Für die Interpretation spricht auch der Vergleich der als Stütze für die Bewertung verwendeten Kriterien der Angemessenheit bei allen Interventionen, also auch den als angemessen bewerteten (vgl. Abb. 34), und den verwendeten Kriterien bei den als unangemessen bewerteten Interventionen (vgl. Abb. 35). Im Vergleich erkennt man, dass das Kriterium „Inhalt“ das vierthäufigste Kriterium bei allen Interventionen ist und bei den unangemessenen Interventionen das zweithäufigste Kriterium darstellt.

Sowohl bei allen als auch bei der Teilmenge der unangemessenen Interventionen stellt die gewählte Art der Hilfe ein zentrales Kriterium zur Bewertung der Angemessenheit dar. Die meisten als unangemessen eingestuften Interventionen werden mit dieser Kategorie begründet. Es ist auch das Kriterium, das am häufigsten exklusiv zur Stütze der Bewertung der Unangemessenheit herangezogen wurde (vgl. Tab. 17). Entsprechend naheliegend ist es, dass auch die Kombination der beiden Kriterien „Inhalt“ und „Art“ die am häufigsten verwendete Kombination zur Stütze der Bewertung der Unangemessenheit bildet (vgl. Tab. 18).

9.2.4 Anzahl der aktiven Schülerinnen und Schüler

Bei der Bewertung der Angemessenheit durch die kleine und die große Expertengruppe wurden jeweils nur ein kurzer Vorlauf vor der Intervention sowie die Intervention selbst beobachtet (vgl. dazu Kap. 8.3.4). Natürlich ist jedoch auch relevant, inwieweit die Intervention dazu beiträgt, die Schülerinnen und Schüler zu aktivieren. Entsprechend kann untersucht werden, wie viele Schülerinnen und Schüler sich vor der Intervention aktiv mit den mathematischen Inhalten der Lernumgebung beschäftigt haben und wie viele dies nach der Intervention tun. In Abbildung 36 ist auf der x-Achse die Anzahl der aktiv arbeitenden Schülerinnen und Schüler abgetragen, auf der y-Achse die Anzahl an Interventionen, bei denen die gleiche Anzahl an Schülerinnen und Schülern aktiv war. In Blau ist das Arbeitsverhalten vor der Intervention und in Hellgrün das Arbeitsverhalten nach der Intervention zu sehen. Auf den ersten Blick wird ersichtlich, dass in den meisten Fällen sowohl vor als auch nach der Intervention ein Großteil der Schülerinnen und Schüler aktiv arbeitet. Betrachtet man die Abbildung genauer, so erkennt man, dass nach der Intervention mehr Schülerinnen und Schüler arbeiten als zuvor. Bei mehreren Gruppen sind nach der Intervention vier Schülerinnen und Schüler aktiv. Insgesamt ist dies bei mehr Gruppen der Fall als vor der Intervention. Dies gilt auch für drei der vier Schülerinnen und Schüler der Gruppe, die sich mit den mathematischen Inhalten der Lernumgebung beschäftigen. Entsprechend gibt es bei zwei Schülerinnen und Schülern, einer bzw. keiner aktiven Person jeweils einen Rückgang. Die Fälle, bei denen keine Beobachtung des Arbeitsverhaltens möglich war (vgl. dazu Kap. 8.3.5), treten vor und nach den Interventionen in etwa gleich oft auf und liefern daher keinen inhaltlichen Mehrwert, sodass diese nachfolgend nicht weiter berichtet werden.

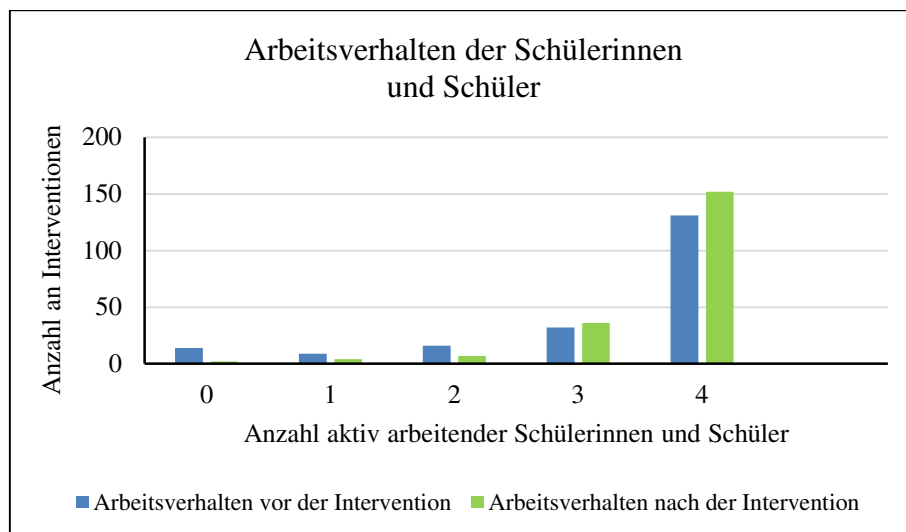


Abbildung 36: Arbeitsverhalten der Schülerinnen und Schüler vor und nach der Intervention der angehenden Lehrkraft

Um die Veränderung noch genauer in den Blick nehmen zu können, wurden die Ergebnisse in Tabelle 19 abgetragen und die Veränderung sowie die zusätzlich aktivierten Schülerinnen und Schüler berechnet. Insgesamt arbeiten nach den Interventionen 73 Schülerinnen und Schüler mehr als davor. Dies ergibt sich aus 96 Schülerinnen und Schülern, die zusätzlich aktiviert wurden und deren positiver Einfluss in den Zeilen mit drei bzw. vier aktiven Schülerinnen und Schülern abgelesen werden kann abzüglich der 23 Schülerinnen und Schüler aus den Gruppen, bei denen zwei oder weniger aktiv waren. Rechnet man die 73 Schülerinnen und Schüler auf

die 208 Interventionen um, könnte man schlussfolgern, dass durchschnittlich etwa jede dritte Intervention dazu führt, dass eine Schülerin bzw. ein Schüler aktiviert wird.

Tabelle 19: Veränderung des Arbeitsverhalten der Schülerinnen und Schüler

Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	131	152	+21	84
3	32	36	+4	12
2	16	7	-9	-18
1	9	4	-5	-5
0	14	2	-12	0

Der Vergleich der vor bzw. nach der Intervention aktiven Schülerinnen und Schüler kann zusätzlich genutzt werden, um etwas über die Angemessenheit der Intervention auszusagen. Daher erscheint es sinnvoll, die 208 Interventionen nochmals differenziert zu betrachten und zwischen den als angemessen bewerteten Interventionen und den als unangemessen bewerteten Interventionen zu unterscheiden (vgl. Tab. 20).

Tabelle 20: Veränderung des Arbeitsverhalten der Schülerinnen und Schüler getrennt nach angemessenen und unangemessenen Interventionen

Angemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	100	118	+18	72
3	23	26	+3	9
2	11	4	-7	-14
1	5	0	-5	-5
0	9	0	-9	0
Unangemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	31	34	+3	12
3	9	10	+1	3
2	5	3	-2	-4
1	4	4	0	0
0	5	2	-3	0

Betrachtet man die Anzahl der aktiven Schülerinnen und Schüler vor und nach den getätigten Interventionen, so erkennt man einerseits, dass in den meisten Interventionen bereits vorher schon der Großteil der Schülerinnen und Schüler aktiv war. Andererseits gibt es einen leichten Zuwachs an aktiven Schülerinnen und Schülern. In Anbetracht der Gesamtanzahl der Interventionen führt jedoch durchschnittlich nur jede dritte Intervention dazu, dass ein Lernender zusätzlich aktiviert wird. Daher könnte man an dieser Stelle festhalten, dass unter dem Aspekt der Schüleraktivität einige Interventionen überflüssig waren.

Differenziert man die Interventionen nach angemessenen und unangemessenen Interventionen, so erkennt man, dass angemessenere Interventionen in höherem Maße zur Schüleraktivität beitragen. In fünf angemessenen Interventionen werden etwa zwei Lernende zusätzlich aktiviert. Bei den unangemessenen Interventionen wird hingegen alle fünf Interventionen ein Lernender zusätzlich aktiviert. Angemessene Interventionen führten im Rahmen der Studie immer dazu, dass anschließend mindestens die Hälfte der Schülerinnen und Schüler einer Gruppe aktiv waren. Bei unangemessenen Interventionen gab es auch Gruppen, in denen niemand oder nur ein Lernender aktiv war. Die Bewertung der Angemessenheit durch die Expertinnen und Experten kann also durch die Beobachtung der Schüleraktivität vor und nach den Interventionen gestützt werden.

9.2.5 Interventionsauslöser

Bisher wurde auf die verschiedenen Hilfen innerhalb der Interventionen, die Angemessenheit der Interventionen und die aktiven Schülerinnen und Schüler vor bzw. nach der Intervention eingegangen. Nun rückt der Interventionsauslöser in den Fokus der Analyse. Hierbei geht es darum, ob die Intervention von den Schülerinnen und Schülern gefordert wurde, also responsiv war, oder ob die Intervention von der angehenden Lehrkraft ausging und somit invasiv war. Abbildung 37 zeigt, dass der überwiegende Teil der Interventionen von den Studierenden ausging. Dies ist insofern nicht verwunderlich, da die Schülerinnen und Schüler die Aufgabe haben, das Arbeitsheft selbstständig, ggf. unter Zuhilfenahme des Hilfehefts, der gegenständlichen Materialien und Simulationen, zu bearbeiten.

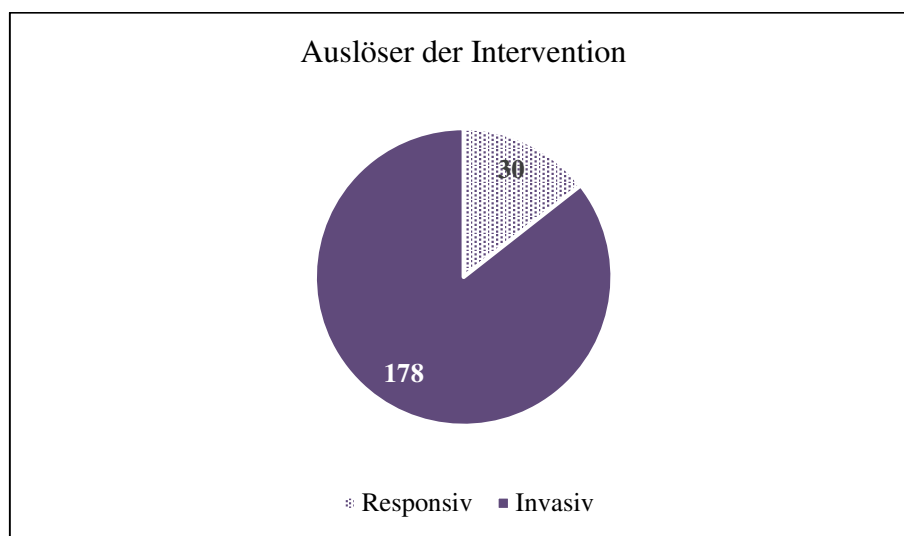


Abbildung 37: Verteilung der responsiven und invasiven Interventionen

Betrachtet man die einzelnen Interventionen genauer, so kann man die Teilgruppe der responsiven bzw. die Teilgruppe der invasiven Interventionen hinsichtlich des Arbeitsverhaltens der Schülerinnen und Schüler untersuchen. Zu diesem Zweck wurden in Tabelle 21 der Interventionsauslöser sowie die Anzahl der vor bzw. nach der Intervention aktiven Schülerinnen und Schüler aufgetragen. Man kann erkennen, dass bei responsiven Interventionen anschließend deutlich mehr Schülerinnen und Schüler aktiv sind. Während in 14 der 30 Gruppen alle Gruppenmitglieder aktiv sind, in deren Gruppenarbeitsprozess responsiv interveniert wurde, so ist dies nach der Intervention bei 23 der 30 Gruppen der Fall. Von den verbleibenden sieben Grup-

pen gibt es vier, in denen lediglich eines der Gruppenmitglieder nicht arbeitet. Vor der Intervention gab es noch einige Gruppen, in denen mindestens die Hälfte der Gruppe nicht aktiv war. Blickt man auf die invasiven Interventionen, so erkennt man, dass auch hier ein Zuwachs stattfindet, dieser jedoch relativ gesehen deutlich geringer ausfällt. Daraus lässt sich schlussfolgern, dass der Großteil der invasiven Interventionen nicht zur Aktivierung der Schülerinnen und Schüler beitragen konnte. An dieser Stelle muss jedoch auch erwähnt werden, dass der Anteil der Gruppen, bei denen bereits alle Schülerinnen und Schüler aktiv waren, recht hoch war.

Hinzu kommt, dass Schülerinnen und Schüler, die sich aktiv darum bemühen, dass sie Hilfe bekommen, vermutlich auch motivierter sind und entsprechend auch eine höhere Bereitschaft aufweisen, sich aktiv mit den Aufgaben bzw. den mathematischen Inhalten zu befassen. Entsprechend trägt dies wahrscheinlich auch dazu bei, dass responsive Interventionen dazu führen, dass die Schüleraktivität hoch bleibt oder sich erhöht. Daher lässt sich nicht direkt ableiten, dass responsive Interventionen absolut gesehen die besseren Interventionen sind. Auch die Voraussetzungen der Schülerinnen und Schüler, beispielsweise deren Motivation, sich mit den mathematischen Inhalten zu beschäftigen, spielen eine maßgebliche Rolle. Diese Voraussetzungen wurden im Rahmen der Studie jedoch nicht erhoben.

Tabelle 21: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern

Auslöser	Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention
Responsiv	4	14	23
	3	3	4
	2	7	1
	1	1	0
	0	3	1
Invasiv	4	117	129
	3	29	32
	2	9	6
	1	8	4
	0	11	1

Berücksichtigt man zusätzlich die Angemessenheit, so erhält man einen noch umfassenderen Einblick in die Gestaltung und Auswirkung der Interventionen. In Abbildung 38 ist auf der x-Achse die Angemessenheit und auf der y-Achse die absolute Anzahl an Interventionen zu sehen. In Lila sind die invasiven Interventionen abgebildet, die responsiven Interventionen in Form der gepunkteten Säulen. Erkennbar ist, dass es kaum responsive Interventionen gibt, die als unangemessen bewertet wurden. Insgesamt war dies nur bei zwei Interventionen der Fall. Alle anderen, insgesamt 52 Interventionen, die als unangemessen eingestuft wurden, waren invasiv. In Anbetracht der 126 invasiven Interventionen, die als angemessen beurteilt wurden, kann man folgern, dass die invasiven Interventionen der Studierenden im Rahmen der Studie zwar überwiegend als angemessen beurteilt wurden, jedoch mehr als jede vierte invasive Intervention unangemessen war.

Dies mag auch daran liegen, dass Interventionen, die von den Schülerinnen und Schülern gefordert werden, fast zwangsläufig dazu führen, dass eine Intervention geboten ist. Natürlich

kann diese aufgrund von verschiedenen Mängeln trotzdem als unangemessen bewertet werden. Jedoch ist es im Regelfall nachvollziehbar, warum an dieser Stelle eine Intervention erfolgt.

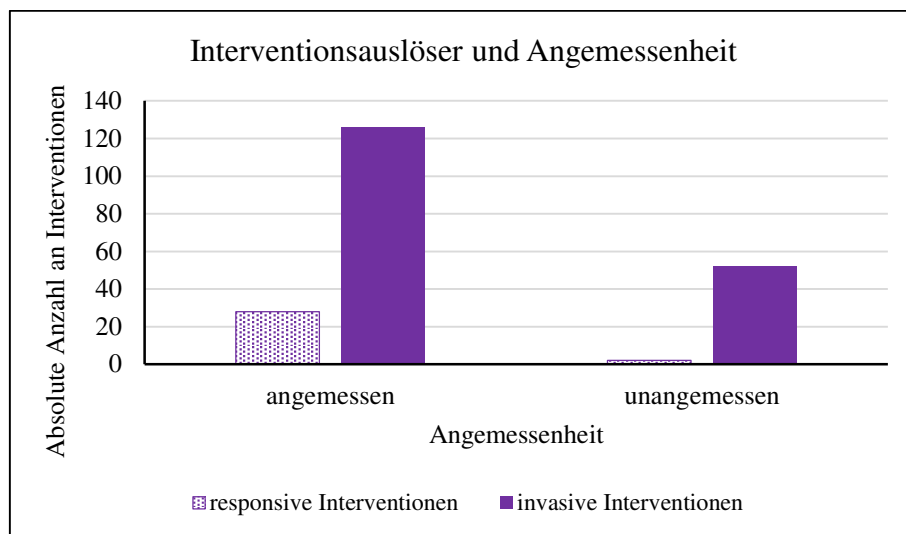


Abbildung 38: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers

Verbindet man den Befund, dass invasive Intervention häufiger unangemessen sind, mit dem Ergebnis, dass invasive Interventionen auch seltener dazu führen, dass anschließend mehr Schülerinnen und Schüler aktiv arbeiten als vor der Intervention, so kann man schlussfolgern, dass invasive Interventionen sparsamer durchgeführt werden sollten. Hingegen werden responsive Interventionen in den meisten Fällen als angemessen eingestuft und führen vergleichsweise häufiger dazu, dass nach der Intervention mehr Schülerinnen und Schüler aktiv sind als vor der Intervention. Entsprechend liegt es nahe, im Kontext des Mathematik-Labors „Mathe ist mehr“ tendenziell eher responsive Interventionen zu tätigen. Natürlich muss in jeder Situation eine Einzelfallentscheidung getroffen werden, die die Umstände, insbesondere auch die Aufgabenschwierigkeit und die Fähigkeiten der Schülerinnen und Schüler sowie deren aktuelle Problemlage, in die Überlegungen miteinbezieht.

9.2.6 Zusammenfassung

Die Studierenden tätigen am häufigsten allgemein-strategische und inhaltsorientiert-strategische Hilfen. Rückmeldehilfen kommen zwar ebenfalls oft vor, sind jedoch meistens nicht dominant. Inhaltliche Hilfen und disziplinarische Hilfen kommen seltener vor, stellen jedoch in den meisten Fällen die dominante Hilfe innerhalb der Intervention dar, sofern sie gegeben werden. Motivationshilfen kommen kaum vor, weder als dominante noch als untergeordnete Hilfe. Dass inhaltsorientiert-strategische Hilfen am häufigsten vorkommen, könnte darin begründet sein, dass die Studierenden wollen, dass die Schülerinnen und Schüler alle Aufgaben in der vorgesehenen Zeit bearbeiten. Entsprechend steigen sie mit einer vergleichsweise inhaltlicheren Hilfe ein, statt die Schülerinnen und Schüler zu motivieren und ihnen eine Rückmeldung zu geben. Ein weiterer Grund für die wenigen Motivationshilfen und Rückmeldehilfen könnte darin bestehen, dass die Schülerinnen und Schüler den Studierenden nicht vertraut sind. Entsprechend besteht in manchen Fällen womöglich eine Hemmschwelle, den Raum, in dem der Studierende sitzt, zu verlassen und den Raum, in dem die Schülergruppe arbeitet, zu betreten, nur um diese zu loben oder um dieser eine Rückmeldung zu geben.

Solche Hilfen dauern jedoch deutlich kürzer und würden den Schülerinnen und Schülern mehr Zeit für die eigenständige Auseinandersetzung mit den mathematischen Inhalten geben als andere Hilfen. Denn je inhaltlicher die Hilfen sind, desto mehr Zeit beanspruchen sie. Inhaltliche Hilfen dauern am längsten, gefolgt von inhaltsorientiert-strategischen Hilfen. Betrachtet man nur die dominanten Hilfen, hat dieser Zusammenhang in gleicher Form Bestand. Betrachtet man nur die untergeordneten Hilfen, zeigt sich, dass die inhaltlichen Hilfen kürzer dauern als die inhaltsorientiert-strategischen Hilfen. Dies liegt vermutlich daran, dass inhaltliche Hilfen schnell dominant werden, wenn sie lange dauern.

Betrachtet man die Reihenfolge der gegebenen Hilfen während der Interventionen, erkennt man eine Tendenz dahingehend, dass zunächst weniger inhaltliche Hilfen gegeben werden. Dies wird auch durch die zuerst verwendeten Hilfen innerhalb der Interventionen bestätigt. Auch dort werden zunächst weniger inhaltliche Hilfen genutzt. Daraus folgt, dass die Studierenden tendenziell eher niederschwelligere Hilfen nutzen und erst im Laufe der Intervention eine vergleichsweise inhaltlichere Hilfe verwenden. Dies ist insofern wünschenswert, als die Schülerinnen und Schüler eigenständig arbeiten sollen und nur dann Interventionen erfolgen sollen, wenn diese nach individueller Einschätzung der Studierenden notwendig ist. Eine solche Intervention sollte den Lernenden nicht alle Lerngelegenheiten nehmen, sondern sie unterstützen, sodass sie im Rahmen ihrer Möglichkeiten maximal selbstständig arbeiten können. Die Studierenden scheinen dies zu befolgen, indem sie zunächst weniger inhaltliche Hilfen geben und dann bei Bedarf darüber hinaus eine Hilfe mit größerem inhaltlichen Bezug erteilen.

Im Hinblick auf die Angemessenheit der Interventionen lässt sich festhalten, dass der überwiegende Teil der Interventionen als angemessen bewertet wurde. Dies gilt jedoch nicht für inhaltliche Hilfen, bei denen nur etwa zwei von fünf Interventionen angemessen durchgeführt wurden. Das kann daran liegen, dass diese Interventionsart den Studierenden aufgrund fehlenden fachlichen Wissens Probleme bereitet. Es könnte aber auch sein, dass inhaltliche Hilfen von den Expertinnen und Experten besonders kritisch beurteilt wurden, da inhaltliche Fehler womöglich häufig dazu führen, dass eine Intervention insgesamt als unangemessen bewertet wird. Betrachtet man die Kriterien, die die Expertinnen und Experten am häufigsten als Stütze für die Bewertung der Angemessenheit angeführt haben, so erkennt man, dass viele Interventionsarten unpassend gewählt wurden, dass die Intensität der Interventionen häufig zu niedrig oder zu hoch war und dass die Studierenden einige inhaltliche Fehler machten. Letztere wurden besonders häufig bei inhaltlichen Hilfen beanstandet.

Die Schüleraktivität vor den getätigten Interventionen war in den meisten Fällen schon recht hoch, beim Großteil der Gruppen arbeiteten alle Schülerinnen und Schüler bereits vor der Intervention. Entsprechend führen die Interventionen insgesamt gesehen nur zu einem kleinen Zuwachs an Schüleraktivität. Bei angemessenen Interventionen ist der Zuwachs etwa doppelt so hoch wie bei unangemessenen Interventionen. Entsprechend kann dies als Stütze für die Bewertung der Angemessenheit der Expertinnen und Experten gesehen werden.

Betrachtet man die Interventionsauslöser für die Interventionen, so erkennt man, dass die Masse an Interventionen invasiv war und nur recht wenige responsive Interventionen getätigt wurden. Dies liegt vermutlich am Setting des Mathematik-Labors. Die Schülergruppe ist dazu angehalten, Probleme in der Gruppe zu besprechen und das Hilfeheft zu nutzen. Erst danach soll eine betreuende Person aufgesucht und befragt werden. Dadurch, dass sie permanent von einem Studierenden per Videoübertragung beobachtet werden, ist es naheliegend, dass dieser interveniert, bevor die Lernenden eine Intervention einfordern. Die Erfahrung im Mathematik-Labor „Mathe ist mehr“ zeigt, dass sich dies anders verhält, wenn die Studierenden sich im gleichen Raum befinden, in dem die Schülerinnen und Schüler arbeiten. Durch die räumliche

Nähe sinkt die Hemmschwelle nachzufragen, wodurch deutlich häufiger responsive Interventionen entstehen.

Responsive Interventionen sind in deutlich höherem Maß angemessen als invasive Interventionen. Daher könnte man folgern, dass im Setting des Mathematik-Labors tendenziell eher responsive Interventionen getätigt werden sollten. Diese tragen auch in höherem Maße dazu bei, dass die Schülerinnen und Schüler nach der Intervention aktiv sind. Angesichts der Prämisse, dass die Schülerinnen und Schüler sich die Lerninhalte nach Möglichkeit selbstständig erarbeiten sollen, wäre es wünschenswert, wenn mehr Motivationshilfen gegeben und Rückmeldehilfen häufiger als dominante Hilfe genutzt werden würden.

Somit wurde das Interventionsverhalten der Studierenden aus verschiedenen Blickwinkeln analysiert und mit einer wünschenswerten Praxis kontrastiert. Im nächsten Teilkapitel wird nun vorgestellt, wie die Studierenden über die getätigten Interventionen sowie über die Erprobung der Laborstation im Allgemeinen reflektieren.

9.3 Reflexion

Die 17 Reflexionsgespräche der 46 Studierenden dauerten zwischen 38 Minuten und 143 Minuten (zwei Stunden und 23 Minuten). Insgesamt dauerten die Reflexionsgespräche 23 Stunden und 6 Minuten. In Abbildung 39 kann man erkennen, dass dabei überwiegend im Bereich „Rückblick“, also auf einer niederschweligen Ebene, reflektiert wurde. Mit 2067 Äußerungen wurden in diesem Bereich mit Abstand die meisten Reflexionsmomente festgestellt. Wesentliche Aspekte werden hingegen kaum als solche reflektiert. Mit lediglich 37 Reflexionsmomenten wurden im Schnitt kaum mehr als zwei wesentliche Aspekte pro Gruppe herausgearbeitet. Dazwischen, aber mit großem Abstand zur Kategorie „Rückblick“, liegt die Kategorie „Alternativen“. Die Studierenden äußerten in 469 Fällen mögliche Alternativen für die Gestaltung der Lernumgebung bzw. für die getätigten Interventionen.

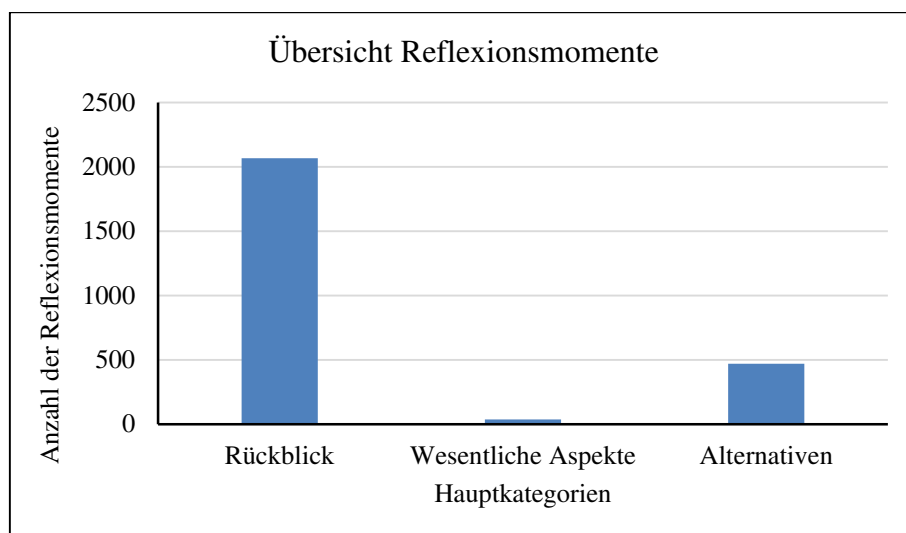


Abbildung 39: Übersicht über die Reflexionsmomente aller Studierenden hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen"

Die Übersicht über die Reflexionsmomente, geordnet nach den drei Hauptkategorien, liefert einen groben Überblick über die Reflexionsgespräche der Studierenden. Da alle Kategorien in

mehrere Unterkategorien aufgeteilt wurden, ist auch ein tieferer Einblick in die Reflexionsgespräche möglich. Die jeweilige Übersicht zu den Reflexionsmomenten innerhalb der Kategorien „Rückblick“, „Wesentliche Aspekte“ und „Alternativen“ findet sich in den Tabellen 22 bis 24 wieder.

Anhand von Tabelle 22 wird ersichtlich, dass der Großteil der Reflexionsmomente in der Kategorie „Rückblick“ im Bereich „Inhaltliches und Organisatorisches“ liegt. Mit dieser Unterkategorie wurden 1380 Reflexionsmomente von den Studierenden kodiert. Die Unterkategorien „Gedanken und Gefühle“ mit 265 Äußerungen, „Grober Rückblick“ mit 311 Äußerungen und „Absichten und Ziele“ mit 111 Äußerungen komplettieren die insgesamt 2067 Reflexionsmomente in der Kategorie „Rückblick“.

Tabelle 22: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“

Unterka- tegorie	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick	Summe
Anzahl	111	1380	265	311	2067

In der Kategorie „Absichten und Ziele“ wurden beispielsweise diese Äußerungen kodiert:

- „An dieser Stelle war es nach wie vor unruhig, es ging um andere Sachen als um Mathematik. Man hat gesehen, dass es nicht an inhaltlichen Dingen gelegen hat. Die Schüler haben nichts hinbekommen, weil sie sich zu sehr haben ablenken lassen. Daher ging es mir in erster Linie darum, disziplinarisch einzugreifen, und nicht fachlich.“
- „Das Ziel der Station war nicht, dass sie das Ziegenproblem lösen können, sondern, da muss ich dir wiederum zustimmen, dass sie mit Baumdiagrammen umgehen können. Es ging eher darum, dass sie die Pfadregeln anwenden können und verstehen.“
- „Das Ziel dieses Heftes ist es ja, dass sie im Sinne des Modellierens die Sachen anwenden.“
- „Zuvor haben sie die Simulation nicht gefunden. Hier hatten sie die Simulation zwar, aber sind inhaltlich nicht weitergekommen. Die Simulation hat ihnen nicht geholfen, da sie die Hilfen nicht verwendet haben, die nebendran noch aufgeführt waren. Das haben sie irgendwie übersehen oder nicht registriert, dass man die noch verwenden kann. Deswegen habe ich hier eingegriffen, um ihnen mitzuteilen, dass sie die Hilfen nutzen können.“

In den Äußerungen der Studierenden wird jeweils ein konkretes Ziel benannt. Entweder bezieht sich dieses auf die Laborstation oder auf die getätigten Interventionen. Es ist zu erkennen, dass die Studierenden ein bestimmtes Ziel verfolgen bzw. eine konkrete Absicht haben. Im Fall der Reflexionsmomente, die sich auf Gedanken und Gefühle beziehen, kann zwischen den Gedanken und Gefühlen der Studierenden und den Gedanken und Gefühlen der Lernenden unterschieden werden. Exemplarisch dafür sind folgende Äußerungen:

- „Die Aufgabe ging dann doch ganz gut. Ich hatte vorher gedacht, dass sie eine totale Katastrophe wird und hatte ein bisschen Angst vor der Durchführung.“
- „Dass die Schüler/innen bei der Aufgabe solche Probleme haben würden, habe ich vorher nicht gedacht. Eigentlich dachte ich, dass sie schnell erledigt wäre.“

- „Ich habe eingegriffen, weil die Schüler/innen die Aufgabe nicht verstanden haben. Sie haben gedacht, dass sie nochmal ziehen müssen.“
- „Da sie so oft das Gleiche machen mussten, waren sie irgendwann davon gelangweilt und genervt.“

Auch hier ist zu erkennen, dass manche Aussagen sich auf die Lernumgebung beziehen, während andere sich auf die Interventionen beziehen. Entsprechend kann die Unterkategorie „Gedanken und Gefühle“ nochmals in vier weitere Unterkategorien unterteilt werden. Bei der Unterkategorie „Grober Rückblick“ wird lediglich zwischen Reflexionsmomenten, die sich auf die Laborstation und Reflexionsmomenten, die sich auf die Interventionen beziehen, unterscheiden. Äußerungen in der Unterkategorie „Grober Rückblick“ waren beispielsweise:

- „Heft eins und zwei haben sie ganz gut hinbekommen, bei Heft drei gab es allerdings Probleme.“
- „Man hat insgesamt gemerkt, dass die Schüler/innen wenig motiviert waren und viel über andere Dinge geredet haben.“
- „Ich würde sagen, bis auf die erste Intervention waren deine Interventionen immer sinnvoll und an der richtigen Stelle.“
- „Ich finde, du hast generell gut eingegriffen. Du hast die Schüler/innen immer selbst draufkommen lassen.“

Man kann erkennen, dass die Äußerungen recht grob ausfallen. So werden alle drei Teilstationen innerhalb kürzester Zeit abgehandelt oder eine Aussage über die Motivation der Schülerinnen und Schüler während der Bearbeitung der gesamten Station getroffen. Die Interventionen einer Kommilitonin bzw. eines Kommilitonen werden einer groben Bewertung unterzogen. Auch die inhaltlichen und organisatorischen Aspekte können sich sowohl auf die Lernumgebung als auch auf die Interventionen beziehen. Es handelt sich hierbei um konkrete Aufgaben („Das Umrechnen bei Aufgabe fünf hat gut geklappt, aber bei der Rechnung hatten sie Probleme.“) oder konkrete Aspekte der Intervention („Da der Ton der Simulation nicht funktioniert hat, musste ich eingreifen. Dadurch hat Aufgabe vier etwas länger gedauert.“). Alle Unterkategorien der Kategorie „Rückblick“ sowie deren Unterkategorien, die eben vorgestellt wurden, bleiben relativ oberflächlich. Tiefere Reflexionen beinhaltet die Kategorie „Wesentliche Aspekte“.

Die Kategorie „Wesentliche Aspekte“ untergliedert sich in die Unterkategorien „Kritisches Hinterfragen“, „Organisatorisches“ und „Inhaltliches“ (vgl. Tab. 23). Das kritische Hinterfragen der eigenen Handlungen kam jeweils nur in Bezug auf die getätigten Interventionen vor, insgesamt war dies 13 Mal der Fall. „Organisatorisches“ wurde acht Mal als wesentlicher Aspekt kodiert, „Inhaltliches“ mit 16 Mal doppelt so oft.

Tabelle 23: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“

Unterkategorie	Kritisches Hinterfragen	Organisatorisches	Inhaltliches	Summe
Anzahl	13	8	16	37

Insgesamt gab es also nur sehr wenige Äußerungen der Studierenden, die der Kategorie „Wesentliche Aspekte“ bzw. einer passenden Unterkategorie zugeordnet wurden. Um einen Ein-

blick in die verschiedenen Aspekte zu erhalten, die von den Studierenden als wesentlich wahrgenommen und im Rahmen des Reflexionsgespräches kundgetan wurden, werden an dieser Stelle einige Äußerungen vorgestellt:

- „Vielleicht hätte ich einfach warten sollen, bis sie zu der nächsten Teilaufgabe gekommen wären. Da hätten sie ja beide verwenden müssen und bestimmt gemerkt, dass es nicht zweimal das Gleiche sein kann und hätten vielleicht nochmal genauer gelesen. Wenn sie es dann noch immer nicht verbessert hätten, wäre ich aber nochmal reingegangen.“
- „Ich würde da nicht mehr eingreifen. Sie hatten ja das Ergebnis raus. Wenn sie es dann nicht mehr in Quadratmeter umrechnen, dann rechnen sie es eben nicht um. Es war ja auch nicht Thema von unserem Heft. So haben wir uns eine Viertelstunde nur an der Umrechnung aufgehalten. Vielleicht würde ich auch in das Hilfeheft schreiben, wie die Umrechnung geht. Ich würde aber auf jeden Fall nicht mehr eingreifen, weil die Umrechnung wirklich nicht so wichtig ist.“
- „Ein wesentliches Problem bei der Bearbeitung von Heft zwei war, dass die Schülerinnen es direkt nach Heft eins bearbeiten mussten. Sie waren dann nicht mehr so konzentriert und hatten noch nicht die Möglichkeit das zuvor Gelernte zu festigen.“
- „Meiner Meinung nach war das fehlende Vorwissen in Bezug auf Terme im Allgemeinen ein wesentlicher Grund für die Probleme, die während der Erprobung aufgetreten sind.“

Bei den kodierten Äußerungen in der Kategorie „wesentliche Aspekte“ handelt es sich unter anderem um Reflexionsmomente, die sich auf das kritische Hinterfragen der eigenen Person beziehen, insbesondere dann, wenn Interventionen künftig nicht mehr getätigt würden. Außerdem wurden teilweise organisatorische Aspekte angesprochen. In dem angegebenen Fall wurde kritisiert, dass die Laborstation für drei Mal 90 Minuten konzipiert wurde und diese im Regelfall an drei unterschiedlichen Tagen durchgeführt werden. Bei dieser Gruppe kam die Schulklasse aber an einem Tag, um die ganze Station zu bearbeiten. Entsprechend wurde die zweite Teilstation direkt im Anschluss an die erste Teilstation bearbeitet. Dies wurde von dem Studierenden als wesentlicher Aspekt formuliert, weshalb die Erprobung nicht wie geplant verlief. Eine weitere Äußerung bezieht sich auf wesentliche Inhalte. Dabei wird bemängelt, dass die Schülerinnen und Schüler nicht über das notwendige Vorwissen verfügten, um mit den Aufgaben zurechtzukommen.

Es zeigen sich also verschiedene Möglichkeiten, wesentliche Aspekte herauszuarbeiten und im Rahmen des Reflexionsgespräches zu äußern. Die niedrige Anzahl an Reflexionsmomenten in diesem Bereich liegt daher weniger an der strengen Kodieranleitung als vielmehr daran, dass die Studierenden größtenteils auf der Oberflächenebene blieben oder direkt alternative Vorschläge zur Gestaltung der Laborstation oder den getätigten Interventionen machten.

Innerhalb der Kategorie „Alternativen“ wurde zwischen der Nennung von Alternativen und der Abwägung von Vor- und Nachteilen unterschieden (vgl. Tab. 24). Man kann erkennen, dass hauptsächlich Alternativen genannt werden (415) und verhältnismäßig selten Vor- und Nachteile abgewogen werden (54).

Tabelle 24: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“

Unterkategorie	Nennung von Alternativen	Abwägung Vor- und Nachteile	Summe
Anzahl	415	54	469

Bei der Nennung möglicher Alternativen kann zwischen Alternativen bezüglich der Lernumgebung, also möglichen Änderungen der Aufgabenstellungen, des Materials oder der Simulationen, und alternativen Interventionsmöglichkeiten unterschieden werden. Gleiches gilt entsprechend auch für die Abwägung von Vor- und Nachteilen. Einen qualitativen Einblick gewähren folgende Äußerungen:

- „Wir hätten im Hilfeheft eine Hilfe ergänzen können und erklären können, was ähnliche Dreiecke überhaupt sind.“
- „Das nächste Mal würde ich sie selbst lesen lassen und ihr die Frage nicht sofort beantworten.“
- „Auf der einen Seite fand ich, dass die Schüler bei Aufgabe zwei zu viele Teilaufgaben derselben Form rechnen mussten und dadurch genervt waren, aber auf der anderen Seite hat man gemerkt, dass sie zunächst viele Fehler gemacht haben und es danach dann saß.“
- „Einerseits finde ich es sinnvoll, wenn die Schüler nicht mehr viel Zeit haben, einzugreifen und ihnen einen kleinen Tipp zu geben, andererseits finde ich es aber generell besser, wenn man ihnen nur dabei hilft, selbst auf den richtigen Weg zu kommen. Dann bekommen sie vielleicht nicht mehr alles hin, aber ihr persönlicher Lernzuwachs ist vielleicht trotzdem besser, weil sie nicht alles vorgesagt bekommen haben und das, was sie letztendlich gelernt haben, auch richtig sitzt.“

Die Kategorien wurden in Unterkategorien unterteilt, wobei diese nochmals zwischen Äußerungen, die sich auf die Lernumgebungen beziehen, und Äußerungen, die sich auf die Interventionen beziehen, unterschieden wurden. In Tabelle 25 wird diese Unterscheidung vorgenommen und die entsprechenden Häufigkeiten der jeweils getätigten Reflexionsmomente werden aufgelistet. Man kann erkennen, dass die Studierenden in allen drei Kategorien häufiger über die Lernumgebung reflektieren als über die Interventionen. Absolut ist der Unterschied in der Kategorie Rückblick mit 1308 zu 759 am größten. Relativ betrachtet ist der Unterschied in der Kategorie „Alternativen“ mit 397 zu 72 am größten. Am geringsten fällt der Unterschied in der Kategorie „Wesentliche Aspekte“ aus. Hier wurden hinsichtlich der Lernumgebung 22 Reflexionsmomente und hinsichtlich der getätigten Interventionen 15 Reflexionsmomente erfasst.

Dass insgesamt mehr Äußerungen zur Lernumgebung getätigt wurden, liegt vermutlich auch daran, dass jeweils zwei der drei Studierenden im Mathematik-Labor die restliche Klasse betreuten, während ein/e Studierende/r für die Videogruppe zuständig war. Entsprechend müssten auch etwa doppelt so viele Aspekte von den zwei Studierenden im Mathematik-Labor wahrgenommen werden als von dem dritten Gruppenmitglied, das für die Videogruppe zuständig ist. Dadurch, dass im Mathematik-Labor je nach Klassengröße zwischen vier und sechs Gruppen arbeiten, könnte der Effekt sogar verstärkt werden, da die Studierenden die Möglichkeit haben, mehrere Gruppen, deren Lernverlauf sowie auftretende Probleme und Schwierigkeiten zu beobachten. Daraus könnte identifiziert werden, wo wesentliche Hürden für die Schülerinnen und Schüler sind, um daraus Handlungsalternativen abzuleiten.

Tabelle 25: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien hinsichtlich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“

Kategorie	Lernumgebung	Intervention
Rückblick	1308	759
Wesentliche Aspekte	22	15
Alternativen	397	72

Tabelle 26 zeigt eine Übersicht über alle Reflexionsmomente, aufgeteilt nach den Studierendengruppen, wobei diese nach der jeweiligen Laborstation bezeichnet wurden, und den drei Hauptkategorien. Darüber hinaus ist der Tabelle zu entnehmen, wie sich die Reflexionsmomente auf die Studierenden innerhalb der Gruppe verteilen. Durch die Aufteilung nach Studierenden und nach Inhalt sind die Reflexionsmomente in der Spalte „Anzahl Codes“ doppelt vertreten. Die Summe über die Reflexionsmomente der Studierenden S1 (Person, die beim Reflexionsgespräch links sitzt), S2 (Person, die rechts neben S1 sitzt) und S3 (Person, die rechts neben S2 sitzt) ist wenig aussagekräftig, da es sich jeweils um unterschiedliche Personen handelt. Betrachtet man die Anzahl der Codes, so sieht man, dass die einzelnen Gruppen zwischen 84 (Zuckerwürfel b) und 276 (Geburtstagsparty b) Reflexionsmomente äußerten.

Auf individueller Ebene kann man erkennen, dass die minimale Anzahl an Reflexionsmomenten (18) von S3 der Gruppe Ziegenproblem a getätigt wurde. Am zweitwenigsten (23) äußerten sich zwei Studierende, S1 der Gruppe Zuckerwürfel b und S1 der Gruppe Aktivurlaub. Die meisten Reflexionsmomente äußerten S1 (106) und S2 (109) der Gruppe Maimarkt 2016 b. Ebenfalls über 100 Äußerungen, die als reflexive Momente kodiert wurden, tätigten S3 der Gruppe Geburtstagsparty b (104) und S1 der Gruppe Maimarkt 2018 b (103).

Betrachtet man die drei Hauptkategorien, so lassen sich ebenfalls Unterschiede zwischen den verschiedenen Gruppen feststellen. Hinsichtlich der Kategorie „Rückblick“ wurden in jedem Gruppenreflexionsgespräch mindestens 69 und maximal 210 Reflexionsmomente identifiziert. Sechs Gruppen ist es nicht gelungen, über einen wesentlichen Aspekt zu reflektieren. In diesen Gruppen wurden folglich keinerlei organisatorische oder inhaltliche Aspekte hinsichtlich der Lernumgebung oder der Interventionen ausfindig gemacht bzw. im Reflexionsgespräch diskutiert. Das eigene Handeln wurde also nicht weiter kritisch hinterfragt. Bei der Gruppe Drei-Türen-Problem konnten mit sieben die meisten wesentlichen Aspekte erfasst werden. Im Vergleich zu den anderen Gruppen ebenfalls viele wesentliche Aspekte wurden von der Gruppe Geburtstagsparty b (6) sowie den Gruppen Jakobsstab b und USA (jeweils 5) angesprochen. Bei der Kategorie „Alternativen“ findet sich eine große Bandbreite an Reflexionsmomenten. Mit fünf Alternativen, die sowohl alle genannten Alternativen mit Blick auf die Lernumgebung als auch auf die getätigten Interventionen beinhalten, nannten die beiden Gruppen Zuckerwürfel a und Maimarkt 2016 a die wenigsten Alternativen. Die Gruppe Maimarkt 2016 b konnte mit 63 die meisten Alternativen erarbeiten. Fast genauso viele Alternativen diskutierte die Gruppe Geburtstagsparty b mit 62 Alternativen.

Mit Blick auf alle drei Kategorien lässt sich erkennen, dass es Gruppen gibt, die keine wesentlichen Aspekte herausgearbeitet haben und trotzdem über 20 Alternativen benennen konnten. Es gibt außerdem Gruppen, die in allen Kategorien viele Reflexionsmomente tätigten bzw. in allen Bereichen verhältnismäßig wenige. Dazwischen existiert eine Bandbreite an Gruppen, die unterschiedliche Reflexionsmuster aufweisen.

Tabelle 26: Übersicht über die Reflexionsmomente aufgeteilt auf die Gruppen von Studierenden und die Hauptkategorien „Rückblick“, „Wesentliche Aspekte“ und „Alternativen“

Station	Anzahl Codes	Codes S1	Codes S2	Codes S3	Rückblick	Wesentliche Aspekte	Alternativen
Tatort Tankstelle	306	74	79	-	132	0	21
Zuckerwürfel a	208	55	49	-	96	3	5
Zuckerwürfel b	168	23	32	29	76	0	8
Aktivurlaub a	278	60	31	48	111	0	28
Aktivurlaub b	264	23	53	56	113	0	19
USA	288	66	52	26	120	5	19
Maimarkt 2016 a	192	36	29	31	90	1	5
Maimarkt 2016 b	430	106	109	-	149	3	63
Drei Türen Problem	230	46	40	29	77	7	31
Jakobsstab b	272	34	47	55	93	5	38
Ziegenproblem a	230	55	42	18	85	1	29
Ziegenproblem b	214	44	63	-	69	2	36
Geburtstagsparty a	432	81	82	53	184	3	29
Geburtstagsparty b	552	85	87	104	208	6	62
Löffelliste	342	92	79	-	156	0	15
Maimarkt 2018 a	244	44	47	31	98	0	24
Maimarkt 2018 b	496	103	83	62	210	1	37
Summe	5146	1027	1004	542	2067	37	469

9.4 Zusammenhänge zwischen Diagnose und Intervention

Nachdem bisher alle drei Bereiche einzeln betrachtet wurden, werden in diesem Teilkapitel die diagnostischen Fähigkeiten der Studierenden mit den getätigten Interventionen in Beziehung gesetzt, um mögliche Zusammenhänge beider Bereiche aufzudecken. Dazu wird die Einteilung der drei Gruppen hinsichtlich der diagnostischen Fähigkeit „Deuten“ aus Teilkapitel 9.1 zugrundegelegt, um die Interventionen der Studierenden zu analysieren.

Zunächst werden die dominanten und die untergeordneten Hilfen betrachtet, auf die die drei Gruppen im Rahmen der Interventionen zurückgegriffen haben. Dabei wird auch der zeitliche Aspekt in die Analysen miteinbezogen. Anschließend richtet sich der Fokus auf die Reihenfolge der Hilfen. Neben den am häufigsten auftretenden Reihenfolgen wird auch die innerhalb der Intervention zuerst genutzte Hilfe-Kategorie vorgestellt. Hinsichtlich der Angemessenheit werden zunächst alle Interventionen der drei Gruppen betrachtet und anschließend nach den einzelnen Hilfe-Kategorien aufgeschlüsselt. Dabei finden zum einen die von den Expertinnen und Experten beanstandeten Kriterien der Angemessenheit Berücksichtigung und zum anderen werden die häufigsten Kriterien für alle Interventionsarten einzeln analysiert. Des Weiteren wird die Anzahl der aktiven Schülerinnen und Schüler vor bzw. nach der Intervention für alle drei Gruppen durchleuchtet. Dabei wird auch zwischen angemessenen und unangemessenen Interventionen unterschieden. Abschließend wird der Interventionsauslöser in die Auswertung der Interventionen miteinbezogen. Dieser wird auch mit der Angemessenheit und der Anzahl der aktiven Schülerinnen und Schüler in Verbindung gebracht, sodass sich insgesamt eine vielschichtige Analyse der Interventionen der einzelnen Gruppen ergibt, anhand der mögliche Zusammenhänge zwischen der diagnostischen Fähigkeit „Deuten“ und den getätigten Interventionen aufgedeckt werden können.

9.4.1 Hilfen innerhalb der Intervention

Im Hinblick auf die genutzten Hilfen innerhalb der Interventionen ist zunächst darauf hinzuweisen, dass die drei Gruppen (mit jeweils 16 Studierenden) eine unterschiedliche Anzahl an Interventionen getätigt haben. Die Gruppe der Studierenden, die im Test eine verhältnismäßig gute prozessdiagnostische Fähigkeit „Deuten“ zeigen konnte, hat 65 Interventionen getätigt. Entsprechend gibt es auch 65 dominante Hilfen bei dieser Gruppe. Die Gruppe, die im Diagnostest verhältnismäßig schlecht abgeschnitten hat, hat mit 64 Interventionen eine Intervention weniger getätigt. Somit wurden bei dieser Gruppe 64 dominante Hilfen ausfindig gemacht. Die mittlere Gruppe hat am häufigsten interveniert, insgesamt 79 Mal.

Zusätzlich zu den dominanten Hilfen, von denen es jeweils nur genau eine pro Intervention gibt, hat die Gruppe der guten Diagnostiker (gemeint ist damit die hohe Übereinstimmung der gegebenen Antworten mit den Antworten der Expertinnen und Experten im Diagnostest, der mittels Videovignetten erhoben wurde) 56 untergeordnete Hilfen innerhalb der 65 Interventionen getätigt (insgesamt 122 Hilfen). Die Gruppe mit niedriger Ausprägung der prozessdiagnostischen Fähigkeit „Deuten“ hat 55 untergeordnete Hilfen in den 64 Interventionen erteilt (insgesamt 120 Hilfen). Somit wären diese beiden Gruppen in absoluten Zahlen gut miteinander zu vergleichen, da die absolute Anzahl der dominanten und untergeordneten Hilfen sich jeweils nur um 1 unterscheidet. Die mittlere Gruppe hat jedoch 64 untergeordnete Hilfen in den 79 Interventionen gegeben (insgesamt 143 Hilfen). Entsprechend würde ein Vergleich der drei Gruppen in absoluten Zahlen ein verzerrtes Bild liefern, da die Gruppe über 20 Hilfen mehr gegeben hat als die anderen beiden Gruppen. Daher werden die Gruppen nachfolgend nicht mit

dem absoluten Anteil, sondern mit dem relativen Anteil verglichen. Bezüglich der verschiedenen Hilfe-Kategorien bedeutet das, dass der Bezugswert jeder Gruppe die Anzahl der getätigten Hilfen innerhalb der eigenen Gruppe ist. Entsprechend ergibt die Summe aller sechs Hilfe-Kategorien bei allen drei Gruppen jeweils 100%. Dadurch wird besser ersichtlich, welche Hilfe-Kategorien innerhalb der jeweiligen Gruppe verhältnismäßig häufig bzw. selten genutzt wurden. Dies gilt sowohl für die dominanten als auch für die untergeordneten Hilfen sowie für die Summe der beiden.

Abbildung 40 zeigt die dominanten Hilfen innerhalb der Intervention (unterteilt nach den drei Gruppen). Auf der x-Achse ist die relative Häufigkeit aufgetragen. Auf der y-Achse sind die sechs verschiedenen Hilfe-Kategorien aufgelistet. Dabei befinden sich rechts von jeder Hilfe-Kategorie jeweils drei Balken, die für den relativen Anteil dieser Hilfe-Kategorie bei den verschiedenen Gruppen stehen. Die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ ist mit vollständig ausgefüllten Balken zu sehen. Die gestreiften Balken bilden die mittlere Gruppe ab, die unausgefüllten Balken die Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“. Hinter jedem Balken ist auch die absolute Anzahl an derartig getätigten Hilfen angegeben.

Betrachtet man die dominanten Hilfen, so ist erkennbar, dass die Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“ am häufigsten inhaltsorientiert-strategische Hilfen nutzt, wohingegen die Gruppe mit hoher Ausprägung am häufigsten allgemein-strategische Hilfen erteilt. Bei der mittleren Gruppe sind diese beiden Hilfe-Kategorien auch am häufigsten die dominanten Hilfen innerhalb der Intervention, wobei die beiden Hilfe-Kategorien in etwa gleich oft als dominante Hilfe erteilt wurden. Auffallend ist, dass die Gruppe mit niedriger Ausprägung verhältnismäßig weniger allgemein-strategische Hilfen gegeben hat als die anderen beiden Gruppen, insbesondere als die Gruppe mit hoher Ausprägung. Bei den inhaltlichen Hilfen ist der Anteil der Gruppe mit hoher Ausprägung etwas höher als bei den anderen, wobei dieser Unterschied so gering ist, dass er auch eine zufällige Schwankung darstellen könnte. Gleiches gilt für die Rückmeldehilfen, bei denen die Gruppe mit niedriger Ausprägung einen leicht höheren relativen Anteil im Vergleich zu den anderen beiden Gruppen aufweist. Motivationshilfen wurden von allen drei Gruppen so selten getätigt, dass darüber keine darüber hinausgehende Aussage getroffen werden kann. Interessant ist, dass die guten Diagnostiker kaum auf disziplinarische Hilfen zurückgreifen, während die anderen beiden Gruppen dies etwas häufiger tun (wobei der relative Anteil bei beiden Gruppen insgesamt noch recht gering ist).

An dieser Stelle wird nun eine erste, (aufgrund der geringen Fallzahl) vorsichtige Interpretation der Verteilung der Hilfen vorgenommen. Dass Studierende mit erhöhter prozessdiagnostischer Fähigkeit „Deuten“ seltener auf disziplinarische Hilfen zurückgreifen müssen, könnte daran liegen, dass sie besser erkennen, wo genau die Probleme und Schwierigkeiten der Schülerinnen und Schüler liegen. Entsprechend können diese (frühzeitig) durch passende Hilfen ausgeräumt werden. Dadurch kommt es seltener dazu, dass die Schülergruppe sich nicht der eigentlichen Aufgabe widmet und diszipliniert werden muss. Der erhöhte Anteil der allgemein-strategischen Hilfen bei der Gruppe mit erhöhter prozessdiagnostischer Fähigkeit „Deuten“ könnte so interpretiert werden, dass die Studierenden es den Schülerinnen und Schüler in folgenden Anforderungssituationen ermöglichen wollen, diese selbstständig zu bewältigen. Dies erfordert jedoch die Kenntnis darüber, was genau das Problem ist bzw. wo die aktuelle Schwierigkeit liegt. Kann dies diagnostiziert werden, reicht ggf. auch eine allgemein-strategische Hilfe aus und es muss keine inhaltsorientiert-strategische Hilfe erteilt werden. Mit der gleichen Argumentation könnte man den hohen Anteil an inhaltsorientiert-strategischen Hilfen bei der Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“ erklären. Dadurch,

dass sie nicht genau feststellen (können), wo die aktuelle Schwierigkeit liegt, greifen sie sicherheitshalber zu einer Hilfe, die die Schülerinnen und Schüler zur Lösung der vorliegenden Aufgabe nutzen können. Dabei wäre es womöglich auch ausreichend (und langfristig zielführender), wenn die Hilfe sich weniger am Inhalt, sondern mehr an einer allgemeinen Vorgehensweise orientieren würde.

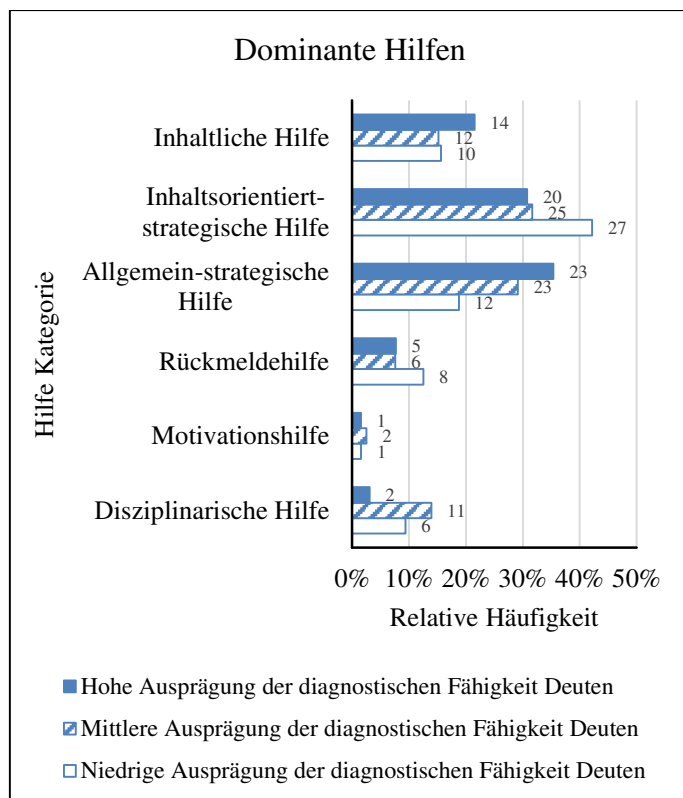


Abbildung 40: Verteilung der dominanten Hilfen innerhalb der Interventionen unterteilt nach Gruppen

Zusätzlich zu den dominanten Hilfen wurden in manchen Interventionen weitere Hilfen gegeben. Die untergeordneten Hilfen der drei Gruppen sind in Abbildung 41 zu sehen. Analog zur vorigen Abbildung sind auch hier die relativen Häufigkeiten (x-Achse) und die Hilfe-Kategorien (y-Achse) abgebildet.

Man kann erkennen, dass alle drei Gruppen am häufigsten Rückmeldehilfen als untergeordnete Hilfen nutzen. Die mittlere Gruppe nutzt inhaltsorientiert-strategische (19) und allgemein-strategische Hilfen (18) ebenfalls ähnlich häufig wie die Rückmeldehilfen (20). Bei der Gruppe der guten Diagnostiker ist der Anteil weniger ausgeglichen. Während etwa 40% der untergeordneten Hilfen Rückmeldehilfen darstellen, machen inhaltsorientiert-strategische mit unter 30% und allgemein-strategische mit unter 20% einen geringeren Anteil aus. Bei den Studierenden mit niedriger diagnostischer Fähigkeit „Deuten“ verhält es sich wiederum etwas ausgeglichener. Rückmeldehilfen und allgemein-strategische Hilfen werden in jeweils gut 30% der Fälle genutzt, inhaltsorientiert-strategische Hilfen in über 20% der Fälle.

An den relativen Anteilen der drei Hilfe-Kategorien lässt sich bereits erkennen, dass auf die anderen drei Hilfe-Kategorien kaum noch Anteile entfallen können. Diese liegen bei allen Gruppen jeweils unter 10%. Im Fall der Motivationshilfen und der disziplinarischen Hilfen verhält es sich in etwa ausgeglichen, wohingegen inhaltliche Hilfen (bis auf eine Ausnahme durch einen Studenten aus der mittleren Gruppe) nur von der Gruppe mit hoher Ausprägung als untergeordnete Hilfe genutzt werden.

Auch hier müssen die vorliegenden Unterschiede vorsichtig interpretiert werden. Rückmeldehilfen sind insbesondere im Setting des Lehr-Lern-Labors sinnvoll, da die Schülerinnen und Schüler grundsätzlich eigenständig arbeiten sollen. Eine Rückmeldehilfe kann ihnen hinsichtlich des Lösungswegs oder der Korrektheit einer Lösung Aufschluss geben. Diese Information nimmt den Schülerinnen und Schülern keine inhaltlichen (Teil-)Schritte vorweg. Dass alle drei Gruppen diese Hilfe am häufigsten als untergeordnete Hilfe genutzt haben, ist zunächst also positiv zu werten. Mit Blick auf die dominanten Hilfen (vgl. Abb. 40) wäre wünschenswert, dass Rückmeldehilfen noch häufiger als dominante Hilfe gegeben würden.

In Bezug auf die allgemein-strategischen Hilfen kann man erkennen, dass die Gruppe mit hoher Ausprägung diese verhältnismäßig seltener als untergeordnete Hilfe nutzt als die beiden anderen Gruppen. Dies liegt vermutlich daran, dass die Gruppe allgemein-strategischen Hilfen häufig als dominante Hilfe nutzt. Hinsichtlich der inhaltsorientiert-strategischen Hilfen lassen sich kaum Unterschiede zwischen den Gruppen feststellen. Dass inhaltliche Hilfen fast ausschließlich von der Gruppe der guten Diagnostiker als untergeordnete Hilfe eingesetzt werden, könnte wiederum daran liegen, dass diese die notwendigen fachlichen Inhalte besser diagnostizieren und entsprechend eine passgenaue kurze inhaltliche Hilfe ergänzend zu einer anderen (dominanten) Hilfe liefern können.

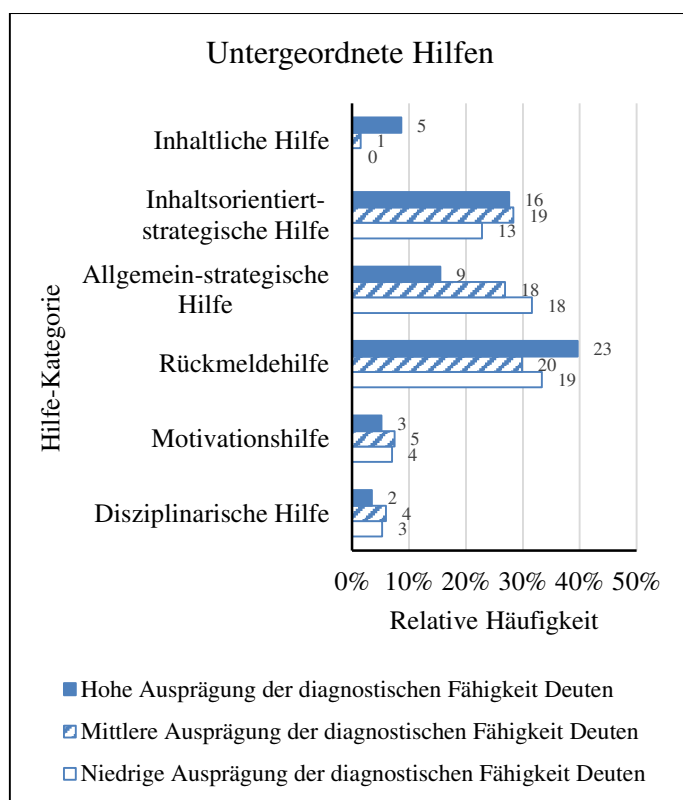


Abbildung 41: Verteilung der untergeordneten Hilfen innerhalb der Interventionen unterteilt nach Gruppen

Nachdem die dominanten Hilfen und die untergeordneten Hilfen einzeln betrachtet wurden, kann man durch die Summe der beiden einen Überblick erhalten, welche Gruppe welche Hilfe-Kategorie wie oft verwendet hat (vgl. Abb. 42). Zunächst fällt dabei auf, dass die drei Gruppen insgesamt recht ähnliche Hilfen gegeben haben. Etwa 30% entfallen auf inhaltsorientiert-strategische Hilfen, etwa 25% auf allgemein-strategische Hilfen und etwa 20% auf Rückmeldehilfen. Der Anteil der Motivationshilfen ist bei allen drei Gruppen sehr gering (jeweils unter 5%).

Kleine Unterschiede gibt es in Bezug auf die disziplinarischen Hilfen und die inhaltlichen Hilfen. Die Studierenden mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ tätigen tendenziell mehr inhaltliche Hilfen und tendenziell weniger disziplinarische Hilfen als die anderen beiden Gruppen.

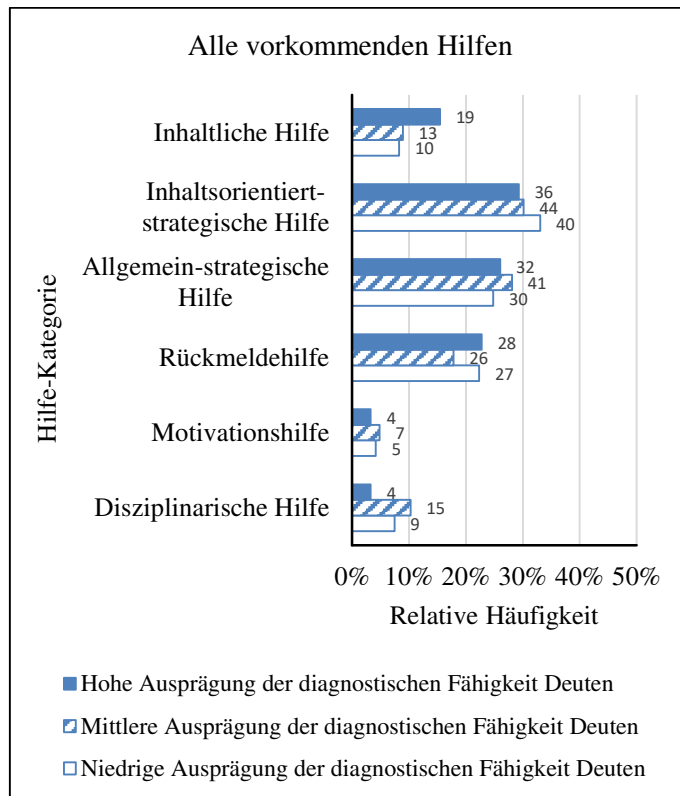


Abbildung 42: Verteilung aller vorkommenden Hilfen innerhalb der Interventionen unterteilt nach Gruppen

Interessant ist an dieser Stelle noch zu erwähnen, dass die Quote an gegebenen Hilfen pro Intervention bei allen drei Gruppen sehr ähnlich ist, alle drei Gruppen geben zwischen 1,8 und 1,9 Hilfen pro Intervention. Während mit Blick auf alle Hilfen kaum Unterschiede festgestellt werden konnten, konnten die Analysen der dominanten und der untergeordneten Hilfen zeigen, dass es unterschiedliche Schwerpunktsetzungen der Gruppen gibt. Um die verschiedenen Hilfen noch weiter zu durchleuchten, wird an dieser Stelle nun die Zeitdauer der jeweiligen Hilfen analysiert.

Tabelle 27 listet die Zeitdauern der verschiedenen Hilfe-Kategorien für die drei Gruppen auf. Neben der durchschnittlichen Dauer kann dabei jeweils auch die kürzeste und die längste Dauer der entsprechenden Hilfe-Kategorie entnommen werden. Zunächst wird auf die einzelnen Gruppen geblickt und die Zeitdauern der verschiedenen Hilfe-Kategorien werden analysiert. Dabei wird einzeln auf die durchschnittliche Dauer sowie auf die längste und kürzeste Dauer eingegangen. Danach werden diese in Beziehung zur Gruppe gesetzt, um ein Gesamtbild für jede Gruppe (einzeln) zu erhalten. Anschließend werden die drei Gruppen hinsichtlich der durchschnittlichen Dauer, der kürzesten und längsten Dauer verglichen und auch der Gesamteindruck der jeweiligen Gruppe wird zu dem Gesamteindruck der anderen Gruppen in Bezug gesetzt.

Tabelle 27: Zeitdauer der verschiedenen Hilfen mit Angabe der jeweilig längsten und der kürzesten Hilfe der jeweiligen Gruppen

Gruppe	Hilfe-Kategorie	Durchschnittliche Dauer (in Sekunden)	Kürzeste Dauer (in Sekunden)	Längste Dauer (in Sekunden)
<i>Hohe Ausprägung</i> der diagnostischen Fähigkeit „Deuten“	Inhaltliche Hilfe	88	5	358
	Inhaltsorientiert-strategische Hilfe	61	10	290
	Allgemein-strategische Hilfe	44	7	115
	Rückmeldehilfe	38	5	295
	Motivationshilfe	13	5	29
	Disziplinarische Hilfe	8	3	14
<i>Mittlere Ausprägung</i> der diagnostischen Fähigkeit „Deuten“	Inhaltliche Hilfe	104	10	290
	Inhaltsorientiert-strategische Hilfe	54	6	230
	Allgemein-strategische Hilfe	36	3	110
	Rückmeldehilfe	18	2	42
	Motivationshilfe	11	3	22
	Disziplinarische Hilfe	25	4	60
<i>Niedrige Ausprägung</i> der diagnostischen Fähigkeit „Deuten“	Inhaltliche Hilfe	63	10	270
	Inhaltsorientiert-strategische Hilfe	51	5	155
	Allgemein-strategische Hilfe	22	4	58
	Rückmeldehilfe	35	5	295
	Motivationshilfe	15	5	34
	Disziplinarische Hilfe	16	5	28

Die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ weist hinsichtlich der durchschnittlichen Länge der Hilfe-Kategorien die erwartete Progression auf. Disziplinarische Hilfen und Motivationshilfen dauern mit acht bzw. 13 Sekunden durchschnittlich am kürzesten. Rückmeldehilfen dauern im Schnitt 38 Sekunden und allgemein-strategische Hilfen 44 Sekunden. Länger dauern inhaltsorientiert-strategische Hilfen mit 61 Sekunden und inhaltliche Hilfen mit 88 Sekunden. Man kann für diese Gruppe also hinsichtlich der durchschnittlichen Dauer festhalten, dass mehr inhaltlicher Bezug mehr Zeit kostet.

Insbesondere bei den kürzesten, aber auch bei den längsten Hilfen findet sich diese Progression nur Ansatzweise. Am kürzesten war eine disziplinarische Hilfe mit einer Dauer von drei Sekunden. Eine Motivationshilfe, eine Rückmeldehilfe und eine inhaltliche Hilfe dauerten mit fünf Sekunden am zweitkürzesten. Mit sieben Sekunden bzw. zehn Sekunden ebenfalls kurz dauerten die kürzeste allgemein-strategische bzw. die kürzeste inhaltsorientiert-strategische Hilfe. Aufgrund der teilweise sehr geringen Fallzahlen sind diese jedoch mit Vorsicht zu interpretieren. Die längste Hilfe war eine inhaltliche Hilfe mit einer Dauer von knapp sechs Minuten.

Ebenfalls lange dauerten eine inhaltsorientiert-strategische Hilfe und eine Rückmeldehilfe (jeweils knapp fünf Minuten). Es folgen eine allgemein-strategische Hilfe mit knapp zwei Minuten, eine Motivationshilfe von 29 Sekunden und eine disziplinarische Hilfe von 14 Sekunden.

Bei der mittleren Gruppe ist derselbe Zusammenhang zwischen Inhalt und Zeitdauer der Hilfen erkennbar. Durchschnittlich dauerten inhaltliche Hilfen mit 104 Sekunden am längsten, gefolgt von inhaltsorientiert-strategischen Hilfen mit 54 Sekunden und allgemein-strategischen Hilfen mit 36 Sekunden. Am kürzesten dauerten Motivationshilfen (elf Sekunden) und Rückmeldehilfen (18 Sekunden). Etwas aus der Reihe liegen die disziplinarischen Hilfen, die durchschnittlich 25 Sekunden dauerten und somit mehr Zeit beanspruchten als Rückmeldehilfen.

Das gleiche Bild ergibt sich im Fall der längsten Hilfen. Eine inhaltliche Hilfe dauerte mit knapp fünf Minuten am längsten, gefolgt von einer inhaltsorientiert-strategischen Hilfe von etwas weniger als vier Minuten und einer allgemein-strategischen Hilfe von etwas weniger als zwei Minuten. Eine disziplinarische Hilfe dauerte eine Minute und war somit länger als die längste Rückmeldehilfe (42 Sekunden) und die längste Motivationshilfe (22 Sekunden). Die kürzesten Hilfen zeichnen wiederum ein undurchsichtigeres Bild, wobei sie zwischen zwei (Rückmeldehilfe) und zehn Sekunden (inhaltliche Hilfe) lagen.

Nimmt man die Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“ in den Blick, so erkennt man, dass es sich bei der längsten Hilfe mit durchschnittlich 63 Sekunden ebenfalls um eine inhaltliche Hilfe handelt. Ähnlich lang dauert in dieser Gruppe die inhaltsorientiert-strategische Hilfe mit 51 Sekunden. Am drittlängsten sind die Rückmeldehilfen (35 Sekunden), die somit länger andauern als die allgemein-strategischen Hilfen (22 Sekunden). Am kürzesten dauern Motivationshilfen mit 15 Sekunden und disziplinarische Hilfen mit 16 Sekunden.

Auch in dieser Gruppe liefert die Analyse der kürzesten Hilfen keinen erheblichen Erkenntnisgewinn. Außer der kürzesten inhaltlichen Hilfe mit einer Dauer von zehn Sekunden dauert keine der anderen kürzesten Hilfen über fünf Sekunden. Die kürzeste Hilfe stellt eine allgemein-strategische Hilfe mit einer Dauer von vier Sekunden dar. Bei den längsten Hilfen gibt es einen Ausreißer in Form einer fünfminütigen Rückmeldehilfe, der längsten Hilfe in dieser Gruppe. Ebenfalls lange dauert eine inhaltliche Hilfe mit einer Dauer von vier Minuten und 30 Sekunden. Am wenigsten lang sind die längste disziplinarische Hilfe mit 28 Sekunden und die längste Motivationshilfe mit 34 Sekunden. Die längste allgemein-strategische Hilfe dauerte knapp eine Minute. Auch bei dieser Gruppe ist grundsätzlich die Koppelung von Inhalt und Zeitdauer erkennbar, wenngleich es insbesondere in Gestalt der langen Rückmeldehilfe einen Ausreißer gibt, der auch dafür sorgt, dass die durchschnittliche Dauer der Rückmeldehilfen länger ist als die der allgemein-strategischen Hilfen.

Über alle drei Gruppen hinweg kann bezüglich der durchschnittlichen Dauer eine klare Tendenz betrachtet werden, die einen Zusammenhang zwischen Inhalt und Zeitdauer der Intervention vermuten lässt. Die längsten Hilfen weisen diese Tendenz ebenfalls auf, wenngleich es in zwei der drei Gruppen auch eine sehr lange Rückmeldehilfe gibt, die bei der Gruppe mit niedriger Ausprägung sogar dazu führt, dass die Rückmeldehilfen durchschnittlich länger dauern als die allgemein-strategischen Hilfen. Ansonsten gibt es einen Ausreißer bei der mittleren Gruppe in Form einer disziplinarischen Hilfe, die mit einer Minute verhältnismäßig lange dauerte. Dies führt bei dieser Gruppe dazu, dass disziplinarische Hilfen länger als Motivationshilfen und sogar länger als Rückmeldehilfen sind. Ohne diese Ausreißer zeichnet sich jedoch ein klares Bild der genannten Tendenz ab, die auch inhaltlich Sinn ergibt. Je mehr die Hilfe inhaltlich gefärbt ist, desto mehr muss erklärt werden (um auf den jeweils vorliegenden Inhalt Bezug

zu nehmen) und entsprechend dauern diese Hilfen länger. Die kürzesten Hilfen können nicht sinnvoll interpretiert werden und ihre Darstellung dient eher der Vollständigkeit. In der vorgenommenen Betrachtung wurde nicht zwischen dominanten und untergeordneten Hilfen unterschieden. Diese Unterscheidung folgt nun in Tabelle 28. Zunächst werden wieder alle Gruppen einzeln betrachtet und anschließend werden diese miteinander verglichen.

Tabelle 28: Durchschnittliche Zeitdauer der verschiedenen Hilfen sortiert nach dominanten und untergeordneten Hilfen der jeweiligen Gruppen

Gruppe	Hilfe-Kategorie	Durchschnittliche Dauer (dominant)	Durchschnittliche Dauer (untergeordnet)
<i>Hohe Ausprägung</i> der diagnostischen Fähigkeit „Deuten“	Inhaltliche Hilfe	168 Sekunden	34 Sekunden
	Inhaltsorientiert-strategische Hilfe	62 Sekunden	59 Sekunden
	Allgemein-strategische Hilfe	50 Sekunden	27 Sekunden
	Rückmeldehilfe	15 Sekunden	43 Sekunden
	Motivationshilfe	10 Sekunden	14 Sekunden
	Disziplinarische Hilfe	5 Sekunden	10 Sekunden
<i>Mittlere Ausprägung</i> der diagnostischen Fähigkeit „Deuten“	Inhaltliche Hilfe	111 Sekunden	20 Sekunden
	Inhaltsorientiert-strategische Hilfe	59 Sekunden	48 Sekunden
	Allgemein-strategische Hilfe	45 Sekunden	25 Sekunden
	Rückmeldehilfe	20 Sekunden	18 Sekunden
	Motivationshilfe	15 Sekunden	10 Sekunden
	Disziplinarische Hilfe	22 Sekunden	32 Sekunden
<i>Niedrige Ausprägung</i> der diagnostischen Fähigkeit „Deuten“	Inhaltliche Hilfe	63 Sekunden	-
	Inhaltsorientiert-strategische Hilfe	53 Sekunden	47 Sekunden
	Allgemein-strategische Hilfe	29 Sekunden	18 Sekunden
	Rückmeldehilfe	60 Sekunden	24 Sekunden
	Motivationshilfe	23 Sekunden	13 Sekunden
	Disziplinarische Hilfe	20 Sekunden	7 Sekunden

Auch an dieser Stelle sei nochmals darauf verwiesen, dass aufgrund der Aufteilung der Hilfen in dominante und untergeordnete Hilfen sowie der Aufteilung in die drei Gruppen eine teilweise geringe Anzahl an Hilfen zugrundeliegt. Da es das Anliegen der Arbeit ist, mögliche Zusammenhänge zwischen den Bereichen aufzudecken, sind diese Analysen trotzdem hilfreich.

Bei der Gruppe mit hoher Ausprägung befinden sich die Zeitdauern der dominanten Hilfen in der erwarteten Reihenfolge. Inhaltliche Hilfen dauern (mit Abstand) am längsten, durchschnittlich 168 Sekunden. Inhaltsorientiert-strategische Hilfen dauern etwas mehr als eine Mi-

nute und allgemein-strategische Hilfen 50 Sekunden. Recht kurz dauern die dominanten Rückmeldehilfen (15 Sekunden), die dominanten Motivationshilfen (zehn Sekunden) und die dominanten disziplinarischen Hilfen (fünf Sekunden). Derartig kurzen Hilfen liegen zwei Aspekte zugrundeliegend, zum einen handelt es sich jeweils um eine geringe Fallzahl, zum anderen bilden die Hilfen überwiegend die einzigen Hilfen innerhalb der kurzen Intervention.

Betrachtet man die untergeordneten Hilfen der Gruppe, kann man erkennen, dass inhaltsorientiert-strategische Hilfen (59 Sekunden) länger dauern als inhaltliche Hilfen (34 Sekunden). Dies liegt vermutlich daran, dass inhaltliche Hilfen recht schnell dominant werden, wenn diese eine gewisse Zeitdauer in der Intervention in Anspruch nehmen. Untergeordnete Rückmeldehilfen (43 Sekunden) dauern in dieser Gruppe ebenfalls länger als dominante Rückmeldehilfen, was jedoch auf die geringe Fallzahl an dominanten Rückmeldehilfen und auf Ausreißer zurückzuführen ist. Untergeordnete disziplinarische Hilfen (zehn Sekunden) und Motivationshilfen (14 Sekunden) dauern durchschnittlich nur recht kurz. Lob oder Ermahnung sind schnell ausgesprochen und abgehandelt, wenn sie nicht der eigentliche Grund der Intervention sind, sondern quasi „nebenbei“ gegeben werden.

Bei der mittleren Gruppe sind die dominanten inhaltlichen Hilfen ebenfalls mit Abstand am längsten (etwas weniger als zwei Minuten). Es folgen inhaltsorientiert-strategische Hilfen (59 Sekunden) und allgemein-strategische Hilfen (45 Sekunden). Am kürzesten sind dominante Motivationshilfen (15 Sekunden), Rückmeldehilfen (20 Sekunden) und disziplinarische Hilfen (22 Sekunden). Bei den untergeordneten Hilfen gibt es zwei Ausreißer. Die inhaltlichen Hilfen sind mit 20 Sekunden sehr kurz (dies wurde schon mehrfach begründet) und die disziplinarischen Hilfen dauern recht lange. Durch einzelne lange untergeordnete disziplinarische Hilfen kommt es dazu, dass diese im Schnitt länger dauern als die dominanten disziplinarischen Hilfen in dieser Gruppe. In Kapitel 9.2.1 konnte gezeigt werden, dass über alle Gruppen hinweg dominante Hilfen länger dauern als untergeordnete (was auch per Definition so sein sollte). Entsprechend sollten solche Verzerrungen nicht überinterpretiert werden. Untergeordnete allgemein-strategische Hilfen (25 Sekunden), Rückmeldehilfen (18 Sekunden) und Motivationshilfen (zehn Sekunden) dauern jeweils kürzer als die entsprechenden dominanten Hilfen. Am längsten sind untergeordnete inhaltsorientiert-strategische Hilfen mit (48 Sekunden), die aber ebenfalls kürzer ausfallen als die dominanten inhaltsorientiert-strategischen Hilfen.

Die Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“ verzeichnet hinsichtlich der dominanten Hilfen einen Ausreißer, nämlich die Rückmeldehilfen, die mit einer Minute am zweitlängsten sind. Länger dauern lediglich die inhaltlichen Hilfen mit 63 Sekunden, beinahe so lange dauern inhaltsorientiert-strategische Hilfen mit 53 Sekunden. Am kürzesten dauern disziplinarische Hilfen (20 Sekunden), Motivationshilfen (23 Sekunden) und allgemein-strategische Hilfen (29 Sekunden). Im Fall der untergeordneten Hilfen tritt die Besonderheit auf, dass es in der Gruppe keine inhaltliche Hilfe gibt (die nicht dominant war). Inhaltsorientiert-strategische Hilfen dauern am längsten mit 47 Sekunden. Etwa halb so lange dauern Rückmeldehilfen mit 24 Sekunden. Allgemein-strategische Hilfen sind im Schnitt 18 Sekunden lang. Am kürzesten sind disziplinarische Hilfen (sieben Sekunden) und Motivationshilfen (13 Sekunden). In dieser Gruppe sind alle dominanten Hilfen im Schnitt länger als die entsprechenden untergeordneten Hilfen.

Einen Unterschied zwischen den Gruppen stellt die Dauer der dominanten inhaltlichen Hilfen dar. Die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit gibt durchschnittlich deutlich längere inhaltliche Hilfen (168 Sekunden) als die anderen beiden Gruppen, wobei die mittlere Gruppe (111 Sekunden) in der Mitte liegt und der Abstand zur Gruppe mit geringer

Ausprägung (63 Sekunden) sehr groß ist. Umgekehrt verhält es sich bei den dominanten Rückmeldehilfen. Diese sind bei der Gruppe mit geringer Ausprägung mit 60 Sekunden deutlich länger als bei den anderen beiden Gruppen, deren dominante Rückmeldehilfen durchschnittlich 15 bzw. 20 Sekunden dauern.

Die durchschnittlich längeren Rückmeldehilfen sind auf eine extrem lange Rückmeldehilfe zurückzuführen. Diese ist derart lange, dass sie die durchschnittliche Dauer deutlich verlängert. Entsprechend ist der Unterschied an dieser Stelle auf diesen Ausreißer zurückzuführen. Bei den inhaltlichen Hilfen ist dieser Effekt durch Ausreißer nicht zu erkennen. Vielmehr scheint es hierfür einen inhaltlichen Grund zu geben. Möglicherweise werden mehrere fachliche Unklarheiten diagnostiziert, sodass auch mehrere fachliche Inhalte erläutert werden und die Interventionen dadurch länger ausfallen. Eine andere Erklärung wäre, dass die Studierenden mit besseren fachlichen bzw. fachdidaktischen Kenntnissen auch besser diagnostizieren und aufgrund des größeren Wissens auch ausführlichere inhaltliche Hilfen geben (können). Da dies jedoch nicht getestet wurde, bleibt es an dieser Stelle bei einer Spekulation.

In allen drei Gruppen werden durchschnittlich ungefähr 1,8 Hilfen pro Intervention geliefert. Hierin unterscheiden sich die Studierenden der drei Gruppen also nicht. Studierende mit höheren diagnostischen Fähigkeiten im Bereich „Deuten“ erteilen etwas mehr inhaltliche Hilfen und etwas weniger disziplinarische Hilfen als die anderen beiden Gruppen, betrachtet man sich nur die dominanten Hilfen. Außerdem geben diese Studierenden prozentual auch mehr dominante allgemein-strategische Hilfen. Mit Blick auf die untergeordneten Hilfen ist erkennbar, dass die Studierenden mit vergleichsweise höherer diagnostischer Fähigkeit „Deuten“ auch inhaltliche Hilfen als untergeordnete Hilfen erteilen, während die anderen beiden Gruppen dies deutlich seltener bzw. gar nicht tun. Insgesamt sieht die Verteilung der Hilfen bei den drei Gruppen ähnlich aus. Die Unterschiede in Bezug auf die allgemein-strategischen Hilfen werden durch die untergeordneten Hilfen ausgeglichen, sodass alle drei Gruppen insgesamt einen ähnlichen Anteil an allgemein-strategischen Hilfen erteilen. Ein ähnlicher Anteil liegt auch bei inhaltsorientiert-strategischen Hilfen, bei Rückmeldehilfen und bei Motivationshilfen vor. Kleine Unterschiede lassen sich bei den inhaltlichen Hilfen und bei den disziplinarischen Hilfen erkennen. Die Gruppe mit höherer diagnostischer Fähigkeit „Deuten“ gibt insgesamt mehr inhaltliche Hilfen und weniger disziplinarische Hilfen als die anderen beiden Gruppen.

Die Zeitdauern der verschiedenen Hilfen ergeben bei den Gruppen ein vergleichbares Muster. Je inhaltlicher die gegebene Hilfe ist, desto länger dauert die Hilfe. Dieser Effekt ist bei allen drei Gruppen erkennbar. Im Fall der dominanten Hilfen gilt dies für alle drei Gruppen. Im Bereich der untergeordneten Hilfen gibt es durch einzelne Ausreißer Unterschiede zwischen den Gruppen, die aufgrund der geringen Fallzahl jedoch nicht überinterpretiert werden sollten. Bei allen drei Gruppen zeigt sich die Tendenz, dass inhaltliche Hilfen vergleichsweise kurz sind, wenn sie untergeordnet sind. Entsprechend stellen inhaltsorientiert-strategische Hilfen bei allen drei Gruppen die längsten untergeordneten Hilfen dar. Hinsichtlich der Zeitdauer sind insgesamt keine nennenswerten Unterschiede zu erkennen. Bei den Hilfen, die die drei Gruppen nutzen, können hingegen kleine Unterschiede ausgemacht werden. Als nächstes wird die Reihenfolge der Hilfen untersucht.

9.4.2 Reihenfolge der Hilfen

Zur Analyse der Reihenfolge der Hilfen wurden die 208 Reihenfolgen entsprechend der getätigten Interventionen auf die drei Gruppen verteilt. Da einige Reihenfolgen nur sehr selten auftreten, werden nachfolgend die zehn häufigsten Reihenfolgen betrachtet. Dabei wird untersucht,

wie häufig welche Reihenfolge in welcher Gruppe auftrat. Die Gruppe mit hoher diagnostischer Fähigkeit wird in Tabelle 29 mit *Gruppe H* bezeichnet. Analog wird die mittlere Gruppe mit *Gruppe M* und die Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“ mit *Gruppe N* bezeichnet.

Tabelle 29: Reihenfolge der auftretenden Hilfen sortiert nach Gruppen

Reihenfolge	Gruppe H	Gruppe M	Gruppe N
Allgemein-strategische Hilfe	14	11	9
Inhaltsorientiert-strategische Hilfe	7	10	8
Allgemein-strategische Hilfe → Inhaltsorientiert-strategische Hilfe	6	6	4
Rückmeldehilfe → Inhaltsorientiert-strategische Hilfe	5	2	8
Disziplinarische Hilfe	2	7	2
Rückmeldehilfe	3	4	3
Inhaltsorientiert-strategische Hilfe → Allgemein-strategische Hilfe	1	2	4
Inhaltsorientiert-strategische Hilfe → Inhaltliche Hilfe	1	2	4
Rückmeldehilfe → Inhaltsorientiert-strategische Hilfe → Allgemein-strategische Hilfe	3	2	1
Rückmeldehilfe → Inhaltliche Hilfe	2	1	2

Die am häufigsten auftretenden Kombinationen von Hilfen sind zunächst gar keine echten Kombinationen, sondern Interventionen, bei denen genau eine Hilfe, nämlich eine allgemein-strategische Hilfe bzw. eine inhaltsorientiert-strategische Hilfe gegeben wurde. Entsprechend handelt es sich dabei um dominante Hilfen, die bereits zuvor in Kapitel 9.4.1 hinsichtlich der Gruppenunterschiede analysiert wurden.

An Stelle 3 und 4 treten die ersten Reihenfolgen mit zwei Hilfen auf. Zunächst eine allgemein-strategische und anschließend eine inhaltsorientiert-strategische Hilfe gaben jeweils sechs Studierende aus den Gruppen H und M sowie vier Studierende aus der Gruppe N. An dieser Stelle lassen sich also kaum Unterschiede zwischen den Gruppen feststellen. Eine Rückmeldehilfe und dann eine inhaltsorientiert-strategische Hilfe treten in dieser Reihenfolge bei acht Studierenden der Gruppe N, bei fünf Studierenden der Gruppe H und bei zwei Studierenden der Gruppe M auf. Die Unterschiede sind jedoch zu gering, um diese interpretieren zu können.

Auch bei den anderen Reihenfolgen mehrerer Hilfen treten keine nennenswerten Unterschiede auf. Deshalb soll diese Analyse nicht weiterverfolgt werden, sondern nun die erste Hilfe innerhalb einer Intervention analysiert werden (vgl. Tab. 30).

Tabelle 30: Anzahl der zuerst genutzten Hilfe innerhalb einer Intervention sortiert nach Gruppen

Erste Hilfe innerhalb der Intervention	Gruppe H	Gruppe M	Gruppe N
Disziplinarische Hilfe	3	14	5
Motivationshilfe	1	3	2
Rückmeldehilfe	20	14	16
Allgemein-strategische Hilfe	25	26	20
Inhaltsorientiert-strategische Hilfe	11	20	20
Inhaltliche Hilfe	5	2	1

Man kann erkennen, dass disziplinarische Hilfen am häufigsten von Gruppe M als erste Hilfe innerhalb der Intervention getätigt wurden. Zwar handelt es sich auch um die Gruppe, die die meisten disziplinarischen Hilfen tätigt, jedoch ist der Abstand zu den anderen beiden Gruppen auffällig groß. In Bezug auf die Motivationshilfen lassen sich keine nennenswerten Unterschiede feststellen, da diese insgesamt selten und somit auch selten an erster Stelle erteilt werden. Rückmeldehilfen scheinen von Gruppe H tendenziell etwas häufiger als erste Hilfe genutzt zu werden als von den anderen beiden Gruppen. Allgemein-strategische Hilfen wurden von allen drei Gruppen am häufigsten zuerst gegeben, wobei die Gruppen H und M diese tendenziell noch häufiger an erster Stelle geben als Gruppe N. Hinsichtlich der inhaltsorientiert-strategischen Hilfen ist Gruppe H die Gruppe, die diese am seltensten zuerst gibt. Inhaltliche Hilfen werden von Gruppe H zwar selten, aber öfter als von Gruppe M und N als erste Hilfe gegeben. Dies liegt vermutlich daran, dass Gruppe H ohnehin die Gruppe ist, die die meisten inhaltlichen Hilfen gibt.

Betrachtet man die drei Gruppen einzeln, so fällt auf, dass Gruppe H überwiegend Rückmeldehilfen und allgemein-strategische Hilfen als erste Hilfe nutzt. Bei den anderen beiden Gruppen sind jeweils die allgemein-strategischen und die inhaltsorientiert-strategischen Hilfen diejenigen Hilfen, die am häufigsten als erste Hilfe genutzt werden. Entsprechend ließe sich schlussfolgern, dass Gruppe H tendenziell eher versucht, eine weniger inhaltliche Hilfe anzubieten. Dass diese Hilfen dann häufig von weiteren, inhaltlicheren Hilfen ergänzt werden, zeigt sich anhand der Verteilung aller Hilfen, die bei allen drei Gruppen recht ähnlich ausfällt. Zumindest scheint Gruppe H zu versuchen, die Schülerinnen und Schüler möglichst eigenständig arbeiten zu lassen. Möglich wäre, dass diese Gruppe im Laufe der Intervention diagnostiziert, dass noch weitere Hilfen notwendig sind. Die anderen beiden Gruppen nutzen von Anfang an etwas inhaltlichere Hilfen.

Hinsichtlich der Reihenfolge der Hilfen lässt sich festhalten, dass bei den am häufigsten auftretenden Kombinationen keine großen Unterschiede zwischen den Gruppen auftreten. Entsprechend werden diese Ergebnisse auch nicht weitergehend vertieft oder interpretiert. Bezüglich der zuerst genutzten Hilfen innerhalb der Interventionen zeigt sich die Tendenz, dass die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ die Interventionen eher mit Hilfen beginnt, die wenig inhaltlich sind. Überwiegend nutzt diese Gruppe Rückmeldehilfen und allgemein-strategische Hilfen. Hingegen nutzen die anderen beiden Gruppen hauptsächlich allgemein-strategische und inhaltsorientiert-strategische Hilfen.

Bisher konnten feine Unterschiede hinsichtlich der dominanten und untergeordneten Hilfen sowie in Bezug auf die innerhalb der Intervention zuerst verwendeten Hilfen festgestellt werden. Als nächstes wird nun untersucht, ob die drei Gruppen sich in Hinblick auf die Angemessenheit der getätigten Interventionen unterscheiden.

9.4.3 Angemessenheit

Bei der Analyse der Angemessenheit der getätigten Interventionen werden zunächst die absoluten Anzahlen der angemessenen und der unangemessenen Interventionen aller drei Gruppen präsentiert. Anschließend wird der relative Anteil betrachtet, da die mittlere Gruppe mehr Interventionen getätigt hat als die anderen beiden Gruppen. Schließlich werden die einzelnen Interventionsarten untersucht und etwaige Unterschiede bezüglich der Angemessenheit herausgearbeitet. Danach wird betrachtet, welche Kriterien bei den Interventionen der drei Gruppen jeweils von den Expertinnen und Experten beanstandet wurden. Dabei wird auch auf die am häufigsten genannten Kriterien bei den verschiedenen Interventionsarten eingegangen. Abschließend werden die beanstandeten Kriterien bei den als unangemessen eingestuften Interventionen beleuchtet.

Abbildung 43 zeigt die absoluten Anzahlen der angemessenen (in grün) und unangemessenen Interventionen (in rot) der drei Gruppen zu sehen. Betrachtet man nur die grünen Balken, so lässt sich erkennen, dass die mittlere Gruppe die meisten angemessenen Interventionen getätigt hat, dicht gefolgt von der Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“. Am wenigsten angemessene Interventionen hat die Gruppe mit niedriger Ausprägung getätigt. Betrachtet man die unangemessenen Interventionen, kann man der Abbildung entnehmen, dass die mittlere Gruppe auch die meisten unangemessenen Interventionen getätigt hat, dicht gefolgt von der Gruppe mit niedriger Ausprägung der diagnostischen Fähigkeit „Deuten“. Die wenigsten unangemessenen Interventionen wurden von der Gruppe mit hoher Ausprägung getätigt.

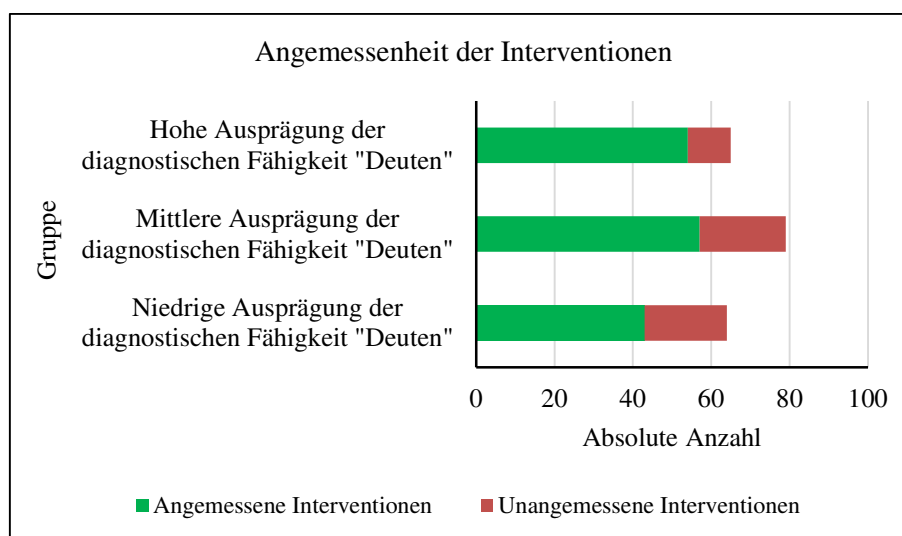


Abbildung 43: Angemessenheit der Interventionen sortiert nach Gruppen

Die Summe der angemessenen und unangemessenen Interventionen, also die Länge des Balkens, gibt Auskunft darüber, wie viele Interventionen insgesamt getätigt und entsprechend auch von den Expertengruppen bewertet wurden. Man kann erkennen, dass die mittlere Gruppe mehr Interventionen getätigt hat als die anderen beiden Gruppen. Entsprechend ist es notwendig, den relativen Anteil zu betrachten, um eine Aussage über die Angemessenheit der Interventionen der jeweiligen Gruppen treffen zu können. Abbildung 44 veranschaulicht, dass die Studierenden mit erhöhter diagnostischer Fähigkeit „Deuten“ in mehr als vier von fünf Fällen eine angemessene Intervention tätigen. Die Gruppe mit niedriger Ausprägung tätigt in etwa zwei von drei Fällen eine angemessene Intervention, die mittlere Gruppe liegt dazwischen. Um den Bezug

zwischen Abbildung 43 und 44 noch besser herzustellen, wurden die absoluten Anzahlen der angemessenen Interventionen bei jeder Gruppe angegeben. Dadurch wird deutlich, dass die Gesamtanzahl der Interventionen eine entscheidende Rolle spielt, wenn man den relativen Anteil der angemessenen Interventionen betrachtet. Die mittlere Gruppe hat zwar insgesamt die meisten angemessenen Interventionen getätigt, aber auch deutlich mehr unangemessene und liegt daher hinsichtlich des relativen Anteils etwas näher an der Gruppe mit niedriger Ausprägung als an der Gruppe mit hoher Ausprägung.

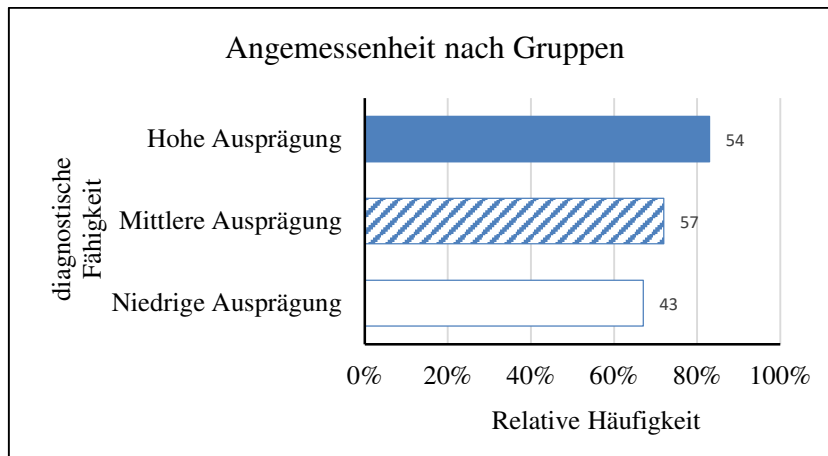


Abbildung 44: Relative Häufigkeit der angemessenen Interventionen sortiert nach Gruppen

Eine Erklärung hierfür könnte sein, dass die Studierenden mit höherer Ausprägung der prozessdiagnostischen Fähigkeit „Deuten“ die Situation besser wahrnehmen und individuell vorliegende Probleme besser deuten können, was es ihnen ermöglicht, eine darauf abgestimmte Intervention zu tätigen, die dann auch von den Expertinnen und Experten als angemessen bewertet wird. Nun stellt sich die Frage, ob dieser Effekt über alle Interventionsarten hinweg besteht oder ob es auch Interventionsarten gibt, bei denen andere Gruppen angemessenere Interventionen tätigen.

In Abbildung 45 sind die relativen Häufigkeiten der angemessenen Interventionen (x-Achse) und die sechs Hilfe-Kategorien (y-Achse) aufgetragen. Dabei wird jeweils zwischen den drei Gruppen unterschieden. Richtet man den Blick auf die disziplinarischen Hilfen, zeigt sich, dass die Gruppe mit hoher Ausprägung der prozessdiagnostischen Fähigkeit „Deuten“ alle Interventionen mit dominanter disziplinarischer Hilfe angemessen durchgeführt hat. Hierbei muss jedoch berücksichtigt werden, dass es sich lediglich um zwei Interventionen handelt. Die Zahl hinter den Balken steht für die absolute Anzahl an angemessenen Interventionen bei der entsprechenden Interventionsart. Die mittlere Gruppe hat also neun angemessene Interventionen getätigt, jedoch auch unangemessene, sodass insgesamt ein relativer Anteil von gut 80% zustande kommt. Die Gruppe mit niedriger Ausprägung tätigte in etwa zwei von drei Fällen eine angemessene Intervention mit dominanter disziplinarischer Hilfe.

In Bezug auf die Motivationshilfen haben alle drei Gruppen genau eine angemessene Intervention durchgeführt. Die mittlere Gruppe hat darüber hinaus auch eine unangemessene Intervention getätigt. Der relative Anteil ist an dieser Stelle nur bedingt aussagekräftig. Bezüglich der Rückmeldehilfen kann man festhalten, dass die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ und die Gruppe mit mittlerer Ausprägung jeweils etwa 80% der Interventionen angemessen gestalten konnten, während die Studierenden mit niedriger Ausprägung in etwa 60% der Fälle positiv beurteilt wurden.

Eine deutlich höhere Fallzahl liegt bei den Interventionen vor, bei denen eine allgemein-strategische oder eine inhaltsorientiert-strategische Hilfe dominant waren. Bei den allgemein-strategischen Hilfen lässt sich erkennen, dass die Gruppe mit hoher Ausprägung fast alle derartigen Interventionen so gestalten konnte, dass die Expertengruppen diese als angemessen beurteilten. Bei der mittleren Gruppe war dies bei knapp 80% der Interventionen und bei der Gruppe mit geringer Ausprägung in gut 60% der Interventionen der Fall. Auch bei den inhaltsorientiert-strategischen Hilfen kommen die verhältnismäßig guten Diagnostiker auf einen höheren Anteil an angemessenen Interventionen. Der Vorsprung zu den anderen Gruppen fällt hier jedoch etwas geringer aus als bei den allgemein-strategischen Hilfen. Mit etwa 90% liegt die Gruppe vor der mittleren mit gut 80% und vor den Studierenden mit geringer Ausprägung, die etwa 75% der Interventionen angemessen durchführen konnten.

Im Fall der inhaltlichen Hilfen konnte keine der drei Gruppen mehr als die Hälfte angemessen gestalten. Genau die Hälfte der Interventionen konnten die Studierenden mit hoher prozessdiagnostischer Fähigkeit „Deuten“ und die Gruppe mit niedriger Ausprägung angemessen durchführen. Der mittleren Gruppe gelang dies nur bei jeder vierten Intervention, wobei hier wieder die geringe Fallzahl für den Ausreißer verantwortlich sein könnte.

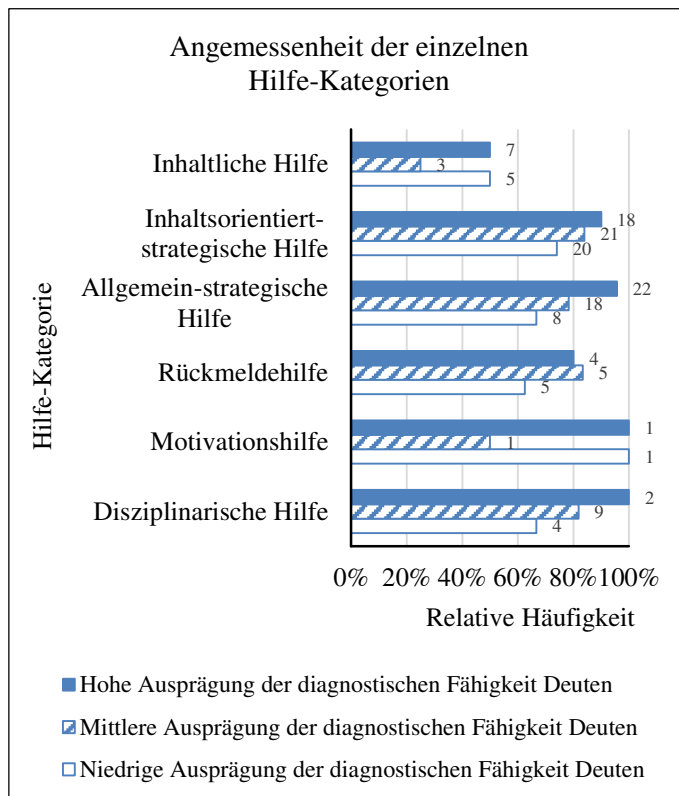


Abbildung 45: Angemessenheit der einzelnen Interventionsarten sortiert nach Gruppen

Über alle Hilfe-Kategorien hinweg konnte die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ auch den größten relativen Anteil an angemessenen Interventionen erzielen. Einzige kleine Ausnahme sind die Rückmeldehilfen, bei denen die mittlere Gruppe einen minimal größeren relativen Anteil aufweisen kann. Bei allen anderen fünf Interventionsarten ist die Gruppe mit hoher Ausprägung der Gruppe mit mittlerer Ausprägung der diagnostischen Fähigkeit „Deuten“ jedoch überlegen. Im Vergleich zu der Gruppe mit niedriger Ausprägung kommt die Gruppe mit hoher Ausprägung bei allen sechs Interventionsarten auf min-

destens denselben prozentualen Anteil an angemessenen Interventionen. Es gibt also keine Interventionsart, bei der die Gruppe mit geringer Ausprägung angemessener agiert hat als die Studierenden, die über eine hohe Ausprägung der diagnostischen Fähigkeit „Deuten“ verfügen. Betrachtet man insbesondere die beiden Interventionsarten, bei denen die Fallzahl am höchsten war, so erkennt man auch die Abstufung, die vorab vermutbar wäre. Die Studierenden, die besser diagnostizieren können, können dieses Wissen anscheinend nutzen, um ihre Interventionen angemessen zu gestalten.

Bevor dieses Ergebnis mit der Anzahl der aktiven Schülerinnen und Schüler sowie mit dem jeweiligen Auslöser der Intervention in Verbindung gebracht wird, wird an dieser Stelle zunächst auf die Kriterien der Angemessenheit eingegangen. Dabei werden zuerst alle beanstandeten Kriterien sortiert nach Gruppen vorgestellt, wobei anschließend nur die unangemessenen Interventionen vertiefend berücksichtigt werden (vgl. Abb. 46).

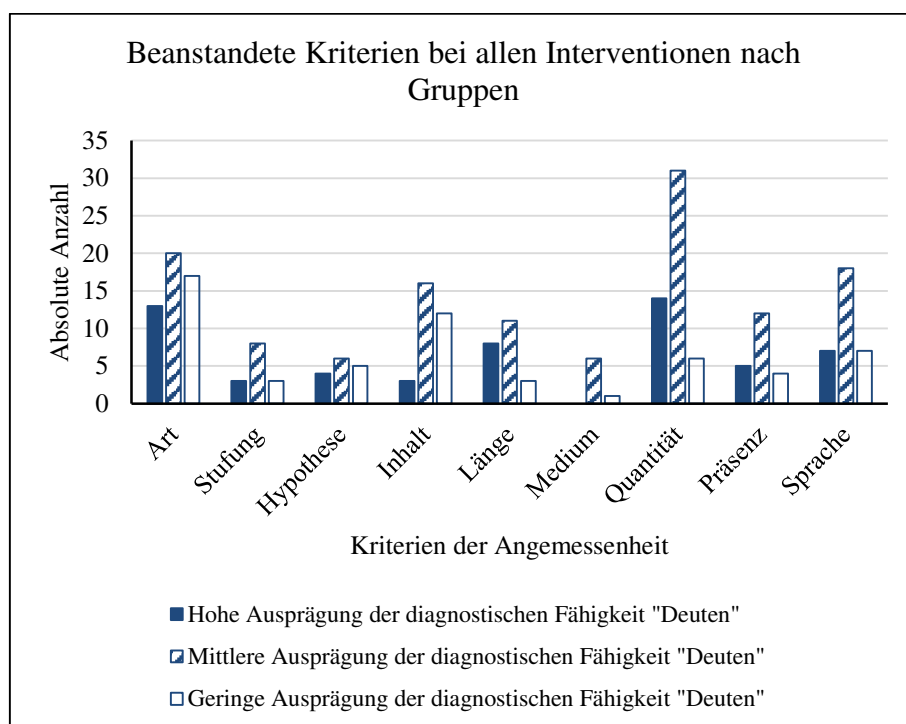


Abbildung 46: Durch die Expertinnen und Experten beanstandete Kriterien innerhalb der Interventionen nach Gruppen

Man kann erkennen, dass das Kriterium „Art“ bei allen drei Gruppen vergleichsweise häufig beanstandet wurde. Dass dies die Studierenden mit erhöhter diagnostischer Fähigkeit „Deuten“ am seltensten betrifft, könnte daran liegen, dass diese besser diagnostizieren und entsprechend auch eine passgenauere Hilfe wählen können. Aufgrund der recht geringen Unterschiede handelt es sich dabei an dieser Stelle jedoch höchstens um ein Indiz für diese Vermutung bzw. um eine leichte Tendenz. Das Kriterium „Intensität“ wurde insbesondere bei der mittleren Gruppe, aber auch bei den Studierenden mit hoher prozessdiagnostischer Fähigkeit am häufigsten kritisiert. Dass die mittlere Gruppe insgesamt die meisten beanstandeten Kriterien aufweist, liegt auch daran, dass diese Gruppe die meisten Interventionen getätigt hat und entsprechend auch mehr „Angriffsfläche“ bietet. Jedoch ist dieser hohe Ausschlag bei der Kategorie nur schwer zu interpretieren.

Ebenfalls in einigen Fällen kritisiert wurden die Kriterien „Inhalt“ und „Sprache“, wobei diese bei der Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ vergleichsweise selten beanstandet wurden. Gerade inhaltlich hatten die Expertinnen und Experten bei Interventionen dieser Gruppe kaum etwas auszusetzen, womöglich weil die Studierenden fachlich besser aufgestellt sind (dies wurde jedoch nicht überprüft) oder weil sie die relevanten Inhalte für ihre Interventionen nutzen. Letzteres könnte in Zusammenhang mit der vorgelagerten Diagnose gebracht werden. In Bezug auf das Kriterium „Länge“ verfügt die Gruppe mit niedriger Ausprägung über die wenigsten Interventionen, die von den Expertinnen und Experten beanstandet wurden. Dies könnte als Indiz dafür gesehen werden, dass die Studierenden innerhalb der Intervention seltener wahrnehmen, wenn die Schülerinnen und Schüler noch akute Probleme oder Schwierigkeiten aufweisen, und die Intervention durch ihren Input als beendet ansehen.

Dieser grobe Überblick über die insgesamt bei allen Interventionen beanstandeten Kriterien kann durch den etwas genaueren Blick auf die je nach Interventionsart am häufigsten beanstandeten Kriterien ergänzt werden. In Tabelle 31 ist für alle drei Gruppen jeweils das am häufigsten beanstandete Kriterium für jede Interventionsart aufgeführt. Die Gruppen werden weiterhin mit den Buchstaben *H*, *M* und *N* abgekürzt.

Tabelle 31: Die häufigsten beanstandeten Kriterien hinsichtlich der verschiedenen Interventionsarten sortiert nach Gruppen

Interventionsart	Gruppe H	Gruppe M	Gruppe N
Inhaltliche Hilfe	Intensität, Länge (je 5)	Inhalt (9)	Inhalt (2)
Inhaltsorientiert-strategische Hilfe	Art, Intensität (je 3)	Intensität (8)	Art (5)
Allgemein-strategische Hilfe	Art, Sprache, Intensität (je 3)	Intensität (8)	Inhalt (2)
Rückmeldehilfe	Sprache, Intensität (je 1)	Art (3)	Inhalt (3)
Motivationshilfe	-	Präsenz, Art, Länge, Intensität, Hypothese, Sprache (je 1)	Präsenz (1)
Disziplinarische Hilfe	Präsenz (2)	Präsenz (7)	Intensität (2)

Bei den inhaltlichen Hilfen ist sowohl bei Gruppe M als auch bei Gruppe N das Kriterium „Inhalt“ das am häufigsten beanstandete Kriterium der Expertinnen und Experten. Das ist insofern zu begründen, als bei inhaltlichen Hilfen ein besonderer Fokus auf dem (mathematischen) Inhalt liegt. Entsprechend wird gerade bei einer inhaltlichen Hilfe darauf geachtet, dass dieser auch (fachlich) korrekt ist. Bei der Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ ist nicht das Kriterium „Inhalt“ am häufigsten problematisch, sondern die Kriterien „Intensität“ und „Länge“. Eine Kritik an der Intensität kann bedeuten, dass zu viel oder zu wenig inhaltliche Hilfe gegeben wurde. Die Länge der Intervention wurde beanstandet, wenn diese deutlich zu lang war. In solchen Fällen wurde vermutlich zu viel Input geliefert, was entsprechend viel Zeit gekostet hat.

Im Fall inhaltsorientiert-strategischer Hilfen wurden häufig die Art der Hilfe sowie die Intensität kritisiert. Eine Beanstandung der Art der Hilfe, wie sie bei Gruppe N häufig auftrat, kann sowohl bedeuten, dass in diesem Fall eine allgemein-strategische Hilfe (oder sogar eine Motivationshilfe oder Rückmeldehilfe) ausgereicht hätte, oder aber, dass an dieser Stelle eine inhaltliche Hilfe notwendig gewesen wäre. Die bei Gruppe M häufig vorkommende Kritik an der Intensität der Hilfe bedeutet, dass die Strategie entweder zu ausführlich oder zu oberflächlich vermittelt wurde.

Interessant ist, dass Gruppe N bei drei Interventionsarten (inhaltliche Hilfe, allgemein-strategische Hilfe und Rückmeldehilfe) am häufigsten aufgrund fachlicher Mängel kritisiert wurde. Während dies bei Gruppe M nur bei inhaltlichen Hilfen der Fall ist, scheint das Kriterium „Inhalt“ bei Gruppe H unabhängig von der Hilfe-Kategorie kein problematischer Faktor zu sein. Bei Hilfen, die keinen inhaltlichen Aspekt aufweisen (disziplinarische Hilfen und Motivationshilfen), scheinen Präsenz und Intensität entscheidend zu sein. Für die Motivationshilfen liegen jedoch zu geringe Fallzahlen vor, um diese sinnvoll zu interpretieren. Bei disziplinarischen Hilfen erscheint die (körperliche) Präsenz der Lehrkraft wichtig zu sein. Die (korrekte) Ausübung einer solchen Hilfe steht jedoch in den genannten Fällen nicht in Zusammenhang mit der diagnostischen Fähigkeit „Deuten“. Dies wäre der Fall, wenn die Art der Hilfe unpassend wäre. Dann wäre nämlich die Wahrnehmung bzw. Deutung der vorliegenden Situation ausschlaggebend. Dass das Kriterium „Art“ bei Gruppe H bei den inhaltsorientiert-strategischen und den allgemein-strategischen mit jeweils drei Mal mit am häufigsten auftritt, erscheint daher zunächst verwunderlich. Die mittlere Gruppe weist jedoch eine höhere Anzahl an Beanstandungen auf, was in Tabelle 31 nicht zu sehen ist, da jeweils das Kriterium „Intensität“ acht Mal aufgelistet ist. Auch Gruppe N hat zumindest bei den inhaltsorientiert-strategischen Hilfen mit fünf Nennungen mehr aufzuweisen als Gruppe H, was die zunächst kontraintuitive Beobachtung relativiert.

Besonders interessant sind die Kriterien, die von den Expertinnen und Experten ins Feld geführt werden, um die Unangemessenheit der Intervention zu stützen. Während in der obigen Betrachtung lediglich aufgeführt wurde, welche Aspekte optimierungswürdig waren (ggf. wurde die Intervention trotzdem als angemessen beurteilt und das beanstandete Kriterium fällt nicht weiter ins Gewicht), werden nun die Interventionen betrachtet, bei denen beanstandete Kriterien dazu geführt haben, dass die gesamte Intervention als unangemessen bewertet wurde (vgl. Abb. 47).

Es zeichnet sich ab, dass die Gruppe mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ in deutlich weniger Interventionen für die getätigte Interventionsart kritisiert wurde. Entsprechend den obigen Überlegungen könnte dies so interpretiert werden, dass erhöhte diagnostische Fähigkeiten dabei helfen, sich für die passende Hilfe-Kategorie zu entscheiden. Bei dem Kriterium „Inhalt“ zeigt sich, dass bei Gruppe H kaum inhaltliche Fehler von den Expertinnen und Experten beanstandet wurden. Da fachliches (und fachdidaktisches) Wissen wichtig ist, um überhaupt erkennen zu können, wo Probleme und Schwierigkeiten bei Schülerinnen und Schülern liegen, könnte es einerseits sein, dass dieses Wissen auch dazu geführt hat, dass die Studierenden die Diagnoseaufträge zu den Videovignetten gut beantworten konnten. Entsprechend wäre auch zu erklären, dass diese Gruppe bei den Interventionen kaum fachliche Fehler macht. Andererseits könnte es auch sein, dass diese Studierenden die Probleme und Schwierigkeiten der Schülerinnen und Schüler relativ gut erfassen und die mathematischen Inhalte innerhalb der Intervention passgenau darauf abstimmen. Eine weitere mögliche Erklärung für den Unterschied zwischen den Gruppen wäre also, dass keine unnötigen Inhalte vermittelt werden und dies entsprechend auch nicht von den Expertinnen und Experten beanstandet wird.

Die Kriterien „Medium“ und „Präsenz“ wurden bei Gruppe H in keiner Intervention beanstandet, die als unangemessen bewertet wurde. Jedoch wurden diese beiden Kriterien auch bei den anderen beiden Gruppen nur selten aufgeführt, weshalb an dieser Stelle nicht genauer darauf eingegangen wird. Ähnliches gilt für die Kriterien „Stufung“, „Hypothese“, „Intensität“ und „Sprache“, bei denen diese Studierenden jeweils am seltensten kritisiert wurden. Keine nennenswerten Unterschiede scheint es hinsichtlich der Kriterien „Hypothese“ und „Länge“ zu geben.

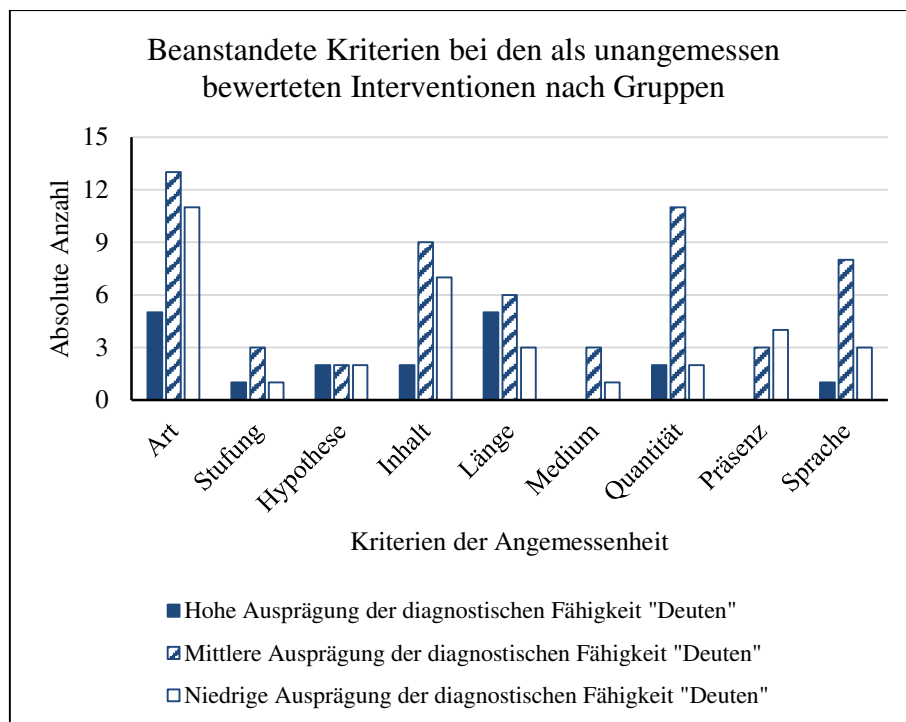


Abbildung 47: Von den Expertinnen und Experten beanstandete Kriterien bei den als unangemessen bewerteten Interventionen sortiert nach Gruppen

Insgesamt lässt sich festhalten, dass die Studierenden mit vergleichsweise höherer diagnostischer Fähigkeit „Deuten“ angemessenere Interventionen tätigen als Studierende mit weniger ausgeprägter Fähigkeit. Dieser Effekt gilt über alle Interventionsarten hinweg, scheint also relativ beständig zu sein. Darüber hinaus scheinen diese Studierenden deutlich weniger inhaltliche Fehler zu machen als ihre Kommilitoninnen und Kommilitonen. Dies kann sowohl aus der Übersicht über die am häufigsten beanstandeten Kriterien über alle Interventionen als auch aus der Übersicht über die beanstandeten Kriterien bei als unangemessen bewerteten Interventionen entnommen werden. Darüber hinaus scheint die prozessdiagnostische Fähigkeit auch einen Einfluss auf die adäquate Auswahl der Interventionsart zu haben, da dieses Kriterium bei den Studierenden mit hoher prozessdiagnostischer Fähigkeit „Deuten“ mit Abstand am seltensten kritisiert wurde.

9.4.4 Anzahl der aktiven Schülerinnen und Schüler

Die Anzahl der aktiven Schülerinnen und Schüler vor bzw. nach einer Intervention kann als weiteres Indiz für die Angemessenheit der Intervention angesehen werden. Entsprechend lohnt es sich, einen genaueren Blick auf das Arbeitsverhalten der Schülerinnen und Schüler zu werfen und dabei die Interventionen der drei Gruppen getrennt voneinander zu untersuchen. Dabei wird

jeweils zunächst ein grober Überblick geliefert und anschließend die jeweilige Veränderung der aktiven Schülerinnen und Schüler analysiert.

Das Arbeitsverhalten der Schülerinnen und Schüler vor bzw. nach den Interventionen der Studierenden mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ (Gruppe H) kann Abbildung 48 entnommen werden. Auf der x-Achse findet sich jeweils die Anzahl der aktiven Schülerinnen und Schüler, auf der y-Achse die absolute Anzahl an Interventionen. Das Arbeitsverhalten der Schülerinnen und Schüler vor der Intervention ist in Blau, das Arbeitsverhalten der Schülerinnen und Schüler nach der Intervention ist in Hellgrün aufgetragen. Auf den ersten Blick lässt sich erkennen, dass in den meisten Gruppen bereits vor der jeweiligen Intervention alle vier Schülerinnen und Schüler aktiv waren. Man kann ebenfalls erkennen, dass es nach den Interventionen noch mehr Gruppen gibt, in denen alle Schülerinnen und Schüler aktiv sind. Dass nur die Hälfte oder weniger Schülerinnen und Schüler aktiv waren, kam kaum vor.

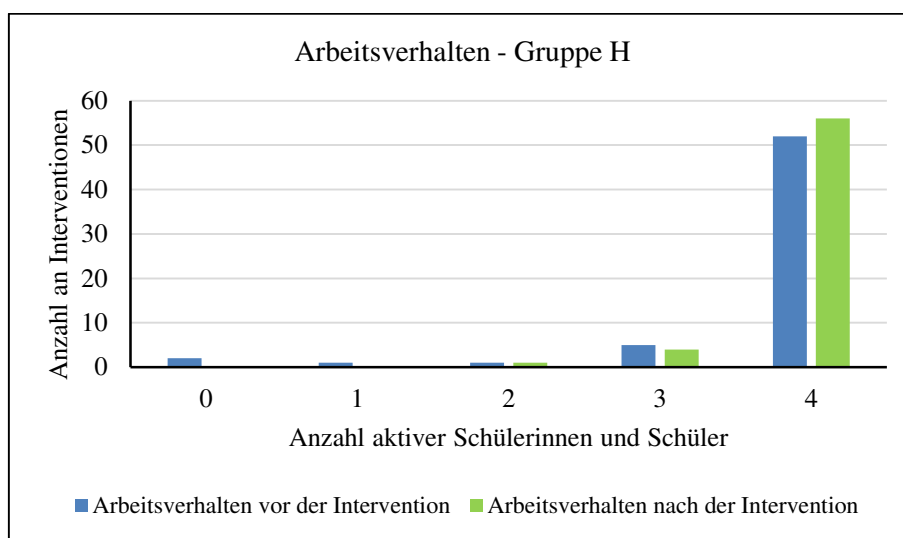


Abbildung 48: Arbeitsverhalten der Schülerinnen und Schüler vor und nach den Interventionen der Gruppe H

Um diese Zahlen noch etwas genauer zu durchleuchten, soll nun Tabelle 32 betrachtet werden, in der zum einen die Anzahl der aktiven Schülerinnen und Schüler vor bzw. nach der Intervention zu entnehmen ist, zum anderen auch die Veränderung und die somit entsprechend zusätzlich aktivierten Schülerinnen und Schüler. Es wird ersichtlich, dass es nach der Intervention keine Gruppe mehr gab, bei der niemand oder nur eine Schülerin bzw. ein Schüler aktiv war. Zuvor traten noch zwei Fälle auf, in denen niemand aktiv war, und ein Fall, in dem nur ein Lernender aktiv war. Die Anzahl der Gruppen, bei denen alle vier Schülerinnen und Schüler aktiv waren, konnte von 52 auf 56 gesteigert werden. Daraus würde sich zunächst ergeben, dass 16 Schülerinnen und Schüler durch die Interventionen aktiviert wurden. Allerdings müssen einmal drei und einmal ein Lernender abgezogen werden, sodass insgesamt zwölf Schülerinnen und Schüler durch die Interventionen zusätzlich aktiviert werden konnten. Bei drei Interventionen konnte weder vor noch nach der Intervention eine Beobachtung durchgeführt werden.

Tabelle 32: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe H

Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	52	56	+4	16
3	5	4	-1	-3
2	1	1	0	0
1	1	0	-1	-1
0	2	0	-2	0

Das Arbeitsverhalten der Schülerinnen und Schüler vor bzw. nach den Interventionen der Studierenden mit mittlerer Ausprägung der diagnostischen Fähigkeit „Deuten“ (Gruppe M) kann Abbildung 49 entnommen werden. Analog zu der Abbildung der Gruppe H befindet sich auf der x-Achse die Anzahl der aktiven Schülerinnen und Schülern und auf der y-Achse die absolute Anzahl an Interventionen. Das Arbeitsverhalten der Schülerinnen und Schüler vor der Intervention ist wiederum in Blau und das Arbeitsverhalten der Schülerinnen und Schüler nach der Intervention in Hellgrün aufgetragen. Auch hier lässt sich auf den ersten Blick erkennen, dass in den meisten Gruppen bereits vor der Intervention alle vier Schülerinnen und Schüler aktiv waren. Allerdings ist der Unterschied zu den anderen Anzahlen an Interventionen, bei denen weniger Schülerinnen und Schüler aktiv waren, dieses Mal etwas geringer. Weiterhin kann eine Steigerung der Gruppen, bei denen drei oder vier Lernende aktiv sind und eine Abnahme der Gruppen, bei denen zwei oder weniger Lernende aktiv sind, beobachtet werden.

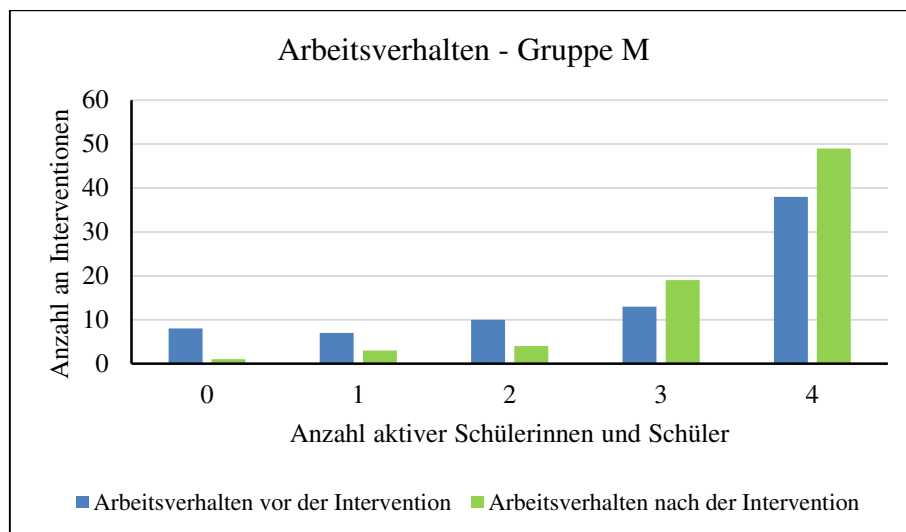


Abbildung 49: Arbeitsverhalten der Schülerinnen und Schüler vor und nach den Interventionen der Gruppe M

Betrachtet man die zusätzlich aktivierten Schülerinnen und Schüler, so lässt sich erkennen, dass insgesamt 46 Schülerinnen und Schüler durch die Interventionen aktiviert werden konnten (vgl. Tab. 33). Bei 49 Gruppen waren in den ersten 30 Sekunden nach der Intervention alle Lernenden aktiv. In 19 Fällen waren immerhin drei Schülerinnen und Schüler aktiv. Der Anteil von zwei aktiven Schülerinnen und Schülern konnte von zehn Gruppen auf vier Gruppen reduziert werden. Nach der Intervention gab es nur drei Gruppen, in denen nur ein Lernender aktiv war. Vor den Interventionen war dies noch in sieben Gruppen der Fall. Auch der schlechteste Fall, bei dem keiner der Schülerinnen und Schüler aktiv ist, konnte von acht

Gruppen auf eine Gruppe reduziert werden. In zwei Gruppen war vor und nach der Beobachtung keine Analyse der Angemessenheit möglich. Insgesamt erscheinen also auch die Ergebnisse der Gruppe M zufriedenstellend. In Gruppen, in denen wenige Lernende aktiv waren, konnten Schülerinnen und Schüler dazu aktiviert werden, sich mit den mathematischen Inhalten auseinanderzusetzen.

Tabelle 33: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe M

Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	38	49	+11	44
3	13	19	+6	18
2	10	4	-6	-12
1	7	3	-4	-4
0	8	1	-7	0

Bevor die Gruppen miteinander verglichen werden, sollen zunächst die Ergebnisse von Gruppe N vorgestellt werden. Das Arbeitsverhalten der Schülerinnen und Schüler vor bzw. nach den Interventionen der Studierenden mit geringer Ausprägung der diagnostischen Fähigkeit „Deuten“ kann Abbildung 50 entnommen werden. Auch bei den Interventionen dieser Gruppe waren die meisten Schülerinnen und Schüler bereits vor der jeweiligen Intervention aktiv. Man erkennt, dass auch diese Gruppe durch die Interventionen dazu beitragen konnte, dass anschließend in mehr Gruppen alle Schülerinnen und Schüler aktiv waren. Der Anteil an Gruppen, in denen drei Lernende aktiv waren, blieb etwa gleich. Fälle, in denen nur die Hälfte der Lernenden oder weniger aktiv waren, kamen kaum vor.

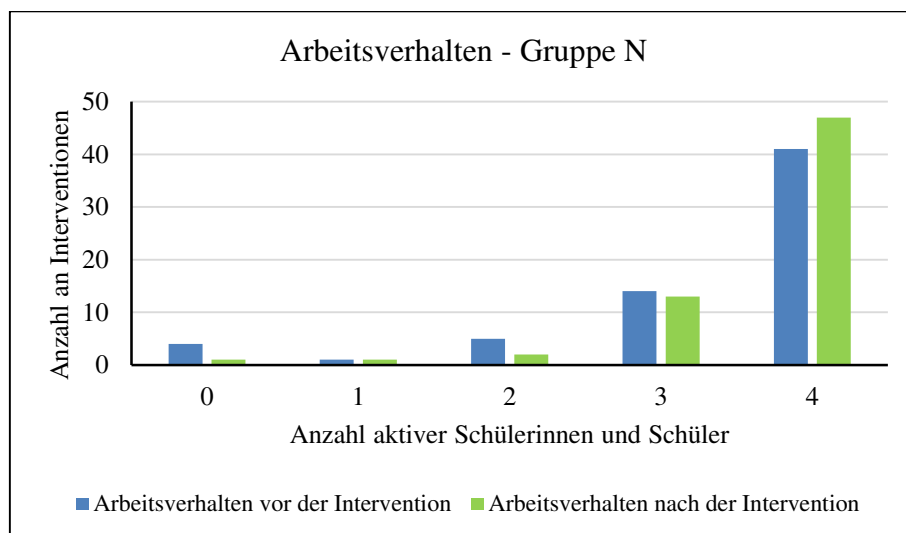


Abbildung 50: Arbeitsverhalten der Schülerinnen und Schüler vor und nach der Interventionen der Gruppe N

Tabelle 34: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe N

Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	41	47	+6	24
3	14	13	-1	-3
2	5	2	-3	-6
1	1	1	0	0
0	4	1	-3	0

Gruppe N konnte durch ihre Interventionen 15 Schülerinnen und Schüler zusätzlich aktivieren (vgl. Tab. 34). Die Anzahl an Gruppen, bei denen alle Lernenden aktiv waren, konnte von 41 auf 47 gesteigert werden. Hingegen blieb die Anzahl an Gruppen, bei denen drei Schülerinnen und Schüler aktiv waren, in etwa gleich. Auch bei dieser Gruppe gab es nach der Intervention noch Gruppen, in denen nur die Hälfte oder weniger Lernende aktiv waren.

Vergleicht man die drei Gruppen miteinander, so kann man zunächst festhalten, dass die mittlere Gruppe die meisten Schülerinnen und Schüler zusätzlich aktivieren konnte. Dies liegt einerseits daran, dass diese Gruppe die meisten Interventionen getätigt hat, andererseits aber auch daran, dass vor den Interventionen weniger Schülerinnen und Schüler aktiv waren als bei den anderen beiden Gruppen und entsprechend auch leichter zusätzliche Schülerinnen und Schüler aktiviert werden konnten. Noch entscheidender als die Veränderung scheint die Anzahl an Lernenden zu sein, die nach der Intervention aktiv sind. Vergleicht man die Werte der drei Gruppen miteinander, so lässt sich erkennen, dass nur bei Gruppe H insofern aktiviert wurde, als nach der Intervention keine Gruppen mehr vorkamen, in denen keiner oder nur ein Lernender aktiv war. Bei den beiden anderen Gruppen war dies jeweils vereinzelt noch der Fall.

Entsprechend scheint auch die Betrachtung der aktiven Schülerinnen und Schüler nach der Intervention für die Angemessenheit der Interventionen von Gruppe H zu sprechen und somit die Bewertung der Expertengruppen zu stützen. Um dies weiter zu untersuchen, werden die Interventionen nun nochmals hinsichtlich ihrer Bewertung der Angemessenheit durch die Expertengruppen betrachtet. Dabei werden abermals alle drei Gruppen zunächst getrennt voneinander analysiert und anschließend miteinander verglichen.

Tabelle 35 führt die aktiven Schülerinnen und Schüler vor und nach der Intervention sowie die daraus resultierende Veränderung der aktiven Schülerinnen und Schüler von Gruppe H getrennt nach angemessenen und unangemessenen Interventionen auf. Betrachtet man zunächst die angemessenen Interventionen, so erkennt man, dass nach den Interventionen in fast allen Gruppen alle Schülerinnen und Schüler aktiv sind. In drei Gruppen sind drei Lernende aktiv. Die drei Gruppen, in denen vor der Intervention nur ein oder kein Lernender aktiv war, konnten also durch die Intervention aktiviert werden. Darüber hinaus scheinen die Interventionen auch keine der aktiven Gruppen dazu gebracht zu haben, nicht weiterhin aktiv zu sein. Nimmt man die 50 angemessenen Interventionen in den Blick, bei denen die Aktivität vor und nach der Intervention analysiert werden konnte, so kann man festhalten, dass dort 197 der 200 Schülerinnen und Schüler und somit nahezu alle Schülerinnen und Schüler aktiv sind.

Tabelle 35: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe H getrennt nach angemessenen und unangemessenen Interventionen

Angemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	44	47	+3	12
3	3	3	0	0
2	0	0	0	0
1	1	0	-1	-1
0	2	0	-2	0
Unangemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	8	9	+1	4
3	2	1	-1	-3
2	1	1	0	0
1	0	0	0	0
0	0	0	0	0

Betrachtet man die unangemessenen Interventionen, von denen es bei Gruppe H deutlich weniger vorkamen als angemessene, so stellt man fest, dass die Interventionen kaum zu einer Veränderung der Aktivität der Lernenden führen. Die elf Interventionen führten dazu, dass ein Lernender aktiviert wurde. Fairerweise muss konstatiert werden, dass der Ausgangswert bereits recht hoch war.

Betrachtet man die Veränderung der aktiven Schülerinnen und Schüler durch die als angemessen bewerteten Interventionen von Gruppe M, so lässt sich feststellen, dass diese in erheblichem Maße dazu beitragen, dass anschließend mehr Schülerinnen und Schüler aktiv sind (vgl. Tab. 36). Während vor den Interventionen in 28 Gruppen alle Gruppenmitglieder aktiv waren, war dies nach den Interventionen in 39 Gruppen der Fall. Außerdem tragen auch die angemessenen Interventionen von Gruppe M dazu bei, dass in allen Gruppen mindestens die Hälfte der Lernenden in den 30 Sekunden nach der Intervention aktiv ist. Dies war vor den Interventionen in acht Gruppen nicht der Fall. Insgesamt wurden 40 Lernende aktiviert. In den 55 Interventionen, bei denen die Aktivität nach der Intervention bewertet werden konnte, waren 201 von 220 Schülerinnen und Schüler aktiv.

Tabelle 36: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe M getrennt nach angemessenen und unangemessenen Interventionen

Angemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	28	39	+11	44
3	10	13	3	9
2	8	3	-5	-10
1	3	0	-3	-3
0	5	0	-5	0

Unangemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	10	10	0	0
3	3	6	3	9
2	2	1	-1	-2
1	4	3	-1	-1
0	3	1	-2	0

Auch bei Gruppe M gab es weniger unangemessene Interventionen als angemessene. Die unangemessenen Interventionen trugen aber verhältnismäßig seltener dazu bei, dass Schülerinnen und Schüler aktiviert wurden. Vor und nach den Interventionen gab es genau zehn Gruppen, in denen vier Lernende aktiv waren. Auch nach den Interventionen gab es eine Gruppe, in der nur die Hälfte der Lernenden aktiv war, und vier Gruppen, in denen weniger als die Hälfte der Lernenden aktiv waren. Insgesamt waren nur 63 der 84 Lernenden aktiv.

Nimmt man die angemessenen Interventionen der Gruppe N genauer in den Blick, so stellt man fest, dass diese in geringem Maße dazu führen, dass nach der Intervention mehr Schülerinnen und Schüler aktiv sind (vgl. Tab. 37). Die Anzahl an Gruppen, in denen alle Lernenden aktiv sind, konnte von 28 auf 32 gesteigert werden. Ebenso gibt es nur eine Gruppe, in der nur die Hälfte der Gruppenmitglieder aktiv ist, und keine Gruppen, in denen niemand oder nur eine Person aktiv ist. Entsprechend sind die angemessenen Interventionen der Gruppe N tendenziell als erfolgreich anzusehen. Nach den Interventionen sind 160 von 172 Schülerinnen und Schülern aktiv.

Tabelle 37: Veränderung der aktiven Schülerinnen und Schüler durch die getätigten Interventionen von Gruppe N getrennt nach angemessenen und unangemessenen Interventionen

Angemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	28	32	+4	16
3	10	10	0	0
2	3	1	-2	-4
1	1	0	-1	-1
0	2	0	-2	0
Unangemessene Interventionen				
Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention	Veränderung	Zusätzlich aktivierte SuS
4	13	15	+2	8
3	4	3	-1	-3
2	2	1	-1	-2
1	0	1	+1	1
0	2	1	-1	0

Im Fall der unangemessenen Interventionen konnte ebenfalls eine leichte Steigerung der Aktivität der Lernenden erreicht werden. Vor den Interventionen gab es 13 Gruppen, in denen alle Gruppenmitglieder aktiv waren. Nach den Interventionen waren es 15 Gruppen. Die Anzahl der

Gruppen, in denen zwei oder weniger Lernende aktiv waren, konnte von vier auf drei gesenkt werden. Dies bedeutet jedoch auch, dass es in den 30 Sekunden nach den entsprechenden Interventionen noch drei Gruppen gab, in denen kein Lernender, nur ein Lernender oder nur zwei Lernende aktiv waren.

Vergleicht man zunächst die angemessenen Interventionen der drei Gruppen miteinander, kann man erkennen, dass es in keiner Gruppe vorkam, dass nach der Intervention kein Lernender oder nur ein Lernender aktiv war. Gruppe H gelang es, dass jeweils mindestens drei Schülerinnen und Schüler aktiv waren. In den Gruppen M und N kam es vor, dass sich nach einer angemessenen Intervention trotzdem nur zwei Schülerinnen und Schüler mit der Aufgabenstellung auseinandersetzten. Die angemessenen Interventionen der Gruppe H sind insgesamt sehr erfolgreich, da anschließend mit 197 von 200 Lernenden fast alle Schülerinnen und Schüler aktiv waren. Zwar war auch bei den anderen beiden Gruppen der Großteil der Schülerinnen und Schüler aktiv, jedoch konnte dieser vergleichsweise extrem hohe Anteil nur von Gruppe H erreicht werden.

Bei den unangemessenen Interventionen erscheinen die Studierenden mit höherer prozessdiagnostischer Fähigkeit „Deuten“ ebenfalls mehr Schülerinnen und Schüler aktivieren zu können. Selbst nach den unangemessenen Interventionen der Gruppe H waren in neun von elf Gruppen alle Gruppenmitglieder aktiv. In einer Gruppe war lediglich ein Gruppenmitglied nicht aktiv und in einer weiteren Gruppe waren zwei Gruppenmitglieder nicht aktiv. Es gab also keine Gruppe, in der nach einer unangemessenen Intervention niemand oder nur ein Lernender aktiv war. Bei den anderen beiden Gruppen verhält es sich anders. Dort gab es nach den Interventionen jeweils eine Gruppe, in der niemand aktiv war, und jeweils eine Gruppe, in der nur ein Lernender aktiv war.

Berücksichtigt man die aktiven Schülerinnen und Schüler vor und nach den Interventionen, so zeigt sich bei allen Gruppen, dass sowohl angemessene als auch unangemessene Interventionen tendenziell dazu führen, dass mehr Schülerinnen und Schüler aktiv sind. Jedoch ist dieser Effekt, gerade im Verhältnis zu der Anzahl der Interventionen, nicht sonderlich groß. Angemessene Interventionen scheinen in größerem Maße dazu beizutragen, dass Schülerinnen und Schüler aktiviert werden. Dies gilt besonders für die Gruppen H und M. Unangemessene Interventionen hingegen tragen etwas weniger dazu bei, dass Schülerinnen und Schüler aktiviert werden. Hervorzuheben ist, dass bei Gruppe H sogar die unangemessenen Interventionen nicht dazu führen, dass keiner oder nur ein Lernender aktiv ist.

Die vorangegangenen Ausführungen machen deutlich, dass die Schülerinnen und Schüler in allen drei Gruppen bereits vor der jeweiligen Intervention überwiegend aktiv waren. Betrachtet man lediglich den Zuwachs an aktiven Schülerinnen und Schülern, so müsste man Gruppe M die erfolgreichsten Interventionen bescheinigen. Berücksichtigt man jedoch, wie viele Schülerinnen und Schüler bereits zuvor aktiv waren, wird recht deutlich, dass Gruppe H die Schülerinnen und Schüler am besten aktiviert. Dies lässt sich daran erkennen, dass nach der jeweiligen Intervention verhältnismäßig mehr Lernende aktiv sind als bei den anderen beiden Gruppen. Dieser Effekt lässt sich sowohl bei den angemessenen als auch bei den unangemessenen Interventionen beobachten. Es scheint, als habe die prozessdiagnostische Fähigkeit „Deuten“ einen positiven Einfluss auf die Aktivierung der Schülerinnen und Schüler. Dies könnte insofern begründet werden, als durch die Abstimmung der Hilfen auf die Bedürfnisse der Lernenden die Motivation steigt, sich mit den mathematischen Inhalten auseinanderzusetzen. Da die Motivation nicht erhoben wurde, ist dies jedoch lediglich eine mögliche Erklärung, die im Rahmen dieser Arbeit nicht überprüft werden kann.

9.4.5 Interventionsauslöser

Als letzte Analyse, um (mögliche) Zusammenhänge zwischen Diagnose und Intervention zu finden, wird der Interventionsauslöser mit in die Überlegungen einbezogen. Dabei wird jeweils die Verteilung der invasiven und responsiven Interventionen dargestellt, wobei diese anschließend nach angemessenen und unangemessenen Interventionen untergliedert werden. Abschließend wird weiterhin die Anzahl der aktiven Schülerinnen und Schüler hinzugezogen, um die Gruppen möglichst detailliert miteinander zu vergleichen.

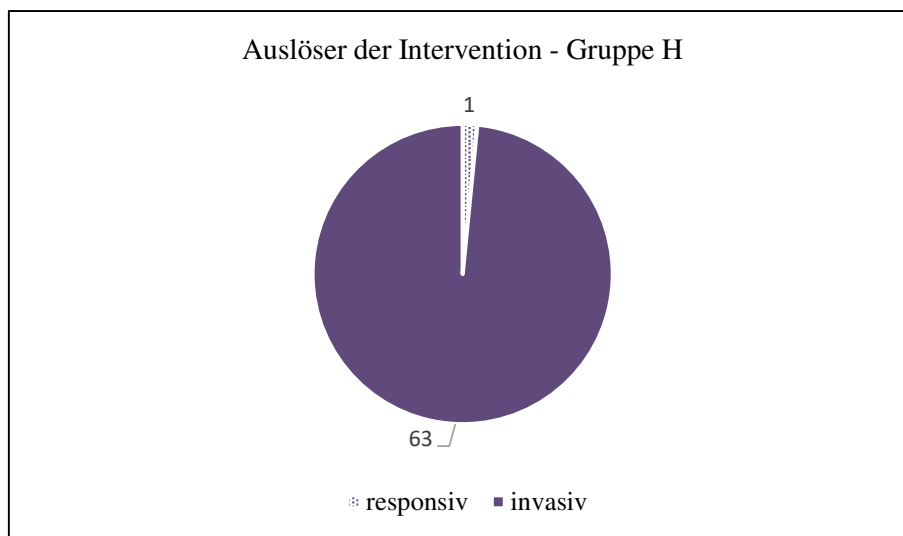


Abbildung 51: Absolute Häufigkeiten der Auslöser der Interventionen der Studierenden mit hoher Ausprägung der diagnostischen Fähigkeit "Deuten"

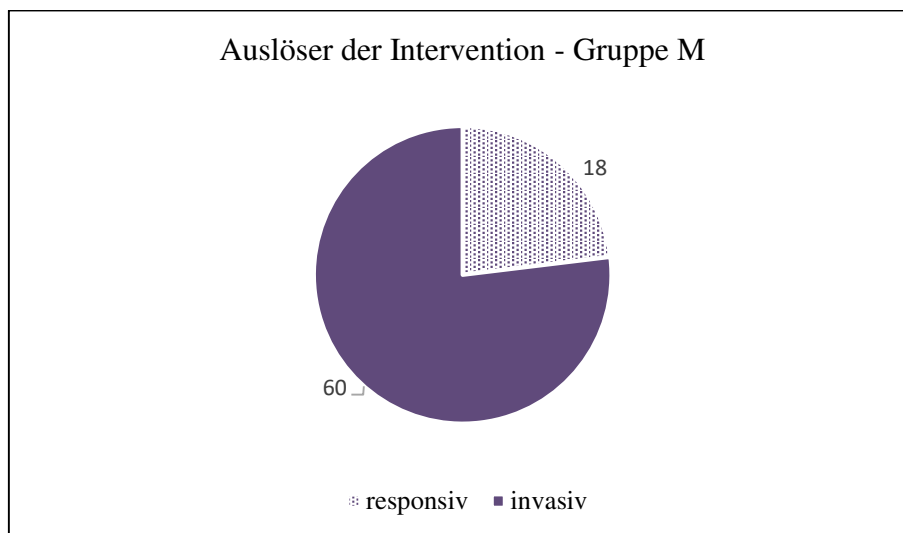


Abbildung 52: Absolute Häufigkeiten der Auslöser der Interventionen der Studierenden mit mittlerer Ausprägung der diagnostischen Fähigkeit "Deuten"

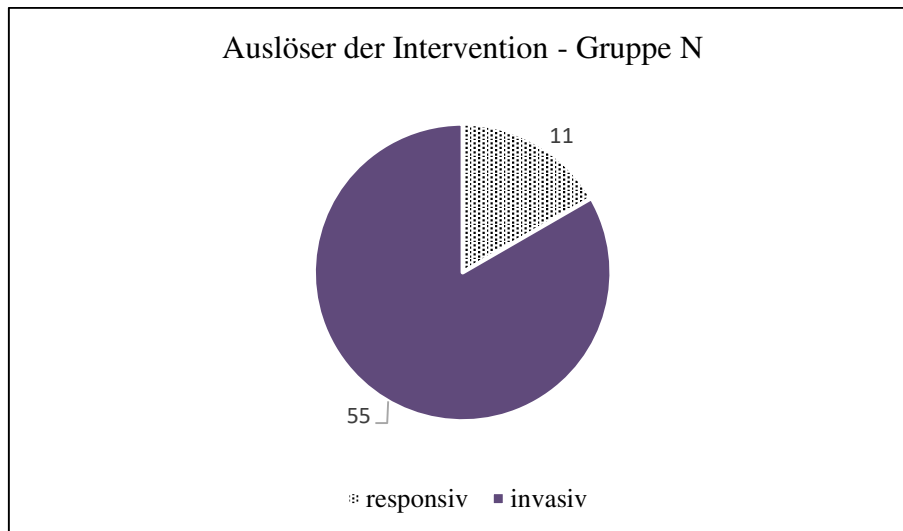


Abbildung 53: Absolute Häufigkeiten der Auslöser der Interventionen der Studierenden mit niedriger Ausprägung der diagnostischen Fähigkeit "Deuten"

Betrachtet man die Interventionen der Studierenden mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“, so erkennt man, dass beinahe alle Interventionen invasiv waren (vgl. Abb. 51). Bei der mittleren Gruppe sind etwas mehr als drei Viertel der Interventionen invasiv (vgl. Abb. 52). Trotzdem ist der Anteil der responsiven Interventionen im Vergleich zu den Studierenden mit hoher Ausprägung der diagnostischen Fähigkeit „Deuten“ deutlich höher. Bei der Gruppe mit niedriger Ausprägung ist ebenfalls der Großteil der Interventionen invasiv (vgl. Abb. 53). Etwas mehr als ein Achtel der Interventionen ist responsiv.

Die insgesamt sehr hohen Anteile der invasiven Interventionen sind unter anderem auch auf das Setting zurückzuführen. Dadurch, dass jeweils vier Schülerinnen und Schüler in einem separaten Raum arbeiten, in dem sich keine Lehrkraft und keine betreuende Person aufhält, ergeben sich nicht allzu viele Möglichkeiten für responsive Interventionen. Darüber hinaus sollen die Lernenden selbstständig arbeiten und aufkommende Fragen zunächst untereinander besprechen. Zudem gibt es ein Hilfeheft, welches die Schülerinnen und Schüler nutzen können, falls sie an einem Punkt nicht weiterkommen sollten. Erst wenn all diese Möglichkeiten erschöpft sind und dann immer noch keine Ansätze vorliegen, mit denen die Schülerinnen und Schüler weiterarbeiten können, sind sie angehalten, sich an eine Betreuerin oder einen Betreuer zu wenden. Da die vorliegenden Gruppenarbeiten jeweils von einer Studentin bzw. einem Studenten in Echtzeit beobachtet wurden, sollte dieser Punkt eigentlich nicht erreicht werden und stattdessen eine invasive Intervention zuvorkommen.

Entsprechend können die Verteilungen der invasiven und responsiven Interventionen auch so interpretiert werden, dass die Studierenden mit hoher diagnostischer Fähigkeit „Deuten“ die Probleme und Schwierigkeiten der Schülerinnen und Schüler frühzeitig wahrnehmen konnten und deswegen womöglich eher invasiv interveniert haben, bevor der Punkt der absoluten Ratlosigkeit und damit verbunden evtl. auch ein Motivationsverlust der Schülergruppe eintrat. Die Anteile an invasiven Interventionen sind bei den anderen beiden Gruppen ebenfalls hoch, jedoch verzeichnen sie im Verhältnis zu Gruppe H einen höheren Anteil an responsiven Interventionen. Dies kann unter der vorigen Prämisse so interpretiert werden, dass teilweise nicht wahrgenommen wurde, wenn gravierende Probleme und Schwierigkeiten vorlagen, die die Schülerinnen und Schüler nicht selbstständig lösen konnten, und die Intervention in diesen Fällen vermutlich zu spät erfolgte.

Neben der reinen Unterscheidung zwischen invasiven und responsiven Interventionen können diese zusätzlich hinsichtlich der Angemessenheit unterschieden werden. Diese zusätzliche Unterscheidung wird zunächst für alle Gruppen einzeln getroffen. Anschließend werden die drei Gruppen miteinander verglichen.

Abbildung 54 führt die Interventionen der Gruppe H, getrennt nach angemessenen und unangemessenen Interventionen und diese wiederum getrennt nach invasiven und responsiven Interventionen (x-Achse), auf. Die jeweilige Anzahl kann der Höhe der Säulen entnommen werden (y-Achse). Die lilafarbenen Säulen stellen die invasiven Interventionen dar, die gepunkteten Säulen die responsiven Interventionen. Es lässt sich erkennen, dass der Großteil der Interventionen von Gruppe H angemessen war. Dabei sind alle responsiven Interventionen angemessen. Zehn invasive Interventionen wurden als unangemessen und 53 invasive Interventionen als angemessen bewertet.

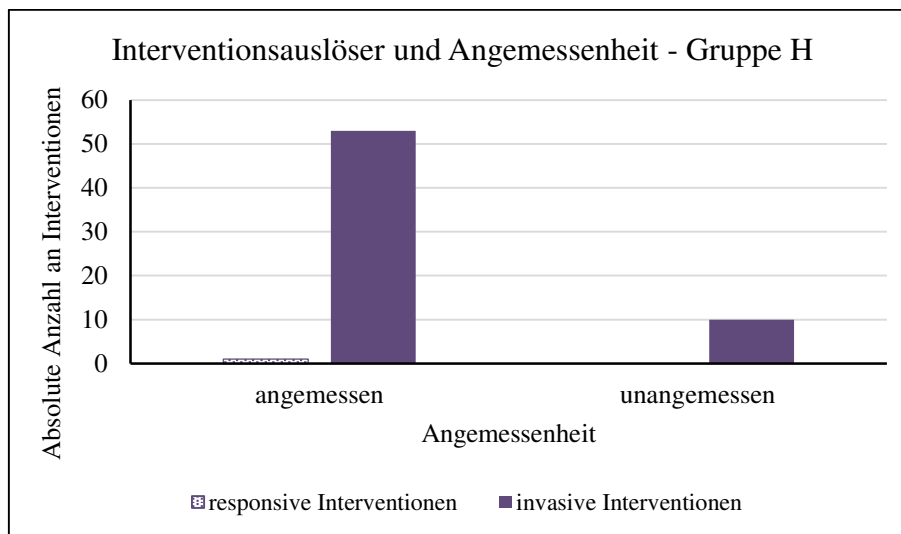


Abbildung 54: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers bei den Interventionen von Gruppe H

Betrachtet man die Interventionsauslöser unter Einbezug der Angemessenheit bei Gruppe M, so stellt man fest, dass sowohl responsive als auch invasive Interventionen als angemessen, aber auch als unangemessen angesehen wurden (vgl. Abb. 55). Im Fall der invasiven Interventionen kann man erkennen, dass 40 invasive Interventionen angemessen waren, während 20 invasive Interventionen als unangemessen angesehen wurden. Bei den responsiven Interventionen wurden 16 Interventionen als angemessen und zwei Interventionen als unangemessen eingestuft.

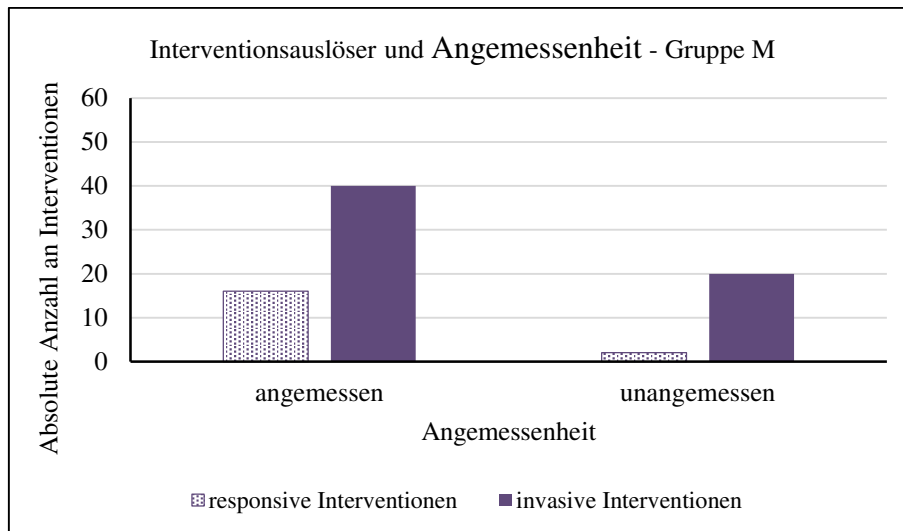


Abbildung 55: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers bei den Interventionen von Gruppe M

Betrachtet man das Verhältnis von angemessenen zu unangemessenen Interventionen hinsichtlich des Interventionsauslösers, ergibt sich ein Verhältnis von acht zu eins bei responsiven Interventionen und ein Verhältnis von zwei zu eins bei invasiven Interventionen. Bei Gruppe M war also ein deutlich höherer Anteil an invasiven Interventionen unangemessen.

Mit Blick auf Gruppe N zeigt sich, dass 33 der invasiven Interventionen angemessen waren und 22 der invasiven Interventionen unangemessen (vgl. Abb. 56). Die elf responsiven Interventionen wurden alle als angemessen bewertet. Der hohe Anteil an unangemessenen invasiven Interventionen könnte daran liegen, dass die Studierenden der Gruppe N nicht in der Lage waren, den Kenntnisstand und den Lernfortschritt der Schülerinnen und Schüler ausreichend gut einzuschätzen, um die Intervention darauf abzustimmen. Womöglich wurden manche Probleme nicht erkannt oder auf vorliegende Schwierigkeiten nicht (in ausreichendem Maße) eingegangen. Bei Interventionen, die von den Lernenden eingefordert wurden, konnte die Gruppe die Interventionen angemessen gestalten.

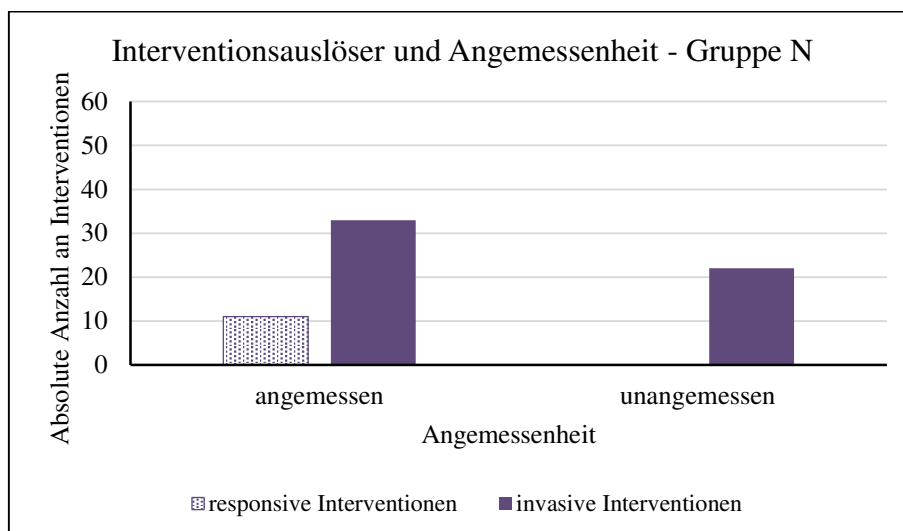


Abbildung 56: Untersuchung der Angemessenheit hinsichtlich des Interventionsauslösers bei den Interventionen von Gruppe N

Vergleicht man die drei Gruppen miteinander, so lässt sich feststellen, dass responsive Interventionen von allen drei Gruppen größtenteils angemessen durchgeführt wurden. Bei Gruppe H und bei Gruppe N waren alle responsiven Interventionen angemessen. Bei Gruppe M wurden immerhin acht von neun responsiven Interventionen als angemessen eingestuft. Bei den invasiven Interventionen liegen deutlichere Unterschiede vor. Hier konnte Gruppe H in mehr als fünf von sechs Interventionen angemessen agieren, während Gruppe M nur auf zwei angemessene Interventionen pro drei Interventionen kommt und Gruppe N lediglich auf drei angemessene Interventionen pro fünf Interventionen. Es scheint also, als würde eine erhöhte Ausprägung der diagnostischen Fähigkeit „Deuten“ dazu führen, dass mehr invasive Interventionen angemessen durchgeführt werden.

Die Bewertung der Angemessenheit wurde von Expertinnen und Experten durchgeführt. Ob diese Einschätzung auch mit der Aktivität der Schülerinnen und Schüler zusammenhängt und wie sich dies bei den drei Gruppen jeweils darstellt, wird im Folgenden untersucht. Dabei werden die responsiven und invasiven Interventionen getrennt voneinander betrachtet.

Tabelle 38 listet die Anzahl der aktiven Schülerinnen und Schüler bei den Interventionen der Gruppe H auf, wobei nach responsiven und invasiven Interventionen unterschieden wird. Bei der einzigen responsiven Intervention von Gruppe H konnte vor der Intervention keine Beobachtung durchgeführt werden. Nach der Intervention sind alle vier Schülerinnen und Schüler aktiv. Da die Intervention von den Schülerinnen und Schülern eingefordert wurde, war es ihnen vermutlich nicht möglich, eigenständig zu einer sinnvollen Lösung zu kommen. Entsprechend ist davon auszugehen, dass vermutlich auch nicht alle Lernenden in den 30 Sekunden vor der Intervention aktiv waren.

Betrachtet man die invasiven Interventionen, so stellt man fest, dass nach den Interventionen fast alle Lernenden aktiv sind. Lediglich in einer Gruppe sind nur zwei Lernende aktiv und in vier Gruppen sind drei Lernende aktiv. In den anderen 55 Gruppen sind alle vier Schülerinnen und Schüler aktiv. Da es vor den Interventionen auch Gruppen gab, in denen wenige oder gar keine Schülerinnen und Schüler aktiv waren, konnten die Interventionen somit dazu beitragen, dass Schülerinnen und Schüler aktiviert wurden. Jedoch ist die Veränderung relativ gering, da der Ausgangswert an aktiven Schülerinnen und Schülern bereits recht hoch war. In 52 der Gruppen arbeiteten bereits vor der Intervention alle Schülerinnen und Schüler und in fünf Gruppen waren jeweils drei Schülerinnen und Schüler aktiv.

Tabelle 38: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern bei den Interventionen der Gruppe H

Auslöser	Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention
Responsiv	4	0	1
	3	0	0
	2	0	0
	1	0	0
	0	0	0
Invasiv	4	52	55
	3	5	4
	2	1	1
	1	1	0
	0	2	0

Gruppe M hat die meisten responsiven Interventionen getätigt. Es lässt sich erkennen, dass diese zu einer Steigerung der Aktivität in den Gruppen führt (vgl. Tab. 39). Nach den responsiven Interventionen gibt es keine Gruppen mehr, in denen nur ein/e Lernende/r oder gar keiner der Schülerinnen und Schüler aktiv sind. Auch die Anzahl an Gruppen, in denen nur die Hälfte der Lernenden aktiv ist, konnte durch die Interventionen von fünf auf eins reduziert werden.

Auch im Fall der invasiven Interventionen der Gruppe M ist anschließend eine erhöhte Aktivität der Schülergruppen erkennbar. Die Steigerung fällt jedoch prozentual gesehen geringer aus als bei den responsiven Interventionen. Außerdem gibt es auch nach den invasiven Interventionen noch Gruppen, in denen weniger als die Hälfte der Lernenden aktiv ist. Die Anzahl der Gruppen, bei denen niemand aktiv war, konnte von sechs auf eins gesenkt werden. Die Anzahl der Gruppen, bei denen nur ein Lernender aktiv ist, konnte von sieben auf drei gesenkt werden. Auf der anderen Seite gibt es jeweils sechs Gruppen mehr, in denen alle Schülerinnen und Schüler arbeiten bzw. in denen drei der vier Gruppenmitglieder aktiv sind.

Tabelle 39: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern bei den Interventionen der Gruppe M

Auslöser	Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention
Responsiv	4	8	13
	3	3	3
	2	5	1
	1	0	0
	0	2	0
Invasiv	4	30	36
	3	10	16
	2	5	3
	1	7	3
	0	6	1

Tabelle 40: Vergleich der responsiven und invasiven Interventionen hinsichtlich der Anzahl an aktiven Schülerinnen und Schülern bei den Interventionen der Gruppe N

Auslöser	Anzahl aktiver SuS	Vor der Intervention	Nach der Intervention
Responsiv	4	6	9
	3	0	1
	2	2	0
	1	1	0
	0	1	1
Invasiv	4	35	38
	3	14	12
	2	3	2
	1	0	1
	0	3	0

Auch Gruppe N konnte die Aktivität der Schülerinnen und Schüler durch responsive Interventionen steigern (vgl. Tab. 40). Bei neun von elf Gruppen waren nach der Intervention alle Gruppenmitglieder aktiv. Bei einer Gruppe waren drei der vier Gruppenmitglieder aktiv. Darüber hinaus gab es jedoch auch eine Gruppe, bei der trotz Intervention keiner der Lernenden aktiv war. Richtet man den Fokus auf die invasiven Interventionen, so tragen diese nur in geringem Maße dazu bei, dass anschließend mehr Schülerinnen und Schüler aktiv sind. Immerhin gibt es nach den Interventionen keine Gruppen mehr, bei denen sich keine Schülerin bzw. kein Schüler mit den mathematischen Inhalten beschäftigt.

Die drei Gruppen konnten mit responsiven Interventionen in größerem Maße dazu beitragen, dass sich anschließend mehr Schülerinnen und Schüler aktiv mit der Aufgabe bzw. den mathematischen Inhalten auseinandersetzten als mit invasiven Interventionen. Dies ist insofern auch logisch, als responsive Interventionen per Definition nur zustande kommen, wenn die Lernenden eine Intervention fordern. Entsprechend sind diese Schülerinnen und Schüler vermutlich auch daran interessiert, sich aktiv mit der Aufgabe auseinanderzusetzen. Insofern sind die Voraussetzungen auf Schülerseite vermutlich tendenziell positiver als bei invasiven Interventionen. Da Gruppe H lediglich eine responsive Intervention getätigt hat, kann hierzu kaum eine Aussage über diese Gruppe getroffen werden. Vergleicht man Gruppe M mit Gruppe N, so lässt sich eine leichte Tendenz zugunsten von Gruppe M erahnen, sprich Gruppe M kann durch die Interventionen mehr Schülerinnen und Schüler dazu bringen, sich mit den mathematischen Inhalten auseinanderzusetzen.

Betrachtet man die invasiven Interventionen, so erkennt man, dass Gruppe M die meisten Lernenden aktivieren konnte. Nimmt man jedoch die Schülerinnen und Schüler in den Blick, die nach den Interventionen aktiv sind, können die Interventionen der Gruppe H als am erfolgreichsten angesehen werden. Dass dort kein großer Zuwachs an aktiven Schülerinnen und Schüler zustande kommt, liegt daran, dass der Ausgangswert schon sehr hoch war. Unter Berücksichtigung der aktiven Schülerinnen und Schüler vor der Intervention und nach der Intervention scheinen also die Studierenden mit erhöhter Ausprägung der diagnostischen Fähigkeit „Deuten“ insofern zielführender zu intervenieren, als im Anschluss die meisten Schülerinnen und Schüler noch aktiv sind oder zusätzlich aktiviert wurden. Die mittlere Gruppe konnte den höchsten Zuwachs erreichen. Mit Blick auf die Schülerinnen und Schüler, die nach der Intervention nicht aktiv sind, ähneln sich die Gruppen M und N. Da Gruppe N weniger Lernende aktivieren konnte, scheinen diese Studierende im Vergleich zu den anderen Studierenden tendenziell weniger zielführende Interventionen durchzuführen. Dieses Ergebnis deckt sich mit der Bewertung der Expertengruppen, die zur gleichen Abstufung der drei Gruppen kommen.

9.4.6 Zusammenfassung

Hinsichtlich des Einflusses der prozessdiagnostischen Fähigkeit „Deuten“ auf die Interventionen der Studierenden lassen sich gemäß der obigen Ausführungen mehrere Indizien für mögliche Zusammenhänge dieser beiden Gebiete feststellen. Andererseits gibt es auch einige Facetten, bei denen kaum oder nur geringe Unterschiede zwischen den Studierenden mit unterschiedlich stark ausgeprägter prozessdiagnostischer Fähigkeit „Deuten“ festgestellt werden konnten.

Betrachtet man die Hilfen innerhalb der Intervention, so lässt sich im Hinblick auf die dominanten Hilfen festhalten, dass Gruppe H eher allgemein-strategische Hilfen nutzt, während die anderen Gruppen eher inhaltsorientiert-strategische Hilfen nutzen. Bei den untergeordneten Hilfen hingegen nutzen die Studierenden in Gruppe H weniger allgemein-strategische Hilfen als die anderen beiden Gruppen. Gruppe H ist die einzige Gruppe, die (abgesehen von einer

Hilfe von Gruppe M) überhaupt inhaltliche Hilfen gibt, die untergeordnet sind. Nimmt man alle Hilfen (also die dominanten und die untergeordneten Hilfen) in den Blick, zeigen sich kaum Unterschiede zwischen den Gruppen. Man kann jedoch feststellen, dass die Studierenden mit hoher Ausprägung der diagnostischen Fähigkeiten „Deuten“ tendenziell mehr inhaltliche Hilfen geben und kaum auf disziplinarische Hilfen zurückgreifen.

Hinsichtlich der Zeitdauer der Interventionen ist es bei allen drei Gruppen so, dass Hilfen umso mehr Zeit in Anspruch nehmen, je mehr Inhalt darin vermittelt wird. Interessant ist hierbei, dass es eine zeitliche Abstufung bei dominanten inhaltlichen Hilfen gibt. Diese dauern bei Gruppe H länger als bei Gruppe M und bei Gruppe M länger als bei Gruppe N. Dies könnte daran liegen, dass höhere fachliche bzw. fachdidaktische Kenntnisse zugrundeliegen und entsprechend auch mehr vermittelt werden kann. Eine alternative Hypothese dazu ist, dass die Studierenden die Schwachstellen der Lernenden dadurch, dass sie diese besser wahrnehmen, durch eine inhaltliche Hilfe schließen wollen. Womöglich stellt Gruppe H mehr Optimierungsbedarf fest und gibt daher ausgedehntere Hilfen. Leider können diese Hypothesen nicht überprüft werden, da weder ein Fachwissenstest durchgeführt noch das fachdidaktische Wissen der Studierenden erhoben wurde. Auch wurden nachträglich keine Interviews mit den Studierenden geführt, um deren Absichten in den jeweiligen Situationen zu hinterfragen.

Hinsichtlich der Reihenfolge mehrerer Hilfen gibt es kaum nennenswerte Unterschiede zwischen den drei Gruppen. Betrachtet man die zuerst verwendete Hilfe, so lässt sich feststellen, dass Gruppe H in Form von Rückmeldehilfen und allgemein-strategischen Hilfen tendenziell eher weniger inhaltliche Hilfen zuerst nutzt. Die anderen beiden Gruppen nutzen allgemein-strategische Hilfen und inhaltsorientiert-strategische Hilfen am häufigsten als erste Hilfe und steigen somit mit Hilfen ein, die inhaltlicheren Charakter haben.

Bezüglich der Angemessenheit der Interventionen zeigen die Analysen, dass Gruppe H durchschnittlich angemessenere Interventionen tätigte als Gruppe M und diese wiederum häufiger positiv beurteilt wurden als die Interventionen von Gruppe N. Hervorzuheben ist an dieser Stelle, dass die Studierenden mit am stärksten ausgeprägter diagnostischer Fähigkeit „Deuten“ in allen Hilfekategorien angemessenere (oder zumindest zu gleichen Teilen angemessene) Interventionen tätigten. Somit besteht dieser Effekt über alle Hilfekategorien hinweg, was dafür spricht, dass die Bewertung bzw. Verteilung kein Zufall ist. Vielmehr kann dies als Indiz zur Bestätigung der Hypothese gewertet werden, dass diagnostische Fähigkeiten Grundlage für adaptive Interventionen sind. Betrachtet man weiterhin die Kriterien, die jeweils zur Begründung der Bewertung der Experten herangezogen wurden, so stellt man fest, dass die Auswahl der Interventionsart von Studierenden mit höherer diagnostischer Fähigkeit „Deuten“ passender zu sein scheint und diese Studierenden auch deutlich weniger inhaltliche Fehler in ihren Interventionen machen.

Darüber hinaus scheint die prozessdiagnostische Fähigkeit „Deuten“ der Studierenden auch einen positiven Einfluss auf die Aktivierung der Schülerinnen und Schüler zu haben. Mit Blick auf Gruppe N zeigt sich, dass diese unter Berücksichtigung der Ausgangslage, also der bereits aktiven Schülerinnen und Schüler vor der Intervention, am wenigsten in der Lage war, die Schülerinnen und Schüler zu aktivieren. Gruppe M konnte in absoluten Zahlen die meisten Lernenden zusätzlich aktivieren. Betrachtet man jedoch die Gruppen, bei denen alle Lernenden im Anschluss an die Intervention aktiv waren, so stellt man fest, dass dies bei Gruppe H der Fall ist (und dies obwohl Gruppe M 14 Interventionen mehr getätigt hat). Relativ betrachtet ist der Anteil von aktiven Schülerinnen und Schülern bei Gruppe H im Vergleich zu den anderen

beiden Gruppen deutlich höher. Gruppe H scheint also die „richtigen Worte“ gefunden zu haben. Übertragen bedeutet dies, dass sie vermutlich aufgrund der besseren Diagnostik adäquatere Hilfen geben konnten, die zu einer höheren Aktivierung der Lernenden geführt haben.

Besonders interessant sind die Unterschiede hinsichtlich des Interventionsauslösers. In Teilkapitel 9.2, in dem alle Interventionen global analysiert wurden, wurde konstatiert, dass responsive Interventionen zielführender erscheinen, da diese in höherem Maße angemessen getätigt werden und dazu beitragen, dass zusätzliche Lernende aktiviert werden. Daher mag es auf den ersten Blick irritierend sein, dass Gruppe H nicht auf responsive Interventionen zurückgreift. In Verbindung mit den zuvor genannten Ergebnissen könnte dies jedoch derart interpretiert werden, dass die Studierenden nicht abwarten, bis die Schülerinnen und Schüler möglicherweise frustriert aufgeben, sondern bewusst invasive Interventionen tätigen, bevor die Frustrationsgrenze der Lernenden erreicht ist. Betrachtet man die Angemessenheit der Interventionen (unter Berücksichtigung des Interventionsauslösers), so erkennt man, dass die Studierenden der Gruppe H alle responsiven Interventionen angemessen durchführen und im Vergleich zu den anderen Gruppen auch invasive Interventionen zu einem höheren Anteil positiv gestalten. Die durchweg positiven Ergebnisse von Gruppe H sollten mindestens Denkanstoß sein, um die Schlussfolgerung aus Teilkapitel 9.4.2, dass überwiegend responsive Interventionen getätigt werden sollten, zu reflektieren. Womöglich ist dies zunächst als globale Strategie sinnvoll und sollte im Laufe der Zeit, mit steigenden diagnostischen Fähigkeiten, schrittweise zurückgenommen werden und durch invasive adaptive Interventionen ersetzt werden.

Eine weiterführende Diskussion der Ergebnisse findet in Kapitel 10 statt. An dieser Stelle wird nun, nachdem mehrere Indizien für konkrete Zusammenhänge zwischen der diagnostischen Fähigkeit „Deuten“ und den Interventionen in Gruppenarbeitsprozesse gefunden wurden, untersucht, ob es auch einen Zusammenhang zwischen der diagnostischen Fähigkeit „Deuten“ und dem gezeigten Reflexionsverhalten der Studierenden gibt.

9.5 Zusammenhänge zwischen Diagnose und Reflexion

Um mögliche Zusammenhänge zwischen diagnostischen Fähigkeiten und dem Reflexionsverhalten zu ergründen, wurden die Studierenden erneut anhand ihrer prozessdiagnostischen Fähigkeit „Deuten“ in drei Gruppen eingeteilt. Im Vergleich zu Teilkapitel 9.4 konnte für das vorliegende Teilkapitel die Aufzeichnung von zwei Studierenden leider aufgrund technischer Probleme nicht ausgewertet werden. Entsprechend können 46 Studierende in den Blick genommen werden. 15 Studierende entfallen auf Gruppe H, 16 Studierende auf Gruppe M und 15 auf Gruppe N.

Abbildung 57 zeigt die Verteilung der Reflexionsmomente auf die drei Hauptkategorien. Dabei hat Gruppe H insgesamt 771 Äußerungen getätigt, die als Reflexionsmomente erfasst wurden. Gruppe M kommt auf 966 und Gruppe N auf 836 Reflexionsmomente. Alle drei Gruppen reflektieren überwiegend auf einer oberflächlichen Ebene, was die Vielzahl an Äußerungen im Bereich „Rückblick“ zeigt. Am meisten Reflexionsmomente äußert dabei Gruppe M (776), gefolgt von 700 Reflexionsmomenten der Gruppe N. Mit 591 kodierten Äußerungen weist Gruppe H die wenigsten Reflexionsmomente im Bereich „Rückblick“ auf.

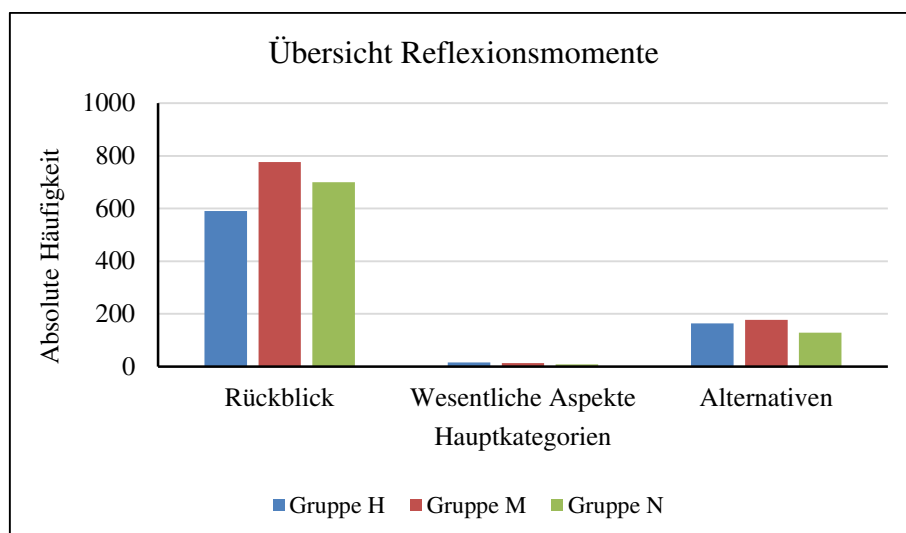


Abbildung 57: Übersicht über die absolute Häufigkeit der Reflexionsmomente aller Studierenden hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen", eingeteilt hinsichtlich der diagnostischen Fähigkeit "Deuten"

In allen drei Gruppen wurden nur sehr selten wesentliche Aspekte reflektiert. Für Gruppe N wurden nur acht Reflexionsmomente derart kodiert. Für Gruppe M waren es 13 und für Gruppe H immerhin 16 Reflexionsmomente. Etwas häufiger wurden mögliche Alternativen für die Lernumgebung bzw. für die getätigten Interventionen reflektiert. Gruppe M zählt hierbei 177 Reflexionsmomente, Gruppe H 164 und Gruppe N 128.

Die Übersicht über die absoluten Zahlen der Reflexionsmomente illustriert, dass die Reflexion in allen drei Gruppen auf einer groben Reflexionsebene stattfindet, kaum wesentliche Aspekte herausgearbeitet werden und nichtsdestotrotz einige mögliche Alternativen genannt werden. Um die Gruppen sinnvoll miteinander vergleichen zu können, muss die Anzahl der Reflexionsmomente in einem Bereich in Bezug zur Gesamtanzahl der Reflexionsmomente in der jeweiligen Gruppe gesetzt werden.

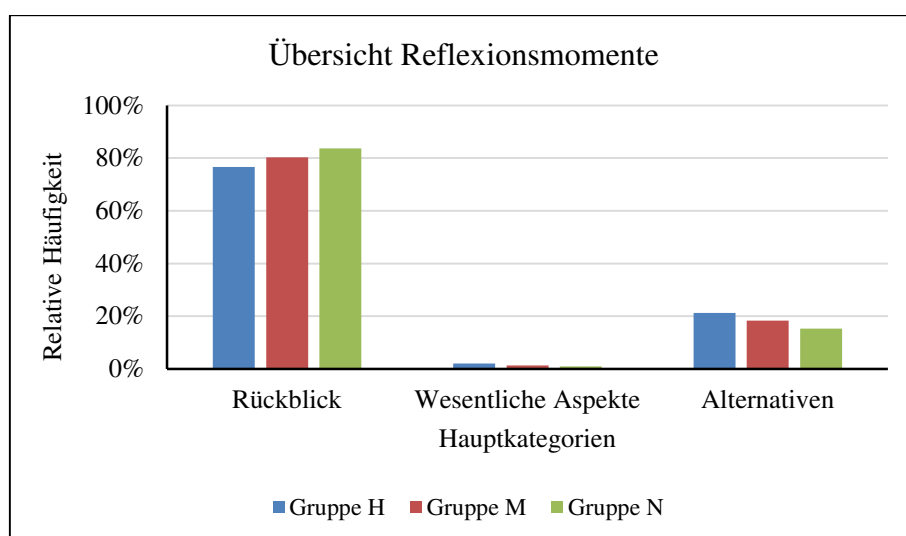


Abbildung 58: Übersicht über die relative Häufigkeit der Reflexionsmomente aller Studierenden hinsichtlich der drei Hauptkategorien „Rückblick“, „Wesentliche Aspekte“ und „Alternativen“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“

Betrachtet man den relativen Anteil der Reflexionsomente, so stellt man fest, dass Gruppe H weniger Reflexionsmomente im Bereich „Rückblick“ äußert als die anderen beiden Gruppen, jedoch vergleichsweise mehr wesentliche Aspekte und mehr Alternativen (vgl. Abb. 58). Aufgrund der geringen Unterschiede zwischen den drei Gruppen wäre es jedoch eine Überinterpretation der Ergebnisse, festzuhalten, dass sich Studierende mit erhöhter diagnostischer Fähigkeit „Deuten“ weniger mit oberflächlichen Aspekten befassen und dafür mehr wesentliche Aspekte ansprechen und Handlungsalternativen diskutieren. Die bestehenden Ergebnisse können maximal als leichte Tendenz in diese Richtung gesehen werden, sollten jedoch ohne weitere Erhebungen und Analysen nicht als gegeben hingenommen werden. An dieser Stelle kann auf oberster Ebene aber zumindest konstatiert werden, dass die Hypothese eines möglichen Zusammenhangs nicht widerrufen werden kann.

Da die Hypothese mit Blick auf die Hauptkategorien nicht vollends abwegig zu sein scheint, lohnt sich ein Blick auf die jeweiligen Unterkategorien. Zunächst werden die Reflexionsmomente innerhalb der Kategorie „Rückblick“ genauer untersucht.

Tabelle 41: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“

Unterkategorie	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick	Summe
Gruppe H	6% (35)	62% (364)	15% (90)	17% (102)	100% (591)
Gruppe M	5% (41)	69% (533)	13% (100)	13% (102)	100% (776)
Gruppe N	5% (35)	69% (483)	11% (75)	15% (107)	100% (700)

In Tabelle 41 sind sowohl die relativen als auch die absoluten Anteile der Reflexionsmomente angegeben. Zur besseren Vergleichbarkeit werden jeweils die relativen Anteile der Gruppen miteinander verglichen. Die absoluten Werte werden aus Gründen der Transparenz mit angegeben.

Vergleicht man die relativen Anteile der Reflexionsmomente miteinander, so erkennt man, dass es kaum Unterschiede zwischen den drei Gruppen gibt. Gruppe H äußert sich in der Kategorie „Rückblick“ etwas seltener zu inhaltlichen und organisatorischen Aspekten als die beiden anderen Gruppen. Hingegen nehmen die anderen drei Unterkategorien einen im Vergleich zu den anderen beiden Gruppen minimal größeren Stellenwert ein. Gruppe H tätigt also relativ gesehen mehr Äußerungen im Bereich „Absichten und Ziele“, „Gedanken und Gefühle“ sowie im Bereich „Grober Rückblick“.

Eine tiefere Interpretation erscheint an dieser Stelle aufgrund der geringen Unterschiede zwischen den Gruppen nicht sinnvoll. Stattdessen wird nun die Kategorie „Wesentliche Aspekte“ mit ihren drei Unterkategorien genauer betrachtet.

Tabelle 42: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“

Unterkategorie	Kritisches Hinterfragen	Organisatorisches	Inhaltliches	Summe
Gruppe H	25% (4)	38% (6)	38% (6)	100% (16)
Gruppe M	38% (5)	8% (1)	54% (7)	100% (13)
Gruppe N	50% (4)	13% (1)	38% (3)	100% (8)

Im Hinblick auf die prozentuale Verteilung der Reflexionsmomente auf die drei Unterkategorien lassen sich deutliche Unterschiede erkennen (vgl. Tab. 42). Für Gruppe N wurden mit 50% verhältnismäßig viele Äußerungen in der Unterkategorie „Kritisches Hinterfragen“ kodiert. An dieser Stelle erscheint jedoch der prozentuale Anteil wenig aussagekräftig, da insgesamt kaum wesentliche Aspekte erörtert wurden. Da dies für alle drei Gruppen gilt, erscheint es hierbei sinnvoller, die absoluten Zahlen zu betrachten und miteinander zu vergleichen. So lässt sich festhalten, dass alle drei Gruppen mit jeweils vier bzw. fünf Reflexionsmomenten in der Unterkategorie „Kritisches Hinterfragen“ in etwa gleich viele Kodierungen vorweisen können. In der Unterkategorie „Organisatorisches“ ist es überwiegend Gruppe H, die hierzu Äußerungen tätigt. Wesentliche inhaltliche Aspekte werden von Gruppe H und Gruppe M häufiger genannt als von Gruppe N.

Leider muss auch an dieser Stelle festgehalten werden, dass die Unterschiede nur minimal sind und entsprechend kaum Aussagekraft besitzen. Daher wird auch an dieser Stelle auf eine weiterführende Interpretation der Ergebnisse verzichtet und stattdessen die Kategorie „Alternativen“ in den Blick genommen.

Tabelle 43: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“

Unterkategorie	Nennung von Alternativen	Abwägung Vor- und Nachteile	Summe
Gruppe H	90% (148)	10% (16)	100% (164)
Gruppe M	87% (154)	13% (23)	100% (177)
Gruppe N	88% (113)	12% (15)	100% (128)

Hinsichtlich der Reflexionsmomente innerhalb der Kategorie „Alternativen“ gibt es keine nennenswerten Unterschiede zwischen den drei Gruppen (vgl. Tab. 43). Alle Gruppen nennen zu etwa 90% Alternativen, ohne dabei Vor- und Nachteile gegeneinander abzuwägen. Dies geschieht bei den drei Gruppen nur in jeweils rund 10% der kodierten Reflexionsmomente.

In Bezug auf eine alternative Gestaltung der Lernumgebung sowie hinsichtlich andersartig gestalteter Interventionen scheint es keine Unterschiede zwischen den drei Gruppen zu geben. Entsprechend scheint die Ausprägung der prozessdiagnostischen Fähigkeit „Deuten“ an dieser

Stelle keinen Einfluss darauf zu haben, in welcher Häufigkeit die Alternativen lediglich genannt werden und in welcher Häufigkeit dabei Vor- und Nachteile gegeneinander abgewogen werden.

Eine letzte Analyse, um einen möglichen Zusammenhang zwischen der prozessdiagnostischen Fähigkeit „Deuten“ und dem Reflexionsverhalten aufzudecken, besteht in der Betrachtung der Reflexionsmomente, die sich auf die Lernumgebung beziehen, und denen, die sich auf die Intervention beziehen.

Tabelle 44: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der diagnostischen Fähigkeit „Deuten“

Kategorie	Gruppe	Lernumgebung	Intervention
Rückblick	H	59% (347)	41% (244)
	M	65% (501)	35% (275)
	N	66% (460)	34% (240)
Wesentliche Aspekte	H	69% (11)	31% (5)
	M	54% (7)	46% (6)
	N	50% (4)	50% (4)
Alternativen	H	87% (142)	13% (22)
	M	85% (151)	15% (26)
	N	81% (104)	19% (24)

Tabelle 44 führt für alle drei Gruppen jeweils die Hauptkategorien auf, wobei dort zwischen Reflexionsmomenten, die sich auf die Lernumgebung beziehen, und Reflexionsmomenten, die sich auf die getätigten Interventionen beziehen, unterschieden wird. Für den Vergleich innerhalb einer Gruppe ist die absolute Zahl ausreichend (dieser Wert befindet sich jeweils in Klammern). Um die drei Gruppen miteinander vergleichen zu können, wird wiederum auf den relativen Anteil zurückgegriffen.

Mit Blick auf die Kategorie „Rückblick“ lässt sich lediglich ein geringer Unterschied zwischen den Gruppen feststellen. Alle drei Gruppen haben mehr Reflexionsmomente in Bezug auf die Lernumgebung als in Bezug auf die Interventionen formuliert. Gruppe H weist hierbei die prozentual gesehen geringste Differenz auf. Durchschnittlich etwa zwei von fünf Äußerungen referieren auf die Interventionen, die anderen drei beziehen sich entsprechend auf die Lernumgebung. Die anderen beiden Gruppen verzeichnen in etwa ein Verhältnis von zwei Reflexionsmomenten hinsichtlich der Lernumgebung zu einem Reflexionsmoment hinsichtlich der getätigten Interventionen.

Im Fall der Kategorie „wesentliche Aspekte“ gibt es prozentual einen größeren Unterschied zwischen Gruppe H (69%) und den anderen beiden Gruppen (54% bzw. 50%). Allerdings ist die Anzahl der Reflexionsmomente in dieser Kategorie so gering, dass es sich auch um einen Zufall handeln könnte. Abgesehen von Gruppe H ist das Verhältnis zwischen Reflexionsmomenten, die sich auf die Lernumgebung beziehen, und Reflexionsmomenten, die sich auf die Interventionen beziehen, in etwa ausgeglichen.

In Bezug auf die genannten bzw. diskutierten Alternativen ist dies nicht der Fall. Alle drei Gruppen äußern deutlich mehr Reflexionsmomente, die sich auf die Lernumgebung beziehen (zwischen 81% und 87%). Es scheint also allen Studierenden, unabhängig von deren diagnostischer Fähigkeit „Deuten“, leichter zu fallen, alternative Gestaltungen für die Lernumgebung zu finden als für die getätigten Interventionen. Womöglich ist der Unterschied aber auch dem

Setting geschuldet, da alle Studierenden bei der Erprobung der kompletten Laborstation anwesend sind. Nur bei einer der drei Teilstationen beobachten sie die Filmgruppe genauer und intervenieren in deren Arbeitsprozess. Das bedeutet, dass sie im Rahmen des Reflexionsgesprächs zunächst in einem zeitlichen Ungleichgewicht zwischen Erprobung der Lernumgebung und Interventionen in den Gruppenarbeitsprozess der Schülerinnen und Schüler sind. Daher erscheint es nachvollziehbar, dass der Großteil der genannten Handlungsalternativen sich auf die Lernumgebung bezieht.

Insgesamt muss konstatiert werden, dass anhand der ausgewerteten Daten kaum erkennbare Zusammenhänge zwischen der prozessdiagnostischen Fähigkeit „Deuten“ und dem Reflexionsverhalten der Studierenden festgestellt werden konnten. Auf Ebene der Hauptkategorien gibt es eine schwache Tendenz dahingehend, dass Studierende mit höherer prozessdiagnostischer Fähigkeit „Deuten“ relativ gesehen etwas mehr wesentliche Aspekte herausarbeiten und etwas mehr Alternativen generieren. Allerdings sind die Unterschiede nicht sonderlich groß und entsprechend auch nur geringfügig aussagekräftig. Der Zusammenhang zwischen Diagnose und Reflexion kann durch die Ergebnisse zumindest nicht verworfen werden. Es sind jedoch weitere Studien notwendig, um mögliche Zusammenhänge zu ergründen. Die Analyse der Unterkategorien erwies sich hierfür als nicht zielführend und zeigte ein gemischtes Bild, aus dem keine Zusammenhänge geschlussfolgert werden können.

9.6 Zusammenhänge zwischen Intervention und Reflexion

In diesem Teilkapitel werden mögliche Zusammenhänge zwischen Interventionsverhalten und Reflexionsverhalten der Studierenden ergründet. Dazu müssen die Studierenden zunächst in Gruppen unterteilt werden, wofür es mehrere Möglichkeiten gibt.

Beispielsweise kann man die Studierenden auf Grundlage der Anzahl der jeweils vorgenommenen Interventionen einteilen. Hierfür erscheint es plausibel, die durchschnittliche Anzahl der Interventionen als Fixpunkt zu bestimmen und die Studierenden dann in eine Gruppe einzuteilen, die weniger Interventionen vorgenommen hat, und eine Gruppe, die mehr Interventionen vorgenommen hat. Da die Studierenden durchschnittlich zwischen vier und fünf Interventionen vorgenommen haben, ließen sich also eine Gruppe bilden, in der die Studierenden mindestens eine und maximal vier Interventionen getätigt haben, und eine Gruppe, in der die Studierenden mindestens fünf Interventionen durchgeführt haben.

Eine weitere Option zur Einteilung basiert auf der Angemessenheit der Interventionen. Dabei wird ähnlich wie bei der zuvor beschriebenen Variante der Mittelwert herangezogen, sprich wie viel Prozent der Interventionen insgesamt angemessen waren. Anschließend betrachtet man alle Studierenden und ordnet diese, je nachdem, ob sie prozentual mehr oder weniger angemessene Interventionen getätigt haben als der Durchschnitt, zu der jeweiligen Gruppe zu.

Eine eher inhaltliche Einteilung kann anhand der verschiedenen Hilfekategorien vorgenommen werden. Theoretisch ist dabei eine Vielzahl an Möglichkeiten denkbar. Es erscheint eine plausible Einteilung zu sein, die beiden Hilfekategorien „inhaltsorientiert-strategische Hilfen“ und „inhaltliche Hilfen“ zu einer Gruppe zusammenzuführen und die anderen vier Hilfekategorien, die weniger inhaltlich sind, ebenfalls zu gruppieren. Dies kann dadurch geschehen, dass die Studierenden je nach dem Maß, in dem ihre dominanten Hilfen inhaltlicher Natur waren, der entsprechenden Gruppe zugeordnet werden.

Da sich alle genannten Gruppeneinteilungen inhaltlich grundlegend unterscheiden und entsprechend eine andere Perspektivierung liefern, wobei dadurch ein breites Spektrum an poten-

ziell denkbaren Zusammenhängen zwischen dem Interventionsverhalten und dem Reflexionsverhalten abgedeckt wird, erscheint es sinnvoll, die Analyse anhand aller genannten Gruppeneinteilungen vorzunehmen. Dadurch wird keiner der verschiedenen Blickwinkel vernachlässigt. Jedoch muss aufgrund der Vielzahl an theoretisch denkbaren Zusammenhängen auch jeweils inhaltlich überprüft werden, ob ein vermeintlicher Zusammenhang inhaltlich zu begründen ist. Denn je mehr Untersuchungen vorgenommen werden, desto wahrscheinlicher ist es, dass auch Zusammenhänge abgebildet werden, die sich in Folgestudien ggf. nicht bestätigen lassen können. Daher ist es wichtig, auftretende Zusammenhänge auch inhaltlich begründen zu können. Zunächst werden mögliche Zusammenhänge unter Einteilung nach der Angemessenheit der Interventionen analysiert (Kap. 9.6.1). Anschließend werden die anderen beiden Einteilungen vorgenommen (Kap. 9.6.2 und Kap. 9.6.3).

9.6.1 Gruppeneinteilung nach Angemessenheit der Interventionen

Die Studierenden in Bezug auf die Angemessenheit ihrer getätigten Interventionen in zwei Gruppen zu unterscheiden, bietet die Möglichkeit, potenzielle Unterschiede im Reflexionsverhalten von Studierenden, die tendenziell mehr angemessene Interventionen durchgeführt haben, und denjenigen, die seltener angemessen interveniert haben, auszumachen. Um dies zu ergründen, werden analog zu dem Vorgehen im vorigen Teilkapitel zunächst die Hauptkategorien untersucht. Anschließend werden die Unterkategorien der verschiedenen Hauptkategorien genauer betrachtet.

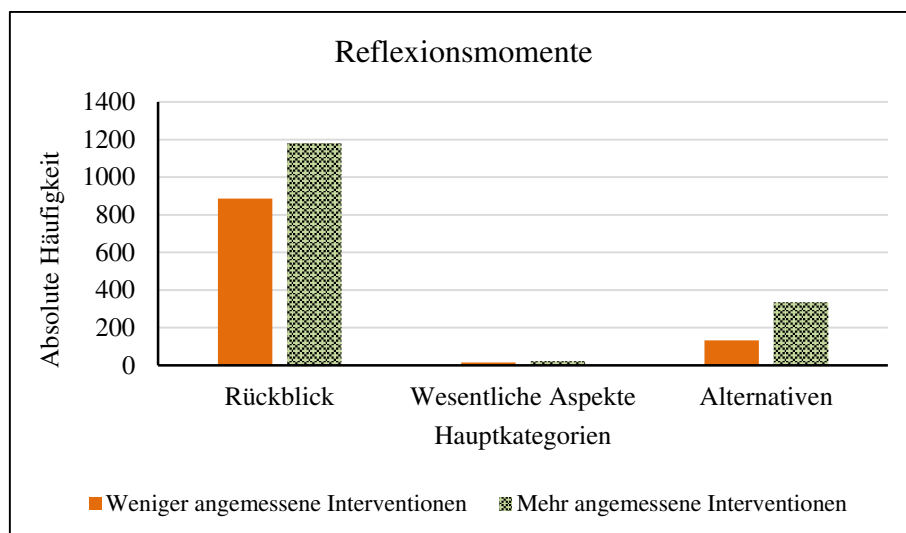


Abbildung 59: Übersicht über die absolute Häufigkeit der Reflexionsmomente der zwei Gruppen von Studierenden auf Grundlage der Einteilung nach der Angemessenheit der getätigten Interventionen

In Abbildung 59 lässt sich erkennen, dass die Studierenden, die mehr angemessene Interventionen getätigt haben als der Durchschnitt, in allen drei Hauptkategorien mehr Reflexionsmomente äußern als die Gruppe von Studierenden, die weniger angemessen interveniert haben. Während die Studierenden mit angemesseneren Interventionen knapp 1200 Reflexionsmomente im Bereich „Rückblick“ und 22 Reflexionsmomente im Bereich „Wesentliche Aspekte“ äußern sowie über 300 Alternativen diskutieren, liegt die Anzahl der Reflexionsmomente der anderen Gruppe bei knapp 900 im Bereich „Rückblick“, 15 im Bereich „Wesentliche Aspekte“ und 133 im Bereich „Alternativen“. Absolut gesehen liegt also der größte Unterschied bei den

Äußerungen, die vergleichsweise oberflächlich sind. Relativ gesehen ist der Unterschied bei den Alternativen am größten.

Um die Ergebnisse nicht verzerrt darzustellen und auf dieser verzerrten Grundlage falsch zu interpretieren, muss die Anzahl der Studierenden in den beiden Gruppen berücksichtigt werden. Es gab 25 Studierende, die durchschnittlich angemessener agiert haben als die durchschnittliche Anzahl an angemessenen Interventionen. Auf der anderen Seite intervenierten 21 Studierende im Vergleich zum Durchschnitt weniger angemessen. Um dem gerecht zu werden, können die Reflexionsmomente einer Gruppe auf die Gruppengröße der anderen Gruppe normiert werden. Indem man die Reflexionsmomente mit der Anzahl der Studierenden der anderen Gruppe multipliziert und durch die Anzahl der eigenen Gruppe teilt sowie das Ergebnis falls nötig auf eine ganze Zahl rundet, erhält man nachfolgende Verteilung von Reflexionsmomenten (vgl. Tab. 45).

Tabelle 45: Normierte Übersicht über die Reflexionsmomente der Studierenden, eingeteilt nach der Angemessenheit der Interventionen

Kategorie	Mehr angemessen	Weniger angemessen
Rückblick	1181	1055
Wesentliche Aspekte	22	18
Alternativen	336	158

Durch die Normierung der Reflexionsmomente zeigt sich, dass der Unterschied in der Kategorie „Rückblick“ nur noch gut 10% beträgt, während der Unterschied in der Kategorie „Alternativen“ am größten ist. Dort besprechen Studierende, die mehr angemessene Interventionen getätigt haben als der Durchschnitt, mehr als doppelt so viele mögliche Handlungsalternativen bzw. alternative Gestaltungen der Lernumgebung als die anderen Studierenden. Dies ist insofern interessant, als man auch vermuten könnte, dass Studierende, die weniger angemessene Interventionen tätigen, eher Änderungen in ihrem Betreuungsverhalten benennen bzw. diskutieren können. Andererseits erscheint es plausibel, dass die Studierenden, die bereits angemessener mit den Schülerinnen und Schülern interagieren, auch über ein größeres Repertoire an potenziell denkbaren Handlungen verfügen könnten. Des Weiteren erscheint es plausibel, dass Studierende, die sich nachträglich intensiver mit den Situationen auseinandersetzen, künftig auch bessere Gestaltungen der Lernumgebung bzw. der Interventionen vornehmen. Entsprechend liegt die Vermutung nahe, dass die Studierenden nicht zum ersten Mal über ihr Verhalten bzw. über Aufgaben reflektieren und aus diesem Grund deutlich mehr Vorschläge erarbeiten.

Bezüglich der wesentlichen Aspekte, die von den Studierenden reflektiert wurden, lässt sich nur ein geringer Unterschied zwischen den beiden Gruppen feststellen, weshalb dieser auch nicht weiter interpretiert wird. Es lässt sich jedoch auch auf Grundlage der normierten Werte festhalten, dass die Gruppe der Studierenden, die eher angemessene Interventionen getätigt haben, in allen drei Bereichen mehr Reflexionsmomente geäußert hat. Nachfolgend werden nun die Unterkategorien der einzelnen Bereiche betrachtet.

Tabelle 46: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen

Angemessenheit der Interventionen	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick	Summe
Weniger angemessen	5% (47)	66% (582)	12% (110)	17% (147)	100% (886)
Mehr angemessen	5% (64)	68% (798)	13% (155)	14% (164)	100% (1181)

Bei den Unterkategorien der Kategorie „Rückblick“ gibt es keine großen Unterschiede zwischen den beiden Gruppen (vgl. Tab. 46). Beide Gruppen reflektieren mit etwa zwei Drittel der getätigten Äußerungen überwiegend über konkrete Inhalte bzw. über organisatorische Aspekte. Am seltensten wurde jeweils die Unterkategorie „Absichten und Ziele“ kodiert. Etwa jede zwanzigste Äußerung innerhalb der Kategorie „Rückblick“ ist dieser Unterkategorie zuzuordnen. Gedanken und Gefühle werden in 12% bzw. in 13% der Fälle angesprochen. Etwas öfter werden grobe Rückblicke über die Erprobung oder die Interventionen gegeben.

Bezüglich des Reflexionsverhaltens können für die Kategorie „Rückblick“ keine Unterschiede zwischen den Studierenden, die angemessener und weniger angemessen intervenieren, festgestellt werden. Entsprechend werden die Ergebnisse an dieser Stelle nicht weiter interpretiert. Stattdessen wird nun die Kategorie „Wesentliche Aspekte“ genauer analysiert.

Tabelle 47: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen

Angemessenheit der Interventionen	Kritisches Hinterfragen	Organisatorisches	Inhaltliches	Summe
Weniger angemessen	67% (10)	13% (2)	20% (3)	100% (15)
Mehr angemessen	14% (3)	27% (6)	59% (13)	100% (22)

Anhand von Tabelle 47 wird ersichtlich, dass es in der Kategorie „Wesentliche Aspekte“ deutliche Unterschiede zwischen den kodierten Reflexionsmomenten der beiden Gruppen gibt. Die Studierenden, die weniger angemessene Interventionen getätigt haben, reflektieren ihr eigenes Verhalten häufiger kritisch als ihre Kommilitoninnen und Kommilitonen. Zwei von drei Reflexionsmomenten aus der Kategorie „Wesentliche Aspekte“ entfallen auf diese Unterkategorie. Den Studierenden wurde zu keinem Zeitpunkt mitgeteilt, ob ihre Intervention als angemessen eingestuft oder eher als unangemessen angesehen wurde. Die Studierenden scheinen teilweise selbst unzufrieden mit ihren Interventionen zu sein und gehen daher auch selbstkritisch damit um. Entsprechend kann das Reflexionsverhalten der Studierenden, die weniger angemessen interveniert haben, als Bestätigung für die Einschätzung der Expertinnen und Experten angesehen werden.

Auch aus einem zweiten Grund kann dieser Befund positiv gewertet werden, denn die Studierenden scheinen (zumindest teilweise) selbst festzustellen, wenn sie unangemessen mit

Schülerinnen und Schülern interagieren. Studierende, die angemessener intervenieren, hinterfragen ihr Handeln in weniger Fällen kritisch. Auch dies spricht für eine gute Wahrnehmung der Situationen durch die Studierenden.

Die Gruppe von Studierenden mit angemesseneren Interventionen äußert hingegen deutlich mehr Reflexionsmomente im Bereich „Inhaltliches“, spricht über mehr konkrete Inhalte heraus, die in ihren Augen für das gezeigte Arbeitsverhalten der Gruppe verantwortlich waren (beispielsweise wurden konkrete fachliche Fähigkeiten benannt, die den Schülerinnen und Schülern fehlten, aber an einer bestimmten Stelle benötigt wurden, um sinnvoll weiterarbeiten zu können). Die Studierenden, die weniger angemessen interveniert haben, weisen in dieser Unterkategorie lediglich drei (bzw. vier, wenn die normierte Anzahl betrachtet wird) Reflexionsmomente auf, während die anderen Studierenden 13 Reflexionsmomente äußern. Dies könnte durch die Erfahrungen aus den Interventionen zustande kommen. Womöglich könnte dies aber auch auf unterschiedliche fachliche bzw. fachdidaktische Kenntnisse zurückzuführen sein. Letzteres lässt sich jedoch im Rahmen dieser Studie nicht überprüfen.

Auch in der Unterkategorie „Organisatorisches“ gibt es kleine Unterschiede zwischen den Studierenden. Studierende, die angemessener intervenieren, reflektieren auch etwas häufiger über organisatorische Aspekte, die aus ihrer Sicht wesentlich waren. Dies können beispielsweise fehlerhafte Materialien oder Simulationen sein, die die Studierenden nicht ordentlich überprüft haben, bzw. Fehler, die bei einer Überprüfung vor Erprobung der Station aufgefallen wären. Da die Unterschiede an dieser Stelle trotz der kleinen Fallzahlen sowohl absolut (bezogen auf die Reflexionsmomente) als auch relativ (bezogen auf den Anteil der Unterkategorie innerhalb der Kategorie) gesehen gering sind, wird an dieser Stelle auf eine tiefere Analyse und Interpretation verzichtet. Es folgt die Analyse der Kategorie „Alternativen“.

Tabelle 48: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen

Angemessenheit der Interventionen	Nennung von Alternativen	Abwägung Vor- und Nachteile	Summe
Weniger angemessen	95% (126)	5% (7)	100% (133)
Mehr angemessen	86% (289)	14% (47)	100% (336)

Mit Blick auf Tabelle 48 lässt sich festhalten, dass die Studierenden, die mehr angemessene Interventionen tätigen, mehr Vor- und Nachteile reflektieren. Dieser Effekt ist umso größer, wenn man berücksichtigt, dass diese Gruppe von Studierenden ohnehin deutlich mehr Reflexionsmomente in diesem Bereich aufweist. In absoluten Zahlen entfallen 47 Reflexionsmomente der Gruppe auf diese Unterkategorie, während die Studierenden mit weniger angemessenen Interventionen sieben (normiert acht) Reflexionsmomente aufweisen. Studierende, die weniger angemessene Interventionen tätigen, beschränken sich überwiegend auf die Nennung von möglichen Alternativen. Dagegen wägt die andere Gruppe häufiger mögliche Vor- und Nachteile einer alternativen Gestaltung der Lernumgebung bzw. einer alternativen Intervention gegeneinander ab.

Die Ergebnisse sind insofern nachvollziehbar, als Studierende, die sich über verschiedene Handlungsalternativen Gedanken machen und diese gegeneinander abwägen, auch durchdachtere Interventionen tätigen. Entsprechend erscheint es sinnvoll, dass diese auch in höherem

Maße angemessen sind. Zwar mussten die Studierenden in Echtzeit auf die Schülerinnen und Schüler reagieren, aber womöglich hat die Auseinandersetzung mit den verschiedenen Aufgaben, dem Material und den Simulationen, bei denen womöglich ebenfalls Vor- und Nachteile möglicher Alternativen gegeneinander abgewogen wurden, dazu beigetragen, dass die Studierenden besser auf die Umstände reagieren konnten als Studierende, die solche Überlegungen nicht anstellten. Hierbei handelt es sich jedoch um eine Vermutung, die anhand der vorliegenden Daten nicht belegt werden kann. Anhand des gezeigten Interventions- und Reflexionsverhaltens der Studierenden erscheint dies aber zumindest plausibel.

Abschließend werden nun die Reflexionsmomente beider Gruppen dahingehend analysiert, ob diese sich auf die Lernumgebung oder die getätigten Interventionen beziehen. Dazu sind in Tabelle 49 in vier Spalten die Kategorien, die Angemessenheit der Interventionen sowie der Anteil an Reflexionsmomenten, die sich auf die Lernumgebung beziehen, und der Anteil an Reflexionsmomenten, die sich auf die Interventionen beziehen, aufgeführt. Neben dem relativen Anteil wird auch der absolute Anteil in Klammern aufgeführt.

Tabelle 49: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Angemessenheit der getätigten Interventionen

Kategorie	Angemessenheit der Interventionen	Lernumgebung	Intervention
Rückblick	Weniger angemessen	63% (557)	37% (329)
	Mehr angemessen	64% (751)	36% (430)
Wesentliche Aspekte	Weniger angemessen	33% (5)	67% (10)
	Mehr angemessen	77% (17)	23% (5)
Alternativen	Weniger angemessen	79% (105)	21% (28)
	Mehr angemessen	87% (292)	13% (44)

Bezogen auf die Kategorie „Rückblick“ sind keine Unterschiede zwischen den beiden Gruppen zu erkennen, betrachtet man die prozentuale Verteilung zwischen Reflexionsmomenten zu der Lernumgebung und Reflexionsmomenten zu den Interventionen. Im Fall der wesentlichen Aspekte hingegen reflektieren Studierende, die weniger angemessen intervenieren, häufiger auf Grundlage der Interventionen, während Studierende, die mehr angemessene Interventionen tätigen, eher wesentliche Aspekte reflektieren, die sich auf die Lernumgebung beziehen. Dieser Unterschied ist einerseits darauf zurückzuführen, dass Studierende mit unangemessenen Interventionen eher kritisch über diese reflektieren. Andererseits reflektieren Studierende, die angemessener intervenieren, häufiger über wesentliche Aspekte, die sich auf die Inhalte der Lernumgebung beziehen, als Studierende, die weniger angemessen intervenieren. Wesentliche Aspekte, die sich auf konkrete Inhalte beziehen, sind in fast allen Fällen Reflexionsmomente, die der Lernumgebung zugeordnet wurden. Lediglich ein Fall wurde im Zuge der Kodierung den

Interventionen zugeordnet. Gleiches gilt für die wesentlichen Aspekte, die sich auf organisatorische Aspekte beziehen. Auch hier gibt es nur einen Fall, der nicht der Lernumgebung zugewiesen wurde. Kritisches Hinterfragen hingegen bezieht sich immer auf die eigene Person und wurde somit auch in allen Fällen der Intervention zugeordnet. Daher ergibt sich die Abweichung aus der verhältnismäßig großen Anzahl an Reflexionsmomenten in der Unterkategorie „Kritisches Hinterfragen“, die die Studierenden mit weniger angemessenen Interventionen tätigten.

In der Kategorie „Alternativen“ beziehen sich die meisten Reflexionsmomente beider Gruppen auf die Lernumgebung. Der prozentuale Anteil ist bei der Gruppe, die angemessener interveniert, noch höher als bei der anderen Gruppe. Hierbei muss jedoch beachtet werden, dass massive Unterschiede in den absoluten Zahlen zugrunde liegen. Trotz des geringeren relativen Anteils werden weit mehr als doppelt so viele Reflexionsmomente getätigt, die sich auf die Lernumgebung beziehen. Dies gilt auch, wenn man die normierten Zahlen betrachtet (indem man die Reflexionsmomente mit der Anzahl der Studierenden der anderen Gruppe multipliziert und durch die Anzahl der Studierenden der eigenen Gruppe dividiert).

Es kann also davon ausgegangen werden, dass sich die Gruppe der Studierenden, die angemessener interveniert, nachträglich deutlich intensiver mit den getätigten Interventionen auseinandersetzt. Man kann vermuten, dass dieser Effekt bereits vor den Interventionen in einer ähnlichen Form vorherrschte und dies einen Einfluss auf die Interventionen hatte. Ebenso lässt sich annehmen, dass die Ergebnisse der Analyse der Angemessenheit der Intervention in einer ähnlich gelagerten Folgestudie mit den gleichen Probanden wieder in einer ähnlichen Form ausfielen, da die Studierenden, die bereits angemessener interveniert haben, diese Interventionen auch noch intensiver reflektiert haben. Daher ist es wahrscheinlich, dass diese Studierenden auch in weiteren, ähnlich gelagerten Situationen wiederum angemessener intervenieren würden.

9.6.2 Gruppeneinteilung nach Anzahl der Interventionen

Anhand der Einteilung der Studierenden auf Basis der Angemessenheit der Interventionen konnten vereinzelt Hinweise auf mögliche Zusammenhänge herausgearbeitet werden. An anderen Stellen gab es keine Unterschiede zwischen den Studierenden. Die Probanden werden nun in zwei andere Gruppen unterschieden. Bei der hier vorgenommenen Einteilung wird die Anzahl der getätigten Interventionen als Unterscheidungsmerkmal herangezogen. Entsprechend wird in diesem Teilkapitel untersucht, ob die Studierenden ein unterschiedliches Reflexionsverhalten zeigen, je nachdem, ob sie öfter bzw. weniger oft interveniert haben. Durchschnittlich wurden zwischen vier und fünf Interventionen getätigt. Daher werden die Studierenden in eine Gruppe mit vier oder weniger und eine Gruppe mit fünf oder mehr Interventionen geteilt.

In Abbildung 60 sind die Reflexionsmomente der beiden Gruppen für die verschiedenen Kategorien zu sehen. Es lässt sich erkennen, dass die beiden Gruppen in allen drei Kategorien in etwa dieselbe Anzahl an Reflexionsmomenten geäußert haben. An dieser Stelle muss jedoch darauf verwiesen werden, dass die Gruppen unterschiedlich groß sind. 25 Studierende haben bis zu vier Interventionen getätigt, 21 Studierende haben mehr als vier (zwischen fünf und neun) Interventionen getätigt. Analog zum vorigen Teilkapitel können die Reflexionsmomente einer Gruppe auch an dieser Stelle auf die Gruppengröße der anderen Gruppe normiert werden. Indem man die Reflexionsmomente mit der Anzahl der Studierenden der anderen Gruppe multipliziert und durch die Anzahl der eigenen Gruppe teilt sowie das Ergebnis falls nötig auf eine

ganze Zahl rundet, erhält man nachfolgende Verteilung von Reflexionsmomenten (vgl. Tab. 50).

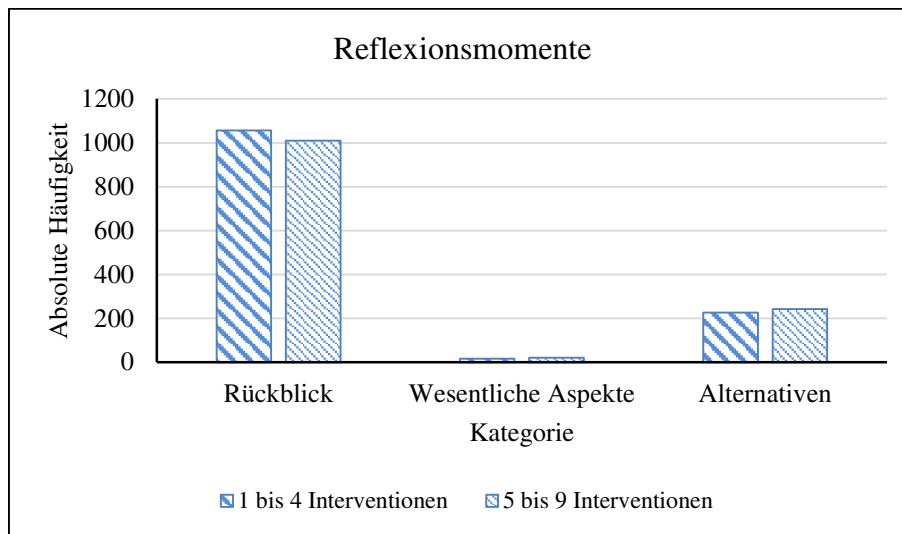


Abbildung 60: Übersicht über die absolute Häufigkeit der Reflexionsmomente der zwei Gruppen von Studierenden auf Grundlage der Einteilung nach der Anzahl der getätigten Interventionen

Tabelle 50: Normierte Übersicht über die Reflexionsmomente der Studierenden, eingeteilt nach der Anzahl der Interventionen

Kategorie	1 bis 4 Interventionen	5 bis 9 Interventionen
Rückblick	1057	1202
Wesentliche Aspekte	16	25
Alternativen	227	288

Es zeigt sich, dass die Studierenden, die häufiger interveniert haben, mehr Reflexionsmomente in allen drei Kategorien äußern. Absolut ist der Unterschied in der Kategorie Rückblick am größten, was jedoch an der Vielzahl an Äußerungen in diesem Bereich liegt. Prozentual gesehen ist der Unterschied dort nämlich am geringsten. In Bezug auf die wesentlichen Aspekte verhält es sich umgekehrt, dort ist der relative Unterschied am größten und der absolute Unterschied am geringsten. Letzteres liegt an den insgesamt wenigen Reflexionsmomenten in dieser Kategorie. Im Fall der Alternativen ist der Unterschied sowohl absolut als auch relativ gesehen jeweils zwischen den beiden anderen Kategorien.

Im Folgenden wird genauer untersucht, wo die Unterschiede der beiden Gruppen in ihrem Reflexionsverhalten liegen. Dazu werden die einzelnen Kategorien unter die Lupe genommen. Anschließend wird untersucht, wie die Reflexionsmomente sich auf die Lernumgebung und die getätigten Interventionen verteilen. Hierbei könnte man vermuten, dass die Studierenden, die öfter interveniert haben, auch mehr Reflexionsmomente tätigen, die sich auf die Interventionen beziehen, als die Gruppe, die seltener interveniert hat.

Tabelle 51: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen

Anzahl an Interventionen	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick	Summe
1 bis 4	5% (57)	64% (675)	13% (138)	18% (187)	100% (1057)
5 bis 9	5% (54)	70% (705)	13% (127)	12% (124)	100% (1010)

Betrachtet man die Unterkategorien der Kategorie „Rückblick“, so sieht man, dass die prozentuale Verteilung der Reflexionsmomente auf die einzelnen Unterkategorien in den beiden Gruppen recht ähnlich ist (vgl. Tab. 51). Bei den beiden Unterkategorien „Absichten und Ziele“ (5%) und „Gedanken und Gefühle“ (13%) ist der prozentuale Anteil in den beiden Gruppen sogar gleich groß. Ein kleiner Unterschied liegt im Fall der anderen beiden Unterkategorien vor. Während sich die Studierenden, die mehr Interventionen getätigt haben, prozentual gesehen noch etwas mehr auf inhaltliche und organisatorische Themen fokussieren, geben die Studierenden, die seltener interveniert haben, etwas häufiger einen groben Rückblick über die Lernumgebung und die Interventionen als die anderen.

Die Studierenden, die insgesamt seltener interveniert haben, äußern sich womöglich deshalb häufiger überblicksartig, weil sie weniger Erfahrungen sammeln konnten als ihre Kommilitoninnen und Kommilitonen. Entsprechend äußern sie sich öfter in Form eines groben Rückblicks als die andere Gruppe. Jedoch sind die Unterschiede zwischen den beiden Gruppen recht gering, sodass die vorliegenden Ergebnisse sowie die mögliche Interpretation mit großer Vorsicht zu betrachten sind. Es handelt sich um einen denkbaren Zusammenhang, der durch gezielte Folgeuntersuchungen weiter beleuchtet werden könnte.

Tabelle 52: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen

Anzahl an Interventionen	Kritisches Hinterfragen	Organisatorisches	Inhaltliches	Summe
1 bis 4	19% (3)	19% (3)	63% (10)	100% (16)
5 bis 9	48% (10)	24% (5)	29% (6)	100% (21)

Bei der Analyse der Kategorie „Wesentliche Aspekte“ gibt es aufgrund der geringen Anzahl an Reflexionsmomenten große Unterschiede in der prozentualen Verteilung bei den beiden Gruppen (vgl. Tab. 52). Studierende, die häufiger intervenieren, hinterfragen ihr eigenes Handeln häufiger kritisch als Studierende, die seltener intervenieren. Durch die höhere Anzahl an getätigten Interventionen konnten die Studierenden womöglich mehr Aspekte in den Blick nehmen, die sie aus ihrer Sicht nicht vollends zufriedenstellend gelöst haben und die sie im Nachgang nun kritisch sehen.

Im vorigen Teilkapitel wurde gezeigt, dass sich Studierende, die unangemessener intervenieren, häufiger kritisch hinterfragen. Daher könnte man vermuten, dass es eine gewisse Schnittmenge zwischen den Studierenden, die häufiger intervenieren, und den Studierenden,

die unangemessener agieren, gibt. Am Ende des Teilkapitels wird diesem Verdacht nachgegangen.

Studierende, die seltener intervenieren, reflektieren prozentual gesehen häufiger über „Inhaltliches“ als ihre Kommilitoninnen und Kommilitonen. Allerdings ist der absolute Unterschied eher gering. Vergleicht man die getätigten Reflexionsmomente, verzeichnet die eine Gruppe zehn und die andere sechs, wobei der normierte Wert bei sieben liegt. Dieser Unterschied erscheint zu gering, als dass dem weiter nachgegangen werden müsste. Gleiches gilt auch für die Unterkategorie „Organisatorisches“.

Tabelle 53: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen

Anzahl an Interventionen	Nennung von Alternativen	Abwägung Vor- und Nachteile	Summe
1 bis 4	89% (203)	11% (24)	100% (227)
5 bis 9	88% (212)	12% (30)	100% (242)

Betrachtet man die beiden Unterkategorien der Kategorie „Alternativen“, so ist festzustellen, dass es keine nennenswerten Unterschiede zwischen den beiden Gruppen gibt (vgl. Tab. 53). Beide Gruppen nennen größtenteils Alternativen, ohne Vor- und Nachteile dieser abzuwägen. Dies geschieht lediglich in elf bzw. zwölf Prozent der Fälle. Insgesamt lässt sich lediglich festhalten, dass Studierende, die häufiger intervenieren, auch etwas mehr Alternativen besprechen. Vergleicht man die normierten Werte miteinander, so stehen den 227 Reflexionsmomenten der Gruppe, die weniger interveniert hat, 288 Reflexionsmomente der Gruppe, die öfter interveniert hat, gegenüber.

Auf die prozentuale Verteilung der Reflexionsmomente auf die beiden Unterkategorien scheint die Anzahl der Interventionen keinen Einfluss zu haben. Ob die höhere Anzahl an Reflexionsmomenten eher Äußerungen entstammt, die sich auf Interventionen beziehen, oder Äußerungen, die auf die Lernumgebung referieren, wird nachfolgend untersucht.

In Tabelle 54 sind die Reflexionsmomente der drei Kategorien für beide Gruppen hinsichtlich der Bereiche Lernumgebung und Intervention unterschieden. Im Hinblick auf die Kategorie „Alternativen“ lässt sich kein Unterschied zwischen den beiden Gruppen erkennen, betrachtet man die prozentuale Verteilung der Reflexionsmomente. Entsprechend ist die höhere Anzahl an Reflexionsmomenten in dieser Kategorie, welche die Studierenden mit mehr Interventionen getätigt haben, nicht auf die höhere Anzahl an Interventionen zurückzuführen. Der Unterschied der beiden Gruppen muss in dieser Kategorie einen anderen Ursprung haben. Ähnliches gilt für die Kategorie „Rückblick“. Auch hier gibt es keine nennenswerten prozentualen Unterschiede zwischen den beiden Gruppen.

Tabelle 54: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Anzahl der getätigten Interventionen

Kategorie	Anzahl der Interventionen	Lernumgebung	Intervention
Rückblick	1 bis 4	65% (688)	35% (369)
	5 bis 9	61% (620)	39% (390)
Wesentliche Aspekte	1 bis 4	75% (12)	25% (4)
	5 bis 9	48% (10)	52% (11)
Alternativen	1 bis 4	84% (191)	16% (36)
	5 bis 9	85% (206)	15% (36)

Anders verhält es sich in Bezug auf die wesentlichen Aspekte. Dort reflektieren die Studierenden, die öfter interveniert haben, auch häufiger wesentliche Aspekte, die sich auf die Interventionen beziehen. Studierende, die seltener intervenieren, reflektieren hingegen in drei von vier Fällen wesentliche Aspekte der Lernumgebung. Wesentliche Aspekte der Interventionen werden kaum herausgearbeitet. Das ist insofern wenig überraschend, als Studierende, die häufiger intervenieren, auch eine höhere Grundlage haben, um darüber zu reflektieren. Sie haben mehr Erfahrungen in diesem Bereich gesammelt und lassen diese entsprechend auch öfter in die Reflexion miteinfließen.

Ein Zusammenhang zwischen der Anzahl der getätigten Interventionen und dem Reflexionsverhalten der Studierenden scheint also im Bereich der wesentlichen Aspekte anhand der Daten plausibel und inhaltlich begründbar. Damit zusammenhängend ist auch die Angemessenheit der Interventionen, denn im vorigen Teilkapitel wurde festgestellt, dass Studierende sich häufiger kritisch hinterfragen, wenn sie unangemessene Interventionen getätigt haben. Aufgrund der bisherigen Ergebnisse liegt der Verdacht nahe, dass es einen Zusammenhang zwischen der Anzahl der Interventionen und der Angemessenheit der Interventionen gibt, welcher das Reflexionsverhalten beeinflusst. Womöglich gibt es auch noch weitere Zusammenhänge, die durch eine Analyse der Daten, die sowohl die Anzahl als auch die Angemessenheit der Interventionen in den Blick nimmt, ergründet werden können. Daher werden die Gruppen nun sowohl anhand der Anzahl der Interventionen als auch anhand der Angemessenheit in Gruppen eingeteilt. Daraus ergeben sich vier Gruppen, die voneinander unterschieden werden können (vgl. Abb. 61).

Man kann erkennen, dass es auf den ersten Blick große Unterschiede zwischen die vier Gruppen zu geben scheint. So äußern die Studierenden, die vergleichsweise seltener und angemessener intervenieren, die meisten Reflexionsmomente. Weniger als die Hälfte der Reflexionsmomente dieser Gruppe äußern die Studierenden, die vergleichsweise selten und unangemessen intervenieren. Die beiden Gruppen, die häufiger intervenieren, unterscheiden sich insofern, als Studierende, die angemessener intervenieren, mehr Reflexionsmomente im Bereich Alternativen aufweisen, wohingegen die Studierenden mit weniger angemessenen Interventionen mehr Reflexionsmomente in der Kategorie „Rückblick“ aufweisen.

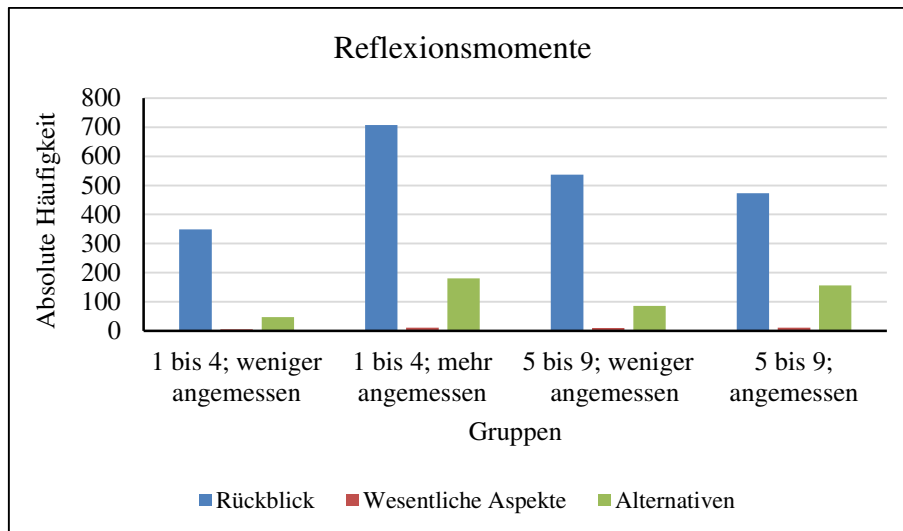


Abbildung 61: Übersicht über die absolute Häufigkeit der Reflexionsmomente der vier Gruppen von Studierenden auf Grundlage der Einteilung nach der Anzahl sowie der Angemessenheit der getätigten Interventionen

Um die Gruppen sinnvoll miteinander vergleichen zu können, muss die jeweilige Gruppengröße berücksichtigt werden. So sind in der Gruppe der Studierenden, die seltener und angemessener intervenieren (16), beinahe doppelt so viele Studierende wie in der Gruppe von Studierenden, die seltener und weniger angemessen intervenieren (9). Zwölf Studierende finden sich in der Gruppe, die häufiger und weniger angemessen interveniert und neun Studierende in der Gruppe, die häufiger und angemessener interveniert. Unter Berücksichtigung der unterschiedlichen Gruppengrößen (durch Normierung der Reflexionsmomente) erhält man folgende Grafik (vgl. Abb. 62).

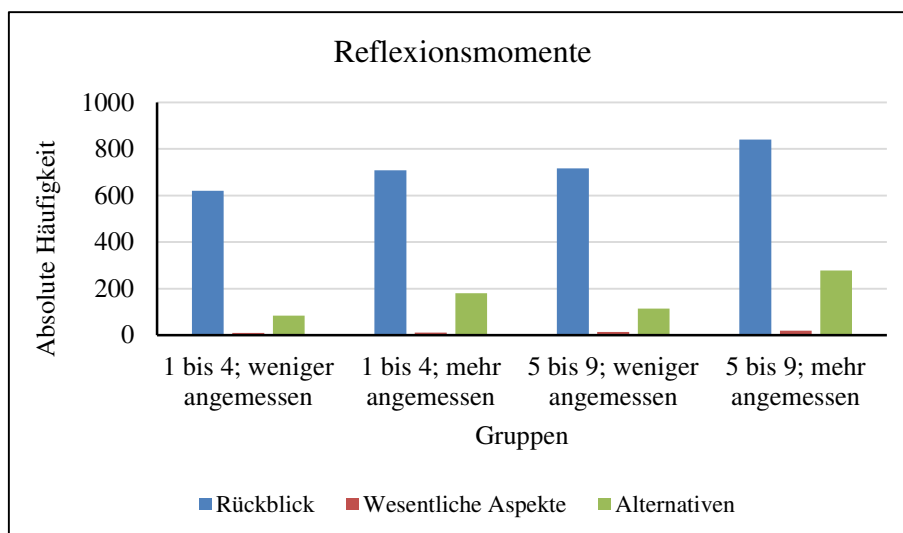


Abbildung 62: Übersicht über die normierten absoluten Häufigkeiten der Reflexionsmomente der vier Gruppen von Studierenden auf Grundlage der Einteilung nach der Anzahl sowie der Angemessenheit der getätigten Interventionen

Es lässt sich erkennen, dass die Studierenden, die zwischen fünf und neun Interventionen getätigt haben und deren Interventionen angemessener waren in allen drei Kategorien mehr Refle-

xionsmomente tätigen. Absolut und relativ gesehen ist der Unterschied in der Kategorie „Alternativen“ am größten, besonders, wenn man die Gruppe mit den Studierenden vergleicht, die zwischen einer und vier Interventionen getätigt haben, die durchschnittlich weniger angemessen waren. Während die eine Gruppe normiert 277 Reflexionsmomente hinsichtlich möglicher Alternativen besprochen hat, erreicht die andere Gruppe nur 84 Reflexionsmomente. Die anderen beiden Gruppen liegen mit 115 Reflexionsmomenten (Studierende mit fünf bis neun Interventionen, die durchschnittlich weniger angemessen waren) bzw. 180 Reflexionsmomenten (Studierende, die eine bis vier Interventionen getätigt haben, die angemessener waren als der Durchschnitt) dazwischen.

In den beiden Kategorien „Rückblick“ und „Wesentliche Aspekte“ findet sich eine ähnliche Befundlage im Hinblick auf die vier Gruppen. Zur besseren Lesbarkeit werden die Gruppen nachfolgend mit Gruppe 1 bis 4 benannt, wobei Gruppe 1 und 2 wenige Interventionen getätigt haben, die bei Gruppe 1 vergleichsweise weniger angemessen waren und bei Gruppe 2 angemessener. Entsprechend beziehen sich Gruppe 3 und 4 auf diejenigen Studierenden, die mehr Interventionen getätigt haben, wobei Gruppe 3 weniger angemessene und Gruppe 4 mehr angemessene Interventionen durchgeführt hat.

Gruppe 4 hat mit 841 Reflexionsmomenten im Bereich „Rückblick“ die meisten derartigen Äußerungen vorgenommen, gefolgt von Gruppe 3 mit 716 und Gruppe 2 mit 708 Reflexionsmomenten. Gruppe 1 verzeichnet mit 620 die wenigsten Äußerungen in dieser Kategorie. In Bezug auf die wesentlichen Aspekte ist es ebenfalls Gruppe 1, die mit neun die wenigsten Reflexionsmomente aufweist. Am meisten hat mit 20 Äußerungen Gruppe 4 getätigt. Dazwischen liegen Gruppe 3 mit 13 und Gruppe 2 mit elf Äußerungen.

Eine Vielzahl angemessener Interventionen scheint also förderlich für das Reflexionsverhalten der Studierenden. Womöglich haben die Studierenden einerseits viele Erfahrungen gesammelt, über die sie sich mit ihren Gruppenmitgliedern austauschen wollen. Andererseits handelt es sich vielleicht auch um eine Teilgruppe, die ohnehin extrovertierter ist und entsprechend kein Problem damit hat, in den Gruppenarbeitsprozess zu intervenieren, und auch keine Hemmnisse dabei hat, über Erfahrungen zu reflektieren. Andererseits weisen Gruppe 2 und 3 in den Kategorien „Rückblick“ und „Wesentliche Aspekte“ sehr ähnliche Werte auf, obwohl sie unterschiedlich oft interveniert haben. Darüber hinaus wurden in der Kategorie „Alternativen“ bei Gruppe 2 deutlich mehr Reflexionsmomente kodiert als bei Gruppe 3. Entsprechend ist die unterschiedliche Anzahl an getätigten Interventionen alleine nicht ausschlaggebendes Kriterium. Vielmehr hat es den Anschein, dass die Kombination aus vielen und angemessenen Interventionen dazu führt, dass mehr reflektiert wird.

An dieser Stelle wäre es interessant, zu untersuchen, ob sich dieser Effekt über die Motivation der Studierenden erklären lässt. Die Motivation könnte einen positiven Einfluss sowohl darauf haben, wie oft interveniert wird, als auch darauf, wie viel Mühe sich die Studierenden anschließend bei der Reflexion geben. Darüber hinaus könnte die Motivation auch dazu beitragen, dass fachliche und fachdidaktische Inhalte, die die Lernumgebung betreffen, vor der Erprobung vertieft wurden und dadurch zu erfolgreichen, sprich angemessenen Interventionen beitragen konnten. Anhand der vorliegenden Daten kann dies leider nicht geklärt werden und bleibt daher lediglich eine denkbare Erklärung.

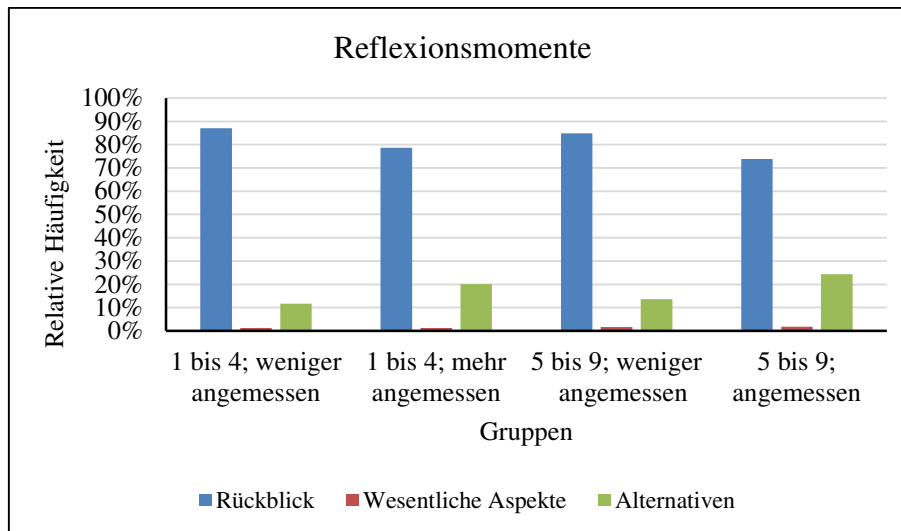


Abbildung 63: Übersicht über die relative Häufigkeit der Reflexionsmomente der vier Gruppen hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen" eingeteilt nach Anzahl und Angemessenheit der getätigten Interventionen

Betrachtet man die Verteilung der Reflexionsmomente innerhalb der vier Gruppen, wobei sich bei jeder Gruppe die Anzahl der Reflexionsmomente der drei Kategorien zu 100% aufsummiert, sieht man, dass Gruppe 1 und Gruppe 3 hauptsächlich Äußerungen in der Kategorie „Rückblick“ tätigen (vgl. Abb. 63). Zwar macht dies bei allen Gruppen den größten Anteil aus, allerdings sind bei Gruppe 2 und 4 verhältnismäßig mehr Reflexionsmomente im Bereich „Alternativen“ vorzufinden. Wesentliche Aspekte machen bei keiner Gruppe mehr als 2% der Reflexionsmomente aus.

Tendenziell scheinen die beiden Gruppen, die angemessener interveniert haben, unabhängig von der Anzahl der getätigten Interventionen, ein wünschenswerteres Verhältnis zwischen den Kategorien „Rückblick“ und „Alternativen“ zu haben. Sieht man einen hohen Anteil an wesentlichen Aspekten und möglichen Alternativen als positiv, so weist Gruppe 4 das beste Verhältnis an Reflexionsmomenten auf. Zwar sind nur 2% der Reflexionsmomente im Bereich wesentliche Aspekte anzusiedeln, was sicherlich ausbaufähig ist, dafür bezieht sich beinahe jede vierte Äußerung auf eine mögliche Alternative.

Somit tätigt Gruppe 4 nicht nur absolut gesehen die meisten Reflexionsmomente in allen Kategorien, sondern zeigt auch die vergleichsweise erstrebenswerteste Verteilung der Reflexionsmomente auf die drei Kategorien. Nachfolgend werden die drei Kategorien genauer unter die Lupe genommen und jeweils für alle vier Gruppen untersucht, welchen Anteil die Unterkategorien einnehmen.

Betrachtet man die Unterkategorien der Kategorie „Rückblick“, so sieht man, dass alle vier Gruppen am häufigsten über „Inhaltliches und Organisatorisches“ sprechen (vgl. Tab. 55). Bei Gruppe 4 ist der Anteil mit 73% besonders hoch. Dafür ist der Anteil bei Äußerungen, die einen groben Rückblick geben, mit 10% vergleichsweise gering. Im Hinblick auf die Unterkategorie „Gedanken und Gefühle“ gibt es keine nennenswerten Unterschiede zwischen den Gruppen. Hinsichtlich der „Absichten und Ziele“ gibt es ebenfalls keine großen Abweichungen. Gruppe 1 hat hier einen Anteil von 3% und Gruppe 3 kommt mit 7% auf den größten Anteil.

Tabelle 55: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Häufigkeit und der Angemessenheit der getätigten Interventionen

Anzahl an Interventionen	Angemessenheit	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick
1 bis 4	Weniger angemessen	3% (12)	64% (223)	13% (46)	19% (68)
	Mehr angemessen	6% (45)	64% (452)	13% (92)	17% (119)
5 bis 9	Weniger angemessen	7% (35)	67% (359)	12% (64)	15% (79)
	Mehr angemessen	4% (19)	73% (346)	13% (63)	10% (45)

Insgesamt lässt sich festhalten, dass Gruppe 4 deutlich seltener grobe Rückblicke liefert, sondern häufiger konkrete Inhalte oder organisatorische Aspekte anspricht. Grundsätzlich gilt das für alle Gruppen, wobei es für die anderen Gruppen in geringerem Maße gilt. In geringstem Maße gilt es für Gruppe 1. Somit gibt es auch innerhalb der Kategorie Rückblick einen deutlichen Unterschied zwischen Gruppe 1 und Gruppe 4, während zwischen den Gruppen 2 und 3 kein Unterschied erkennbar ist. Falls sich dieser Trend auch in den anderen beiden Kategorien wiederfinden lässt, spricht Vieles dafür, dass es einen Zusammenhang zwischen der Kombination aus Anzahl und Angemessenheit der Interventionen und dem Reflexionsverhalten der Studierenden gibt.

Tabelle 56: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Häufigkeit und der Angemessenheit der getätigten Interventionen

Anzahl an Interventionen	Angemessenheit	Kritisches Hinterfragen	Organisatorisches	Inhaltliches
1 bis 4	Weniger angemessen	40% (2)	20% (1)	40% (2)
	Mehr angemessen	9% (1)	18% (2)	73% (8)
5 bis 9	Weniger angemessen	80% (8)	10% (1)	10% (1)
	Mehr angemessen	18% (2)	36% (4)	45% (5)

Betrachtet man die Verteilung der Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, werden prozentual eklatante Unterschiede zwischen den Gruppen deutlich (vgl. Tab. 56). Allerdings gilt es hierbei zu berücksichtigen, dass aufgrund der geringen Anzahl an Reflexionsmomenten Vorsicht bei der Interpretation geboten ist.

Gruppe 3 weist mit acht Reflexionsmomenten in der Unterkategorie „Kritisches Hinterfragen“ mit Abstand die meisten Äußerungen in diesem Bereich auf. Somit lässt sich die Vermutung bestätigen, dass viele Interventionen, von denen weniger angemessen sind als der Durchschnitt, dazu führen, dass im Rahmen der Reflexion kritisch auf das eigene Handeln geblickt

wird. Aus Gruppe 2 hingegen entfällt nur eine Äußerung auf diese Unterkategorie. Dies ist insoweit schlüssig, als Studierende, die selten intervenieren und diese Interventionen zudem angemessen gestalten konnten, wenig Anlass dazu haben, ihr eigenes Handeln zu hinterfragen.

Diese Studierenden weisen hingegen mit acht Reflexionsmomenten die meisten inhaltlichen Aspekte auf. Am zweitmeisten Reflexionsmomente brachte Gruppe 4 in die Diskussion mit ein. Somit scheinen Studierende, die angemessener intervenieren, unabhängig von der Anzahl der getätigten Interventionen, vergleichsweise oft wesentliche inhaltliche Aspekte herauszuarbeiten. Auffällig ist darüber hinaus auch, dass Gruppe 1 in keiner Unterkategorie mehr als zwei Reflexionsmomente aufweist. Bei der Betrachtung muss man sich jedoch darüber bewusst sein, dass es sich um die tatsächlichen (und nicht um die normierten) Werte handelt. Gleiches gilt auch für die Reflexionsmomente innerhalb der Kategorie „Alternativen“, die nachfolgend analysiert werden.

Tabelle 57: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Häufigkeit der getätigten Interventionen

Anzahl an Interventionen	Angemessenheit	Nennung von Alternativen	Abwägung Vor- und Nachteile
1 bis 4	Weniger angemessen	98% (46)	2% (1)
	Mehr angemessen	87% (157)	13% (23)
5 bis 9	Weniger angemessen	93% (80)	7% (6)
	Mehr angemessen	85% (132)	15% (24)

Tabelle 57 führt die beiden Unterkategorien der Kategorie „Alternativen“ für die vier Gruppen auf. Man kann erkennen, dass Gruppe 1 mit 98% fast ausschließlich Alternativen nennt. Gruppe 4 hingegen wägt in 15% der Fälle Vor- und Nachteile der Alternativen ab. Die anderen beiden Gruppen liegen mit 13% (Gruppe 2) und 7% (Gruppe 3) dazwischen. Somit lässt sich einerseits festhalten, dass auch in der Kategorie „Alternativen“ ein deutlicher Unterschied zwischen Gruppe 1 und Gruppe 4 erkennbar wird. Andererseits zeigt sich, dass unabhängig von der Anzahl der durchgeführten Interventionen die Angemessenheit ein Faktor zu sein scheint, wie häufig Vor- und Nachteile der Alternativen abgewogen werden.

Es scheint also einen Zusammenhang zwischen der Kombination aus Anzahl und Angemessenheit der Interventionen und dem Reflexionsverhalten der Studierenden zu geben, wobei mehr Interventionen, die größtenteils angemessen durchgeführt wurden, dazu führen, dass die Studierenden ein wünschenswerteres Reflexionsverhalten zeigen. Die Kategorie „Wesentliche Aspekte“ muss dabei ausgeklammert werden, da zu wenige Reflexionsmomente vorliegen, um dies folgern zu können. Zwar handelt es sich auch im Fall der anderen beiden Kategorien nur um einen vermuteten Zusammenhang, jedoch erscheint dieser sich immer mehr abzuzeichnen, und daher wäre es durchaus erstrebenswert, diesen in einer Folgeuntersuchung noch weiter zu erforschen. Die letzte Untersuchung in diesem Teilkapitel nimmt nun die Unterscheidung zwischen Reflexionsmomenten, die sich auf die Lernumgebung und Reflexionsmomenten, die sich auf die Interventionen beziehen, in den Blick.

Tabelle 58: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Anzahl und der Angemessenheit der getätigten Interventionen

Kategorie	Anzahl der Interventionen	Angemessenheit	Lernumgebung	Intervention
Rückblick	1 bis 4	Weniger angemessen	63% (221)	37% (128)
		Mehr angemessen	66% (467)	34% (241)
	5 bis 9	Weniger angemessen	63% (336)	37% (201)
		Mehr angemessen	60% (284)	40% (189)
Wesentliche Aspekte	1 bis 4	Weniger angemessen	60% (3)	40% (2)
		Mehr angemessen	82% (9)	18% (2)
	5 bis 9	Weniger angemessen	20% (2)	80% (8)
		Mehr angemessen	73% (8)	27% (3)
Alternativen	1 bis 4	Weniger angemessen	74% (35)	26% (12)
		Mehr angemessen	87% (156)	13% (24)
	5 bis 9	Weniger angemessen	81% (70)	19% (16)
		Mehr angemessen	87% (136)	13% (20)

In Bezug auf die Kategorie „Rückblick“ lassen sich kaum Unterschiede zwischen den vier Gruppen feststellen (vgl. Tab. 58). Alle Gruppen reflektieren häufiger über die Lernumgebung als über die Interventionen, wobei der Anteil zwischen 60% (Gruppe 4) und 66% (Gruppe 2) variiert. Hinsichtlich der wesentlichen Aspekte gibt es eine größere Divergenz zwischen den Gruppen. Während sich die wesentlichen Aspekte, die von Gruppe 3 erarbeitet wurden, überwiegend auf die Interventionen beziehen (80%), reflektiert Gruppe 2 zum größten Teil über die Lernumgebung (82%). Bei Gruppe 1 sind die Reflexionsmomente am gleichmäßigsten verteilt, wobei Gruppe 1 auch die wenigsten Reflexionsmomente aufweist. Gruppe 4 arbeitet tendenziell eher wesentliche Aspekte heraus, die sich auf die Lernumgebung beziehen.

Aufgrund der geringen Fallzahlen müssen diese Unterschiede jedoch vorsichtig interpretiert werden. So lässt sich vermuten, dass angemessene Interventionen unabhängig von der Anzahl an getätigten Interventionen eher dazu führen, dass wesentliche Aspekte tendenziell in Bezug auf die Lernumgebung reflektiert werden. Der hohe Anteil an wesentlichen Aspekten in Bezug auf die Interventionen liegt bei Gruppe 3 an der Vielzahl der unangemessenen Interventionen. Insofern sind die Unterschiede in der Kategorie „Wesentliche Aspekte“ inhaltlich nachvollziehbar.

Alle vier Gruppen nennen häufiger Alternativen, ohne dabei Vor- und Nachteile gegeneinander abzuwägen, als dass sie Alternativen nennen und deren Vor- und Nachteile diskutieren. Verhältnismäßig oft werden Vor- und Nachteile von Gruppe 1 abgewogen, wobei an dieser Stelle auf die geringe Anzahl an Reflexionsmomenten bei dieser Gruppe hingewiesen werden muss. Gruppe 4 wägt beinahe doppelt so oft Vor- und Nachteile gegeneinander ab, hat jedoch aufgrund der Vielzahl an Alternativen, die ohne Abwägung genannt werden, ein schlechteres prozentuales Verhältnis als Gruppe 1. Die beiden Gruppen sind insofern gut miteinander zu vergleichen, als beide Gruppen aus neun Studierenden bestehen. Entsprechend kann beim Vergleich dieser beiden Gruppen auf eine Normierung der Zahlen verzichtet werden.

In diesem Teilkapitel wurde gezeigt, dass Studierende, die mehr Interventionen vorgenommen haben, tendenziell auch mehr Reflexionsmomente äußern. Dies gilt für alle drei Kategorien. Studierende, die häufiger intervenieren und deren Interventionen weniger angemessen sind, hinterfragen ihr eigenes Handeln häufiger als ihre Kommilitoninnen und Kommilitonen. Studierende, die seltener intervenieren, arbeiten mehr wesentliche inhaltliche Aspekte heraus als Studierende, die häufiger interveniert haben. Das Reflexionsverhalten der Gruppe, die häufiger und vergleichsweise angemessener interveniert hat, ist ausgeglichener als das der anderen Gruppen. Insgesamt scheint diese Kombination einen Einfluss auf das Reflexionsverhalten der Studierenden zu haben.

9.6.3 Gruppeneinteilung nach Art der getätigten Interventionen

Nachdem bereits zwei verschiedene Gruppeneinteilungen vorgenommen wurden und das Reflexionsverhalten daraufhin in Abhängigkeit zu den getätigten Interventionen untersucht wurde, wird nun eine weitere Unterscheidung vorgenommen. Dieses Mal werden die Studierenden anhand der Art der getätigten Interventionen in Gruppen eingeteilt, wobei es eine Gruppe gibt, die eher inhaltlichere dominante Hilfen in den Interventionen genutzt hat, eine Gruppe, die eher weniger inhaltliche dominante Hilfen in den Interventionen verwendet hat, und eine Gruppe, bei der diese Unterscheidung nicht getroffen werden kann, da inhaltliche und weniger inhaltliche Hilfen gleich oft als dominante Hilfen innerhalb der Interventionen eingesetzt wurden.

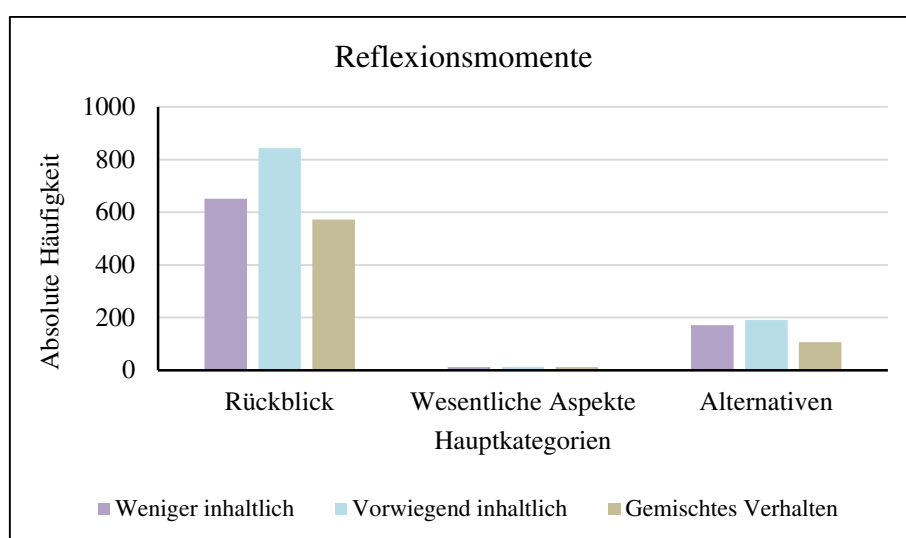


Abbildung 64: Übersicht über die absolute Häufigkeit der Reflexionsmomente der drei Gruppen von Studierenden auf Grundlage der Einteilung nach der Art der getätigten Interventionen

Betrachtet man die Übersicht über die kodierten Reflexionsmomente der drei Gruppen, so stellt man fest, dass die Gruppe, die vorwiegend inhaltlich interveniert hat, in allen drei Kategorien mehr Äußerungen getätigt hat (vgl. Abb. 64). Interessanterweise ist dies der Fall, obwohl die Gruppe, die weniger inhaltlich interveniert hat, eine Person mehr zählt (17) als die Gruppe, die vorwiegend inhaltlich interveniert hat (16). Ein gemischtes Verhalten zeigten 13 Studierende. Da dieses jedoch schwierig zu interpretieren ist und es bei dieser Einteilung hauptsächlich um die Unterscheidung zwischen der vorwiegend inhaltlichen und der weniger inhaltlichen Gruppe geht, wird auf eine Normierung verzichtet.

Insbesondere in der Kategorie „Rückblick“ gibt es einen deutlichen Unterschied zwischen den beiden Gruppen. Studierende, die vorwiegend inhaltlich interveniert haben, äußern hierbei deutlich mehr Reflexionsmomente. Der Unterschied beträgt etwa 200 Reflexionsmomente und somit mehr als zehn pro Person. Auch in Bezug auf Reflexionsmomente, die der Kategorie „Alternativen“ zugeordnet wurden, erreicht diese Gruppe die größte Anzahl, der Unterschied beträgt hierbei etwa 20 Reflexionsmomente. In der dritten Kategorie ist kein Unterschied erkennbar. Alle drei Gruppen reflektieren über zwölf bzw. 13 wesentliche Aspekte.

Inhaltlichere Hilfen scheinen die Studierenden dazu zu veranlassen, intensiver über die Erprobung und die Interventionen zu reflektieren. Allerdings gilt diese erste grobe Tendenz hauptsächlich für die Kategorie „Rückblick“ und in geringem Maße auch noch für die Überlegungen hinsichtlich möglicher Alternativen. Bei der Herausarbeitung wesentlicher Aspekte ist kein Unterschied zwischen den Studierenden, die vorwiegend inhaltlich interveniert haben, und denen, die weniger inhaltlich interveniert haben, erkennbar.

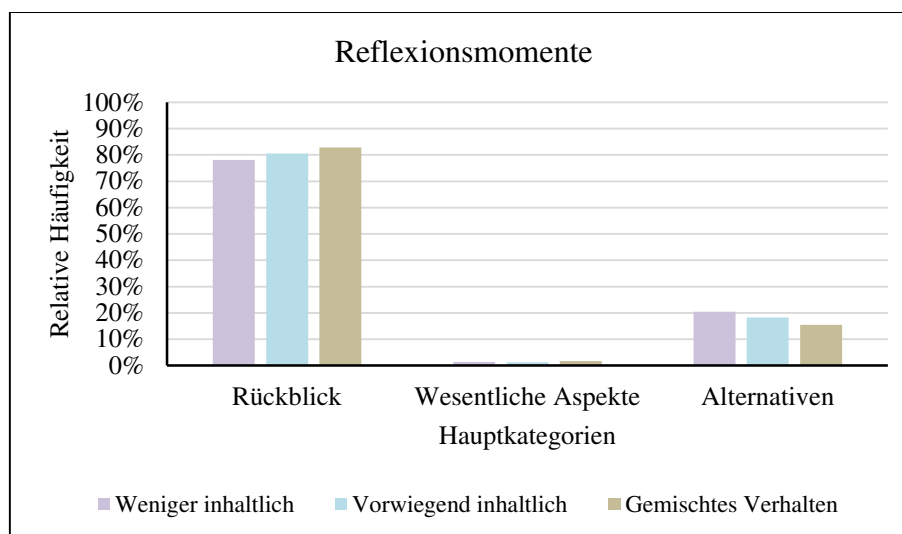


Abbildung 65: Übersicht über die relative Häufigkeit der Reflexionsmomente der drei Gruppen hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen" eingeteilt nach Anzahl und Angemessenheit der getätigten Interventionen

Bei der Betrachtung der Verteilung der Reflexionsmomente innerhalb der drei Gruppen ergeben sich nur geringfügige Unterschiede zwischen den Studierenden, die vorwiegend inhaltlich interveniert haben, und denen, die weniger inhaltlich interveniert haben (vgl. Abb. 65). Entsprechend kann an dieser Stelle festgehalten werden, dass hierbei keine Unterschiede festgestellt werden können, die eine gezielte Folgeuntersuchung dieses Teilausschnittes nach sich ziehen sollten. Stattdessen werden nun zusätzlich innerhalb der einzelnen Kategorien mögliche Zusammenhänge zwischen dem Interventions- und dem Reflexionsverhalten ergründet.

Tabelle 59: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Interventionsart	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick	Summe
Weniger inhaltlich	6% (38)	64% (418)	13% (84)	17% (111)	100% (651)
Vorwiegend inhaltlich	5% (40)	70% (590)	13% (107)	12% (107)	100% (844)
Gemischtes Verhalten	6% (33)	65% (372)	13% (74)	16% (93)	100% (572)

Die Verteilung der Reflexionsmomente auf die Unterkategorien der Kategorie „Rückblick“ ist in Tabelle 59 aufgelistet. Man kann erkennen, dass es kaum Unterschiede zwischen den Gruppen in den beiden Unterkategorien „Absichten und Ziele“ und „Gedanken und Gefühle“ gibt. Ein kleiner Unterschied liegt in den Unterkategorien „Inhaltliches und Organisatorisches“ und „Grober Rückblick“ vor. Zwar reflektieren alle Gruppen überwiegend über inhaltliche und organisatorische Punkte, jedoch ist der prozentuale Anteil bei der Gruppe, die vorwiegend inhaltlich interveniert hat, noch etwas höher als bei den anderen beiden Gruppen. Umgekehrt verhält es sich in der Unterkategorie „Grober Rückblick“. Dort verzeichnet die Gruppe, die weniger inhaltlich interveniert hat, prozentual gesehen mehr Äußerungen, die sich auf diese Unterkategorie beziehen als die Gruppe, die vorwiegend inhaltlich interveniert hat.

Die leichte Tendenz in diese Richtung erscheint insofern nachvollziehbar, als der Fokus bei den Studierenden, die vorwiegend inhaltlich interveniert haben, sowohl während der Interventionen als auch im Nachgang bei der Reflexion etwas stärker auf inhaltlichen Aspekten zu liegen scheint. Womöglich gibt es also einen Zusammenhang zwischen der Art der getätigten Interventionen und dem Reflexionsverhalten im Bereich „Rückblick“. Diese leichte Tendenz ergibt sich jedoch aus nur geringen Unterschieden zwischen den Gruppen und müsste in Folgestudien weiter untersucht werden. Im Rahmen dieser Untersuchung kann dies nicht detaillierter geleistet werden, weshalb nun auf die beiden anderen Kategorien eingegangen wird.

Tabelle 60: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Interventionsart	Kritisches Hinterfragen	Organisatorisches	Inhaltliches	Summe
Weniger inhaltlich	17% (2)	33% (4)	50% (6)	100% (12)
Vorwiegend inhaltlich	46% (6)	23% (3)	31% (4)	100% (13)
Gemischtes Verhalten	42% (5)	8% (1)	50% (6)	100% (12)

Hinsichtlich der wesentlichen Aspekte, die die Gruppen reflektiert haben, gibt es aufgrund der geringen Anzahl an Reflexionsmomenten in den einzelnen Unterkategorien große Abweichungen in der prozentualen Verteilung innerhalb der Kategorie (vgl. Tab. 60). Der Anteil an Äußerungen, bei denen die Studierenden ihr Verhalten kritisch hinterfragen, ist bei der Gruppe,

die vorwiegend inhaltlich interveniert hat, deutlich höher als bei der Gruppe, die weniger inhaltlich interveniert hat. Die Studierenden, die weniger inhaltlich intervenieren, reflektieren hingegen vergleichsweise häufiger über „Inhaltliches“.

Dass sich Studierende, die vorwiegend inhaltlich intervenieren, häufiger kritisch hinterfragen, könnte damit zusammenhängen, dass diese auch öfter unangemessen agieren, da inhaltliche Hilfen in höherem Maße als unangemessen angesehen wurden als andere Hilfen. Ob dies der Fall ist, wird im Laufe des Teilkapitels noch untersucht werden, indem die Angemessenheit als zusätzliches Unterscheidungsmerkmal berücksichtigt wird. Zunächst werden jedoch die Alternativen betrachtet, die die Studierenden im Rahmen ihrer Reflexionsgespräche besprochen haben. Anschließend wird außerdem zwischen Reflexionsmomenten, die sich auf die Lernumgebung beziehen, und solchen, die sich auf die Interventionen beziehen, unterschieden.

Tabelle 61: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Interventionsart	Nennung von Alternativen	Abwägung Vor- und Nachteile	Summe
Weniger inhaltlich	88% (150)	12% (21)	100% (171)
Vorwiegend inhaltlich	93% (177)	7% (14)	100% (191)
Gemischtes Verhalten	82% (88)	18% (19)	100% (107)

Nimmt man die Kategorie „Alternativen“ in den Blick, zeigt sich, dass alle Gruppen größtenteils Alternativen nennen und nur in seltenen Fällen deren Vor- und Nachteile abwägen (vgl. Tab. 61). Vergleicht man die beiden Gruppen, die im Fokus stehen, miteinander, so kann man einen recht geringen Unterschied zwischen den beiden Gruppen feststellen. Die Studierenden, die vorwiegend inhaltlich intervenieren, wägen im Rahmen ihres Reflexionsgespräches seltener die Vor- und Nachteile möglicher Alternativen ab als Studierende, die weniger inhaltlich intervenieren. Die Studierenden, die ein gemischtes Verhalten zeigen, wägen prozentual gesehen am häufigsten Vor- und Nachteile der Alternativen gegeneinander ab.

Aufgrund der geringen Unterschiede werden die Ergebnisse aus der Kategorie „Alternativen“ an dieser Stelle nicht weiter interpretiert. Stattdessen liegt das Augenmerk im Folgenden auf der Unterscheidung der Reflexionsmomente in solche, die sich auf die Lernumgebung, und solche, die sich auf die Interventionen beziehen.

Im Fall der beiden Kategorien „Rückblick“ und „Alternativen“ gibt es kaum einen Unterschied zwischen den Studierenden, die vorwiegend inhaltlich interveniert haben, und denen, die weniger inhaltlich interveniert haben (vgl. Tab. 62). In Bezug auf die reflektierten wesentlichen Aspekte wurden für die Gruppe, die weniger inhaltlich interveniert hat, deutlich mehr Reflexionsmomente kodiert, die sich auf die Lernumgebung bezogen. In der Gruppe, die vorwiegend inhaltlich interveniert hat, waren die Reflexionsmomente in ihrem Bezug auf die Lernumgebungen oder die Interventionen in etwa gleichverteilt. Die Gruppe, die ein gemischtes Interventionsverhalten zeigte, liegt hinsichtlich der reflektierten wesentlichen Aspekte jeweils zwischen den beiden anderen Gruppen.

Tabelle 62: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Kategorie	Interventionsart	Lernumgebung	Intervention
Rückblick	Weniger inhaltlich	64% (416)	36% (235)
	Vorwiegend inhaltlich	62% (526)	38% (318)
	Gemischtes Verhalten	64% (366)	36% (206)
Wesentliche Aspekte	Weniger inhaltlich	75% (9)	25% (3)
	Vorwiegend inhaltlich	46% (6)	54% (7)
	Gemischtes Verhalten	58% (7)	42% (5)
Alternativen	Weniger inhaltlich	86% (147)	16% (24)
	Vorwiegend inhaltlich	87% (167)	13% (24)
	Gemischtes Verhalten	78% (83)	22% (24)

Die Analyse der vorgenommenen Kodierung der Reflexionsgespräche bringt in Bezug auf die Unterschiede zwischen Äußerungen, die sich auf die Lernumgebung beziehen, und Äußerungen, die sich auf die Interventionen beziehen, kaum neue Erkenntnisse hervor. Während es in den Kategorien „Rückblick“ und „Alternativen“ keine Unterschiede zwischen den beiden im Fokus stehenden Gruppen gibt, lassen sich Unterschiede im Reflexionsverhalten im Hinblick auf reflektierte wesentliche Aspekte feststellen, die aufgrund der geringen Anzahl an Reflexionsmomenten nur vorsichtig interpretiert werden können. Der Unterschied fügt sich jedoch in die vorherigen Befunde, nach denen sich Studierende, die vorwiegend inhaltlich intervenieren, öfter kritisch hinterfragen. Ob dies daran liegt, dass diese auch häufiger unangemessene Interventionen getätigt haben, wird nachfolgend untersucht, indem die drei Gruppen nochmals anhand der Angemessenheit unterteilt werden.

Abbildung 66 zeigt eine Übersicht über die Reflexionsmomente, wobei die Gruppeneinteilung nun sowohl nach der Art als auch nach der Angemessenheit der getätigten Interventionen vorgenommen wurde. Zur einfacheren Handhabung werden die Gruppen von A bis F durchnummeriert, wobei die Gruppen A und B aus den Studierenden bestehen, die weniger inhaltlich interveniert haben. Die Gruppen C und D sind diejenigen, die vorwiegend inhaltlich interveniert haben, und die Gruppen E und F bestehen aus Probanden, die ein gemischtes Interventionsverhalten gezeigt haben. Die Gruppen A, C und E haben jeweils vergleichsweise weniger angemessen interveniert, während die Gruppen B, D und F prozentual mehr angemessene Interventionen durchgeführt haben als der Durchschnitt. Gruppe A, C, D und F bestehen aus jeweils acht, Gruppe B aus neun und Gruppe E aus fünf Studierenden. Da insbesondere die Gruppen A bis D verglichen werden, ist keine Normierung der Zahlen notwendig.

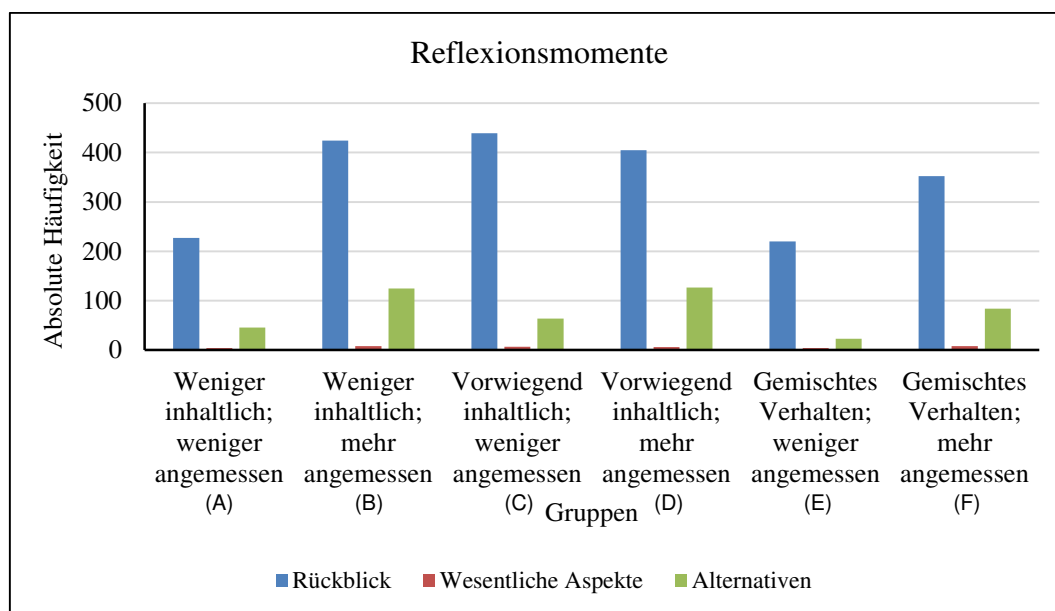


Abbildung 66: Übersicht über die absolute Häufigkeit der Reflexionsmomente der sechs Gruppen von Studierenden auf Grundlage der Einteilung nach der Art sowie der Angemessenheit der getätigten Interventionen

Es lässt sich erkennen, dass Gruppe A deutlich weniger Reflexionsmomente äußert als die Gruppen B, C und D. Besonders groß sind die Unterschiede in den Kategorien „Rückblick“ und „Alternativen“. Dass die Gruppe, die weniger inhaltlich interveniert hat und deren Interventionen dabei weniger angemessen waren, derart wenig reflektiert (etwa die Hälfte im Vergleich zu den Gruppen B, C und D), ist interessant und wird in den Analysen der jeweiligen Unterkategorien noch genauer beleuchtet.

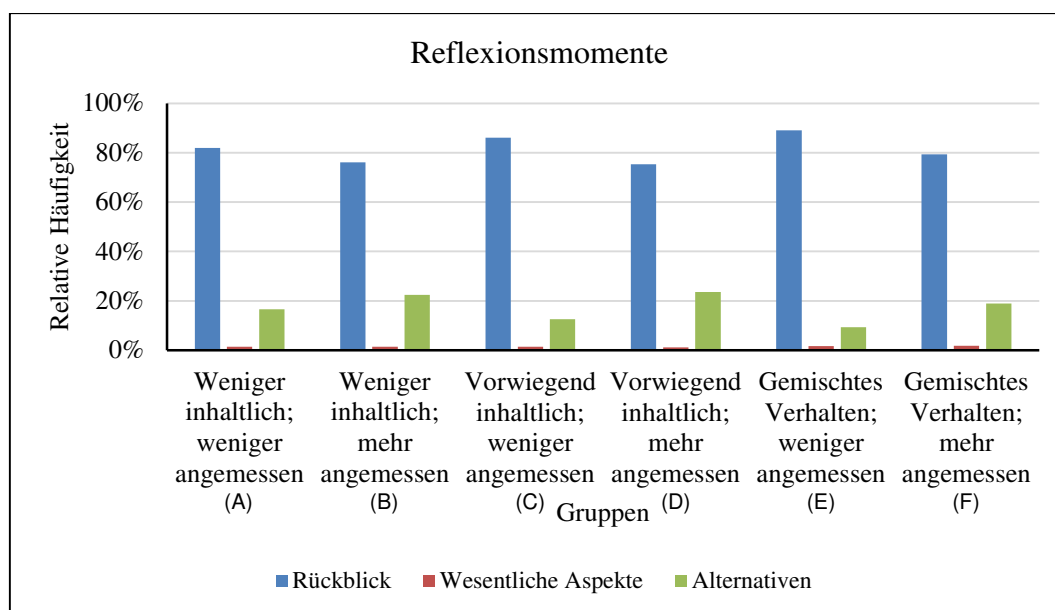


Abbildung 67: Übersicht über die relative Häufigkeit der Reflexionsmomente der sechs Gruppen hinsichtlich der drei Hauptkategorien "Rückblick", "Wesentliche Aspekte" und "Alternativen" eingeteilt nach Art und Angemessenheit der getätigten Interventionen

Weiterhin ist auffällig, dass die Gruppen B und D mit Abstand die meisten Alternativen in die Gruppendiskussion einbringen. Am drittmeisten Alternativen werden von Gruppe F geäußert. Bei den drei Gruppen handelt es sich um die Studierenden, die angemessenere Interventionen durchgeführt haben als ihre Kommilitoninnen und Kommilitonen. Bevor auch hierfür nach einer Erklärung durch die Analyse der Unterkategorien gesucht wird, steht nun zunächst die Verteilung der Reflexionsmomente innerhalb der Gruppen im Analysefokus.

Bezüglich der Kategorie „Rückblick“ gibt es nur geringe Unterschiede zwischen den sechs Gruppen (vgl. Abb. 67). Die Gruppen A, C und E bleiben etwas häufiger auf der vergleichsweise oberflächlichen Reflexionsebene als die anderen drei Gruppen. Richtet man den Blick auf die Kategorie „Alternativen“, wird deutlich, dass die Gruppen B, D und F relativ gesehen häufiger über mögliche Alternativen reflektiert haben als die anderen drei Gruppen. Der Anteil der reflektierten wesentlichen Aspekte liegt bei allen Gruppen bei einem bzw. zwei Prozent.

Anhand der Verteilungen der Reflexionsmomente wurde sichtbar, dass die Studierenden, die mehr angemessene Interventionen getätigt haben, nicht nur insgesamt häufiger über Alternativen reflektieren als die anderen Gruppen, sondern dass die Reflexion von Alternativen auch einen höheren prozentualen Anteil an den Reflexionsgesprächen ausmacht als in den anderen Gruppen. Der Effekt besteht über alle Einteilungen nach der Art der Hilfe hinweg, sprich unabhängig davon, welche Interventionsarten bevorzugt verwendet wurden. Die Angemessenheit der getätigten Interventionen scheint also ausschlaggebender dafür zu sein, in welcher Häufigkeit über Alternativen nachgedacht wird, als die Art der getätigten Interventionen. Ob dies auch für weitere potenzielle Zusammenhänge der Fall sein könnte, wird nachfolgend anhand der Analyse der Unterkategorien untersucht.

Tabelle 63: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Rückblick“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Interventionsart	Angemessenheit	Absichten und Ziele	Inhaltliches und Organisatorisches	Gedanken und Gefühle	Grober Rückblick
Weniger inhaltlich	Weniger angemessen	7% (15)	57% (130)	12% (27)	24% (55)
	Mehr angemessen	5% (23)	68% (288)	13% (57)	13% (56)
Vorwiegend inhaltlich	Weniger angemessen	4% (19)	69% (305)	13% (57)	13% (58)
	Mehr angemessen	5% (21)	70% (285)	12% (50)	12% (49)
Gemischtes Verhalten	Weniger angemessen	6% (13)	67% (147)	12% (26)	15% (34)
	Mehr angemessen	6% (20)	64% (225)	14% (48)	17% (59)

Die erste Kategorie, die genauer beleuchtet wird, ist die Kategorie „Rückblick“ (vgl. Tab. 63). Es lässt sich erkennen, dass Gruppe A im Vergleich zu den anderen Gruppen deutlich häufiger einen groben Rückblick über die Erprobung der Lernumgebung bzw. die getätigten Interventionen gibt. Hinsichtlich der Gedanken und Gefühle, die die Studierenden im Rahmen ihrer Reflexionsgespräche geäußert haben, gibt es keine großen Unterschiede zwischen den Grup-

pen. Gleiches gilt für die Unterkategorie „Absichten und Ziele“. In der Unterkategorie „Inhaltliches und Organisatorisches“ weist Gruppe A entsprechend einen geringeren Anteil an Reflexionsmomenten auf als die anderen Gruppen.

Die Studierenden, die weniger inhaltliche Interventionen getätigt haben, die tendenziell weniger angemessen waren, äußern also nicht nur 200 Reflexionsmomente weniger als die Gruppen B, C und D, sondern beschränken sich dabei anteilig deutlich häufiger auf einen groben Rückblick. Eine mögliche Erklärung dafür ist es, dass diese Studierenden über geringere fachliche bzw. fachdidaktische Kenntnisse verfügen als ihre Kommilitoninnen und Kommilitonen. Dies würde erklären, weshalb sie weniger inhaltliche Interventionen getätigt haben und anschließend derart wenig inhaltliche Aspekte in die Gruppenreflexion mit einbringen.

Eine weitere denkbare Erklärung wäre, dass diese Studierenden weniger motiviert waren als ihre Kommilitoninnen und Kommilitonen. Dies könnte dazu geführt haben, dass sie sich auf kürzere Hilfen beschränken. In Teilkapitel 9.2 wurde herausgearbeitet, dass Hilfen länger dauern, je inhaltlicher sie sind. Entsprechend könnten die Studierenden sich aus mangelnder Motivation auf zeitlich kürzere Hilfen beschränkt haben. Die fehlende Motivation wäre auch eine Erklärung dafür, weshalb die Interventionen von den Expertinnen und Experten in höherem Maße als unangemessen eingestuft wurden als bei anderen Studierenden. Ebenso könnte damit begründet werden, weshalb einerseits so wenige Reflexionsmomente geäußert wurden und weshalb sich andererseits bei diesen wenigen Reflexionsmomenten beinahe jede vierte Äußerung auf einen groben Rückblick beschränkt.

Leider können beide Hypothesen in dieser Arbeit nicht untersucht werden. Da die Facetten Fachwissen, fachdidaktisches Wissen und Motivation jedoch bereits an anderen Stellen als mögliche Erklärung dienten, erhärtet sich die Tendenz, dass diese in Folgestudien mit erhoben werden sollten. Nachfolgend werden die beiden Kategorien „Wesentliche Aspekte“ und „Alternativen“ genauer untersucht.

Tabelle 64: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Wesentliche Aspekte“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Interventionsart	Angemessenheit	Kritisches Hinterfragen	Organisatorisches	Inhaltliches
Weniger inhaltlich	Weniger angemessen	25% (1)	50% (2)	25% (1)
	Mehr angemessen	13% (1)	25% (2)	63% (5)
Vorwiegend inhaltlich	Weniger angemessen	86% (6)	0% (0)	14% (1)
	Mehr angemessen	0% (0)	50% (3)	50% (3)
Gemischtes Verhalten	Weniger angemessen	75% (3)	0% (0)	25% (1)
	Mehr angemessen	25% (2)	13% (1)	63% (5)

Die Kategorie „Wesentliche Aspekte“ umfasst die wenigsten Reflexionsmomente und ist durch die Aufteilung in sechs Gruppen kaum noch interpretierbar (vgl. Tab. 64). Die folgenden Beschreibungen und Interpretationen müssen also mit größter Vorsicht gesehen werden.

Gruppe A äußert in keiner der Unterkategorien mehr als zwei Reflexionsmomente, was die Vermutungen minimal stützt, dass diese Studierenden womöglich nicht motiviert waren oder über mangelndes Fachwissen bzw. fachdidaktisches Wissen verfügen, um wesentliche Aspekte herauszuarbeiten. Die Gruppen B, D und F arbeiten vermehrt inhaltliche und organisatorische Aspekte heraus, die sie als wesentlich empfunden haben. Dabei hinterfragen sie ihr eigenes Handeln eher selten kritisch. Da es sich hierbei um die Studierenden handelt, die angemessener interveniert haben, ergibt diese Verteilung durchaus Sinn.

Am häufigsten hinterfragt Gruppe C ihr eigenes Handeln kritisch. Bei dieser Gruppe handelt es sich um diejenigen Studierenden, die vorwiegend inhaltlich interveniert haben und deren Interventionen weniger angemessen waren. Da inhaltliche Hilfen von den Expertinnen und Experten häufiger als unangemessen angesehen wurden als andere Hilfen und diese Teilmenge von Studierenden auch tendenziell weniger angemessene Interventionen durchgeführt hat, erscheint es plausibel, dass diese ihr Interventionsverhalten kritisch hinterfragen.

Tabelle 65: Übersicht über die Reflexionsmomente innerhalb der Kategorie „Alternativen“, eingeteilt hinsichtlich der Art der getätigten Interventionen

Interventionsart	Angemessenheit	Nennung von Alternativen	Abwägung Vor- und Nachteile
Weniger inhaltlich	Weniger angemessen	91% (42)	9% (4)
	Mehr angemessen	86% (108)	14% (17)
Vorwiegend inhaltlich	Weniger angemessen	95% (61)	5% (3)
	Mehr angemessen	91% (116)	9% (11)
Gemischtes Verhalten	Weniger angemessen	100% (23)	0% (0)
	Mehr angemessen	77% (65)	23% (19)

Die letzte Kategorie, die untersucht wird, ist die Kategorie „Alternativen“ (vgl. Tab. 65). Hierbei wird zwischen der reinen Nennung von Alternativen und dem Abwägen von Vor- und Nachteilen unterschieden. Es zeigt sich, dass die Gruppen, bei denen die Studierenden angemessener interveniert haben, anteilig jeweils häufiger über Vor- und Nachteile der Alternativen reflektieren als die Gruppen, die weniger angemessen interveniert haben. Von Gruppe E wurden keinerlei Vor- und Nachteile diskutiert.

Es lässt sich festhalten, dass die Angemessenheit das Reflexionsverhalten unabhängig von der überwiegend genutzten Interventionsart derart beeinflusst, dass häufiger über Vor- und Nachteile reflektiert wird, wenn die Interventionen hauptsächlich angemessen waren. Während die drei Gruppen, die angemessener reflektiert haben, 47 Reflexionsmomente äußern, die sich auf die Diskussion von Vor- und Nachteilen möglicher Alternativen bezieht, äußern die anderen drei Gruppen nur insgesamt sieben. Abschließend wird nun zwischen Lernumgebung und Intervention unterschieden (vgl. Tab. 66).

Tabelle 66: Überblick über die Verteilung der Reflexionsmomente in den Hauptkategorien bezüglich der beiden Reflexionsgegenstände „Lernumgebung“ und „Intervention“, eingeteilt hinsichtlich der Art und der Angemessenheit der getätigten Interventionen

Kategorie	Interventionsart	Angemessenheit	Lernumgebung	Intervention
Rückblick	Weniger inhaltlich	Weniger angemessen	63% (144)	37% (83)
		Mehr angemessen	64% (272)	36% (152)
	Vorwiegend inhaltlich	Weniger angemessen	63% (275)	37% (164)
		Mehr angemessen	62% (251)	38% (154)
	Gemischtes Verhalten	Weniger angemessen	63% (138)	37% (82)
		Mehr angemessen	65% (228)	35% (124)
Wesentliche Aspekte	Weniger inhaltlich	Weniger angemessen	75% (3)	25% (1)
		Mehr angemessen	75% (6)	25% (2)
	Vorwiegend inhaltlich	Weniger angemessen	14% (1)	86% (6)
		Mehr angemessen	83% (5)	17% (1)
	Gemischtes Verhalten	Weniger angemessen	25% (1)	75% (3)
		Mehr angemessen	75% (6)	25% (2)
Alternativen	Weniger inhaltlich	Weniger angemessen	80% (37)	20% (9)
		Mehr angemessen	88% (110)	12% (15)
	Vorwiegend inhaltlich	Weniger angemessen	84% (54)	16% (10)
		Mehr angemessen	89% (113)	11% (14)
	Gemischtes Verhalten	Weniger angemessen	61% (14)	39% (9)
		Mehr angemessen	82% (69)	18% (15)

In der Kategorie „Rückblick“ lassen sich keine Unterschiede im Reflexionsverhalten zwischen den sechs Gruppen erkennen. Dies spricht dafür, dass es auf die Verteilung der Reflexionsmomente hinsichtlich Lernumgebung und Intervention keinen Einfluss hat, welche Interventions-

arten überwiegend genutzt wurden oder wie angemessen interveniert wurde. Des Weiteren erscheint die Anzahl an getätigten Reflexionsmomenten groß genug zu sein, um keine zufälligen Effekte in dieser Kategorie zu produzieren. Dies bestärkt auch die vorhergehenden Ergebnisse.

Anders verhält es sich in Bezug auf die wesentlichen Aspekte. Hierbei könnte die geringe Anzahl an Reflexionsmomenten durchaus gravierende Auswirkungen auf die Verteilung haben. Daher wurde bisher immer darauf verwiesen, dass die Interpretationen in diesem Bereich auf einer schwachen Basis fußen. Nimmt man die Ergebnisse, wie sie vorliegen, so lässt sich festhalten, dass die Gruppen A und B in jeweils drei von vier Fällen über die Lernumgebung reflektiert haben. Dagegen unterscheiden sich die anderen beiden Gruppenpaare deutlich. Während Gruppe C fast ausschließlich über die Interventionen reflektiert, liegt der Fokus bei Gruppe D deutlich auf der Lernumgebung. Gruppe E reflektiert in 75% der Fälle über die Interventionen, Gruppe F in 75% der Fälle über die Lernumgebung. Diese Unterschiede sind über die Angemessenheit der getätigten Interventionen begründbar. Studierende, die unangemessener agieren, reflektieren ihre Interventionen eher kritisch. Entsprechend kommt die höhere Anzahl an wesentlichen Aspekten, die sich auf die Interventionen beziehen, zustande.

Da Gruppe A insgesamt sehr wenige Reflexionsmomente geäußert und insbesondere kaum wesentliche Aspekte herausgearbeitet hat, könnte man die Hypothese aufstellen, dass es bei dieser Gruppe zu einem verzerrten Bild kommt. Folgt man dieser Annahme, ließe sich ableiten, dass Studierende, die weniger angemessen interveniert haben, deshalb anteilig mehr wesentliche Aspekte herausarbeiten, die sich auf die Interventionen beziehen, da sie ihr eigenes Handeln häufiger kritisch hinterfragen. Stützt man sich eher auf die vorliegenden Daten, so könnte man diese derart interpretieren, dass die Studierenden, die weniger motiviert sind, entsprechend auch weniger über ihre eigenen Interventionen reflektieren und sich seltener hinterfragen, da dies mit vergleichsweise hohem Arbeitsaufwand verbunden ist. Einfacher ist es, in der Gruppe mitzudiskutieren und sich zu einzelnen inhaltlichen Aspekten zu äußern, die ein Gruppenmitglied eingebracht hat.

Leider wurde bei der Kodierung nicht berücksichtigt, ob sich Reflexionsmomente auf Äußerungen von Gruppenmitgliedern beziehen, oder ob diese sich auf ein Thema beziehen, das bisher noch nicht angesprochen wurde. Im letzten Teilkapitel werden die Ergebnisse des Kapitels nochmals zusammengefasst, bevor diese im anschließenden Kapitel diskutiert werden.

9.6.4 Zusammenfassung

Studierende, die angemessener interveniert haben, äußern in allen Bereichen mehr Reflexionsmomente als ihre Kommilitoninnen und Kommilitonen, die weniger angemessen interveniert haben. Dabei ergibt sich der Unterschied nur zu einem geringen Teil aus mehr Äußerungen, die sich auf die Interventionen beziehen. Ein Großteil der Reflexionsmomente bezieht sich auf die Lernumgebung. Dies könnte daran liegen, dass die Studierenden über mehr Fachwissen bzw. fachdidaktisches Wissen verfügen. Andererseits könnte auch die Motivation ausschlaggebend sein.

Studierende, die angemessener intervenieren, äußern darüber hinaus auch mehr Gedanken über mögliche Vor- und Nachteile verschiedener Alternativen. Studierende, die weniger angemessen intervenieren, wägen nur sehr selten mögliche Vor- und Nachteile gegeneinander ab. Daher könnte man vermuten, dass dieses Verhaltensmuster auch vor den Interventionen vorherrschte und die Interventionen womöglich aus diesem Grund seltener angemessen waren.

Immerhin hinterfragen Studierende, die weniger angemessen intervenieren, ihr Verhalten häufiger kritisch. Dies spricht dafür, dass sie einen Optimierungsbedarf in ihren eigenen Interventionen sehen, und deckt sich somit mit der Einschätzung der Expertinnen und Experten.

Betrachtet man die Häufigkeit der Interventionen, so lässt sich festhalten, dass Studierende, die häufiger intervenieren, mehr Reflexionsmomente äußern als ihre Kommilitoninnen und Kommilitonen. Dies gilt für alle drei Kategorien. Das könnte daran liegen, dass sie aufgrund der höheren Anzahl an Interventionen mehr prägende Erfahrungen sammeln konnten, über die sie im Rahmen des Reflexionsgespräches mit ihren Gruppenmitgliedern diskutieren wollen. Untersucht man jedoch die Verteilung der Reflexionsmomente hinsichtlich des enthaltenen Bezugs auf die Lernumgebung oder die Interventionen, so lässt sich lediglich in der Kategorie „Wesentliche Aspekte“ feststellen, dass dort häufiger über Interventionen reflektiert wird. Dies ist jedoch auf eine höhere Anzahl an kritischen Äußerungen dem eigenen Interventionsverhalten gegenüber geschuldet und liegt vermutlich daran, dass diese Studierenden auch häufiger unangemessen interveniert haben. Entsprechend scheint dieser Unterschied eher auf die Angemessenheit der Interventionen zurückzuführen zu sein und weniger auf die Anzahl der Interventionen.

Studierende, die häufig intervenieren und deren Interventionen angemessener sind als der Durchschnitt, reflektieren in allen drei Kategorien am häufigsten. Besonders auffällig ist hierbei der Unterschied bei der Reflexion über mögliche Alternativen. Dabei werden nicht nur mehr Alternativen genannt (als von Studierenden, die weniger angemessen agiert haben), sondern auch anteilig am meisten Vor- und Nachteile abgewogen. Mögliche Begründungen dafür könnten auch an dieser Stelle höheres fachliches bzw. fachdidaktisches Wissen oder eine höhere Motivation sein.

Studierende, die wenige Interventionen getätigt haben, die überdurchschnittlich angemessen waren, arbeiten hauptsächlich inhaltliche Aspekte heraus, die sie als wesentlich empfinden. Studierende, die wenig interveniert haben und dabei weniger angemessene Interventionen durchgeführt haben, weisen den größten Anteil an Reflexionsmomenten auf, die sich auf die Interventionen beziehen. Absolut gesehen besprechen sie jedoch die wenigsten Alternativen.

In der Kategorie „Rückblick“ gibt es hinsichtlich der Unterscheidung zwischen Reflexionsmomenten, die sich auf die Lernumgebung bzw. die Interventionen beziehen, keine nennenswerten Unterschiede zwischen den Studierenden, unabhängig davon, ob diese anhand der Angemessenheit, der Anzahl oder der Art der getätigten Interventionen in Gruppen eingeteilt und miteinander verglichen werden. Es scheint also für die Verteilung zwischen Äußerungen, die sich auf die Lernumgebung beziehen und Äußerungen, die sich auf die Interventionen beziehen, keinen Unterschied zu machen, wie, wie oft oder wie angemessen interveniert wurde.

Anders verhält sich dies in Bezug auf die wesentlichen Aspekte. Studierende, die vorwiegend inhaltlich intervenieren, hinterfragen ihr eigenes Verhalten bei den Interventionen häufiger als Studierende, die weniger inhaltlich interveniert haben. Nimmt man zusätzlich die Angemessenheit der Interventionen in den Blick, so lässt sich festhalten, dass Studierende, die vorwiegend inhaltlich intervenieren und deren Interventionen weniger angemessen waren, ihr Interventionsverhalten deutlich häufiger hinterfragen als andere Studierende. Dies ist vermutlich darauf zurückzuführen, dass inhaltliche Hilfen häufiger unangemessen waren als andere Hilfen. Daher haben die Studierenden, die vorwiegend inhaltliche Hilfen getätigt haben, auch öfter unangemessene Interventionen getätigt. Im Rahmen der Reflexionsgespräche haben sie diese Interventionen dann teilweise auch selbst kritisch beleuchtet. Auch wenn die Tendenz an dieser Stelle zunächst zufriedenstellend erscheint, so wurde nur ein Bruchteil der als unangemessen eingestuften Interventionen von den Studierenden kritisch hinterfragt.

Studierende, die weniger inhaltlich intervenieren und deren Interventionen weniger angemessen waren, äußern im Rahmen der Reflexionsgespräche deutlich weniger Reflexionsomente als ihre Kommilitoninnen und Kommilitonen. Dies könnte einerseits auf mangelnde Motivation, andererseits auf mangelnde fachliche bzw. fachdidaktische Kenntnisse zurückzuführen sein. Letzteres wäre in Anbetracht der Tatsache, dass diese Studierenden kurz vor ihrem Masterabschluss stehen, ein gravierender Befund. Eine Diskussion der Ergebnisse folgt nun im letzten Kapitel.

10 Diskussion

Durch die vielfältigen Ergebnisse in den verschiedenen Bereichen konnten mögliche Zusammenhänge gefunden werden, wohingegen an anderen Stellen vermutete Zusammenhänge zwischen Bereichen nicht erhärtet werden konnten. In diesem Kapitel werden die Ergebnisse zunächst diskutiert und dabei in Zusammenhang mit den empirischen Befunden anderer Studien gebracht (Kap. 10.1). Anschließend werden zugrundeliegende Limitationen der Studie benannt (Kap. 10.2) sowie Forschungsdesiderate vorgestellt (Kap. 10.3).

10.1 Diskussion der Ergebnisse

Hinsichtlich der diagnostischen Fähigkeiten einzelner Lehrkräfte stellten Hoge und Colardaci (1989) beträchtliche Unterschiede fest. Auch in dieser Studie zeigte sich eine heterogene Verteilung bzgl. der Übereinstimmung der Antworten der Studierenden mit denen der vorab befragten Expertinnen und Experten. Keiner der Studierenden war in der Lage, mehr als elf der 14 Diagnoseaufträge im Sinne der Expertinnen und Experten zu beantworten. Entsprechend erstreckt sich das Leistungsspektrum nicht bis in den höchsten Bereich. Dies ist insofern verständlich, als die Studierenden insgesamt über recht wenig praktische Erfahrung verfügen. Daher wäre es verwunderlich, wenn einige der Studierenden in der Lage gewesen wären, annähernd alle Diagnoseaufträge adäquat zu beantworten. In den Freitextantworten der Lehrkräfte zeigt sich, dass diese deutlich differenzierter diagnostizieren bzw. ihre wahrgenommenen Aspekte besser einordnen und wiedergeben. Insbesondere in Bezug auf den Diagnoseauftrag, der sich für die Studierenden als am schwierigsten erwies, konnten zwei Drittel der Lehrkräfte eine schlüssige Argumentation für ihre Einschätzung liefern, die sich mit der der Expertengruppe deckte.

Auch über alle drei Vignetten hinweg waren zwei Drittel der Lehrkräfte in der Lage, mindestens zehn der 14 Fragen im Sinne der Expertinnen und Experten zu beantworten. Jeweils drei Lehrkräfte konnten sogar zwölf bzw. 13 Diagnoseaufträge richtig beantworten. Diagnostische Fähigkeiten scheinen sich also im Laufe des Lehrerdaseins aufzubauen. Dies würde dafür sprechen, dass diagnostische Fähigkeiten im Rahmen des Studiums nicht unbedingt angebahnt und aufgebaut werden müssen. Jedoch sollte berücksichtigt werden, dass es sich bei den Lehrkräften um eine Positivauswahl an Lehrkräften handelt. Entsprechend erscheint es richtig und wichtig, dass diagnostische Fähigkeiten bereits in der ersten Phase der Lehramtsausbildung geschult werden.

Ein weiteres Argument hierfür ist der festgestellte Zusammenhang zwischen der Ausprägung der prozessdiagnostischen Fähigkeit „Deuten“ und der Angemessenheit der Interventionen. Studierende, die besser diagnostizierten, tätigten anschließend bei der Betreuung der Schülergruppe auch angemessenere Interventionen. Dies erhärtet den ohnehin in der Literatur angenommenen Zusammenhang zwischen Diagnose und Intervention, bei dem davon ausgegangen wird, dass es wichtig sei, dass Lehrkräfte den Kenntnisstand ihrer Schülerinnen und Schüler gut einschätzen können, um adäquate Interventionen tätigen zu können (van de Pol et al. 2010; Kleber 1976).

Hinsichtlich der Interventionen ist die Verteilung der gegebenen Hilfen innerhalb der Interventionen im Einklang mit den Befunden von Stender (2016, 198), welcher in seiner Studie am häufigsten strategische Hilfen auswies. Wie in Kapitel 9.2 beschrieben, waren die häufigsten Interventionen allgemein-strategischer oder inhaltsorientiert-strategischer Art. Dies steht konträr zu den Ergebnissen von Leiss (2007, 281), der befand, dass strategische Interventionen so

gut wie nicht vorkommen (bzw. im Repertoire der Lehrkräfte vorhanden sind). Besonders häufig treten die beiden strategischen Hilfen (allgemein-strategische Hilfen und inhaltsorientiert-strategische Hilfen) gemeinsam auf. Im Rahmen dieser Studie handelt es sich dabei um die häufigste Kombination von Hilfen, gefolgt von der Kombination aus inhaltsorientiert-strategischen Hilfen und Rückmeldehilfen.

Weiterhin beschreibt Stender (2016, 198), dass kaum Motivationshilfen und Rückmeldehilfen gegeben werden. Dies kann teilweise bestätigt werden. Auch in der vorliegenden Studie gaben die Studierenden kaum Motivationshilfen. Ebenso kamen kaum dominante Rückmeldehilfen vor, was sich mit der Studie von Stender (2016) deckt. Jedoch gab es viele untergeordnete Rückmeldehilfen. Eine solche Unterscheidung zwischen dominanter und untergeordneter Hilfe nahm Stender (2016) allerdings nicht vor.

Diese Rückmeldehilfen haben nach Seidel und Shavelson (2007) einen positiven Einfluss auf den Lernfortschritt der Schülerinnen und Schüler. In der vorliegenden Studie wurde nicht untersucht, inwieweit einzelne Hilfen einen Einfluss auf den Lernfortschritt der Schülerinnen und Schüler haben. An dieser Stelle kann daher nur festgehalten werden, dass Rückmeldehilfen selten als dominante Hilfe genutzt werden, jedoch häufig als untergeordnete Hilfe vorkommen.

Ähnliches gilt für die Feststellung von Dekker & Elshout-Mohr (2004, 62), dass weniger inhaltliche Hilfen bei Lehrerinterventionen in Kleingruppen zu höherem Lernerfolg der Schülerinnen und Schüler führen. Inhaltsorientiert-strategische Hilfen kamen in der vorliegenden Studie am häufigsten und inhaltliche Hilfen kamen am dritthäufigsten vor. Inwieweit diese den Lernerfolg der Schülerinnen und Schüler möglicherweise gehemmt haben, wurde nicht untersucht. Außerdem könnte die Verteilung auch durch das gewählte Studiendesign beeinflusst worden sein, da das Verlassen des Beobachtungsraumes sowie das Betreten des Arbeitsraumes der Schülerinnen und Schüler als zusätzliche Hemmschwelle angesehen werden kann, die insbesondere dazu führen könnte, auf weniger inhaltliche Hilfen zu verzichten.

Betrachtet man die erste Hilfe innerhalb der Interventionen, so sieht man, dass allgemein-strategische Hilfen am häufigsten genutzt wurden, gefolgt von Rückmeldehilfen. Dies spricht dafür, dass die Studierenden zunächst weniger inhaltliche Hilfen nutzen wollten. Vermutlich ergab es sich teilweise im Laufe der Interventionen, dass eine inhaltliche Hilfe nachgelegt wurde. Unklar ist hierbei, ob diese intendiert waren oder ad hoc vorgenommen wurden.

Hinsichtlich des Interventionsauslösers lässt sich festhalten, dass der absolute Großteil der Interventionen invasiv war. Dies ist im Einklang mit bisherigen Befunden (vgl. dazu u.a. Seifried & Klüber 2006, 157ff.; Fürst & Haag 1998, 66). Aufgrund des Studiendesigns, mit dem unter anderem eine räumliche Trennung der Schülergruppe und der beobachtenden Person einhergeht, war dies auch in der Form zu erwarten. Interessant ist, dass responsive Interventionen fast alle als angemessen eingestuft wurden und vergleichsweise häufiger dazu führten, dass mehr Schülerinnen und Schüler nach der Intervention aktiv waren. Gegaart mit dem Grundprinzip im Mathematik-Labor „Mathe ist mehr“, dass Schülerinnen und Schüler möglichst selbstständig arbeiten sollen, erscheint es in diesem Kontext sinnvoll, eher responsive Interventionen vorzunehmen. Invasive Interventionen sollten also sparsamer durchgeführt werden, da sie weniger aktivierend sind als responsive Interventionen und auch seltener als angemessen eingestuft wurden.

Kontraintuitiv ist jedoch der Befund, dass Studierende mit erhöhter prozessdiagnostischer Fähigkeit „Deuten“ genau dies nicht tun, denn sie intervenierten fast ausschließlich invasiv. Dies könnte mehrere Gründe haben. Vermutlich erkennen sie Probleme frühzeitig und greifen daher bereits ein, bevor die Schülerinnen und Schüler sich melden und eine Hilfe einfordern

können. Außerdem muss festgehalten werden, dass diese Studierenden es schaffen, obwohl invasive Interventionen im Allgemeinen seltener angemessen und weniger aktivierend sind, in diesen beiden Facetten trotzdem erfolgreich zu sein. Sie tätigten angemessenere und aktivierendere Interventionen als ihre Kommilitoninnen und Kommilitonen, obwohl sie sich für invasive Interventionen entschieden. Mit diesem Wissen ist es nun schwierig, an der allgemeinen Empfehlung für responsive Interventionen festzuhalten. Grundsätzlich scheinen responsive Interventionen geeigneter, jedoch sind Studierende mit ausgeprägter prozessdiagnostischer Fähigkeit „Deuten“ in der Lage, invasive Interventionen angemessener und aktivierender zu tätigen als ihre Kommilitoninnen und Kommilitonen. Entsprechend sollten invasive Interventionen nicht grundsätzlich ausgeschlossen oder als minderwertigere Interventionen angesehen werden. Vielmehr können diese zur Produktivität der Lernenden und somit vermutlich auch zu deren Lernerfolg beitragen. Daher sollte die Handlungsanweisung von Diegritz (1999, 346), dass invasive Lehrerinterventionen vermieden werden sollen, kritisch hinterfragt werden.

Nimmt man die Zeitdauer der Interventionen in den Blick, so lässt sich festhalten, dass die Interventionen desto länger dauern, je eher sie inhaltlichen Charakter haben. Dass inhaltliche Hilfen am längsten dauern und am häufigsten als unangemessen angesehen wurden, fügt sich in die Ergebnisse von Fürst und Haag (1998, 61), die befanden, dass zeitlich längere Interventionen zu schlechteren Arbeitsergebnissen führen. Jedoch muss auch an dieser Stelle erneut eine differenzierte Betrachtung erfolgen. Studierende mit höheren diagnostischen Fähigkeiten intervenieren nämlich länger als ihre Kommilitoninnen und Kommilitonen. Trotzdem liefern sie angemessenere und auch aktivierendere Interventionen. Die Zeitdauer scheint also nicht per se zu schlechteren Interventionen zu führen. Die zusätzliche Zeit könnte beispielsweise dadurch bedingt sein, dass die Studierenden mehr Probleme und Schwierigkeiten identifizieren konnten und aus diesem Grund innerhalb ihrer Interventionen auch mehr unterschiedliche Themen ansprechen bzw. mit den Schülerinnen und Schülern erarbeiten. Eine andere Erklärung könnte sein, dass diese Studierenden fachlich oder fachdidaktisch besser aufgestellt sind als ihre Kommilitoninnen und Kommilitonen und entsprechend mehr Wissen vermitteln bzw. dieses besser anbahnen und die zusätzliche Zeit dafür nutzen.

Bezüglich der Angemessenheit der Interventionen kann festgehalten werden, dass 154 der 208 Interventionen als angemessen eingestuft wurden. Auch bei vergleichbaren Studien, wie etwa der Studie von Stender (2016, 199), wurden insgesamt deutlich mehr Interventionen als erfolgreich bewertet als als nicht erfolgreich. Stender (2016) definiert Erfolg in seiner Studie jedoch damit, dass die Aufgabe durch die Schülerinnen und Schüler richtig bearbeitet wurde und diese möglichst selbstständig arbeiten konnten. Somit ist der Bewertungsmaßstab ein anderer als in dieser Studie, da in dieser Studie Kriterien der Angemessenheit von Interventionen definiert wurden (vgl. dazu Kapitel 8.3.4). Diese berücksichtigen nicht, ob die Schülerinnen und Schüler am Ende zu einer richtigen Lösung kommen oder nicht. Auch das selbstständige Arbeiten der Schülerinnen und Schüler ist nicht direkt in den Kriterien enthalten, sondern findet sich indirekt, beispielsweise im Kriterium „Auswahl der Hilfe-Kategorie“ wieder.

Des Weiteren kommt Stender (2016, 199) zu dem Schluss, dass inhaltliche Interventionen tendenziell weniger angemessen sind. Dieses Ergebnis wird bei ihm aufgrund der geringen Stichprobe allerdings nicht signifikant. Es deckt sich jedoch mit den Ergebnissen in dieser Studie. Bei allen drei Gruppen (gruppiert nach der prozessdiagnostischen Fähigkeit „Deuten“) sind inhaltliche Hilfen in geringerem Maße angemessen als andere Hilfen. Unklar ist, inwieweit dies einen Bezug zum Lernerfolg hat. Zudem könnten die Kriterien der Angemessenheit sowie die Expertengruppe einen Einfluss auf das Ergebnis haben. So gibt es Kriterien, die zwar für inhaltliche Hilfen gelten, jedoch nicht für alle anderen Hilfen anwendbar sind. Darüber hinaus

könnte es sich dabei auch um die Hilfe-Kategorie handeln, die die Expertinnen und Experten einerseits sehr gut beurteilen können, andererseits aber auch recht streng ahnden, da ein inhaltlicher Fehler vermutlich schneller dazu führt, dass eine Intervention als unangemessen eingestuft wird, als ein Fehler in einer anderen Kategorie.

Mit kritischer Betrachtung der Ergebnisse kann man hinterfragen, ob die Expertinnen und Experten in der Lage waren, trennscharf zwischen den verschiedenen Hilfen zu unterscheiden. Zwar gibt es einen deutlichen Unterschied in der Bewertung der Angemessenheit der inhaltlichen Hilfen und der Angemessenheit der anderen fünf Hilfe-Kategorien. Jedoch werden diese jeweils mit ungefähr 80% als angemessen bewertet. Entsprechend sollten die damit zusammenhängenden Ergebnisse vorsichtig interpretiert und in Folgestudien weiter untersucht werden.

Gore und Zeichner (1991) stellten fest, dass Reflexion oftmals wenig kritisch ist. Auch im Rahmen dieser Studie wurde die Unterkategorie „Kritisches Hinterfragen“ nur selten kodiert. In einem weiteren Verständnis kann man auch anführen, dass die Studierenden hauptsächlich auf der oberflächlichen Ebene „Rückblick“ reflektiert und sehr wenige wesentliche Aspekte besprochen haben. Studierende, die weniger angemessen interveniert haben, hinterfragen ihr eigenes Handeln häufiger kritisch als ihre Kommilitoninnen und Kommilitonen. Da die Studierenden nicht wussten, wie ihre Interventionen eingestuft werden, scheint die eigene Einschätzung bzw. Unzufriedenheit dazu geführt zu haben, dass sie sich im Rahmen des Gruppenreflexionsgespräches kritisch dazu äußern. Dies bestärkt die Einschätzung der Expertinnen und Experten und lässt hoffen, dass die Studierenden ihr Handeln auch künftig hinterfragen. Für die Lehramtsausbildung, insbesondere aber für den späteren Berufsalltag, ist es wichtig, dass eine Lehrperson in der Lage ist, ihr eigenes Handeln kritisch zu hinterfragen, um sich stetig weiterzuentwickeln. Gemäß von Felten (2005, 165) gilt Reflexion als hilfreich für die (Weiter-)Entwicklung des eigenen Handelns.

Die untersuchten Gruppenreflexionsgespräche unterschieden sich massiv in der Zeitdauer, welche zwischen 38 und 143 Minuten lag. Unterschieden nach der diagnostischen Fähigkeit fällt hinsichtlich der Verteilung der Reflexionsmomente auf, dass es kaum Unterschiede in Bezug auf die Hauptkategorien gab, lediglich die minimale Tendenz, dass Studierende mit höherer prozessdiagnostischer Fähigkeit „Deuten“ mehr Alternativen und mehr wesentliche Aspekte nennen. Dies könnte jedoch auch mit anderen Konstrukten wie Intelligenz oder Motivation zusammenhängen und sollte daher vorsichtig interpretiert werden. Insgesamt gibt es kaum erkennbare Zusammenhänge zwischen Diagnose und Reflexion. Untersucht wurde allerdings nur ein möglicher Einfluss von der diagnostischen Fähigkeit „Deuten“ auf das Reflexionsverhalten. Eventuell hat die Reflexionsfähigkeit einen Einfluss auf das Diagnoseverhalten von Studierenden. Dies wurde jedoch nicht untersucht, da die Diagnosehandlung vor den Reflexionsgesprächen stattfand und eine anders gelagerte Implikation entsprechend nicht zulässig wäre.

Nimmt man das Interventionsverhalten als Grundlage, um das Reflexionsverhalten der Studierenden zu untersuchen, lassen sich durchaus mögliche Zusammenhänge erkennen. Studierende, die angemessener interveniert haben, setzen sich deutlich intensiver mit den getätigten Interventionen auseinander. Auch hier müssten sonstige Einflussfaktoren wie Intelligenz oder Motivation kontrolliert werden. Ähnliches gilt für den Befund, dass Studierende, die mehr Interventionen tätigen, anschließend auch ausführlicher reflektieren. Hierbei könnte es eine Rolle spielen, ob es sich um eine eher extrovertierte oder eine eher introvertierte Person handelt. Zusätzlich ist an dieser Stelle anzumerken, dass durch eine erhöhte Anzahl an Interventionen auch eine größere Grundlage zur Reflexion vorhanden ist. Dadurch, dass zunächst derjenige, der die Intervention getätigt hat, darlegen soll, weshalb er interveniert hat, ob er wieder intervenieren

würde und falls ja, ob er die Intervention anders gestalten würde, ergibt sich automatisch eine größere Reflexionsgrundlage.

Studierende, die sowohl häufiger als auch angemessener interveniert haben, zeigen ein wünschenswerteres Reflexionsverhalten, sprich sie diskutieren anteilig mehr über wesentliche Aspekte und mögliche Alternativen. Optimalerweise basieren die alternativen Handlungsmöglichkeiten und alternativen Ideen für die Gestaltung der Lernumgebung auf den herausgearbeiteten wesentlichen Aspekten. Aufgrund der geringen Anzahl an Probanden, die für diese Untersuchung gleich zweimal untergliedert wurde, kann diese Feststellung jedoch nur als Indiz dienen, welchem es weiter nachzugehen lohnt. Gleiches gilt für Studierende, die inhaltlich intervenieren und deren Interventionen weniger angemessen waren. Diese Studierenden hinterfragen ihr eigenes Handeln nämlich häufiger als ihre Kommilitoninnen und Kommilitonen, was grundsätzlich als positiv anzusehen ist, da dies ein erster Schritt zur Optimierung des eigenen Handelns sein kann. Die Studierenden scheinen wahrzunehmen, ob ihre Interventionen zufriedenstellend waren (als angemessen angesehen werden können) oder nicht (als unangemessen eingestuft werden müssen). Dies lässt hoffen, dass sie dies auch im Klassenzimmer können und dass sie bestrebt sind, aus ihren Fehlern zu lernen. Das Hinterfragen der eigenen Handlungen sowie das stetige Weiterentwickeln und Optimieren führt schließlich dazu, dass Lehrkräfte die Lernprozesse der Schülerinnen und Schüler adäquater unterstützen können.

Davis (2006, 290) fand heraus, dass es deutliche Unterschiede im Reflexionsverhalten gibt. Auf Grundlage der Ergebnisse kann dieser Aussage insofern zugestimmt werden, als es in einigen Unterkategorien zu teilweise erheblichen Unterschieden zwischen den betrachteten Gruppen kam. Auf oberster Ebene zeichnete sich jedoch häufig ein ähnliches Bild ab. In der Regel wurden überwiegend Reflexionsmomente im Bereich „Rückblick“ getätigt. Alternativen wurden vergleichsweise deutlich seltener genannt und am wenigsten wurde über wesentliche Aspekte gesprochen. Dies gilt für alle beobachteten Gruppen, weshalb der Aussage von Davis (2006) nur teilweise zugestimmt werden kann. Vollumfänglich kann jedoch folgender Aussage von Davis zugestimmt werden: Studierende müssen kontinuierlich dazu angeregt werden, produktiv zu reflektieren (Davis 2006, 290).

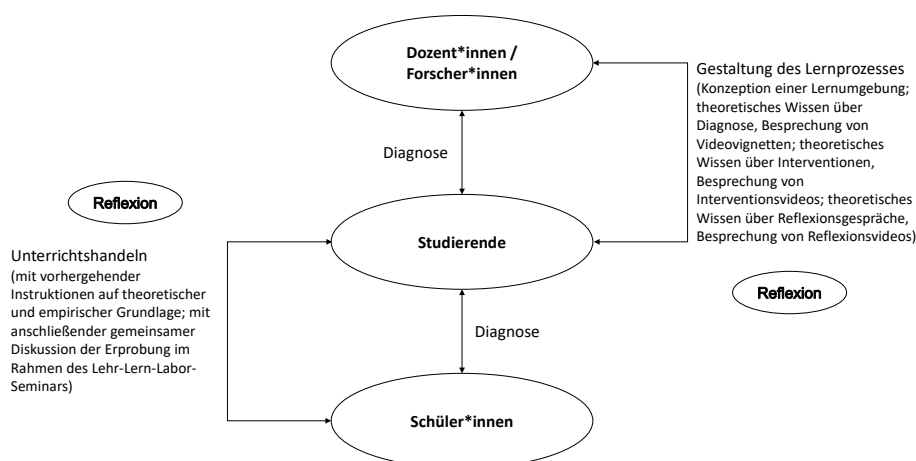


Abbildung 68: Übersicht über die drei zu unterscheidenden Ebenen hinsichtlich der Forschungsvorhaben im Mathematik-Labor "Mathe ist mehr" ergänzt durch auf den Forschungsergebnissen aufbauende Optimierungen

Dazu sollte Studierenden künftig im Lehr-Lern-Labor-Seminar auch theoretisches Wissen über Diagnose vermittelt werden. Die Bearbeitung der Videovignetten sollte nachbesprochen werden (vgl. Abb. 68). Ähnliches gilt für die Interventionen. Auch hier erscheint es sinnvoll, dass die Studierenden zunächst theoretisches Wissen über mögliche Interventionsarten erwerben, um ihr Handeln bewusster durchzuführen. Eine Besprechung ausgewählter Interventionen im Seminar könnte die Studierenden dazu befähigen, sich kritisch mit ihren Interventionen auseinanderzusetzen. Des Weiteren könnten Positivbeispiele die Studierenden in ihrem Handeln bestärken. Das Briefing der Studierenden bereits vor der Erprobung könnte sich auch positiv auf die Betreuung der Schülerinnen und Schüler im Mathematik-Labor auswirken. Bei den Reflexionsgesprächen könnte neben der Vermittlung theoretischen Inhalts auch darüber nachgedacht werden, ob der Dozierende an dem Reflexionsgespräch teilnimmt, um ggf. andere Facetten mit in die Gespräche einbringen zu können und die Reflexion dadurch zu bereichern.

10.2 Limitationen

Zur Überprüfung der prozessdiagnostischen Fähigkeit „Deuten“ wurden thematisch nur „Terme“ dargeboten. Da diagnostische Fähigkeiten vermutlich bereichsspezifisch sind, insofern sie mit Fachwissen und fachdidaktischem Wissen korrelieren, hätten Fachwissen und fachdidaktisches Wissen entweder kontrolliert, sprich miterhoben werden müssen, oder es hätten Videovignetten aus verschiedenen Bereichen verwendet werden müssen.

Bei den Interventionen hatten jeweils nur zwei bzw. drei Studierende die gleiche Schülergruppe, die sie beobachten sollten. Andere Studierende haben jedoch andere Schülerinnen und Schüler betreut. Aufgrund der ungleichen Voraussetzungen kann dies zu einem unterschiedlichen Bedarf an Interventionen führen. Möglich ist auch, dass manchen Lernenden weniger inhaltliche Hilfen ausreichen, während andere verstärkt inhaltliche Hilfen benötigen. Selbst innerhalb der Kleingruppen kann es sein, dass sich die Schülerinnen und Schüler bei der Bearbeitung unterschiedlicher Teilstationen (an verschiedenen Tagen) in verschiedenen Merkmalen (wie beispielsweise der Motivation) unterscheiden.

Es wurde nicht kontrolliert, zu welcher Tageszeit die Schülerinnen und Schüler das Mathematik-Labor besucht haben. Auch wenn alle Schulklassen vormittags zu Besuch waren, kann es einen Unterschied machen, ob diese zu den ersten beiden Stunden kommen oder zur fünften und sechsten Stunde. Vereinzelt gab es auch Klassen, die aufgrund der räumlichen Distanz alle drei Teile der Laborstation an einem Tag erarbeitet haben. Dies könnte sich auf die Leistungsfähigkeit und die Motivation der Lernenden, insbesondere bei der dritten Teilstation, auswirken. Gerade jüngere Schülerinnen und Schüler haben erfahrungsgemäß Probleme, sich über einen solch langen Zeitraum zu konzentrieren.

Hinzu kommt, dass manche Klassen aus organisatorischen Gründen nicht die vollen 90 Minuten an einem Stationsteil arbeiten konnten, wodurch die Zeit, die die Studierenden hatten, um die Videogruppe zu beobachten, leicht variiert. Es wurde jedoch keine Auswertung in Abhängigkeit der Zeitdauer des Laborbesuches vorgenommen. Es wäre möglich, dass eine erhöhte Arbeitszeit der Schülerinnen und Schüler bzw. Beobachtungsdauer der Studierenden dazu führt, dass öfter interveniert wird. Gleichmaßen führt eine verminderte Beobachtungszeit vermutlich dazu, dass weniger (kritische) Situationen zu beobachten sind, bei denen in den Lernprozess eingegriffen wird.

Es wurde nicht geprüft, ob Situationen auftraten, in denen die Studierenden aufgrund der vorliegenden Situation hätten intervenieren müssen, stattdessen wurden lediglich die getätigten

Interventionen analysiert. Da unterschiedliche Schülergruppen unterschiedliche Kenntnisstände besitzen und unterschiedliche Schwierigkeiten bei der Bearbeitung auftreten können, könnte es sein, dass durch die Analyse dieses Ausschnittes ein verzerrtes Bild entsteht.

Außerdem könnte es hinsichtlich der Bewertung der Angemessenheit schwierig für die Expertinnen und Experten gewesen sein, den Kriterienkatalog zur Bewertung der Angemessenheit direkt umzusetzen. Womöglich war es den Expertinnen und Experten auch nur bedingt möglich, alle Feinheiten in den Videosequenzen zu erkennen. Letzteres gilt sowohl für die Interventionen als auch für die Videovignetten im Bereich Diagnose.

10.3 Forschungsdesiderate

Auf Grundlage der vorliegenden Studie könnte untersucht werden, inwieweit einzelne Interventionsarten zum Lernerfolg der Schülerinnen und Schüler beitragen. Beispielsweise fanden Seidel und Shavelson (2007) heraus, dass Rückmeldehilfen einen positiven Einfluss auf den Lernfortschritt von Schülerinnen und Schülern haben. Dekker und Elshout-Mohr (2004, 62) kommen zudem zu dem Ergebnis, dass weniger inhaltliche Hilfen bei Lehrerinterventionen in Kleingruppen zu höherem Lernerfolg der Schülerinnen und Schüler führen. Inspiriert durch diese beiden Arbeiten und die Ergebnisse der vorliegenden Studie könnten die einzelnen Interventionsarten noch genauer beleuchtet werden. Insbesondere der Zusammenhang mit dem Lernerfolg, der beispielsweise über anschließende Leistungstests mit Kontrollgruppen überprüft werden könnte, wäre hierbei interessant.

Mit Blick auf die Studierenden könnte man folgendem Befund weiter nachgehen: Studierende, die angemessener intervenieren, scheinen sich auch deutlich intensiver mit den getätigten Interventionen auseinanderzusetzen. Dieser Befund ergibt sich aus den Ergebnissen der vorliegenden Studie, müsste jedoch noch kontrolliert werden, indem die Erhebung erneut durchgeführt würde und dabei Konstrukte, die einen Einfluss darauf haben könnten, miterhoben würden. Dadurch könnte man mithilfe einer Kovarianzanalyse feststellen, welcher Teil des Effekts wirklich auf das Interventionsverhalten entfällt und welcher Teil bereits durch die anderen Konstrukte erklärt wird.

Hinsichtlich der Angemessenheit der Interventionen wurden in dieser Studie kaum Unterschiede zwischen disziplinarischen Hilfen, Motivationshilfen, Rückmeldehilfen, allgemein-strategischen Hilfen und inhaltsorientiert-strategischen Hilfen festgestellt. An dieser Stelle könnten Expertinnen und Experten für die einzelnen Hilfe-Kategorien definiert und befragt werden, um belastbare Aussagen über die (womöglich doch unterschiedliche) Angemessenheit treffen zu können.

Studierende, die häufiger intervenieren und diese Interventionen größtenteils angemessen gestalten, reflektieren mehr über wesentliche Aspekte und denkbare Alternativen. Auch hier Bedarf es noch weiterer Forschungsarbeit. Es müsste anhand einer größeren Stichprobe untersucht werden, ob der Zusammenhang weiterhin Bestand hat oder sich aufgrund der geringen Stichprobe zufällig ergeben hat.

11 Literatur

- Aebli, H. (1997). *Zwölf Grundformen des Lehrens. Eine allgemeine Didaktik auf psychologischer Grundlage*. [9. Auflage] Stuttgart: Klett-Cotta.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), pp. 317-332.
- Akinwunmi, K. (2012). *Zur Entwicklung von Variablenkonzepten beim Verallgemeinern mathematischer Muster*. Wiesbaden: Springer Spektrum.
- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 3, S. 175-193.
- von Aufschnaiter, C., Cappell, J., Dübbelde, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J., Sträßer, R. & Wolgast, A. (2015). Diagnostische Kompetenz. Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 61, S. 738-758.
- Bartel, M.-E. & Roth, J. (2017). Diagnostische Kompetenz von Lehramtsstudierenden fördern – Das Videotool ViviAn. In J. Leuders, T. Leuders, S. Prediger & S. Ruwisch (Hrsg.), *Mit Heterogenität im Mathematikunterricht umgehen lernen – Konzepte und Perspektiven für eine zentrale Anforderung an die Lehrerbildung* (S. 43-52). Wiesbaden: Springer Spektrum.
- Bartel, M.-E. & Roth, J. (2020). Video- und Transkriptvignetten aus dem Lehr-Lern-Labor – Die Wahrnehmung von Studierenden. In B. Priemer & J. Roth (Hrsg.), *Lehr-Lern-Labore*. Wiesbaden: Springer Spektrum.
- Barzel, B. und Herget, W. (2006). Zahlen, Symbole, Variablen – abstrakt und konkret. Plädoyer für einen lebendigen Umgang mit Termen. *Mathematik lehren*, 136, S. 4-9.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9. Jg., Heft 4, S.469-520.
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M. & Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation und Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV) – Ein Forschungsprogramm. In: M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 7-26). Münster: Waxmann.

- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In: M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S.29-54). Münster: Waxmann.
- Becker, G. B. (2008): *Unterricht durchführen*. [9. Auflage] Weinheim: Beltz.
- Beretz, A.-K., Lengnink, K., & Aufschnaiter, C. v. (2017). Diagnostische Kompetenz gezielt fördern – Videoeinsatz im Lehramtsstudium Mathematik und Physik. In C. Selter, S. Hußmann, C. Höhle, C. Knipping, K. Lengnink, & J. Michaelis (Hrsg.), *Diagnose und Förderung heterogener Lerngruppen: Theorien, Konzepte und Beispiele aus der MINT-Lehrerbildung* (S. 149-168). Münster: Waxmann.
- Bertalan, D. (2007). Buchstabenrechnen? In: B. Barzel, T. Berlin, D. Bertalan & A. Fischer (Hrsg.), *Algebraisches Denken. Festschrift für Lisa Hefendehl-Hebeker* (S. 27-34). Hildesheim: Franzbecker.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223(1), S. 3–13.
- Boud, D., Keogh, R., Walker, D. (1985). Promotion Reflection in Learning: a Model. In: D. Boud, R. Keogh, D. Walker (Ed.): *Turning Experience into Learning* (pp. 18-40). London: Routledge Falmer.
- Bromme, R. (1992): *Der Lehrer als Experte: zur Psychologie des professionellen Wissens*. Huber-Psychologie Forschung. Bern-Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert, N. Birbaumer & C. F. Graumann (Hrsg.), *Psychologie des Unterrichts und der Schule* (Bd. 3, S. 177–212). Göttingen: Hogrefe Verl. für Psychologie.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen: Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler*. Münster: Waxmann.
- Brunner, M., Anders, Y., Hachfeld, A. & Krauss, S. (2011): Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Clarà, M. (2015). What is Reflection? Looking for Clarity in an Ambiguous Notion. *Journal of Teacher Education*, Vol. (66(3)), pp. 261-271.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. [2nd edition]. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Davis, Elisabeth A. (2006). „Characterizing productive reflection among preservice elementary teachers: Seeing what matters“. *Teaching and Teacher Education* 22, pp. 281-301.
- Dekker, R. & Elshout-Mohr, M. (2004). Teacher interventions aimed at mathematical level raising during collaborative learning. *Educational Studies in Mathematics* 56, pp. 39–65.
- Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA: D.C. Heath & Co Publishers.
- Diegritz, T., Rosenbusch, H. S., & Dann, H.-D. (1999). Neue Aspekte einer Didaktik des Gruppenunterrichts. In H.-D. Dann, T. Diegritz, & H. S. Rosenbusch (Hrsg.), *Gruppenunterricht im Schulalltag. Realität und Chancen* (S. 331–356). Erlangen: Universitätsverbund Erlangen-Nürnberg.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. [5. Auflage] Heidelberg: Springer.
- Drijvers, P. H. M. (2003). *Learning algebra in computer algebra environment. Design research on the understanding of the concept of parameter*. Utrecht: Freudenthal Institute.
- Dubs, R. (2009): Lehrerverhalten. *Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht*. [2. Auflage] Stuttgart: Franz Steiner Verlag.
- Duebbelde, G. (2013). *Diagnostische Kompetenzen angehender Biologie-Lehrkräfte im Bereich der naturwissenschaftlichen Erkenntnisgewinnung*. Kassel: Universität Kassel.
- Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th edition.). Englewood Cliffs, NJ: Prentice Hall.
- Enenkiel, P. & Roth J. (2017). Diagnosekompetenz mit Videovignetten fördern – Der Einfluss von Feedback. In U. Kortenkamp & A. Kuzle (Hrsg.), *Beiträge zum Mathematikunterricht 2017* (S. 1351-1354). Münster: WTM-Verlag.
- Von Felten, R. (2005). *Lernen im reflexiven Praktikum. Eine vergleichende Analyse*. Münster: Waxmann.
- Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. London: Sage.
- Franke, M. & Wynands, A. (1991). Zum Verständnis von Variablen – Testergebnisse in 9. Klassen Deutschlands. *Mathematik in der Schule*, 29 (10), S. 674-691.
- Freudenthal, H. (1983). *Didactical Phenomenology of Mathematical Structures*. Dordrecht: D. Reidel Publishing Company.
- Früh, W. (2017). *Inhaltsanalyse. Theorie und Praxis*. [9. Auflage] Stuttgart: UTB.

- Fürst, C. (1999). Die Rolle der Lehrkraft im Gruppenunterricht. In: H.-D. Dann, T. Diegritz & H.S. Rosenbusch (Hrsg.), *Gruppenunterricht im Schulalltag: Realität und Chancen* (107-150). Erlangen: Universitätsverbund Erlangen-Nürnberg.
- Fürst, C. & Haag, L. (1998). *Lehrer-Schüler-Interaktion im Gruppenunterricht. Ergebnisse der Außensicht*. Nürnberg: Sozialwissenschaftliches Forschungszentrum (SFZ) und Institut für Psychologie II.
- Gamer, M., Lemon, J., Fellows, I. & Singh, P. (2019). Irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1 [Computer Software]. Verfügbar unter <https://CRAN.R-project.org/package=irr>
- Gibbs, G. (1988). *Learning by Doing. A guide to teaching and learning methods. Further Education Unit*. Oxford Polytechnic: Oxford.
- Gore, J. M., & Zeichner, K. M. (1991). Action research and reflective teaching in preservice teacher education: A case study from the United States. *Teaching and Teacher Education*, 7(2), pp. 119-136.
- Greefrath, G., Oldenburg, R., Siller, H.-S., Ulm, V. & Weigand, H.-G. (2016). *Didaktik der Analysis. Aspekte und Grundvorstellungen zentraler Begriffe*. Heidelberg: Springer Spektrum.
- Griesel, H. (1982). Leerstellenbezeichnung oder Bedarfsname – Anmerkungen zur Didaktik des Variablenbegriffs. *Mathematische Semesterberichte*, 29 (1), S. 68-81.
- Guest, G., MacQueen, K. & Namey, E. (2012). *Applied Thematic Analysis*. Thousand Oaks, CA.: Sage Publications.
- Hänze, M., Schmidt-Weigand, F. & Blum, S. (2007). Mit gestuften Lernhilfen im naturwissenschaftlichen Unterricht selbständig lernen und arbeiten. In K. Rabenstein & S. Reh (Hrsg.), *Kooperatives und selbständiges Arbeiten von Schülern: Zur Qualitätsentwicklung von Unterricht* (197-208). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartkens, J. (2016). *Mathematische Reflexion in argumentativ geprägten Unterrichtsgesprächen. Eine empirisch-interpretative Untersuchung im 3. Und 4. Grundschuljahr*. Wiesbaden: Springer Spektrum.
- Hascher, T. (2011). Diagnostizieren in der Schule. In: Bartz, A., Dammann, M. & Huber, G. (Hrsg.), *PraxisWissen SchulLeitung Grundschule*. Kronach: Carl Link.
- Hatton, N. & Smith, D. (1995). Reflection in teacher education: towards definition and implementation. *Teaching & Teacher Education*, 11(1), pp. 33-49.

- Heinrichs, H. (2015). *Diagnostische Kompetenz von Mathematik-Lehramtsstudierenden: Messung und Förderung*. Heidelberg, Germany: Springer.
- Helmke, A. & Schrader, F.-H. (1987). Interactional Effects of Instructional Quality and Teacher Judgement Accuracy on Achievement. *Teaching & Teacher Education*, 3 (2), pp. 91-98.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Hohengehren: Schneider.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer.
- Hesse, I. & Latzko, B. (2011). *Diagnostik für Lehrkräfte*. [2. Auflage] Opladen: Barbara Budrich.
- Hilzensauer, W. (2008). Theoretische Zugänge und Methoden zur Reflexion des Lernens. Ein Diskussionsbeitrag. In: *bildungsforschung, Jahrgang 5, Ausgabe 2*. Verfügbar unter: <http://www.bildungsforschung.org/Archiv/2008-02/lernvermoegen/>
- Hofmann, R. & Roth, J. (2017). Fähigkeiten und Schwierigkeiten im Umgang mit Funktionsgraphen erkennen – Diagnostische Fähigkeiten von Lehramtsstudierenden fördern. In U. Kortenkamp & A. Kuzle (Hrsg.). *Beiträge zum Mathematikunterricht 2017* (S. 445-448). Münster: WTM-Verlag.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), pp. 297–313.
- Horstkemper, M. (2004). Diagnosekompetenz als Teil pädagogischer Professionalität. *Neue Sammlung*, 44, S. 201-214.
- Hoth, J. (2016). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften. Eine Vertiefungsstudie zur TEDS-Follow-Up-Studie*. Wiesbaden: Springer Spektrum.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), S. 1-55.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik*. Weinheim, Basel: Beltz.
- Jennrich, R. I. & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, 31(3), pp. 313-323.
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule*. Weinheim: Beltz.

- Karing, C., Pfost, M. & Artelt, C. (2011): Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for Educational Research Online*, Vol. 3, No. 2, pp. 119-147.
- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* [2. Auflage] (S. 75–102). Berlin: Springer.
- Kindt, M. (1980). Als een kat om de hete algebrij. *De Wiskrant, Tijdschrift voor Nederlands Wiskundeonderwijs*, 5(21), S. 155-157.
- King, P. M. & Kitchener, K. S. (1994). *Developing reflective judgement. Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.
- Kiper, H. & Mischke W. (2009): *Unterrichtsplanung*. Weinheim: Beltz.
- Kleber, E. W. (1976): Die Beurteilerfunktion des Lehrers. In E. W. Kleber, H. Meister, C. Schwarzer & R. Schwarzer (Hrsg.), *Beurteilung und Beurteilungsprobleme*. Weinheim: Beltz, S. 11-38.
- Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, 30, pp. 38-46.
- Kobarg, M. (2009): *Unterstützung unterrichtlicher Lernprozesse aus zwei Perspektiven. Eine Gegenüberstellung*. Münster: Waxmann.
- Körner, H., Lergenmüller, A., Schmidt, G. & Zacharias, M. (Hrsg.) (2016). *Mathematik 7 Neue Wege. Arbeitsbuch für Gymnasien Rheinland-Pfalz*. Braunschweig: Schroedel.
- Korthagen, F. A. J. (1985). Reflective Teaching and Preservice Teacher Education in the Netherlands. *Journal of Teacher Education*, 36(5), pp. 11-15.
- Korthagen, F. A. J. & Wubbels, T. (2002). Aus der Praxis lernen. In F. A. J. Korthagen, J. Kessels, B. Koster, B. Lagerwerf & T. Wubbels (Hrsg.), *Schulwirklichkeit und Lehrerbildung. Reflexion der Lehrertätigkeit* (S. 41 – 54). Hamburg: EB-Verlag.
- Krammer, K. (2009): *Individuelle Lernunterstützung in Schülerarbeitsphasen. Eine videobasierte Analyse des Unterstützungsverhaltens von Lehrpersonen im Mathematikunterricht*. Münster: Waxmann.
- Küchemann, D. (1978). Children's Understanding of Numerical Variables. *Mathematics in School*, 7(4), pp. 23-26.

- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. [3. Auflage] Weinheim: Beltz Juventa.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011): Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55-68). Münster: Waxmann.
- Kutscher, J. (1979). *Pädagogische Diagnostik. Zum Problem der Schülerbeurteilung*. Regensburg: Forum Academicum.
- Kvale, S. (2007). *Doing Interviews*. London: Sage.
- Lazarevic, C. (2015). *Professionelle Wahrnehmung und Analyse von Unterricht durch Mathematiklehrkräfte. Eine fallrekonstruktive Studie*. Wiesbaden: Springer.
- Leiss, D. (2007). „Hilf mir es selbst zu tun“. *Lehrerinterventionen beim mathematischen Modellieren*. Hildesheim: Franzbecker.
- Leiss, D. & Tropper, N. (2014). *Umgang mit Heterogenität im Mathematikunterricht. Adaptives Lehrerhandeln im Modellieren*. Heidelberg: Springer Spektrum.
- Leuders, T., Dörfler, T., Leuders, J. & Philipp, K. (2018): Diagnostic Competence of Mathematics Teachers. Unpacking a Complex Construct. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic Competence of Mathematics Teachers* (pp. 3-32). Cham: Springer International Publishing AG.
- Lindmeier, A. (2011). *Modeling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model for mathematics*. Münster, Germany: Waxmann.
- Link, F. (2011). *Problemlöseprozesse selbstständigkeitsorientiert begleiten. Kontexte und Bedeutungen strategischer Lehrerinterventionen in der Sekundarstufe I*. Wiesbaden: Vieweg und Teubner.
- Malle, G. (1993). *Didaktische Probleme der elementaren Algebra*. Wiesbaden: Vieweg.
- Malle, G. (2004). Grundvorstellungen zu Bruchzahlen. *Mathematik lehren*, 123, S. 4-8.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. [12. Auflage] Weinheim: Beltz.
- Mayring, Philipp (2016). *Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken*. [6. Aufl.] Weinheim: Beltz.

- Münster, C. & von Aufschnaiter, C. (2019). Theoriebezüge von Lehramtsstudierenden beim Diagnostizieren. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Kiel 2018*, S. 823-826.
- Praetorius, A.-K. & Südkamp, A. (2017): Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 13-18). Münster: Waxmann.
- Prediger, S., Tschierschky, K., Wessel, L. & Seipp, B. (2012). Professionalisierung für fach- und sprachintegrierte Diagnose und Förderung im Mathematikunterricht. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 17(1), S. 40–58.
- Randi, J. & Corno, L. (2005). Teaching and learner variation. *Pedagogy – Learning for Teachers. BJEP Monograph Series II*, 3 (pp. 47-69). The British Psychological Society.
- Rath, V. (2017). Diagnostische Kompetenz von angehenden Physiklehrkräften. Modellierung, Testinstrumententwicklung und Erhebung der Performanz bei der Diagnose von Schülervorstellungen in der Mechanik. In N. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Reinhold, S. (2018). Revealing and Promoting Pre-service Teachers' Diagnostic Strategies in Mathematical Interviews with First-Graders. In T. Leuders, K. Philipp & J. Leuders (Hrsg.), *Diagnostic Competence of Mathematics Teachers* (pp. 129-148). Cham: Springer International Publishing AG.
- Rießinger, T. (2016). *Gleichungen, Umformungen, Terme. Umgang mit Formeln leicht gemacht*. Heidelberg: Springer Spektrum.
- Robitzsch, A., Kiefer, T. & Wu, M. (2020). TAM: Test Analysis Modules. R package version 3.4-26. Verfügbar unter <https://CRAN.R-project.org/package=TAM>
- Roters, B. (2012). *Professionalisierung durch Reflexion in der Lehrerbildung. Eine empirische Studie an einer deutschen und einer US-amerikanischen Universität*. Münster: Waxmann.
- Roth, J. (2013a). Mathematik-Labor „Mathe ist mehr“ – Forschendes Lernen im Schülerlabor mit dem Mathematikunterricht vernetzen. *Der Mathematikunterricht*, 59(5), S. 12-20.
- Roth, J. (2013b). Vernetzen als durchgängiges Prinzip – Das Mathematik-Labor „Mathe ist mehr“. In A. S. Steinweg (Hrsg.), *Mathematik vernetzt - Tagungsband des AK Grundschule in der GDM 2013* (S. 65-80). Bamberg: University of Bamberg Press.
- Roth, J. (2020). Theorie-Praxis-Verzahnung durch Lehr-Lern-Labore – Das Landauer Konzept der mathematikdidaktischen Lehramtsausbildung. In B. Priemer & J. Roth (Hrsg.), *Lehr-Lern-Labore – Konzepte und deren Wirksamkeit in der MINT-Lehrpersonenbildung*. Heidelberg: Springer Spektrum.

- Schäfer, S. & Seidel, T. (2015). Noticing and reasoning of teaching and learning components by pre-service teachers. *Journal for Educational Research Online*, 7(2), pp. 34–58.
- Scherres, C. (2013). *Niveauangemessenes Arbeiten in selbstdifferenzierenden Lernumgebungen. Eine qualitative Fallstudie am Beispiel einer Würfelnetz-Lernumgebung*. Wiesbaden: Springer.
- Schoenfeld, A. H. & Arcavi, A. (1988). On the Meaning of Variable. *Mathematics Teacher*, 81(6), pp. 420-427.
- Schön, D. A. (1983). *The Reflective Practitioner. How Professionals Think in Action*. New York: Basic Books.
- Schrader, F.-W. & Helmke, A. (2001): Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45-58), Weinheim: Beltz.
- Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 68–71). Weinheim: Beltz.
- Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), pp. 461-464.
- Schwarz, B., Wissmach, B. & Kaiser, G. (2008). “Last curves not quite correct”: diagnostic competences of future teachers with regard to modelling and graphical representations. *ZDM*, 40(5), pp. 777–790.
- Seidel, T. & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen bei Lehrpersonen mit Hilfe von Videosequenzen. In M. Prenzel, I. Gogolin, H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaften, Sonderheft 8* (S. 201-218). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Seidel, T. & Shavelson, R. J. (2007): Teaching Effectiveness Research in the Last Decade: Role of Theory and Research Design in Disentangling Meta Analysis Results. *Review of Educational Research*, 77(4), pp. 454-499.
- Seidel, T., Blomberg, G. & Stürmer, K. (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. Zeitschrift für Pädagogik, Beiheft 56* (S. 296-306). Weinheim: Beltz.
- Seidel, T., & Sturmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, 51(4), pp. 739–771.

- Seifried, J. & Klüber, C. (2006). Lehrerinterventionen beim selbstorganisierten Lernen. In P. Gonon (Hrsg.), *Bedingungen beruflicher Moralentwicklung und beruflichen Lernens* (S. 153-164). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Seifried, J. & Wuttke, E. (2010). Student errors. How teachers diagnose them and how they respond to them. *Empirical Research in vocational education and training*, 2(2), pp. 147–162.
- Shulman, L. S. (1986): Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), pp. 4-14.
- Shulman, L. S. (1987): Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review* 57(1), pp. 1-23.
- Siebel, F. (2005). *Elementare Algebra und ihre Fachsprache: eine allgemein-mathematische Untersuchung*. Mühlthal: Verl. Allg. Wiss. – HRW.
- Siebert, H. (1991). Aspekte einre reflexiven Didaktik. In W. Mader (Hrsg.), *Zehn Jahre Erwachsenenbildung* (S. 19-26). Bad Heilbrunn: Julius Klinkhardt.
- Siller, H.-S. & Roth, J. (2016). Herausforderung Heterogenität. Grundvorstellungen als Basis und Bezugsnorm – das Beispiel Terme. *Praxis der Mathematik in der Schule*, 70, S. 2-18.
- Sjuts, J. (2007). Kompetenzdiagnostik im Lernprozess – auf theoriegeleitete Aufgabengestaltung und –auswertung kommt es an. *Mathematica didactica* 30(2), S. 33-52
- Specht, B. J. (2009). *Variablenverständnis und Variablen verstehen: empirische Untersuchungen zum Einfluss sprachlicher Formulierungen in der Primar- und Sekundarstufe*. Hildesheim: Franzbecker.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer/innen und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, S. 85–95.
- Stahnke, R., Schueler, S. & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: a systematic review of empirical mathematics education research. *ZDM Mathematics Education*, 48, pp. 1-27.
- Stender, P. (2016): *Wirkungsvolle Lehrerinterventionsformen bei komplexen Modellierungsaufgaben*. Wiesbaden: Springer Spektrum.
- Steindorf, G. (2000): *Grundbegriffe des Lehrens und Lernens*. [5. Auflage] Bad Heilbrunn: Verlag Julius Klinkhardt.

- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), pp. 743–762.
- Treich, F. (2018). Die Entwicklung der Professionellen Unterrichtswahrnehmung im Lehr-Lern-Labor Seminar. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik und Chemielernen. Band 261*. Berlin: Logos.
- Ursini, S. & Trigueros, M. (1997). Understanding of different uses of variable: A study with starting college students. In E. Pekhonen (Ed.), *Proceedings of the Twenty-first International PME Conference*, 4 (pp. 254-261). Lahti, Finland: PME.
- Usiskin, Z. (1999). Conceptions of school algebra and uses of variables. In B. Moses (Ed.), *Algebraic thinking, grades K-12: Readings from NTCM's school-based journals and other publications* (pp. 7-13). Reston, VA: National Council of Teachers of Mathematics.
- Van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educational Psychology Review*, 22(3), pp. 271-296.
- van Es, E.A., & Sherin, M.G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10, pp. 571–596.
- van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz - eine theoretische und empirische Annäherung. In N. B. W. Berkemeyer, H.-G. Holtappels, N. McElvany & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung* (S. 204–234). Weinheim: Beltz.
- Vom Hofe, R. (1995). *Grundvorstellungen mathematischer Inhalte*. Heidelberg: Spektrum.
- Vollrath, H.-J. (1989). Funktionales Denken. *Journal für Mathematikdidaktik* 10, S. 3-37.
- Vollrath, H.-J. & Weigand, H.-G. (2007). *Algebra in der Sekundarstufe*. [3.Auflage] München: Elsevier GmbH.
- Warren, E. (1999). The concept of a variable: gauging students' understanding. In O. Zaslavsky (Ed.), *Proceedings of the Twenty-third International PME Conference*, 4 (pp. 313-320). Haifa, Israel.
- Walz, M. & Roth, J. (2017). Professionelle Kompetenzen angehender Lehrkräfte erfassen – Zusammenhänge zwischen Diagnose-, Handlungs- und Reflexionskompetenz. In U. Kortenkamp & A. Kuzle (Hrsg.), *Beiträge zum Mathematikunterricht 2017* (S. 1277-1280), Münster: WTM-Verlag.

- Weinert (2001). Qualifikation und Unterricht zwischen gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In W. Melzer & U. Sandfuchs (Hrsg.): *Was Schule leistet. Funktionen und Aufgaben von Schulen* (S. 65-85). Weinheim: Juventa.
- Weinsheimer, J. (2016). *Diagnostische Fähigkeiten von Mathematiklehrkräften bei der Begleitung von Lernprozessen im arithmetischen Anfangsunterricht. Theoretische Konzeptualisierung, empirische Erfassung und Analyse*. Hildesheim: Franzbecker.
- Wiater, W. (2015): *Unterrichtsplanung*. [3. Auflage] Augsburg: Auer.
- Wildgans, A., Obersteiner, A., Fischer, F. & Reiss, K. (2018). Analyse der Diagnosekompetenzen von Studierenden des Grundschullehramtes in simulationsbasierten Lernumgebungen. In Fachgruppe Didaktik der Mathematik der Universität Paderborn (Hrsg.), *Beiträge zum Mathematikunterricht 2018. Vorträge zur Mathematikdidaktik und zur Schnittstelle Mathematik/Mathematikdidaktik auf der gemeinsamen Jahrestagung GDM und DMV 2018 (52. Jahrestagung der Gesellschaft für Didaktik der Mathematik)* (S. 1979–1982). Münster: WTM Verlag für wissenschaftliche Texte und Medien.
- Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17(2), pp. 89-100.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, Australien: ACER Press.
- Wygotski, L. (1987). *Ausgewählte Schriften. Band 2: Arbeiten zur psychischen Entwicklung der Persönlichkeit*. Köln: Pahl-Rugenstein.
- Zech, F. (2002): *Grundkurs Mathematikdidaktik. Theoretische und praktische Anleitungen für das Lehren und Lernen von Mathematik*. [10. Auflage] Weinheim: Beltz.

12 Anhang

A Diagnoseaufträge

B Kodierleitfaden Diagnoseaufträge

C Kodierleitfaden Interventionen

D Kodierleitfaden Reflexionsgespräche

Anhang A - Diagnoseaufträge

Der Anhang ist auf der beiliegenden CD-Rom ersichtlich

Anhang B - Kodierleitfaden Diagnoseaufträge

Der Anhang ist auf der beiliegenden CD-Rom ersichtlich

Anhang C - Kodierleitfaden Interventionen

Der Anhang ist auf der beiliegenden CD-Rom ersichtlich

Anhang D - Kodierleitfaden Reflexionsgespräche

Der Anhang ist auf der beiliegenden CD-Rom ersichtlich