

Empirische Überprüfung eines Verfahrens zur Messung von Reflexion bei Lehramtsstudierenden

DISSERTATION

zur Erlangung des Doktorgrades
im Fach Erziehungswissenschaft

vorgelegt an der
Philosophischen Fakultät der

Universität Passau

von

Martina Riel

Passau

2022

Erstgutachterin: Prof. Dr. Jutta Mägdefrau

Zweitgutachter: Prof. Dr. Detlef Urhahne

Tag der Disputation: 13.07.2022

Zusammenfassung

Reflexionsfähigkeit zu schulen, gilt als wichtiges Ziel der Lehrkräftebildung. Die Frage nach Befunden zur Ausprägung von Reflexionsfähigkeit bei Studierenden, deren Förderbarkeit und Wirkung steht deshalb im Zentrum des Interesses vieler Forschungsbemühungen, die allerdings mit dem Problem der schwierigen Messbarkeit von Reflexion und Mängeln bestehender Messmethoden hinsichtlich der Gütekriterien konfrontiert sind.

Davon ausgehend überprüft die vorliegende Studie in einem Mixed-Methods-Ansatz die Validität eines häufig eingesetzten Verfahrens zur Messung von Reflexion, nämlich die qualitative Inhaltsanalyse studentischer Reflexionstexte mithilfe des Stufenmodells von Hatton und Smith (1995). Dabei wird zwei Forschungsfragen nachgegangen: 1. Inwiefern wird das Ergebnis der Messung vom Reflexionsanlass in der Aufgabenstellung beeinflusst? 2. Welchen Einfluss haben verschiedene weitere Faktoren, die bisher selten oder nie kontrolliert wurden, auf das Messergebnis? Untersucht werden situationale Faktoren wie aufgabenbezogene Motivation bzw. wahrgenommener Nutzen der Aufgabe sowie das Verständnis des Begriffs Reflexion und verschiedene Persönlichkeitsmerkmale.

Lehramtsstudierende im Praktikum erhielten in einem Between-Subjects-Design drei verschiedene Reflexionsaufgaben, die sich im Grad des persönlichen Bezugs zum Reflexionsanlass unterschieden. Zu diesen verfassten sie reflektierende Texte und füllten im Anschluss einen Online-Fragebogen aus, der die zu untersuchenden weiteren Einflussfaktoren erhob. Die Ergebnisse der qualitativen Textanalyse wurden in zwei unterschiedliche quantitative Reflexionsmaße überführt und mit den Fragebogendaten trianguliert.

Die Aufgabenstellung hatte sowohl einen Einfluss auf die Reflexionsleistung als auch auf weitere qualitative Aspekte der Reflexionstexte. Die Reflexionsleistung zeigte sich unabhängig von den gemessenen Persönlichkeitsmerkmalen und der aufgabenbezogenen Motivation, wurde aber vom Grad der wahrgenommenen Nützlichkeit prädiziert, die der Aufgabe zugeschrieben wurde. Die Studierenden zeigten ein sehr stark divergierendes Verständnis des Begriffs Reflexion, welches

sich auch in den Texten niederschlug. Innerhalb des Studiendesigns und der Forschungsfragen lieferte die qualitative Inhaltsanalyse mit dem Stufenschema von Hatton und Smith (1995) plausible Ergebnisse, die Validität des Verfahrens wird aber von den Studienergebnissen zumindest teilweise in Frage gestellt.

Auf die Aufgabenstellung und die Identifikation möglicher einflussnehmender Drittvariablen muss bei der Messung von Reflexionsleistungen zukünftig noch mehr Aufmerksamkeit gerichtet werden. Die Suche nach einem Verfahren zur Messung von Reflexion, das die Gütekriterien zufriedenstellend erfüllt, stellt weiterhin ein Forschungsdesiderat dar.

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS	7
TABELLENVERZEICHNIS	8
1 EINLEITUNG	9
2 THEORIE UND FORSCHUNGSSTAND	11
2.1 Das Konstrukt Reflexion	11
2.1.1 Der Begriff Reflexion	11
2.1.1.1 Wortursprung und Begriffsgeschichte	11
2.1.1.2 Definitionsversuche	13
2.1.2 Aspekte von Reflexion	19
2.1.2.1 Anlass und Gegenstand von Reflexion	20
2.1.2.2 Ziel von Reflexion	22
2.1.2.3 Zeitpunkt der Reflexion	24
2.1.2.4 Ablauf eines Reflexionsprozesses	27
2.1.2.5 Reflexion als sozialer oder solitärer Prozess	29
2.1.2.6 Voraussetzungen für Reflexion	30
2.1.2.7 Reflexion und Emotion	37
2.1.2.8 Wirkungen von Reflexion	39
2.1.2.9 Nebenwirkungen von Reflexion	46
2.1.3 Zusammenfassung und Bewertung	52
2.2 Messung von Reflexion	54
2.2.1 Messinstrumente	54
2.2.1.1 Fragebögen	54
2.2.1.2 Inhaltsanalytische Verfahren	70
2.2.1.3 Weitere Messversuche	80
2.2.1.4 Zusammenfassung und Bewertung	82
2.2.2 Messergebnisse	85
2.2.2.1 Befunde zu Qualität und Häufigkeit von Reflexion	85
2.2.2.2 Entwicklung von Reflexionsfähigkeit	88
2.2.2.3 Förderbarkeit von Reflexionsfähigkeit	90
2.2.2.4 Zusammenfassung und Bewertung	95
2.2.3 Schwierigkeiten bei der Messung	98
2.2.3.1 Reflexionsanlass	98
2.2.3.2 Aufgabenformulierung und Prompts	101
2.2.3.3 Sichtbarmachung über ein Medium der Reflexion	104
2.2.3.4 Rahmenbedingungen	107
2.2.3.5 Zusammenfassung und Bewertung	110

3	FRAGESTELLUNG UND HYPOTHESEN	112
3.1	Fragestellung	112
3.2	Hypothesen	114
3.2.1	Aufgabenbezogene Hypothesen.....	114
3.2.2	Prädiktorbezogene Hypothesen	116
4	METHODE	119
4.1	Stichprobe	119
4.2	Instrumente	120
4.2.1	Aufgaben	120
4.2.2	Fragebogen	124
4.3	Untersuchungsdesign	128
4.4	Vorbereitende Auswertung und Datenanalyse	129
4.4.1	Auswertung der Reflexionstexte.....	129
4.4.1.1	Qualitative Inhaltsanalyse	129
4.4.1.2	Quantifizierung der Ergebnisse	135
4.4.2	Auswertung des Kreativitätstests	137
5	ERGEBNISSE	139
5.1	Reflexionsmaße	139
5.2	Aufgabenbezogene Hypothesen	140
5.3	Prädiktorbezogene Hypothesen	146
6	DISKUSSION	157
6.1	Zusammenfassung und Einordnung der Ergebnisse	157
6.2	Grenzen	168
6.3	Ausblick	176
6.3.1	Folgerungen für zukünftige Forschung	176
6.3.2	Folgerungen für die Lehrkräftebildung	181
7	LITERATURVERZEICHNIS	186
	ANHANG	222

Abbildungsverzeichnis

Abbildung 1. Formulierung der drei eingesetzten Reflexionsaufgaben.....	121
Abbildung 2. Untersuchungsdesign.	129
Abbildung 3. Kategoriensystem 1 (nach Hatton und Smith, 1995) mit Punktwertungen der Kategorien.	133
Abbildung 4. Höchste erreichte Reflexionsstufe in Abhängigkeit von der Reflexionsaufgabe. Die Fehlerbalken indizieren einfache Standardabweichungen.....	141
Abbildung 5. Durchschnittliche Reflexionsleistung in Abhängigkeit von der Reflexionsaufgabe. Die Fehlerbalken indizieren einfache Standardabweichungen.....	141
Abbildung 6. Boxplots der in den Reflexionstexten hergestellten Theoriebezüge in Abhängigkeit von der Reflexionsaufgabe.	144
Abbildung 7. Boxplots der Bezüge zu Gefühlen in Abhängigkeit von der Reflexionsaufgabe.....	145
Abbildung 8. Einschätzung der Proband*innen zu Reflexionsaufgaben: Boxplots zu den „Wie wichtig ist... im Rahmen einer Reflexion“-Items.	152
Abbildung 9. Boxplots zum Vorkommen der inhaltsanalytischen Kategorien (Teil 1 und 2) in den Texten.....	154

Tabellenverzeichnis

Tabelle 1. Überblick über die vorgestellten quantitativen Messinstrumente.....	55
Tabelle 2. Überblick über die vorgestellten qualitativen Messinstrumente.....	71
Tabelle 3. Vorkommen der jeweils höchsten im Text erreichten Reflexionsstufe in Prozent.....	140
Tabelle 4. Aufgabenbezogene Mittelwerte und Standardabweichungen der Variablen Motivation und Nutzen.....	143
Tabelle 5. Mittelwerte, Standardabweichungen und Interkorrelationen der Prädiktoren für die Reflexionsmaße (N = 93)	147
Tabelle 6. Regressionsanalyse für die Vorhersage des Maßes „Höchste erreichte Reflexionsstufe“	147
Tabelle 7. Regressionsanalyse für die Vorhersage des Maßes „Durchschnittliche Reflexionsleistung“	148
Tabelle 8. Mittelwerte, Standardabweichungen und Korrelationen der Prädiktoren für die Reflexionsmaße (N = 93)	149
Tabelle 9. Mittelwert, Standardabweichung und Korrelationen des Prädiktors Problemgehalt der Aufgabe für die Reflexionsmaße (N = 91).....	151

1 Einleitung

„It is good to reflect, but reflection itself also requires reflection.“ (Akbari, 2007)

Eines der größten Ziele der Lehrerbildung – und darüber herrscht erstaunlich große Einigkeit – ist die Förderung der Reflexionsfähigkeit der angehenden Lehrkräfte. Der „reflective practitioner“ als sich hinterfragender und sich lebenslang selbstständig weiterentwickelnder Professioneller gilt spätestens seit Donald Schöns Veröffentlichungen als Idealbild (Schön, 1983). Reflexion wird häufig gehandelt als „undoubtedly the right way of coping most successfully with the challenges of teaching“ (Penso et al., 2001, S. 336). Auch die Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Kultusministerkonferenz bzw. KMK) fordert in verschiedenen Kompetenzbereichen, dass Lehrkräfte sowohl theoretische Inhalte als auch ihr Wissen, ihre Haltungen und ihr Handeln regelmäßig reflektieren sollen (Kultusministerkonferenz, 2004, i.d.F. von 2019). Zahlreiche von der Qualitätsoffensive Lehrerbildung geförderte Projekte befassen sich mit der Förderung und Einbindung von Reflexion in die Lehrerbildung (Auge, 2019).

Vor allem hinsichtlich der Verknüpfung von Theorie und Praxis und damit auch bezüglich des Kompetenzerwerbs in schulpraktischen Phasen wird viel Hoffnung in das Konzept der Reflexion gesetzt. Auch in der zunehmend an Bedeutung gewinnenden Bildung für nachhaltige Entwicklung nimmt Reflexionskompetenz eine Schlüsselrolle ein (Fuertes-Camacho et al., 2021; Haan, 2010; Hager, 2020).

Der Terminus ist so allgegenwärtig, dass man mit Neuweg auch kritisch fragen kann, ob der Stellenwert des Reflektierens im Lehramtsberuf so zugenommen hat, „dass uns der Meister seines Faches dort gar nicht mehr durch die Kunstfertigkeit seiner Praxis, sondern fast nur mehr durch seine Reflexivität beeindrucken darf“ (Neuweg, 2017, S. 89).

Seit den 1990er Jahren und dem von Donald Schön bezeichneten „reflective turn“ (Schön, 1991), im Rahmen dessen Reflexion eine zentrale Größe beim Erwerb von Professionalität wurde, werden vielerorts Interventionen entwickelt, um Reflexionsfähigkeit von Lehramtsstudierenden zu fördern und diese Interventionen auch evaluiert, um ihren Erfolg zu belegen. Es gibt also eine immense Menge an Studien

und Publikationen, die – auf verschiedene Arten und mit verschiedenen Zielen – Reflexion beforschen und insofern auch zu messen versuchen. Über die Frage, wie Reflexion objektiv, reliabel und valide gemessen werden kann, herrscht bisher aber noch keine Einigkeit – genauso wenig wie über die Frage, wie der Begriff zu definieren und zu operationalisieren ist (Beauchamp, 2006; Rodgers, 2002; Williams & Grudnoff, 2011).

Damit Reflexionskompetenz in Zukunft vergleichbar und zuverlässig erforscht werden kann, setzt es sich die vorliegende Arbeit zunächst zum Ziel, die bisher vorliegenden Erkenntnisse zur Messung von Reflexion systematisch darzustellen und auszuwerten. Dafür wird der Begriff Reflexion beleuchtet sowie Aspekte dieses Konstrukts untersucht, die im Rahmen einer Messung relevant erscheinen. Es werden die verschiedenen bisher eingesetzten Messinstrumente vorgestellt, der Forschungsstand zur Messung von Reflexion dargelegt und mögliche Schwierigkeiten bei der Messung herausgearbeitet.

Diese Vorüberlegungen münden in einer eigenen empirischen Untersuchung, in der ein häufig verwendetes Messverfahren von Reflexion, nämlich die qualitative Inhaltsanalyse studentischer Reflexionstexte, eingesetzt und hinsichtlich verschiedener Aspekte auf seine Validität geprüft wird. Dafür werden die Ergebnisse der qualitativen Analyse in quantitative Reflexionsmaße überführt und der Einfluss der Aufgabenstellung, situationaler Faktoren bei der Aufgabenbearbeitung und genereller Persönlichkeitsmerkmale der Studierenden auf die Reflexionsmaße und die Reflexionstexte untersucht.

2 Theorie und Forschungsstand

2.1 Das Konstrukt Reflexion

„Although there is broad agreement that reflection is crucial for teacher education and teaching improvement and change, there is also, at the same time, similarly broad agreement that there is no clarity on what reflection is.“ (Clarà, 2015, S. 261)

Trotz der Tatsache, dass es etliche Definitionsversuche und Abhandlungen über Reflexion gibt, besteht weder in der Wissenschaft noch in der Alltagssprache Klarheit oder Einigkeit darüber, was Reflexion bedeutet. Gleichzeitig ist es natürlich schwer möglich, Reflexion, ohne die Basis eines gemeinsamen Begriffsverständnisses zu erforschen. Eine Einschätzung, die die meisten Forscher*innen teilen, ist, „dass Reflexion ein multidimensionales Konstrukt darstellt, wobei hinsichtlich der Anzahl und Natur der Dimensionen wenig Einigkeit besteht“ (Gutzwiller-Helfenfinger et al., 2017, S. 134).

Im ersten Teil dieses Kapitels wird der Begriff Reflexion kurz geschichtlich hergeleitet und auf die Schwierigkeiten bei der Definition eingegangen. Im zweiten Teil werden systematisch verschiedene Teilaspekte des Konstrukts vertiefter betrachtet.

2.1.1 Der Begriff Reflexion

2.1.1.1 Wortursprung und Begriffsgeschichte

Das Wort „Reflexion“ geht zurück auf den lateinischen Begriff ‚reflectere‘ und bedeutet „zurückbeugen“ oder „zurückwerfen“. Ursprünglich als ein Terminus der Optik verwendet, ging der Begriff später auch in den philosophischen Sprachgebrauch ein, blieb aber aufgrund seiner Herkunftsgeschichte „häufig mit der Metapher des Spiegels und des Sich-Spiegeln verbunden“ (Zahn, 2007, S. 396). Die Reflexion wurde in ähnlichem Sinn, wie sie von uns heute verstanden wird, also als eine Rückwendung des Denkens auf vergangenes Denken, Handeln und Fühlen, als

ein Wahrnehmen des eigenen Wahrnehmens bereits bei Platon und Aristoteles thematisiert (ebd.). Reflexion wird zum Teil bis heute gleichgesetzt mit der aristotelischen Tugend „phronesis“, der Klugheit des situationsangemessenen Handelns, vor allem auch im Sinne eines moralischen Urteilsvermögens (Birmingham, 2004). Descartes legte in der Neuzeit mit seiner Philosophie den Grundstein für die Selbstreflexion, indem er den Gedanken verfolgte, dass es möglich sei, von sich zurückzutreten und sich selbst als Objekt zu betrachten.

Die Verwendung des Begriffs und seine Bedeutung veränderten sich im Laufe der Philosophiegeschichte ständig. Das zeigt einerseits, dass der Terminus so offen ist, dass Unterschiedliches darunter verstanden werden kann und andererseits, dass der Begriff auch zur Projektionsfläche für den Zeitgeist werden kann.

Der Begriff Reflexion gelangte schließlich von der Philosophie auch in weitere Wissenschaftsbereiche. So wurde sie unter anderem zentrales Thema in der Soziologie, beispielsweise bei Pierre Bourdieu oder Niklas Luhmann (Forster, 2014), aber auch, seit John Deweys Werk „How We Think“ (1910), in der Erziehungswissenschaft. In diesem Bereich verstärkte sich die Auseinandersetzung noch um ein Vielfaches nach der Veröffentlichung von Donald Schöns Werken „The Reflective Practitioner“ (1983) und „Educating the Reflective Practitioner“ (1987). Infolge des dadurch ausgelösten „reflective turns“ setzte man sich schließlich auch in weiteren Professionen sehr intensiv mit dem Konzept auseinander, zum Beispiel in verschiedenen Gesundheitsberufen wie Medizin und Pflege (Mann et al., 2009) und in den Sozialwissenschaften, speziell in der Sozialen Arbeit, aber auch im Management (Fook et al., 2008).

Dass das Thema Reflexion in so vielen verschiedenen Wissenschaftsdomänen eine Rolle spielt, ist ein Fakt, den man bei der Auseinandersetzung mit dem Phänomen nicht aus den Augen verlieren darf. Thompson und Pascal (2012) kritisieren deshalb, dass Reflexion in der Erziehungswissenschaft zu einseitig betrachtet wird und beispielsweise soziologische und politische Elemente zu wenig Beachtung finden. Immer wieder melden sich auch Stimmen zu Wort, die zu einem Einbezug ganzheitlicher fernöstlicher Sichtweisen in die Diskussion mit Begriffen wie beispielsweise „mindfulness“ (Brockbank & McGill, 2000; Tremmel, 1993) anregen wollen.

Bestrebungen, den Fokus mehr zu erweitern und Reflexion als sehr komplexes Konstrukt mit Bedeutung in unterschiedlichen Kontexten und Perspektiven zu betrachten, stehen wiederum dem Versuch gegenüber, Reflexion einzugrenzen, so dass eine handhabbare Definition entwickelt werden kann.

Akbari (2007) stellt fest, dass gerade die vielfältige Geschichte des Begriffs eine einheitliche Definition heute schwierig werden lässt: „In fact, the term is traceable to so many influences that it has lost its real, core meaning and it means whatever academics want it to mean“ (S. 196).

Man hat es bei Reflexion also mit einem Thema zu tun, das herkömmliche Disziplinengrenzen überschreitet, und, als sei das nicht genug, ist der Begriff auch noch in der Alltagssprache populär, was im folgenden Kapitel genauer beleuchtet wird.

2.1.1.2 Definitionsversuche

Annäherung über ein Alltagsverständnis. Der Begriff Reflexion gelangte schon vor vier Jahrhunderten aus der Wissenschaft in die Alltagssprache und entwickelte sich dort eigenständig weiter: „Vom 17. Jh. an ist er [der Begriff „Reflexion“; M.R.] auch im Englischen und Französischen alltagssprachlich im weiteren Sinne von ‘Überlegung’, ‘nachdenkende Betrachtung’ gebräuchlich“ (Zahn, 2007, S. 396). In diesem Sinne wird er auch heute noch primär verwendet; so gibt die Dudenredaktion (o.J.) als Bedeutung von Reflexion „das Nachdenken; Überlegung, prüfende Betrachtung“ an. Darüber hinaus gibt es aber auch noch breitere und teils vage Verständnisweisen des Begriffs. Berndt und Häcker ermittelten im Rahmen einer Studierendenbefragung weitere Facetten des Verständnisses. So verwendeten die Proband*innen der Studie zur Umschreibung von „Reflexion“ Begriffe wie „Vergegenwärtigen (von Gedanken), Hinterfragen, Analysieren, Auswerten, Nachdenken, Wider- und Zurückspiegeln, Nachsinnen sowie rückwirkende Eigenbeobachtung“ (Berndt & Häcker, 2017, S. 246). Es zeigte sich eine „semantische Breite“ (ebd.), die auf ein heterogenes Verständnis hindeutet. Ähnliche Befunde erbrachte eine Studie von Wyss und Mahler (2021), die in einer qualitativen Untersuchung ebenfalls nach der Bedeutung des Begriffs Reflexion gefragt hatten und sehr vage Umschreibungen erhielten (vgl. S. 18). Auch die Abgrenzung zu anderen Begriffen,

beispielsweise der Analyse, fiel schwer. Eine Studienteilnehmerin beschrieb beispielsweise: „Also ja, ich finde Analysieren passt als Begriff auch, wenn man an das Reflektieren denkt“ (S. 19).

Sich mit dem Alltagsverständnis auseinanderzusetzen, in dem der Begriff sowohl breit ausgelegt als auch individuell unterschiedlich definiert wird, ist deshalb von Bedeutung, weil Forschung zur Reflexion nahezu ausschließlich an Personen durchgeführt wird, die zum Begriff keinen wissenschaftlichen, sondern einen alltäglichen Zugang haben. Deren Verständnis von Reflexion spielt also auch für die Wissenschaft eine wesentliche Rolle, weil es die Ergebnisse der Studien beeinflussen könnte.

Wissenschaftliche Definitionsversuche. Relevant sind diese Überlegungen außerdem, weil es im Interesse der Wissenschaft liegen muss, den Begriff so weit einzugrenzen, dass er handhabbar wird. Es scheint deshalb wichtig zu sein, sich auch die Frage zu stellen, was Reflexion *nicht* ist. Das betrifft sowohl das grundsätzliche „Nachdenken“ als auch die bereits erwähnte Analyse. Häcker (2017) beschreibt Reflexion als eine spezielle Form des Denkens „und zwar als eine rekursive, referenzielle bzw. selbstreferenzielle“ (S. 23). John Dewey macht ebenfalls deutlich, dass Reflexion nicht mit jeder Form des Nachdenkens gleichzusetzen ist, beschreibt solche verschiedenen Formen des Nachdenkens und grenzt Reflexion davon zusammenfassend ganz klar durch zwei Charakteristika ab. Reflexion, im Gegensatz zum einfachen Nachdenken, „(1) involves a state of doubt, hesitation, perplexity, mental difficulty, in which thinking originates and (2) [is] an act of searching, hunting, inquiring, to find material that will resolve the doubt, settle and dispose of the perplexity“ (Dewey, 1933, S. 12).

Gibt es keinen externen Anlass im Sinne eines Problems oder einer Schwierigkeit, kann laut ihm nicht von Reflexion gesprochen werden. Außerdem gehört zum Reflektieren die abwägende, kritische Suche nach einer Lösung des Problems. Es handelt sich also um einen zielgerichteten Akt im Gegensatz zu einem zufälligen, umherschweifenden, ungerichteten Nachdenken.

Eine Abgrenzung von Reflexion zum Begriff „Analyse“ legen beispielsweise Kempin et al. (2017) oder auch Mühlhausen (2006) nahe. Eine Gleichsetzung bezeichnet Letzterer als Etikettenschwindel. Analyse ist nach Mühlhausen die Zerlegungen eines komplexen Sachverhalts in Einzelaspekte. Reflexion dagegen meine

ein Nachdenken, das im Gegensatz zur Analyse immer auch mit der eigenen Person zu tun hat und auf zukünftiges Handeln wirken soll (vgl. S. 19).

Was eine Abgrenzung allerdings erschwert, ist die Tatsache, dass Tätigkeiten wie die Analyse durchaus Teilkompetenzen von Reflexion darstellen können, auch wenn sie nicht mit dem Begriff selbst gleichzusetzen sind.

Weitere solche Teilkompetenzen oder „Skills“ im Rahmen einer Reflexion beschreiben beispielhaft Atkins und Murphy (1993): „self-awareness“, „description“, „critical analysis“, „synthesis“ und „evaluation“ (vgl. S. 1190). Auch Beauchamp (2006) legt auf der Basis einer umfassenden Literaturanalyse eine Übersicht vor, welche Prozesse im Rahmen einer Reflexion erwartet werden. Sie listet auf: Untersuchen, Denken und Verstehen, Problemlösen, Analysieren, Evaluieren, Konstruieren, Entwickeln und Transformieren. Dem hinzuzufügen wäre nach Korthagen beispielsweise auch noch der Begriff des Strukturierens, da er Reflexion definiert als „mental process of structuring or restructuring an experience, a problem or existing knowledge or insights“ (Korthagen, 1999, S. 193). Diese Begriffe werden zum Teil als Aspekte von Reflexion aufgeführt, teilweise aber auch implizit oder explizit mit dieser gleichgesetzt. Eine klare Abgrenzung ist bisher nicht gelungen. Das liegt auch daran, dass es sich bei Reflexion um ein normatives Konstrukt handelt, wodurch es möglich wird, Unterschiedliches in den Begriff hineinzulegen. In diesem Sinne könnte man Reflexion als ein „Plastikwort“ (Pörksen, 2011) bezeichnen, das zwar den Anschein erweckt, einen Fachbegriff darzustellen, aber eigentlich eine leere Worthülse ist. Wie man Reflexion definiert, hängt also davon ab, was man selbst als wichtig erachtet, was man beispielsweise für die Lehrerbildung oder zukünftige Lehrkräfte als wichtige Fähigkeiten und Kompetenzen betrachtet und was man unter Professionalität versteht. Dabei wird ersichtlich, dass bezüglich Reflexion häufig nicht nur ungeklärt ist, um welche Tätigkeit es sich dabei genau handelt, sondern auch, zu welchem Zweck und in welchen Kontexten sie eingesetzt werden soll:

Today's discourse of reflection incorporates an array of meanings: a demonstration of self-consciousness, a scientific approach to planning for the future, a tacit and intuitive understanding of practice, a discipline to become more professional, a way to tap into one's authentic inner voice, a means to become a more effective teacher, and a strategy to redress injustices in society. (Fendler, 2003, S. 20)

Genauso unklar ist dabei, an welchen Kriterien eine erfolgreiche Reflexion festgemacht werden soll. Berndt und Häcker (2017) beispielsweise machen den Erfolg alltäglicher Reflexion daran fest, ob sie subjektiv nützlich und hilfreich war, professionelle Reflexion hingegen gelte nur als erfolgreich, wenn sie sich auf theoretisches Wissen beziehe (vgl. S. 252).

Je mehr sich mit dem Konzept Reflexion auseinandergesetzt wird und in je mehr Kontexten davon gesprochen wird, desto mehr verwässert letztendlich der Begriff: „Reflection has suffered from a loss of meaning. In becoming everything to everybody, it has lost its ability to be seen” (Rodgers, 2002, S. 843).

Zusammenfassend kann man sagen, dass bisher keine konsensfähige Definition formuliert werden konnte (Korthagen et al., 2001, S. 51) und sich hinter dem Begriff eine große Variation verschiedener Auffassungen verbirgt (Williams & Grudnoff, 2011, S. 281). Das ist, nach Beauchamp (2015), das größte Defizit der bisherigen Forschung zu Reflexion und daraus leitet sie als wichtigste Forderung für die Zukunft ab, an einer einheitlichen Definition zu arbeiten (vgl. S. 133). Damit diese Arbeit das Forschungsfeld nicht mit noch einer weiteren individuell formulierten Definition verkompliziert, wird in der vorliegenden Arbeit mit der **Definition des Begriffs Reflexion von John Dewey** (1910; 1933) gearbeitet. Diese wurde gewählt, weil Dewey als der erste und wichtigste Theoretiker zum Begriff Reflexion in der Erziehungswissenschaft gelten kann und unser heutiges Verständnis des Konzepts maßgeblich geprägt hat.

Reflexion ist laut ihm eine „active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends” (Dewey, 1933, S. 9). Reflexives Denken sei sowohl induktiv als auch deduktiv und springe hin und her zwischen der Betrachtung der Situation als ganzer und ihrer Einzelaspekte. Er bezeichnet Reflexion als eine Kette von zusammenhängenden Gedankengängen, von denen einer zum anderen führt und auf eine investigative Art und Weise einer Lösung zustrebt. Dabei sind die Aktivierung von Vorerfahrungen, genaue Beobachtung und der Einbezug alternativer Erklärungen von Bedeutung. Auch Theoriewissen und wissenschaftliche Erkenntnisse können in diesen Prozess einfließen, müssen aber nicht zwangsläufig, um zu einem erfolgreichen Abschluss zu kommen.

Begriffe im Kontext der Messung von Reflexion. Die Komplexität des Begriffs Reflexion wurde bereits erörtert, stellt aber in der Forschung zu diesem Konstrukt nicht das einzige Problem dar. Bei der Messung von Reflexion tauchen nämlich noch diverse andere Begrifflichkeiten auf, die teilweise synonym, teilweise aber doch mit unterschiedlichen Schwerpunkten und Konnotationen verwendet werden, ohne dass diese klar definiert würden. Das Problem verstärkt sich zusätzlich durch die Unterschiedlichkeit der Begrifflichkeiten, die im deutschen und anglo-amerikanischen Sprachraum verwendet werden. In Letzterem sind es die Begriffe „reflection, reflective practice, critical reflection, reflexivity, and critical thinking“ (van Beveren et al., 2018, S. 3), die dem Konzept Reflexion zuzuordnen sind und die nicht trennscharf verwendet werden. Im deutschen Sprachraum werden zu messen versucht: Reflexivität (Auge, 2019; Freisler-Mühlemann et al., 2021), Reflektierfähigkeit (Hager, 2020), Reflexionsfähigkeit (Lambrecht & Bosse, 2020; Volmer et al., 2021), Reflexionsbereitschaft (Neuber & Göbel, 2016) und Reflexionsperformanz (Kempin et al., 2019; Stender et al., 2021) als Ausdruck der Reflexionskompetenz (Aufschnaiter et al., 2019). Wie sich die englischen und deutschen Begriffe zueinander verhalten, ist dabei genauso unklar, wie die Bedeutungen der Begriffe einzeln betrachtet. Häufig werden sie zudem synonym verwendet. Vor allem der neue Begriff der Reflexionskompetenz vermittelt den Eindruck, dass damit eine neue Qualität des Konzepts ausgedrückt würde, obwohl nicht immer definiert wird, inwiefern hier etwas anderes gemeint oder erfasst wird, als wenn beispielsweise von Reflexionsfähigkeit oder schlicht von Reflexion gesprochen wird. Einige Studien versuchen aber durchaus, den Begriff der Reflexionskompetenz zu definieren und deren Teilkompetenzen und Entwicklungsstufen zu systematisieren (Hilzensauer, 2017; Larrivee, 2008; Meyer-Siever & Levin, 2018; Stender et al., 2021).

Stender, Watson, Vogelsang und Schaper (2021) beispielsweise definieren Reflexionskompetenz als situationsspezifische (kognitive) Fähigkeit, die beeinflusst wird von reflexionsbezogenen Dispositionen (z.B. Selbstwirksamkeitserwartungen) und sich ausdrückt in einer reflexionsbezogenen Performanz, die wiederum Rückschlüsse auf die Fähigkeit zulässt (vgl. S. 232). Was hier beispielsweise fehlt, sind volitionale und motivationale Komponenten und auch emotionale Aspekte. Hilzensauer (2017) listet in einer „Taxonomie für Reflexionskompetenz“ (S. 81ff)

sechs Merkmale einer kompetenten Reflexion auf: Bezug zu unterrichtsrelevanten Kriterien, Bezug zu handlungsleitenden subjektiven Überzeugungen, Bezug zu situationsbezogenen Gefühlen, das Adressieren mehrerer Beurteilungsperspektiven, die Artikulation mehrerer Handlungsalternativen und die Entstehung eines Kompetenzentwicklungsplans. Bei der Arbeit mit einer solchen Taxonomie ist zu bedenken, dass auch kontextspezifische Rahmenbedingungen und Dispositionen einen wesentlichen Einfluss auf das Reflektieren nehmen und die ermittelte Kompetenzstufe beeinflussen können:

Inwiefern und in welchem Ausmaß die kognitiven, pädagogischen, motivationalen und selbstregulativen Dispositionen in einer Situation zum Tragen kommen und eine kompetente Handlung in Form ihrer sichtbaren Performanz „erzeugen“, ist weiterhin noch eher ungeklärt. Menschen werden in ihren psychischen Dispositionen permanent von äußeren Rahmenbedingungen beeinflusst. (Kufner, 2012, S. 192)

Auch beim Begriff Reflexionskompetenz werden also wieder unterschiedliche Schwerpunktsetzungen deutlich, sodass in Kombination mit den weiteren erwähnten Begriffen das ohnehin bereits schwierige Konzept „Reflexion“ noch weiter ausdifferenziert und verkompliziert wird. Während sich die Forschungslandschaft dahingehend wandelt, dass immer häufiger versucht wird, Reflexion als komplexe, mehrdimensionale Kompetenz zu messen, ist unklar, wie diese zu modellieren ist und welche Teildimensionen erhoben werden sollen.

Im Folgenden sollen solche relevanten Dimensionen und Aspekte von Reflexion, die bereits angerissen wurden, noch weiter vertieft werden, um das unklare Konstrukt zu klären. Dabei wird jeweils auch die Haltung Deweys dargestellt, weil in nahezu allen Texten über Reflexion zwar immer wieder auf ihn als ersten und wichtigsten Ideengeber und Urvater der Reflexion in der Erziehungswissenschaft Bezug genommen wird, seine Werke und Ansichten dann aber selten tatsächlich dargestellt und einbezogen werden. Die vorliegende Arbeit möchte deshalb noch einmal umfassender die Wurzeln des Konzepts von Dewey in den Fokus rücken.

2.1.2 Aspekte von Reflexion

Im vorausgegangenen Kapitel wurde herausgearbeitet, dass das Formulieren einer einheitlichen Definition schwierig ist und es sich bei der Reflexion um ein komplexes Konstrukt handelt. Herzog spricht noch einmal die wesentlichen Aspekte an, die im Kontext von Reflexion eine Rolle spielen und anhand derer sich die zu vertiefenden Aspekte für dieses Kapitel ergeben:

Einige meinen damit [mit dem Reflektieren, M.R.] das schlichte Nachdenken über irgend etwas [sic], wozu auch die eigenen Erziehungsvorstellungen gehören können, andere grenzen die Reflexion auf ein Nachdenken über Verhaltensweisen ein. Einige sehen als Themen von Reflexion zeitlich enge Einheiten (wie Handlungen), andere denken an zeitlich ausgedehnte Einheiten (wie biographische Abläufe, Lebensphasen etc.). Einige konzentrieren Reflexion auf (akute) Probleme, andere vertreten einen problemunabhängigen Ansatz, bei dem auch über unproblematische Verhaltensweisen nachgedacht wird. Einige meinen mit Reflexion einen Prozess, der solitär durchgeführt wird, andere verstehen darunter ein kommunikatives Geschehen, das nur im Gespräch mit anderen stattfinden kann. Einige grenzen die Reflexion auf das Feld des Bewusstseins ein, andere denken auch an die Aufdeckung unbewusster oder (bisher) unartikulierter Wissensbestände oder Handlungsmotive. Einige sind eher deskriptiv ausgerichtet und beschränken den reflexiven Prozess auf die möglichst genaue Analyse, andere verstehen unter Reflexion ein normatives Unterfangen, das kritisch orientiert ist und unreflektierte Werte hinterfragen soll. (Herzog, 1995, S. 253)

Betrachtet werden im Folgenden Anlass und Gegenstand von Reflexion, das Ziel dieses Prozesses, der Zeitpunkt, also wann Reflexion stattfindet, und der Ablauf des Prozesses. Ferner wird die Frage beleuchtet, ob Reflexion ein soziales oder solitäres Unterfangen ist, welche Voraussetzungen es für ihre Durchführung gibt und inwiefern Reflexion und Emotion zusammenhängen. Abschließend wird auf mögliche positive und negative Folgen und Wirkungen von Reflexion eingegangen.

2.1.2.1 Anlass und Gegenstand von Reflexion

Zunächst soll die Frage betrachtet werden, wodurch Reflexion ausgelöst werden kann und welche Situationen oder Themen Gegenstand einer Reflexion werden können.

Den Anlass für eine Reflexion bildet nach Dewey immer eine „perplexed, troubled, confused situation“ (1933, S. 106), aus der sich ein Problem ergibt, für das die betroffene Person nicht unmittelbar eine Lösung hat. Diese reflexionsauslösende Situation ist gekennzeichnet durch eine gewisse Ambiguität, eine Uneindeutigkeit und die Möglichkeit verschiedener alternativer Lösungen oder Bewältigungsmöglichkeiten. Es gibt nach Dewey also keine Reflexion, wenn es keine ungelösten Fragen oder Schwierigkeiten gibt, die überwunden werden müssen. Deren Auflösung erfolgt dann über den Reflexionsprozess (vgl. Dewey, 1910, 10f).

Auch Donald Schön argumentiert in die gleiche Richtung und setzt das Vorhandensein einer gewissen irritierenden Situation als Anlass für eine Reflexion voraus: „There is some puzzling, or troubling, or interesting phenomenon with which the individual is trying to deal“ (Schön, 1983, S. 50). Während bei Dewey die beschriebenen Auslöser zum größten Teil eher negativ konnotiert sind, schwingen bei Schön mit „interesting phenomenon“ auch positivere Aspekte mit. Boud, Keogh und Walker beschreiben, dass Reflexion auch von positiven Ereignissen ausgelöst werden kann, beispielsweise der unerwarteten Bewältigung einer unlösbar geglaubten Aufgabe (2013, S. 19).

Artmann, Herzmann, Hoffmann und Proske (2013) ergänzen die möglichen Auslöser um eine weitere Kategorie und unterscheiden einerseits zwischen Problemstellungen und andererseits Differenzerfahrungen. Mit ersteren stützen sie die bisherige Argumentation. Mit Differenzerfahrungen fügen sie einen neuen Begriff hinzu und definieren ihn als eine Diskrepanz beispielsweise zwischen Vorstellungen über Unterricht und der tatsächlichen Beobachtung (vgl. S. 137). Diese Differenzerfahrungen könnten auch die von Boud et al. (2013) beschriebenen überraschend positiven Erlebnisse miteinschließen.

Auch im studentischen Alltagsverständnis von Reflexion, wie es von Berndt und Häcker untersucht wurde, spiegeln sich diese Arten von Auslösern wider. Hier werden als Initiatoren für einen Reflexionsprozess beschrieben: „die Wahrnehmung von Diskrepantem, etwas zwischen Gewusstem und Ungewusstem („Neugier“), zwischen Gewolltem und Gekonntem bzw. Gemachtem oder zwischen Vorhaben und Resultat („Unzufriedenheit“) und zum anderen die Abwägung zwischen verschiedenen Handlungsmöglichkeiten“ (Berndt & Häcker, 2017, S. 247).

Artmann et al. (2013) erweitern den Gegenstand von Reflexion über die Dimension des Erlebten hinaus auf den Bereich des Wissens (vgl. S. 137). Dieses Wissen kann entweder akademisches, fachwissenschaftliches Wissen oder Wissen in Form von subjektiven Theorien sein. Auch Jenert (2008) „vertritt die Auffassung, dass reflexives Lernen [...] auch Grundannahmen zum Lernen und individuelle Bildungsziele thematisieren muss“ (S. 1). Ähnliche Forderungen finden sich auch in den eingangs erwähnten KMK-Standards. Beauchamp (2006) arbeitet im Rahmen einer Sichtung verschiedener Ansätze heraus, dass Reflexion oft nicht nur auf eigene Praxis oder Erlebtes, sondern auch auf soziales Wissen, Informationen, Theorien, Bedeutungen, Überzeugungen oder Interessen bezogen wird.

Im Gegensatz dazu steht Deweys Position, dass Reflexion nur aus einer direkt erlebten Situation entstehen kann, „some situation that is directly experienced, something undergone, done, enjoyed, or suffered, and not just thought of“ (Dewey, 1933, S. 99).

Daraus ergibt sich die Frage, ob Reflexion wirklich nur nach einem solchen Anlass im Erleben erfolgen kann oder ob sie auch didaktisch induziert werden kann. Häcker unterteilt Reflexionen gemäß dieser beiden Kategorien, abhängig von der Frage, ob sie einen natürlichen Anlass hatten oder veranlasst wurden (Häcker, 2017, 26f). Demzufolge wäre eine Reflexion auch auslösbar, ohne dass zuvor selbst etwas erlebt wurde. Deweys Ansatz deckt sich an dieser Stelle nicht mit der didaktischen Praxis und der Forschungspraxis, bei der nur zum Teil auf eigene Erlebnisse zurückgegriffen wird, häufig aber auch zu ausgewählten Texten, Fallvignetten in Form von Videos und Fotos oder theoretischen Fragestellungen reflektiert werden soll.

Nach Hahn (1987) ist Reflexion sogar weniger eine natürliche, aus dem Selbst herauskommende Tätigkeit, sondern es ist eher das Vorhandensein einer bestimmten Umgebung, die Menschen dazu zwingt oder es ihnen nahelegt bzw. ermöglicht, zu reflektieren. Das spräche wiederum für das Potential bewusst gestalteter Lernumgebungen bei der Auslösung von Reflexion. Auch Boud, Keogh und Walker (2013) unterteilen Reflexionsanlässe in eher offizielle „didaktische“ Anlässe wie Seminare, Vorlesungen oder Workshops und informelle Anlässe wie Alltagssituationen. „It could be provoked by an external agent or it could be an internal experience, arising out of some discomfort with one’s present state“ (S. 18).

Es ist also nicht ganz abschließend geklärt, ob Reflexion bewusst von außen ausgelöst werden kann oder nur von innen. Ebenfalls offen ist die Frage, ob sich die Reflexion auf etwas richten kann, das nichts mit der eigenen Biografie und dem eigenen Erleben zu tun hat oder nicht. Es ist nicht auszuschließen, dass Reflexion auch didaktisch induziert werden kann, denn auch diese Situationen werden ja vom Subjekt erlebt und können zu einer Irritation, Überraschung oder Differenzerfahrung führen. Allerdings ist es dafür nötig, dass eine „aktive Übernahme und Bedeutungs-zuweisung durch das Subjekt“ erfolgt (Häcker, 2017, S. 26). Es wäre also essenziell, dass das Außenliegende, das zunächst nichts mit der Person und ihrer Biografie zu tun hat, internalisiert und zu einem eigenen Problem gemacht wird. Das kann letztlich nur durch das Subjekt selbst geschehen. Reflexion, wenn sie veranlasst wird, bleibt also optional.

2.1.2.2 Ziel von Reflexion

“[...] reflection is not an end in itself but a tool or vehicle [...]” (Rodgers, 2002, S. 863)

Von Reflexion wird häufig so positiv und optimistisch gesprochen, dass der Eindruck entstehen kann, sie wäre ein Selbstzweck und hätte einen Wert an sich, der nicht mehr belegt werden müsste. Dem ist nicht so, denn Reflexion ist eigentlich ein Mittel zum Zweck, ein Werkzeug, um ein bestimmtes Ziel erreichen zu können: „The function of reflective thought is, therefore, to transform a situation in which there is experienced obscurity, doubt, conflict, disturbance of some sort, into a

situation that is clear, coherent, settled, harmonious” (Dewey, 1933, 100f). Das Ziel von Reflexion besteht also nach Dewey in der Auflösung des Problems, das sich als Ausgangspunkt der Reflexion gestellt hat.

Dass Reflexion überhaupt ein Ziel haben muss, wird auch von Dewey an verschiedenen Stellen nachdrücklich betont. Es geht also nicht darum, dass sich Gedanken nur ziellos im Kreis drehen – „reflective thinking aims at a conclusion“ (ebd., S. 5). Es gilt, dieses Ziel der finalen Problemlösung zu erreichen und den Prozess nicht vorher abubrechen: „Think it *out*“ (ebd., S. 6; Hervorhebung im Original). Handelt es sich nur um eine gedankliche Auflösung oder soll diese in eine Tätigkeit, also in verändertes Handeln übergehen? Dazu schreibt er an anderer Stelle: „Reflection converts action that is merely appetitive, blind, and impulsive into intelligent action” (ebd., S. 13). Hier wird auch auf den handlungswirksamen Charakter von Reflexion hingewiesen. Es geht bei der Reflexion also darum, natürliche menschliche Schwächen durch kontrollierte Denkprozesse auszugleichen und dadurch zu intelligentem Handeln zu gelangen. Das Tier handelt nur im Sinne einer Reaktion auf äußere Stimuli, der Mensch hingegen kann intentional handeln, sein Verhalten auf ein erdachtes Ziel hin planen. Er kann Vorbereitungen treffen und Dingen eine Bedeutung zuweisen und erhält so ein gewisses Maß an Kontrolle über die Außenwelt. Gewisse Neigungen des Menschen stehen dieser Fähigkeit entgegen. Dazu führt Dewey beispielsweise den Hang zum Aberglauben auf, die leichtfertige Übernahme der Meinung anderer, Wahrnehmungsfehler oder das Leitenlassen von Lustgefühlen oder Launen (ebd., S. 26). Das reflektierende Denken soll also mit dem Ziel eingesetzt werden, diese menschlichen Schwächen auszugleichen und sich nicht von diesen fehlerhaften Denk- und Verhaltensweisen in die Irre führen zu lassen. Die Stärken des menschlichen Intellekts sollen durch Reflexion zum Tragen kommen und eine bessere und bewusstere Sicht auf die Dinge erzeugen. Dabei soll es also zu einer tatsächlichen Veränderung kommen, die sich auch in verändertem Handeln widerspiegelt: „Application is as much an intrinsic part of genuine reflective inquiry as is alert observation or reasoning itself.” (ebd., S. 213). Es scheint also wichtig zu sein, die Handlungsebene mitzudenken. Häufig liegt gerade in der wissenschaftlichen Diskussion der Fokus primär auf Kognition und einer intellektuellen Betätigung, die sich nicht unbedingt im Handeln und Verhalten widerspiegeln müssen. „Teacher educators have separated reflection from the world of action

and experience” (Russell, 2013, S. 87). Reflexion und Aktion können sogar dialektisch erscheinen und als Gegenpole wahrgenommen werden. Das wird allerdings von einigen kritisch gesehen, die, ähnlich wie Dewey, ein Handeln und Aktivwerden in den Reflexionsprozess miteinschließen: „Reflection is not seen as being static; implicit in its meaning is action“ (Knowles, 1993, S. 83). Knowles betont, dass Reflexion verändertes Verhalten und veränderte Praxis zur Folge haben sollte, wenn sie als erfolgreich gelten soll. Das könnte sich im pädagogischen Handlungsfeld beispielsweise in einer Verbesserung der Unterrichtsqualität und einer Entwicklung des Unterrichts manifestieren (Wyss, 2011).

Was damit auch deutlich wird, ist, dass Reflexion nicht nur zurückgerichtet stattfindet, sondern im Wesentlichen auf die Zukunft ausgerichtet ist. Nach Akbari hat sich das Denken über Reflexion bisher primär auf die Klärung der Vergangenheit oder des bisher Gewesenen gerichtet, während zu wenig Fokus auf den Zukunftsaspekt von Reflexion gelegt wurde: „Most of the stress has been on retrospective accounts of reflection, not on the prospective, creative aspects of the concept“ (Akbari, 2007, S. 192). Dies wurde auch von Studierenden so gesehen, die in der Befragung von Berndt und Häcker (2017) als Ziel und Aufgabe von Reflexion unter anderem nannten, „Handlungsoptionen und –alternativen zu identifizieren und damit das eigene zukünftige Handeln zu optimieren“ (S. 246).

Genannt werden zum Teil noch weitere Ziele, deretwegen Reflexion eingesetzt werden kann und sollte. Diese Ziele sind allerdings mehr als positive Wirkungen von Reflexion zu verstehen und werden in Kapitel 2.1.2.8 ausführlicher beleuchtet.

2.1.2.3 Zeitpunkt der Reflexion

Der potenzielle Reflexionsanlass kann nach Dewey dadurch verstreichen, dass die Person, der sich das Problem stellt, unmittelbar dazu übergeht, mit einer Handlung darauf zu reagieren. Diese Tendenz ist Dewey zufolge natürlicherweise vorhanden. Das würde Donald Schön mit dem Begriff „knowing-in-action“ beschreiben: Ein Reflexionsprozess wird gar nicht initiiert, da eine unmittelbare Reaktion in Form einer Handlung möglich ist und dafür aus implizitem oder explizitem Wissen geschöpft wird (Schön, 1983, 50ff).

Für den Beginn eines Reflexionsprozesses muss nach Dewey eine Art „Lücke“ im Handlungsfluss entstehen, ein Unterbrechen des Tuns zugunsten eines Denkprozesses. Wenn es sich dabei nur um eine sehr kurze Lücke handelt und die Situation nicht wesentlich unterbrochen wird, würde hier Schön zufolge ein Reflexionsakt erfolgen, den er als „reflection-in-action“ bezeichnet. Diese Form der Reflexion unterbräche den Handlungsfluss nur sehr kurz und nähme ihn dann sofort wieder auf (vgl. ebd., S. 54ff).

Schön beschreibt bei seinem Konzept „reflection-in-action“ allerdings beispielhaft Berufsfelder wie Psychotherapie oder Stadtplanung, ohne auf das Arbeitsfeld der Lehrkraft zu fokussieren, worauf nur von wenigen Forscher*innen eingegangen wird (Russell, 2018). Obwohl Schön in einer späteren Arbeit zusammen mit verschiedenen Kolleg*innen Reflexion auch im pädagogischen Handlungsfeld betrachtet hat (Schön, 1991), wird zum Teil bezweifelt, ob sich das Konzept der „reflection-in-action“ tatsächlich auf den Lehrberuf übertragen lässt (Richieri, 2017).

Leonhard und Abels (2017) äußern ebenfalls begründete Zweifel an diesem Konzept und beziehen sich dabei argumentativ auf das Modell des „intuitiv-improvisierenden Handelns“ (Volpert, 2003) und die evolutionäre Handlungstheorie von Herzog (Herzog, 2006). Letztere bezeichnet die drei Phasen „Planen“, „Handeln“ und „Reflexion“ als unvereinbar und Reflexion damit nur als möglich, nachdem die Handlung bereits abgeschlossen ist. Reflexion hat dann den Zweck, die Differenz, die in der Praxis zwischen der Planung und der Handlung entstanden ist, zum Thema zu machen und auszugleichen. Reflexion ist damit zwar immer bezogen auf ein Handeln, findet aber „in der Regel handlungsentlastet“ (Herzog, 1995, S. 254) statt. Eine Reflexion während des Handelns wird nicht kategorisch ausgeschlossen, die Fähigkeit dazu müsse sich aber erst durch Erfahrung entwickeln. In Volperts Konzept tritt an die Stelle der „reflection-in-action“ zwar ein hohes Maß an Konzentration und geistiger Aufmerksamkeit, aber eben nicht die Distanzierung vom eigenen Selbst und der Situation, die mit Reflexion einhergehen würde. Häcker verweist darauf, dass der Begriff Reflexion ohnehin bereits uneindeutig sei und durch Gebrauch für verschiedene Formen von Kognition oder auch Intuition noch zusätzlich verwässert werde (Leonhard & Abels, 2017, S. 49). Auch andere Autor*innen äußern sich kritisch: „To pause for contemplation during instruction could disrupt the pace of classroom events and almost surely would result in a loss

of momentum” (Borich 2008, S. 13). In diesem Argument kommt noch der Aspekt hinzu, dass „reflection-in-action“ vielleicht nicht grundsätzlich unmöglich, aber doch eigentlich, bedenkt man die Folge des Verlustes von Unterrichtsfluss und Schwung, zumindest in einigen Situationen nicht wünschenswert sein kann.

Für Dewey wäre es nur dann ein echter Reflexionsprozess, wenn eine Unterbrechung des Handelns stattfindet, ein „stop and think“; [...] reflection involves, at some point, stopping external observations and reactions so that an idea may mature” (Dewey, 1910, S. 209).

Ein solches nachträgliches Reflektieren in Form eines Rückblicks auf eine Situation oder Handlung wird nach Schöns Kategorisierung als „reflection-on-action“ bezeichnet.

Obwohl die Möglichkeit einer „reflection-in-action“ nicht zweifelsfrei ausgeschlossen werden kann, so spricht doch einiges dagegen, das Konzept in der Messung von Reflexion zu berücksichtigen, zumal Leonhard und Abels (2017, S. 53) auch zu dem Schluss kommen, dass sich eine Förderung von Reflexionsfähigkeit nur auf „reflection-on-action“ beziehen könne und der Kern der universitären Bildung in einer kritischen Distanz liege und nicht in der Unmittelbarkeit der Praxis. Situative Formen von Reflexion „entziehen sich systematisch der Instruierbarkeit“ (Leonhard, 2020, S. 23) und auch der Messbarkeit. Auf der Basis dieser Argumentation wird im weiteren Verlauf der Arbeit unter Reflexion nur noch eine „reflection-on-action“ verstanden.

Im Anschluss an Schöns Kategorisierung wurden von anderen Autor*innen noch weitere Begriffe eingeführt (Edwards, 2017). Dem im vorausgegangenen Kapitel angesprochenen Zukunftsaspekt von Reflexion Rechnung tragend, ist ein solcher Begriff beispielsweise „reflection-for-action“ (Moghaddam et al., 2020; Richieri, 2017), der sich ursprünglich bei Killion und Todnem (1991) findet. Bei dieser Art der Reflexion geht es darum, Probleme auf der Basis vorangegangener Erfahrungen vorherzusehen und durch die Reflexion der vergangenen Erlebnisse für die zukünftigen die besten Entscheidungen zu treffen. Genau genommen wird mit diesem und anderen Begriffen aber keine neue Form der Reflexion eingeführt oder bezeichnet, sondern lediglich ein Aspekt hervorgehoben, der auch schon Teil des ursprünglichen Reflexionsbegriffs nach Dewey ist. Auch dieser betont immer wieder die Wichtigkeit sowohl einer Rückschau als auch einer Vorausschau: “Projection and reflection, going directly ahead and turning back in scrutiny, should alternate”

(Dewey, 1933, S. 217). Auch hier zeigt sich wieder die Vielschichtigkeit des Begriffs Reflexion, der scheinbar Gegensätzliches in sich vereint.

2.1.2.4 Ablauf eines Reflexionsprozesses

Zur Frage, wie ein Reflexionsprozess genau abläuft, gibt es verschiedene Ansichten, die sich insgesamt zwar nicht stark unterscheiden, aber doch unterschiedliche Schwerpunkte setzen. Einige Ansätze zu dieser Frage werden im Folgenden betrachtet.

Dewey beschreibt fünf Phasen, die den reflexiven Denkprozess ausmachen (Dewey, 1933, 107ff). Wie im vorausgegangenen Kapitel bereits dargestellt, ist es in einer ersten Phase, von ihm als „suggestions“ bezeichnet, zunächst nötig, ein direktes Handeln zu verschieben und Raum zu schaffen für das Nachdenken. Es entstehen meistens unmittelbar erste spontane Ideen, wie das Problem gelöst werden könnte. Diese automatisch entstehenden Ideen und dieses antizipierende Vorausdenken sind der erste Schritt einer Reflexion. Sie werden an dieser Stelle noch nicht kontrolliert, sondern entstehen einfach. Wird eine solche erste „suggestion“, also Idee, sofort ohne weiteres Nachdenken akzeptiert, dann liegt unkritisches Denken vor.

In einer zweiten Phase, „intellectualization“, geht es darum, eben nicht sofort diesen ersten Ideen zu folgen, sondern das Problem genauer zu betrachten. Es wird von einer emotionalen Ebene (Überraschung, Erschrecken, Irritation, Zweifel, Freude) auf eine intellektuelle Ebene geholt, verortet und genauer analysiert.

In der dritten Phase, „the guiding idea“, wird eine dieser ersten Ideen, die entstanden ist und sich auch nach der zweiten Phase noch als mögliche Lösung anbietet, genauer betrachtet und als Hypothese in den Raum gestellt.

Was auf den ersten Blick plausibel wirkt, stellt sich bei genauerer Betrachtung auch oft als nicht praktikabel oder als Irrweg heraus. Deshalb werden in der vierten Phase, „reasoning“, auf der Basis dieser Hypothese noch genauere Betrachtungen des Problems angestellt und die Hypothese durchdacht. Sollte an dieser Stelle eine Hypothese bereits verworfen werden, beginnt der Prozess erneut.

Phase fünf, „testing the hypothesis by action“, bildet das vorläufige Ende des Reflexionsprozesses, indem die zuvor gebildete Hypothese, sofern sie die Überprüfung in Phase vier erfolgreich bestanden hat, getestet wird. Das kann entweder durch tatsächliches Handeln geschehen oder auch nur in der Vorstellung. Es wird also eine Art Verifizierung oder Falsifizierung unternommen. Im Falle, dass die Realität dem Gedachten entspricht und das Problem gelöst werden kann, endet die Reflexion mit erfolgreichem Handeln. Für den Fall, dass das Problem nicht gelöst werden kann, werden aus der dadurch neu entstehenden Situation wiederum Schlüsse gezogen und der Reflexionsprozess beginnt von vorne.

Die letzte auflösende Phase eines Reflexionsprozesses bezeichnet er als „post-reflective“ (Dewey, 1933, S. 106). Sie geht im Normalfall einher mit einer „experience of mastery, satisfaction, enjoyment“ (ebd., S. 107), also positiven Gefühlen des Erfolgs und der Zufriedenheit.

Einige Eigenschaften dieses Ablaufs von Reflexion sind als bemerkenswert festzuhalten: Erstens handelt es sich bei diesen fünf Phasen um eine komplexitätsreduzierte Darstellung der Wirklichkeit. Nicht jede Reflexion läuft zwangsweise in genau dieser Reihenfolge und genau diesen Phasen ab, die Phasen gehen ineinander über und vermischen sich, wie auch Dewey beschreibt. Eine Phase kann stark ausgeweitet, andere fast übersprungen werden. Er bemerkt außerdem: „The sequence of the five phases is not fixed“ (ebd., S. 115). Es lässt sich dadurch also kein Stufenschema in dem Sinne ableiten, dass man festlegen könnte, wie weit eine Person bereits fortgeschritten ist. Reflexion stellt sich als zirkulärer Prozess dar: Bei jeder verworfenen Lösung beginnt der Prozess von Neuem. Ebenfalls als Zyklus beschreiben Gutzwiller-Helfenfinger, Aeppli und Lötscher (2017) Reflexion in ihrem Modell EDAMA: Erfahrung machen, Erfahrung darstellen, Erfahrung analisieren, Maßnahmen planen, Maßnahmen umsetzen und im Anschluss beginnt der Zyklus von vorne.

Korthagen (1985) beschreibt mit dem ALACT-Modell ebenfalls einen Reflexionsprozess. Lernen aus Erfahrung braucht ihm zufolge eine Abwechslung zwischen Handeln und Reflexion. Der Reflexionsprozess gliedert sich in fünf Phasen: A (action), L (looking back on the action), A (awareness of essential aspects), C (creating alternative methods of action), T (trial). Nach einer Tätigkeit oder Aktion wird auf diese noch einmal zurückgeschaut und zum Inhalt einer Reflexion

gemacht, zentrale Aspekte werden ins Bewusstsein geholt, alternative Handlungsmöglichkeiten abgewogen und schließlich gehandelt. Das Ende des Modells ist gleichzeitig wieder der Anfang eines neuen Zyklus. Auch dieses Modell geht also davon aus, dass Handeln und Reflexion untrennbar verbunden sind.

Die Modelle weisen sowohl Ähnlichkeiten, zum Beispiel die Verknüpfung zum Handlungsaspekt, als auch Unterschiede auf. Problematisch ist, dass der Ablauf von Reflexionsprozessen empirisch (noch?) nicht gezeigt werden konnte, sodass diese Modelle alle normativ sind und damit im Grunde vorerst gleichberechtigt nebeneinanderstehen.

2.1.2.5 Reflexion als sozialer oder solitärer Prozess

Ein immer wieder auftauchendes Problem ist außerdem die Frage, ob es sich bei Reflexion um einen Prozess handelt, den ein Individuum für sich allein durchläuft, oder ob es sich dabei um eine soziale und damit geteilte Tätigkeit handeln kann, die in einer Art Ko-Konstruktion abläuft.

In der Schule gibt es im Normalfall keine formalen Gelegenheiten zur Reflexion, wie sie beispielsweise in Unternehmen über Post-Mortem-Meetings, Action-Reviews oder Retrospektiven institutionalisiert sind (Knipfer et al., 2014). Von Lehrkräften wird verlangt, dass sie Reflexion für sich selbst und in Eigenverantwortung, beispielsweise im Rahmen der Unterrichtsnachbereitung durchführen, wenn sie es überhaupt tun. Reflexion wird eher als isolierend und als Gespräch mit der Situation gesehen und weniger als sozialer Prozess, beispielsweise als Dialog von Kolleg*innen (Zeichner & Liston, 2014). Bei Poom-Valickis und Mathews (2013) findet sich eine kurze Zusammenfassung, warum eine gemeinsame Reflexion bei Lehrkräften so wenig verbreitet ist (vgl. S. 422): Isolation im Klassenzimmer, kaum kollegiale Hospitationen oder Teamarbeit, kaum Mentors- oder Coachingsituationen neben weiteren strukturellen und logistischen Gründen. Es fehlen sowohl eine gewisse Kultur von Feedback und Reflexion als auch die Zeit im Alltag.

Auch unabhängig von Lehrberuf wird Reflexion häufig eher als etwas für andere Personen Unzugängliches, Unbeobachtbares betrachtet, das im Inneren eines Menschen abläuft, als ein „intra-personal process“ (Knowles, 1993, S. 83). Reflexion

wurde auch in der Forschung bisher primär als individueller Prozess, als etwas Privates gesehen, findet aber unzweifelhaft auch in sozialen Kontexten statt, sodass die Berücksichtigung dieser Dimension von Reflexion eine Bereicherung darstellen könnte. Der Versuch einer Konzeptualisierung dieser kollektiven Komponente findet sich beispielsweise bei Collin und Karsenti (2011). In diesem Sinne wird Reflexion eben nicht „als intellektuelles Vermögen, als geistige Leistung von weltlosen Subjekten gesehen, sondern als gelungene Form der Partizipation in einer bestimmten Praxisgemeinschaft, die reflexive Praktiken einführt und vollzieht“ (Rosenberger, 2017, S. 193). Boud (2010) fordert eine weniger individualistische Sicht auf Reflexion, da er das gesamte professionelle Team in der Verantwortung sieht, das Beste für Patienten, Klienten oder Lernende zu erreichen und es dafür nötig ist, Reflexion nicht nur im Sinne einer persönlichen Weiterentwicklung, sondern als Teamaufgabe zu betrachten (vgl. S. 31).

Reflexion kann also sowohl ein individueller als auch ein sozialer Akt sein und beide Seiten hätten sicher das Potential, sich gegenseitig zu befruchten. Für eine sinnvolle Einbindung von Reflexion in die Lehrkräftebildung muss Reflexion vermutlich sogar in beiden Dimensionen betrachtet werden. Aktuell herrscht aber eine Dysbalance vor, die die individualistische Sicht bevorzugt und deshalb scheint es wichtig zu sein, sich zukünftig mehr der sozialen Seite von Reflexion zu öffnen. Schön beschreibt, dass die Isolation der Lehrkräfte, sowohl im Klassenzimmer als auch in der Vor- und Nachbereitung, wie man sie heute größtenteils immer noch vorfindet, für eine gelingende Reflexion hinderlich ist. Denn eine Lehrkraft „needs to communicate her privat puzzles and insights, to test them against the views of her peers“ (1983, S. 333).

2.1.2.6 Voraussetzungen für Reflexion

Der folgende Abschnitt befasst sich mit der Frage, ob jede*r gleichermaßen reflektieren kann oder ob es spezielle Voraussetzungen gibt, die dafür erfüllt sein müssen. Bei Dewey und auch anderen Forscher*innen wird Reflexion übereinstimmend als „*conditio humana*“ (Häcker, 2017, S. 24), also als etwas, das dem Menschen eigen ist, gesehen. Als Eigenschaft der menschlichen Gattung *per se* grenzt sie uns, so

Dewey, vom instinktgeleiteten Tier ab (Dewey, 1933, 17ff). Die Grundvoraussetzungen für Reflexionsfähigkeit bringt also jeder normal entwickelte Mensch natürlicherweise mit. Die Fähigkeit, von sich selbst einen Schritt zurückzutreten und sich aus einer gewissen Distanz zu betrachten, hängt mit der Entwicklung eines Ich-Bewusstseins zusammen. Auch weitere wichtige Komponenten, wie beispielsweise die Fähigkeit, Vergangenheit und Zukunft zum Gegenstand des Denkens zu machen, bilden zusammen mit dem Selbst-Bewusstsein die Basis der menschlichen Reflexionsfähigkeit.

Obwohl die grundsätzliche Fähigkeit also im Menschen angelegt ist und bei normaler Entwicklung im Laufe der Zeit reift, kann daraus aber nicht geschlossen werden, dass jeder Mensch in gleichem Maße reflexionsfähig oder auch -willig ist. Die tatsächliche Reflexionskompetenz ist nach Dewey erstens abhängig von bestimmten (Persönlichkeits-)Voraussetzungen, aber auch bis zu einem gewissen Grad trainierbar und erlernbar. Er beschreibt, dass es Aufgabe einer guten Lehrkraft ist, die „original native tendencies“ (siehe oben), die jeder Schüler und jede Schülerin mitbringt, zu (er-)kennen und diese dann in gewinnbringende Bahnen zu lenken. Das Endprodukt eines solchen erfolgreichen Lernprozesses wäre dann ein „general habit of reflecting“, also eine reflexive Grundhaltung (Dewey, 1933, S. 35). Die so erlernten Methoden des reflexiven Denkens müssen aber gepaart mit bestimmten Voraussetzungen, oder auch „attitudes“, wie Dewey sie bezeichnet, auftreten (vgl. für die folgenden Ausführungen: ebd., S. 28ff). Ohne eine bestimmte Haltung kann Reflexion also nicht erlernt und angewandt werden und umgekehrt reicht die bloße geistige Haltung noch nicht, wenn das Denken nicht systematisch in einer reflexiven Haltung geschult wurde.

Voraussetzungen nach Dewey. Welche Voraussetzungen sind es nach Dewey also, die erfüllt sein sollten, um reflektieren zu können und vor allem auch zu wollen? Zunächst nennt er zu diesem Thema „open-mindedness“, „whole-heartedness“ und „responsibility“ als grundsätzlich nötige Einstellungen. Diese sollen gepaart sein mit einer grundsätzlichen „readiness“ sich auf reflexive Denkprozesse einzulassen. Diese „readiness“ beschreibt das Gegenteil eines Denkens, das von vorschnellen Urteilen auf der Basis von Vorurteilen, Gewohnheit oder Tradition ge-

prägt ist. Es muss der Wille da sein, sich Zeit zu nehmen für eine intensive Auseinandersetzung mit den Dingen. Mit „open-mindedness“ bezeichnet er keine beliebige Offenheit für alles, sondern den aktiven Wunsch, die Dinge multiperspektivisch zu betrachten, verschiedene Standpunkte zu hören und einzunehmen. Es geht dabei auch darum, die Möglichkeit eigener Fehlannahmen und -konzepte in Betracht zu ziehen und dafür offen zu sein, diese zu überwerfen und zu korrigieren. „Whole-heartedness“ erinnert in der Beschreibung Deweys sehr stark an das Konzept der intrinsischen Motivation (Deci & Ryan, 1985). Etwas mit „ganzem Herzen“ zu tun, beschreibt Dewey mit den folgenden Worten: „When anyone is thoroughly interested in some object and cause, he throws himself into it“ (Dewey, 1933, S. 31). Sich einer Sache ganz hinzugeben, ihr volle Aufmerksamkeit zu schenken und sich vielleicht sogar in einem Zustand von Flow mit ihr zu beschäftigen, ist eine Voraussetzung für ausdauernde und intensive Auseinandersetzung: „When a person is absorbed, the subject carries him on“ (ebd.). Insofern erfordert Reflexion auch Zeit, damit eine derart intensive Auseinandersetzung erst möglich werden kann. „Responsibility“ stellt, im Gegensatz beispielsweise zu Offenheit, eher einen einschränkenden Faktor dar, der für eine erfolgreiche Reflexion aber nicht minder bedeutsam ist. Bei der Verantwortungsbereitschaft geht es darum, auch Konsequenzen eigener Ideen, Überzeugungen und eigenen Handelns in Betracht zu ziehen und vollumfänglich dafür einzustehen.

In einem weiteren Abschnitt beschreibt Dewey noch andere spezielle Qualitäten, die herausgebildet werden müssen und die ebenfalls Voraussetzungen für erfolgreiches Reflektieren darstellen. Diese sind „curiosity“, „suggestion“ und „orderliness“. „Curiosity“, also Neugier, beschreibt er als grundsätzlich angeborene Fähigkeit, die für das Lernen und Überleben eines Kindes existenziell ist. Insofern geht es bei dieser Qualität nicht in erster Linie um eine Ausbildung der Neugier, sondern darum, diese natürliche Neugier zu fördern und nicht zu unterdrücken. Neugier kann sehr leicht verloren gehen, wenn nicht auf sie eingegangen wird, aufkommende Fragen nicht beantwortet werden oder den Subjekten Themen aufgezwungen werden, die zum aktuellen Zeitpunkt für diese nicht von Interesse sind, wie er es am Schulsystem seiner Zeit kritisiert. Monotonie und Uniformität wirken ebenfalls einer natürlichen Neugier entgegen. Eine aktive Neugier ist aber eine wichtige Voraussetzung für die Aufnahme reflektierender Gedanken zu einem Thema; eine

Erweiterung von Erfahrungen kann nur durch Neugier erreicht werden. „Suggestion“ meint verschiedene Ideen, die im Kontext des Nachdenkens aufkommen, Antworten des Geistes auf erlebte Situationen. Im Reflexionsprozess sind damit verschiedene Lösungsansätze oder Einfälle gemeint, die in unterschiedlicher Schnelligkeit, Menge und Qualität von selbst auftauchen. Dabei geht es auch um eine Tiefe dieser Ideen, nicht um oberflächliche Auseinandersetzung. Nur wenn einem reflektierenden Subjekt verschiedene Aspekte, Lösungen oder Argumente einfallen, kann es zwischen diesen abwägen. Was Dewey hier beschreibt, ist eigentlich Kreativität, eine Fähigkeit, die Guilford (1967) mit dem Akt des Problemlösens, dem Reflexion ja nach Dewey sehr nahesteht, annähernd gleichsetzt: „The two topics [problem solving and creative production; M.R.] are treated together because they have so much in common that they are basically the same phenomenon“ (S. 312). Auch von anderen Forscher*innen wird die Bedeutung von Kreativität im Kontext von Reflexion hervorgehoben (Akbari, 2007). „Orderliness“ bezeichnet zu guter Letzt, dass diese Einfälle systematisch geordnet, gewichtet und verarbeitet werden müssen. Viele Einfälle zu haben, bedeutet noch nicht, reflexiv zu denken. Die Ideen zu systematisieren und dadurch zu einer Schlussfolgerung oder Lösung zu kommen, ist der letzte wichtige Schritt. Man könnte „orderliness“ also als die Fähigkeit des systematischen Denkens und Abwägens bezeichnen.

Es gibt noch zwei weitere Voraussetzungen, die nicht explizit in diesen Kapiteln thematisiert werden, aber trotzdem an verschiedenen Stellen genannt und als essenziell beschrieben werden. Das sind einerseits Vorerfahrungen und Wissen: „No one can think about anything without experience and information about it“ (ebd., S. 34). Ohne ein gewisses Vorwissen können Situationen nicht sinnvoll eingeordnet werden. Je mehr Wissen vorhanden ist, desto komplexer können sich die Wahrnehmung und das Nachdenken über eine Situation gestalten. Mit mehr Erfahrung gehen außerdem eine detailliertere Beobachtung und Wahrnehmung einher. Es ist nicht unbedingt vorauszusetzen, dass Anfänger ein Problem überhaupt wahrnehmen und erkennen können, abgesehen davon, dass sie es für die Reflexion ja auch noch detailliert beschreiben und analysieren können sollen (Akbari, 2007). Den positiven Einfluss sowohl von bildungswissenschaftlichem und fachdidaktischem Wissen als auch von Fachwissen auf die Reflexions- und Analysefähigkeit von Unterricht konnten bereits einige Studien zeigen (Kempin et al., 2019; König et al., 2014;

Linninger, 2016; Schmelzing et al., 2010; Stender et al., 2021). Zum zweiten ist eine gewisse Ausdauer erforderlich, die auch als Aushalten von Unsicherheit bezeichnet werden kann: „One can think reflectively only when one is willing to endure suspense and to undergo the trouble of searching. [...] To be genuinely thoughtful, we must be willing to sustain and protract that state of doubt which is the stimulus to thorough inquiry“ (ebd., S. 16). Es geht also um die Bereitschaft, sich nicht mit dem ersten Gedanken oder einer schnellen Lösung zufriedenzugeben, sondern längere Zeit einen Zustand des Suchens und der Ungewissheit aushalten zu können. Dem steht die Tendenz entgegen, ein gestörtes Äquilibrium möglichst schnell wieder ins Gleichgewicht zu bringen (Piaget, 1976). Personen mit einer höheren Ungewißheitstoleranz suchen eher neue und ungewisse Situationen auf und ertragen ungewisse Situationen länger. Weniger unsicherheitstolerante Menschen empfinden solche offenen, uneindeutigen Situationen als eher bedrohlich und neigen dazu, sich ihnen schneller zu entziehen (Dalbert, 1999). Dewey beschreibt, dass Unsicherheit – vor allem nach außen sichtbare – von vielen Menschen abgewehrt wird, unter anderem, weil dies als Schwäche ausgelegt werden könnte. Es erfordert also einigen „Mut, Ungewisses zuzulassen“ und einen Reflexionsprozess nicht vorschnell abubrechen oder generell abzuwehren (Kosinár, 2017, S. 157). „Suspense is likely to be somewhat painful“ (Dewey, 1910, S. 13). Es entspricht unserer alltäglichen Gewohnheit, schnell und häufig unbewusst zu reagieren, einen Aufschub sind wir oft nicht gewohnt.

Dass eine größere Überwindung nötig ist, genau hinzuschauen und womöglich auch Unangenehmes zu entdecken, gilt nach Dewey vor allem für Menschen, die sich nicht schon eine grundsätzlich reflexive Haltung angewöhnt haben, sondern noch in der Anfangsphase der Ausbildung einer solchen stehen: „Until the habit of thinking is well formed, facing the situation to discover the facts requires an effort.“ (Dewey, 1933, S. 103). Für wie bedeutend er diese Fähigkeit hält, zeigt sich in dieser Aussage: „To maintain the state of doubt and to carry on systematic and protracted inquiry – these are the essentials of thinking“ (Dewey, 1910, S. 13).

Weitere Voraussetzungen. Unabhängig von Deweys Ausführungen wurden in der Forschung noch weitere Voraussetzungen von Reflexion diskutiert. Hier sei zunächst der Faktor Selbstwirksamkeit erwähnt. Selbstwirksamkeit bezeichnet die

Überzeugung, Anforderungen oder Probleme erfolgreich bewältigen zu können. Sie beeinflusst wesentlich die Anstrengungen, die zur Problemlösung unternommen werden, und ob und wie lange diese auch angesichts von Widerständen fortgeführt werden (vgl. Bandura, 1977, S. 191). Die Relevanz für die Reflexionsfähigkeit zeichnet sich schon dahingehend ab, als es sich beim Reflektieren, wie bereits dargestellt, um einen Akt des Problemlösens in Anbetracht einer Schwierigkeit handelt. Deshalb lässt sich aus der Theorie Banduras ein Einfluss von Selbstwirksamkeitserwartungen auf das Reflektieren ableiten. Zusätzlich wurde die Selbstwirksamkeit als Voraussetzung für Reflexionsfähigkeit aber auch schon in anderen Studien untersucht. Bei Runhaar, Sanders und Yang (2010) zeigte sich Selbstwirksamkeit als Prädiktor von Reflexionsbereitschaft. Von Lohse-Bossenz, Schönknecht und Brandtner (2019) wurde sogar ein *Fragebogen zur Erfassung Reflexionsbezogener Selbstwirksamkeit von Lehrkräften im Vorbereitungsdienst* entwickelt.

Selbstwirksamkeit steht eng in Verbindung mit dem Faktor Erwartung in sogenannten Erwartungs-Wert-Theorien (z.B. Wigfield & Eccles, 2000). Diese sehen die Motivation abhängig von den zwei Faktoren Erwartung und Wert, also der erwarteten Wahrscheinlichkeit, ein Ziel zu erreichen, und dem Wert, der diesem Ziel zugeschrieben wird. Beide Komponenten sollten möglichst hoch sein, um eine hohe Anstrengungsbereitschaft zu evozieren, welche wiederum in besserer Leistung und größerem Erfolg münden kann. Folglich wird auch der Wert, der der Reflexion zugeschrieben wird, im Sinne einer Nützlichkeits einschätzung, eine Rolle spielen. Die Einschätzung der Nützlichkeits einschätzung von Reflexionsbemühungen wurde deshalb bereits in einzelnen Studien als Kontextfaktor erhoben (Göbel & Gösch, 2019). Es ist allein schon aufgrund der Häufigkeit und positiven Konnotation, mit der Reflexion in Lehrerbildungskontexten erwähnt und eingesetzt wird, zu erwarten, dass ihr generell nicht nur von Institutionen und Dozierenden, sondern auch von den Lehramtsstudierenden und Lehrkräften ein hoher Wert beigemessen wird (Lohse-Bossenz et al., 2019, S. 177). Die Studie von Göbel und Gösch (2019) kam sogar zu dem Ergebnis, dass der im Pretest abgefragte erwartete Nutzen einer Reflexionsgelegenheit als höher eingeschätzt wurde, als der im Posttest ermittelte Wert des tatsächlich empfundenen Nutzens (S. 283).

Auch das Professionalisierungsverständnis und das Berufsbild können Einfluss auf die Bereitschaft zur Reflexion haben. In einer qualitativen Studie von Kosinár (2017) wurden mithilfe von Fallanalysen Typen mit bestimmten Professionalisierungsverständnissen entwickelt: das einschränkende Professionalisierungsverständnis, das eher mit Vermeidung und Abwehr auf Unerwartetes reagiert und das entwicklungsbezogene, bei dem Misserfolge auch positiv gewendet werden können (Kosinár, 2017, S. 148). Ob man gewillt ist, sich reflektierend mit den eigenen Erlebnissen auseinanderzusetzen, hängt also auch davon ab, ob man den Lehrberuf so versteht, dass Misserfolge unvermeidlich sind und Chancen zur Entwicklung darstellen oder ob man der Auffassung ist, dass keine Fehler passieren dürfen und man Unsicherheiten folglich keinen Raum lassen will: „Teachers may not get engaged in reflection since it might highlight some features in their selves and profession which are negative“ (Akbari, 2007, S. 202). Der Grund für diese Abwehr von Neuem und Unerwünschtem oder Unerwartetem kann also auch in dem Versuch begründet liegen, den eigenen Selbstwert zu schützen. Landes und Steiner gehen davon aus, dass die meisten Menschen eher zu einer selbstwertdienlichen Interpretation der Ereignisse neigen und so kognitive Dissonanz abwehren. Sie sehen die „Tendenz, bei einmal getroffenen Entscheidungen zu bleiben, da es in vielen Fällen fälschlich als Eingeständnis eines Fehlers gedeutet wird, getroffene Entscheidungen [...] zu überdenken oder zu verändern.“ (2014, S. 5). Wenn also das Bestreben, das eigene Selbstbild zu bewahren und getroffene Entscheidungen zu bestätigen, eine Veränderungsbereitschaft überwiegt, so wird das auch einer ehrlichen Reflexion im Wege stehen, da diese ergebnisoffen wäre und Dissonanz erzeugen könnte. Das kann sogar noch weiter gehen, wenn Reflexion abgelehnt wird, weil der emotionale Druck der Situation und die damit einhergehende Angst zu groß werden und das Selbst zum Schutz vor diesen zu intensiven Emotionen abgeschirmt wird (Ferguson, 2018).

Weitere mögliche Zusammenhänge mit Voraussetzungen wie beispielweise Intelligenz, Kontrollüberzeugungen oder, wie zuletzt angedeutet, dem Umgang mit Emotionen wären denkbar und plausibel, wurden bisher aber noch nicht untersucht. Während bisher in erster Linie personenbezogene Voraussetzungen betrachtet wurden, sind auch situationale Voraussetzungen oder Kontextfaktoren nicht außer Acht zu lassen (vgl. Beauchamp, 2015, 129ff). Der Faktor „Zeit“ wurde bereits erwähnt.

Fehlende Zeit und zu wenig etablierte Strukturen sind besonders auch im Umfeld Schule wesentliche Faktoren, die eine Reflexion im Arbeitsalltag erschweren. Meist wird Reflexion auch in Kontexten von Hierarchien und Machtgefällen gefordert, sowohl in der Universität als auch am späteren Arbeitsplatz Schule, was ebenfalls hinderlich für den Prozess ist. Transaktionale Führung, bei der die persönliche Weiterentwicklung und Mitsprache aller Teammitglieder im Mittelpunkt steht, stellte sich in einer Studie von Runhaar, Sanders und Yang (2010) beispielsweise als Prädiktor für Reflexionsbereitschaft im Team heraus. Wichtig scheint also, falls Reflexion in einem sozialen Setting wie beispielsweise in einem Seminar, einem Supervisions- oder einem Peergespräch stattfindet, ein vertrauensvolles und offenes Klima (Berndt & Häcker, 2017, S. 247). Damit man ehrlich reflektieren kann, ist eine Umgebung nötig, die eine „psychological safety“ (Edmondson, 1999), also eine emotionale Sicherheit bietet. Auch wenn die Reflexion zunächst schriftlich stattfindet und von Korrektor*innen, Forscher*innen oder Dozierenden später eingesehen wird, ist ein Vertrauensverhältnis unabdingbar. Damit in diesen Kontexten ehrlich mit Unsicherheiten, Wissenslücken, Hilflosigkeit und Scheitern umgegangen werden kann, bedarf es insgesamt auch einer anderen Fehlerkultur, als sie bisher in unserem Bildungssystem vorzufinden ist (vgl. Kosinár, 2017, S. 157).

2.1.2.7 Reflexion und Emotion

Wie im vorangegangenen Kapitel bereits gezeigt wurde, besteht ein möglicher Zusammenhang von Reflexion und Emotion darin, dass Ängste oder Scham verhindern, dass man sich auf einen Reflexionsprozess überhaupt einlässt. Auch wurde schon darauf hingewiesen, dass „Diskrepanzerfahrungen“ (Rihm, 2017, S. 229) und Emotionen wie Überraschung oder Irritation als Auslöser von Reflexionsprozessen dienen können. In Kapitel 2.1.2.9, welches unerwünschte Nebenwirkungen von Reflexion beleuchtet, wird noch erläutert, inwiefern Emotionen auch eine Folge von Reflexionsprozessen sein können. Auch können die eigenen Emotionen selbst Thema von Reflexion sein (Beauchamp, 2015, S. 132).

Diesen Erkenntnissen steht die Tatsache gegenüber, dass Reflexion in der bisherigen Forschung häufig primär als eine kognitive Tätigkeit verstanden und untersucht wurde (Artmann et al., 2013, S. 135). Oft wird sie als rein intellektuelle, distanziert-

rationale Tätigkeit gesehen, als etwas, das eine Metaebene voraussetzt: „Reflection often remains solely an intellectual exercise“ (Beauchamp, 2015, S. 127).

Folglich wurden auch in erster Linie Denkopoperationen analysiert, um Schlüsse auf die Reflexionsfähigkeit und -performanz ziehen zu können. Der Einfluss von Emotionen wurde bisher kaum untersucht.

Auch bei Dewey scheint zunächst eine starke Fokussierung auf dem Denkprozess zu liegen, worauf schon der Titel des Hauptwerkes „*How We Think*“ hinweist. Boud und Kolleg*innen formulieren als Antwort darauf, dass es trotz aller Wertschätzung für diese Leistung wichtig sei, Reflexion weiter zu denken und die affektive Komponente stärker einzubinden: „Although we acknowledge a great debt to Dewey, we do not fully follow his path: in particular we give much greater emphasis to the affective aspects of learning, the opportunities they provide for enhancing reflection and the barriers which they pose to it“ (Boud et al., 2013, S. 21).

Auch in Psychologie und Hirnforschung ist unbestritten, dass Kognition und Emotion nicht zu trennen sind (Damasio, 2015) und dass Emotionen und Affekte eine wichtige Rolle beispielsweise für Entscheidungen (George & Dane, 2016) und Erinnerungen spielen (Christianson, 2014), welche wiederum auch im Kontext von Reflexion relevant sind. Für viele Forschende ist deshalb klar, dass Reflexion auch „emotionale und motivationale Dimensionen“ enthalten muss (Gutzwiller-Helfenfinger et al., 2017, S. 134). Shoffner (2011) kritisiert, dass die persönliche Dimension in der Lehrkräftebildung zu kurz komme, Ängste, Sorgen und Wünsche würden zu wenig thematisiert. Über den Weg der Reflexion könnten diese integriert werden. Auch der Möglichkeit, soziale Transformation zu erzeugen, kann Reflexion in der auf Kognition reduzierten Form nicht gerecht werden: “Reflection, in its purely cognitive sense, will not be responsive to the social dilemmas the global community is faced with and can not contribute to the improvement of human society, as envisioned by Dewey” (Akbari, 2007, S. 197).

Einen Überblick über verschiedene Haltungen zur Beziehung von Kognition und Emotion bzw. Affekt gibt Beauchamp (2006). Sie fasst zusammen, dass die von ihr betrachteten Autor*innen alle Schwierigkeiten haben, die Beziehung zwischen kognitiven und affektiven Anteilen klar zu definieren (vgl. S. 124f). Es herrsche zwar bereits unter einigen Forschenden Einigkeit darüber, dass Emotionen einen

Einfluss auf Reflexionsprozesse haben und dahingehend auch bei Forschungsvorhaben zu berücksichtigen seien. Welche Emotionen aber in welchen Anteilen zu welchen Zeitpunkten auf welche Aspekte von Reflexion Einfluss nehmen und wie diese mess- oder sichtbar gemacht werden können, ist weitgehend unklar. Erwähnenswert scheint außerdem, dass die von Beauchamp untersuchten Texte zum Großteil in den 90er Jahren veröffentlicht wurden. Aktuell gibt es durchaus einzelne Bestrebungen, dieses Defizit zu beseitigen und Emotionen auch in Kompetenzmodellen von Reflexion zu positionieren (Hilzensauer, 2017). Große Fortschritte, vor allem darin, diese Rolle von Emotion auch empirisch nachzuweisen, gibt es allerdings bisher kaum.

2.1.2.8 Wirkungen von Reflexion

Reflexion ist, wie bereits erläutert, kein Selbstzweck; sie muss sich an ihrem Erfolg und ihren positiven Wirkungen messen lassen, welche im Folgenden vorgestellt werden. Postulierte Wirkungen gibt es viele, es scheint fast, als wäre Reflexion eine Art Allheilmittel in der Lehrkräftebildung: „Reflection in education is a field full of promises: promises for improving professional proficiency, for fostering personal growth, and for increasing social justice. This promising character makes reflection very attractive“ (Procee, 2006, S. 252). Kann Reflexion diesen Erwartungen gerecht werden?

Die Bedeutung von Reflexion wird gerade im Kontext des Lehrberufes so stark hervorgehoben, weil es sich um einen Beruf handelt, der mit Besonderheiten verbunden ist, die eine gezielte Vorbereitung für diesen schwierig gestalten.

Um diese Besonderheit des Lehrberufs zu spezifizieren, bedient sich Leonhard (2019) des Vergleichs zum Arztberuf. Während es bei diesem möglich ist, eine Diagnose über den gegenwärtigen Zustand des Patienten zu stellen („Was ist los?“, Leonhard, 2019, S. 1), kann es aufgrund der Komplexität, Intransparenz und Flüchtigkeit des Unterrichtsgeschehens erst im Nachhinein gelingen, sich die Frage zu stellen: „Was war eigentlich los?“. Diese Rückschau könnte im Rahmen einer Reflexion geschehen. Auch nach Helsper ist das Arbeitsfeld der Lehrkraft von Dilemmata, Paradoxien und Antinomien geprägt. Die Folgen von Handlungen sind nicht absehbar und es besteht immer eine Möglichkeit des Scheiterns. Er spricht von

„konstitutive[n], nicht aufhebbare[n], sondern lediglich reflexiv zu handhabende[n] Antinomien des Lehrerhandelns“ (S. 75). Er bezieht sich dabei auch auf Luhmann und Schorr und leitet daraus ein Technologie-, Struktur- und Verstehensdefizit im pädagogischen Handlungsfeld ab (Helsper, 2002, S. 71). Besonders die Begründungsantinomie ist in diesem Kontext hervorzuheben: Handeln ist geprägt von einem hohen Entscheidungsdruck bei einer gleichzeitig gesteigerten Begründungspflicht, das heißt, dass pädagogisches Handeln immer legitimiert sein muss, indem es auf abgesicherten, also fachwissenschaftlichen Begründungen basiert. Andererseits ist man natürlich auf routinisiertes Handeln und schnelle Entscheidungen angewiesen. An dieser Stelle kann die Reflexion ansetzen: Diese Entscheidungen kann man zumindest im Nachhinein fragwürdig machen und einer Reflexion unterziehen, um wiederum damit auf zukünftiges Handeln positiv einwirken zu können (vgl. ebd., S. 77). Durch Reflexion könnten auch die Entscheidungsfreiräume, die sich für Lehrkräfte ergeben, begründet und unter Rückbezug auf die fachwissenschaftliche Theorie gefüllt werden. So kann es durch Reflexion zu einer Integration von Wissen und einer komplexeren Sicht auf verschiedene berufliche Aspekte kommen (Davis, 2006). Gerade dieser nachträgliche Einbezug der Theorie über die Reflexion könnte einen sinnvollen Umgang mit dem erwähnten Technologiedefizit (Herzog, 1995) darstellen, da die Wissenschaft in diesem Fachbereich keine mit Sicherheit „richtigen“ Vorgehensweisen für die Praxis anbieten kann. So kann es gelingen, dass nur routinemäßiges Handeln ohne theoretische Fundierung durch Reflexion wieder auf wissenschaftliche Füße gestellt werden kann, denn sonst besteht die Gefahr, dass „bloße Praxis eigentlich nur Schlendrian [sei]“ (Herbart, 1997), wenn sie nicht hinterfragt, geprüft und theoriegestützt werde.

Diese Forderung richtet sich also primär an bereits länger tätige Lehrkräfte, aber auch für Studierende und Noviz*innen hat die Verknüpfung von Theorie und Praxis eine besondere Bedeutung, da es ihnen meist noch sehr wenig gelingt, diese Verknüpfung herzustellen (Penso et al., 2001). Pädagogisches Wissen kann nicht einfach in pädagogisches Handeln überführt werden, das ist in der Professionsforschung Konsens (vgl. Herzmann & Proske, 2014, S. 33). Patry definiert Praxis als zielgerichtetes Handeln, das sich immer nur auf einen Einzelfall beziehen kann (Patry, 2014, S. 33). Daraus ergeben sich ihm zufolge spezifische Probleme, die auftreten, wenn man versuchen will, Theorie in die Praxis zu übersetzen: die Allgemeinheit-Konkretheit-Antinomie, die Theorienvielfalt, die Zielvielfalt und die

Situationsspezifität. Das heißt, Theorien müssen allgemein gehalten werden, was sie schwer auf einen konkreten Einzelfall anwendbar macht. In der Wissenschaft stehen sich häufig auch Theorien gegenüber, die sich zum Teil widersprechen, sie verfolgen unterschiedliche Ziele. Die Frage der Verknüpfung von theoretischem Wissen und praktischem Tun nimmt folglich in der Lehrerbildung eine zentrale Stellung ein (Artmann, 2019). Auch Schön beschreibt die Kluft „between the universities and the professions, research and practice, thought and action“ (Schön, 1983, S. viii). Er geht mit der Einschätzung konform, dass es sich bei Theorie und Praxis um unterschiedliche Dimensionen mit eigenen Regeln handelt und dass theoretisches Wissen nicht einfach in einer positivistischen Denkweise auf Praxisprobleme übertragbar ist. Trotzdem schlägt er in seinen Texten einen Ansatz vor, die Kluft zwischen den genannten Dimensionen zu schließen, und zwar Reflexion. Reflexion wird dabei als Kitt gehandelt, der den Bruch zwischen Theorie und Praxis reparieren kann (Herzog, 1995, S. 256). Mit ihrer Hilfe ließe sich eine Verbindung von der Praxis zur Theorie und umgekehrt herstellen. Sie wäre demnach „an iterative, forward-moving spiral from practice to theory and theory to practice“ (Rodgers, 2002, S. 863).

Ob Reflexion tatsächlich erfolgreich als Brücke zwischen Theorie und Praxis funktioniert und ob sich dadurch auch bessere Ergebnisse einstellen, ist empirisch noch nicht nachgewiesen. Leonhard (2020) kommentiert sogar einige Studien, die zeigten, dass Unterrichtsnachbesprechungen meist keine theoretischen Bezüge aufwiesen, mit den Worten, dass „Reflexion in der beruflichen Praxis [...] die theoretische Rahmung schlicht nicht zu brauchen [scheint]“ (S. 21). Eine Studie von Kasztner (2009) befasste sich mit der Fragestellung, ob theoriebezogene Praxisreflexion die Anhäufung trügerischer Wissensreduzieren oder vermeiden kann, konnte aber eine solche Wirkung von Reflexion nicht nachweisen.

Doch nicht nur durch die Verknüpfung von Theorie und Praxis bietet sich die Möglichkeit der Weiterentwicklung, sondern auch durch das Lernen aus eigener Erfahrung. Mit dem Technologiedefizit und der Unbestimmtheit des Arbeitsfeldes sowie dessen ständiger Veränderung und Weiterentwicklung geht die Notwendigkeit einher, auch aus der eigenen Erfahrung zu lernen und sich anzupassen: „Wenn es gelingt, durch eine gedankliche Analyse der bisherigen Arbeitserfahrungen zukünftige Herausforderungen besser zu meistern, sprechen wir von Lernen durch Reflexion“ (Knipfer et al., 2014, S. 19).

Lebenslanges Lernen ist gegenwärtig in „Zeiten vielfältiger, zum Teil stark beschleunigter gesellschaftlicher Entwicklungsprozesse (z.B. Digitalisierung, Erosion von Traditionen und Werten, Wissensexplosion) und einem damit einhergehenden Verlust an Sicherheit und Orientierung“ (Deing, 2019, S. 320) wichtiger denn je. Das Gelingen lebenslangen Lernens ist stark abhängig von der Selbstregulationskompetenz des Individuums. Reflexion wiederum ist eine zentrale Komponente erfolgreicher Selbstregulation (Arrastia-Chisholm et al., 2017; Jenert, 2008; Linninger, 2016) und dient als metakognitive Strategie (Cavilla, 2017). Durch Reflexion kann es gelingen, Lernfelder zu identifizieren, sinnvolle Ziele und Vorgehensweisen auszuwählen und den eigenen Fortschritt und Erfolg zu kontrollieren und zu bewerten. Gerade bei Untersuchungen zu Lern-, Arbeits- und Organisationsstrategien von Lehramtsstudierenden zeigten sich größtenteils enttäuschende Ergebnisse, sodass hier Potential und Notwendigkeit für Förderung bestünde (Rauin & Meier, 2007). Reflexion kann Bewusstsein über die eigenen Ressourcen fördern und helfen, Emotionen und Gedanken auszudrücken, womit sie Resilienz und Wohlbefinden fördern kann (McKay & Barton, 2018). Auch einzelne andere Wissenschaftler*innen erwähnen einen Zusammenhang von Reflexion und einer höheren Berufszufriedenheit (Korthagen & Wubbels, 2000) oder eine Minderung von Stress und Angst bei steigender Kompetenz- und Selbstwahrnehmung (Contreras et al., 2020).

Die Notwendigkeit der lebenslangen Weiterentwicklung gilt einerseits für gesamtgesellschaftliche Veränderungen und andererseits im Speziellen für das Arbeitsfeld der Lehrkraft. Das betrifft sowohl die bereits angesprochenen Digitalisierungsprozesse als auch die zunehmende Diversität der Lernenden:

The reality of teacher's work in the 21st century is changing rapidly. Bringing technology into schools, the increasing complexity of learning and teaching in diverse classrooms, the growing societal expectations of raising students' achievements and the need for tailoring and implementing innovative teaching practices are just some of the factors driving this change. (Kowalczyk-Walędziak et al., 2019, S. 15)

Besonders die immer weiter zunehmende Diversität von Lernenden erfordert ein kritisches Bewusstsein von Lehrkräften, die diese mithilfe von Reflexion, beispielsweise unter Einbezug historischer Standpunkte, besser bewältigen und adäquater auf sie reagieren könnten (Carter Andrews et al., 2019; K. Liu & Ball,

2019; Richieri, 2017). Die nötige berufliche Weiterentwicklung kann also durch Reflexion ermöglicht werden, wenngleich diese nicht notwendigerweise auch zu einem besseren Unterricht führen muss (Roters, 2011).

Eine distanzierte Betrachtung im Nachgang kann trotzdem dabei helfen, das Geschehen zu verstehen und zu interpretieren. Reflexion kann dazu dienen, Muster im scheinbaren Chaos des Alltags zu erkennen und das eigene Handeln flexibel an die Gegebenheiten anzupassen. Handlung und Reflexion würden sich in diesem Sinne abwechseln und so könnte man sich immer wieder neu auf das einstellen, was man gerade vorfindet und die relevanten Probleme identifizieren. Damit bekommt Reflexion sowohl eine Orientierungs- als auch eine Legitimationsfunktion: Das Handeln in ungewissen und schwer fassbaren Situationen wird dadurch nicht einfach beliebig oder gesteuert von unbewussten Einflüssen wie Persönlichkeit oder Gefühlslage, sondern wird immer wieder hinterfragt und neu ausgerichtet. Eine solche reflexive Haltung wird oft als Kern der Professionalität bezeichnet, die in allen Phasen der Lehrkräftebildung eingeübt und habitualisiert werden sollte (Boer & Reh, 2012, S. 285). Eine Studie von Hernandez zeigte beispielsweise, dass sich eine Reflexionsintervention positiv auf die Entwicklung einer professionellen Identität auswirkte (Hernandez, 2020).

Ein Teil der genannten Professionalität besteht auch darin, die Einflüsse der eigenen Prägungen und des eigenen Hintergrundes (z.B. soziokultureller, sprachlicher, ethisch-moralischer Natur) auf das berufliche Handeln zu kennen (van Beveren et al., 2018, S. 4). Solche unbewussten Anteile sichtbar und damit auch veränderbar zu machen, bezeichnet auch Dewey als wichtige Leistung von Reflexion (vgl. Dewey, 1910, S. 215). "Personal theories need to be revealed [...] so that they can be scrutinized, challenged, compared to public theories, and then conformed or reconstructed" (Griffiths & Tann, 1992, S. 71). Dies scheint besonders wichtig vor dem Hintergrund, dass immer wieder der Verdacht geäußert wird, dass Berufsbiographie und Professionalisierung zu sehr geprägt sind von eigenen Schuleindrücken und Lernerfahrungen (Dick, 1999; Rauin & Meier, 2007). Es besteht die Gefahr, dass diejenigen, die nicht ihre eigene Person und die professionellen Handlungen hinterfragen „teach as they were taught“ (Braun & Crumpler, 2004, S. 61) und so womöglich ineffiziente Strategien und unbewusste Muster aufrecht erhalten. Reflexion hingegen kann hier neue Perspektiven und Blickwinkel eröffnen und damit sogar eine Veränderung bestehender Glaubenssysteme ermöglichen (van Beveren

et al., 2018, S. 5). Das ist vor allem auch deshalb wichtig, weil Forschungsarbeiten zeigen konnten, dass die Auseinandersetzung mit dem eigenen Selbstkonzept bei Lehrpersonen sehr gering ausgeprägt ist (Gehrmann, 2007, S. 90).

Dass sich (zukünftige) Lehrkräfte über ihr eigenes Denken, ihre Vorannahmen und Weltbilder Klarheit verschaffen, hat auch deshalb eine große Bedeutung, weil ihre Berufstätigkeit einen politischen Akt darstellt (Harry & Salvador, 2021). „Teaching is not a neutral activity, and it is inextricably related to broader social issues, issues such as justice, race, gender, poverty, discrimination...etc.“ (Akbari, 2007, S. 197). Mithilfe von Reflexion können und müssen sich (zukünftige) Lehrkräfte also auch mit der sozialen, moralischen und politischen Dimension von Unterricht auseinandersetzen (Jay & Johnson, 2002, S. 75). Gorski und Dalton (2020) betonen, dass Lehrkräfte im Besonderen dann auf eine kritische Reflexion angewiesen sind und diese auch beständig selbst durchführen müssen, wenn sie Aspekte wie Gleichheit, soziale Gerechtigkeit und einen positiven Umgang mit Multikulturalität fördern wollen. Auch bereits kleinere Unterrichtsentscheidungen wie zum Beispiel die Bildung von Gruppen können in dieser Hinsicht auch weitreichende und oft zu wenig bedachte Folgen haben (Liston & Zeichner, 1990, S. 243). Professionelles pädagogisches Handeln hat immer, selbst wenn es nicht so intendiert ist, auch „macht-, kontroll- und überwachungsformige Züge“ und diese lassen es „umso erforderlicher erscheinen, dass Professionelle gegenüber ihrem eigenen Handeln eine habitualisierte Skepsis, ein habitualisiertes Misstrauen ausbilden, das zugleich als reflexive Irritation der eigenen Praxis gegenüber institutionalisiert werden muss“ (Helsper, 2002, S. 74).

Liston und Zeichner verknüpfen den Begriff der Reflexion mit dem der Verantwortung und sehen die Chance, dass durch sie das Geschehen im Klassenzimmer mit dem größeren sozialen und politischen Kontext in Verbindung gebracht werden kann (Liston & Zeichner, 1987, S. 3).

Dieser kritisch-reflexive Fokus sollte sich über das Klassenzimmer hinaus ausweiten und Schulleben, Schulverwaltung und Schulsystem in den Blick nehmen, damit Lehrkräfte nicht zu Spielbällen von politischen oder gesellschaftlichen Entscheidungen und Bedingungen, sondern zu aktiven Gestaltern ihres Arbeitsfeldes werden:

Teachers who are unreflective about their work uncritically accept [...] everyday reality in schools and concentrate on finding the most effective and efficient

means to achieve ends and to solve problems that have been defined for them by others. These teachers lose sight of the fact that their everyday reality is only one of many possible alternatives. (Grant & Zeichner, 1984, zitiert nach Noffke & Brennan, 1988, S. 4)

Dass viele Studierende für solche transformatorischen Gedanken prinzipiell offen sind oder dazu angeregt werden können, zeigt eine Studie von Carrington und Selva (2010) in der sich veranschaulichen ließ, „how the students critique their world and imagine how they could contribute to a better world in their work as a beginning teacher“ (S. 1). Auch Schön denkt sein Konzept des „reflective practitioner“ bis zur Ebene des Systems und der Gesellschaft. Er beschreibt das Schulsystem als ein „game of control and evasion“, in dem Leistung oft nur vorgetäuscht wird oder nur ein Minimum abgeliefert wird, um Strafe zu vermeiden und in dem Anpassung belohnt wird. Reflexion sieht er als Möglichkeit, diese Logik des Systems sichtbar zu machen und aufzubrechen, was natürlich auf Widerstand des Systems stoßen kann: „What happens in such an educational bureaucracy when a teacher begins to think and act not as technical expert but as reflective practitioner? Her reflection-in-action poses a potential threat to the dynamically conservative system in which she lives“ (Schön, 1983, S. 332). Dass Lehrkräfte reflektieren und sich als Folge davon begründet entscheiden, ihre Profession anders zu erfüllen, als es Lehrpläne, Verwaltung oder Lehrerbildung vielleicht vorgeben, sieht er als eine der großen Errungenschaften dieses Konzepts. Studierende und Lehrkräfte können und sollen also eine Metaebene einnehmen und sich so „in der Reflexion auf ihr Fach, also indem sie bedenken, was sie vollbringen, und auch in der Reflexion auf sich selbst, [...] über den Umkreis des tatsächlich angeeigneten erheben“ (Adorno, 2017, S. 31).

Indem Lehrkräfte in Unterricht und Schulleben Reflexion bewusst offen praktizieren und Einblick in diese geben, können sie eine Vorbildfunktion für Lernende einnehmen und diese beim kritischen Denken unterstützen. Denn neben den häufig betonten intellektuellen und leistungsbezogenen Aspekten von Schule darf die Erziehung zu Mündigkeit, Selbstbestimmung und kritischer Partizipation nicht vergessen werden: „We state emphatically that, upon its intellectual side education consists in the formation of wide-awake, careful, thorough habits of thinking“ (Dewey, 1933, S. 78). Die Fähigkeit des kritischen Denkens als ein Aspekt von Re-

flexion hat in letzter Konsequenz emanzipatorischen Charakter und führt zum Erlangen einer gewissen Freiheit von äußeren Vorgaben und Einflüssen oder, um aus aktueller Perspektive zu sprechen, von Fake News und Verschwörungstheorien. Um sicherzustellen, dass „Ausschwitz nicht noch einmal sei“, konstatiert Adorno sogar, dass „Erziehung [...] überhaupt nur [sinnvoll wäre] als eine zu kritischer Selbstreflexion“ (Adorno, 2017, S. 90).

Zusammenfassend ist festzuhalten, dass es sehr viele denkbare positive Wirkungen von Reflexion gibt, die sich allerdings empirisch bisher kaum oder gar nicht nachweisen ließen (Akbari, 2007; Roessger, 2014; Russell, 2013) oder auch von ihrer Natur her schwer nachweisbar sind. Es lässt sich nicht verlässlich aus der bisherigen Forschung ableiten, „whether reflective professionals are necessarily more effective than non-reflective professionals“ (Sumsion & Fleet, 1996, S. 121).

Natürlich besteht die Gefahr, Reflexion in der Hoffnung auf positive Wirkungen einzufordern und zu finanzieren, obwohl es am Ende gar keine positiven Effekte gibt: „Thanks to the addition of *reflection* to the teacher education vocabulary, classes and assignments have changed in wording. Yet there is little evidence to suggest profound changes in results“ (Russell, 2013, S. 86).

2.1.2.9 Nebenwirkungen von Reflexion

„There is something called reflection and it has value, but it also has limits, which need to be understood.“ (Ferguson, 2018, S. 416)

Im vorherigen Abschnitt wurde ausführlich herausgearbeitet, welche positiven Wirkungen der Reflexion zugeschrieben werden. Während sich ein Großteil der Lehrerbildner*innen einig ist, dass Reflexion uneingeschränkt erstrebenswert ist und sich zum Teil emphatisch für diese aussprechen, gibt es auch einige kritische Stimmen, die die Grenzen der Reflexion betonen, vor deren Übereinsatz warnen oder sogar negative Effekte befürchten: “Has reflective practice done more harm than good in teacher education?” (Russell, 2013). Dieser Frage soll in diesem Abschnitt nachgegangen werden. Dabei werden zunächst Routine und Reflex als Formen unreflektierten, aber für Lehrkräfte auch durchaus wertvollen Handelns betrachtet. Dann

wird untersucht, welche Folgen ein übermäßiger Einsatz von Reflexion nach sich ziehen kann und schließlich auf einige unerwünschte, aber mögliche Auswirkungen eingegangen: Verunsicherung, Stress, Handlungsunfähigkeit und Grübeln. Abschließend wird Reflexion als potenzielle Form internalisierter Kontrolle kritisch betrachtet.

Reflexion wurde im vorangegangenen Kapitel als Mittel herausgestellt, mit den spezifischen Anforderungen des Lehrberufs, der Ungewissheit, dem Technologiedefizit und der Komplexität zurechtzukommen. Nun gibt es aber auch noch andere Möglichkeiten, erfolgreich mit diesen umzugehen und dabei werden häufig auch Gegenpole von Reflexion genannt, nämlich Routine und Reflex. Helsper weist in diesem Kontext auf den strukturtheoretisch doppelten Habitus der Lehrkraft hin: einerseits wissenschaftlich-reflexiv und andererseits routinisiert-praktisch (vgl. Helsper, 2002, S. 13). Idel und Schütz betonen, dass es speziell im Modus des Handelns unter Druck nötig und selbstverständlich ist, sich gewisser Routinen und Schemata zu bedienen (2017, S. 201). In ihrem Beruf sind Lehrkräfte in ihrer Aufmerksamkeit so gefordert und haben häufig so wenig Zeit, dass einer gewissen Automatisierung des Handelns und Routinen eine wesentliche Rolle zukommt (vgl. Prieto et al., 2020, S. 15). Gleichzeitig werden genau diese oft kritisch gesehen und das, obwohl auch unreflektiertes Handeln von Intelligenz und Rationalität geprägt sein kann (Weichold, 2014). Sich vom Nachdenken zu befreien und der spontanen Eingebung und der Intuition zu vertrauen, sei zu Unrecht abgewertet. Eine zunehmende Meisterschaft bringe natürlicherweise Routinen mit sich. Ein echter Experte ist nach Dreyfus und Dreyfus (1986) jemand, der holistisch wahrnimmt und Übung und Erfahrung in ein implizites Wissen überführt hat, der sich also intuitiv angemessen verhalten kann. Auch Neuweg weist darauf hin, dass bei der Meisterschaft planvolles Handeln und Routine verschmelzen und nicht als eindeutige Kategorien zu trennen sind (Neuweg, 2017, S. 90). Nicht zuletzt sieht auch Dewey die Vorteile eines unreflektierten Handelns in bestimmten Situationen: „Unconsciousness gives spontaneity and freshness“ (Dewey, 1910, S. 217). Dieses Argument spricht nicht per se gegen Reflexion, sondern nur dafür, diese nicht einseitig als die beste Option für jede Problemlage zu betrachten und uneingeschränkt auf ein Podest zu stellen. Durch diese einseitige Darstellung können Routine und Auto-

matisierung eine Abwertung erfahren. Situations- und problemspezifisch muss abgewogen werden, ob es eher Reflexion ist, die zielführend ist, oder ob einem routinemäßigen Vorgehen der Vorzug zu gewähren ist, bzw. inwiefern diese beiden sich sinnvoll ergänzen können.

Dass Reflexion in der Lehrkräftebildung allgegenwärtig ist und eben nicht differenziert, sondern in verschiedenen Kontexten fachübergreifend und phasenübergreifend eingefordert wird, kann dazu führen, dass der Sinn von Reflexion durch den übermäßigen Einsatz letztendlich verloren geht. Reflexion wird dadurch schließlich selbst zur gedankenlosen, oberflächlichen Routine, wie Beauchamp die Kritik verschiedener Autor*innen zusammenfasst: „The result may be reflective output that is perfunctory, conducted at a surface level” (Beauchamp, 2015, S. 127). Ähnliches stellt Rosenberger in ihrer Studie fest, nämlich dass 37 % der Studierenden Reflexionsaufgaben keine relevante Bedeutung zuwiesen und sie im Modus reiner Pflichterfüllung abarbeiteten (vgl. Rosenberger, 2017, S. 197). In einer Befragung von Dozierenden in der Lehrkräftebildung beschreibt ein Studienteilnehmer, dass die Überdosierung dazu führen könne, dass die Studierenden „sich nicht mehr darauf einlassen können“ (Wyss & Mahler, 2021, S. 21). Das verweist darauf, dass erzwungene, verordnete Reflexion keinen Mehrwert hat (Idel & Schütz, 2017, S. 208). Daraus folgt, dass es von Bedeutung ist, den Sinn und Mehrwert von Reflexion besser zu kommunizieren, aber auch, Reflexion nur gezielt und punktuell und nicht im Überfluss einzusetzen. Ansonsten kann es zur „Ermüdung vieler Studierender und Referendarinnen und Referendare gegenüber der dauerhaften Anforderung sich zu reflektieren [kommen]“ (Kosinár, 2017, S. 147). Russel verweist in diesem Kontext nicht nur auf die übermäßige Menge von Reflexionsaufgaben, sondern auch auf fragwürdige Vorgaben („write a one-page reflective paper“ Russell, 2013, S. 81), die den Modus des Abarbeitens und Erfüllens formaler Vorgaben geradezu provozieren. Auch er bestätigt noch einmal die Notwendigkeit, den Sinn von Reflexion zu kommunizieren und mit dem Reflektieren um des Reflektierens willen aufzuhören:

I hear many complaints from students that they are weary of so much reflection, particularly when so many different assignments call for reflection and when they have been given little guidance in terms of what reflection involves, what the results of reflection should look like, and how reflection can help them learn to teach. (Ebd.)

Wenn kein Sinn oder Nutzen in der Erfüllung der Reflexionsaufgaben gesehen wird, fehlt es folglich auch an der Motivation, diese zu bearbeiten. Das mangelnde Interesse an Reflexion, das Studierenden häufig unterstellt wird (Korthagen, 2001, S. 67), könnte letztlich auf den übermäßigen und nicht immer nutzbringenden Einsatz zurückzuführen sein.

Diese Argumente sprechen für einen abwägenden und durchdachten Einsatz von Reflexion. Es gibt aber auch Stimmen, die der Reflexion an sich negative Folgen zuweisen. Während im vorangegangenen Kapitel zu den positiven Wirkungen von Reflexion gezeigt wurde, dass sie die Berufszufriedenheit steigern und Angst vermindern kann, gibt es gleichzeitig Forschungsergebnisse, die nahe legen, dass Reflexion auch Unsicherheit und Stress auslösen kann:

These results highlight two seemingly contradictory findings: that a higher level of reflective capacity is on one hand positively associated with positive outcomes such as confidence and job satisfaction; while on the other hand is also positively associated with negative outcomes such as uncertainty and stress. (Priddis & Rogers, 2018, S. 96).

Reflexion kann zunächst unbequem und verunsichernd sein und Ausdauer und Resilienz erfordern, wie schon im Kapitel 2.1.2.7 erläutert wurde. Damit eine Person ein Problem trotzdem mithilfe von Reflexion angehen will, muss dieses als positive Herausforderung gesehen und nicht als Überforderung wahrgenommen werden (vgl. Keller-Schneider, 2021, S. 71). Dass sie als Überforderung wahrgenommen werden, könnte aber vor allem zu Beginn der ersten Phase der Lehrerbildung der Fall sein, wenn Studierende noch nicht einmal eine basale Handlungssicherheit gewonnen haben. Zu früh in Reflexion einzusteigen, könnte kontraproduktiv sein, da es zu diesem Zeitpunkt noch darum geht, Überlebensstrategien zu erwerben (Fuller, 1969). Eine innere Spaltung, wie sie durch Reflexion erzeugt werden kann, könnte zu einer dauernden Unzufriedenheit mit dem eigenen Selbst führen und lähmend wirken (Frankfurt, 2016). Auf die Wahrnehmung von Verunsicherung, Wandel und Auflösung dann mit einer „deflexiven Abwehrreaktion“ zu antworten (Jain, 2013, S. 186), sie also nicht zuzulassen oder abubrechen, wäre durchaus eine natürliche menschliche Reaktion.

Dass man durch Reflexion wahrnimmt, dass das Gewohnte und Bekannte nicht mehr funktioniert, aber auch etwas Neues noch nicht etabliert ist, kann dazu führen, dass man sich noch mehr auf das Alte zurückorientiert, weil man dieses unsichere

Zwischenstadium vermeiden will (Larrivee, 2000, S. 304). Dadurch könnten bestehende Meinungen und Ansichten bestärkt und verfestigt, statt aufgebrochen werden. So können Reflexionen auch das werden, was sie eigentlich genau nicht sein sollen, nämlich „exercises in justifying or rationalizing preconceived ideas“ (Cushion, 2018, S. 6; Fendler, 2003). Anstatt sich ehrlich zu hinterfragen und die Komfortzone zu verlassen, was zu einer echten Weiterentwicklung und Veränderung führen könnte, wird auf gewohnte Einstellungen, Haltungen und Meinungen zurückgegriffen und diese zur Rechtfertigung, Begründung oder Erläuterung von Situationen oder Handlungen herangezogen. Dafür bieten natürlich gerade „erzwungene“ Reflexionen einen Nährboden. Zwar wird eine echte Reflexion abgewehrt, es muss aber zumindest formal reflektiert werden, wodurch es zu einer Verstärkung von Vorurteilen, Routinen und Mustern kommen kann, die dann für Begründungen herangezogen werden.

Ein weiterer möglicherweise problematischer Punkt ist die Nähe von Reflexion und Grübeln. Aus psychologischen Studien ist bekannt, dass zwischen Reflexion und Grübeln, also einem unsystematischen, negativen Kreisen von Gedanken, ein enger Zusammenhang besteht (Junkins & Haeffel, 2017). Die Grenzen zwischen Reflexion und Grübeln sind fließend. Für Grübeln wiederum gibt es einen starken positiven Zusammenhang mit dem Auftreten von Depressionen (ebd.).

Eine zusätzliche Gefahr, auf die Neuweg (2017) hinweist, ist die Verdichtung eines ständigen Reflektierens zum Persönlichkeitszug. Er beschreibt damit die Möglichkeit, dass sich das übermäßige Nachdenken über Vergangenes, erlebte Situationen und das eigene Selbst gewissermaßen verselbstständigt und zu einem dauerhaften und relativ stabilen Zustand wird. Er bezeichnet die Reflexion als „janusköpfig“ insofern sie zwar das Lernen voranbringe, aber gleichzeitig auch das stoppe, was sich im Fluss befinde. Das könne letztlich zu einer dauerhaften Erstarrung führen und damit dazu, dass Planungshandlungen Ausführungshandlungen ersetzen, oder, wie er beispielhaft Reinhard Sprenger (2004) zitiert: „Sie wissen genau, jetzt müssen Sie handeln. Aber bevor Sie handeln, lesen Sie lieber noch mal ein Buch ... oder besuchen ein Seminar ... oder gehen zum Therapeuten. [...] Die Energie fließt ins Analysieren und Reflektieren“ (zitiert nach Neuweg, 2017, S. 96). Aus dieser Erstarrung folgt ein Verlust von Spontaneität und Frische. Denn es ist nicht möglich, sich gleichzeitig aus der Situation herauszunehmen und ganz im Moment zu

sein: Ein Flow-Erleben wird durch die Reflexion unmöglich gemacht und die Verbindung zur Situation bewusst für eine gewisse Zeit gekappt (ebd.). Deshalb kritisieren Zahn zufolge beispielsweise Rousseau, Fénelon oder Kleist Reflexion als unnatürlichen Zustand, bei dem Unmittelbarkeit und Spontaneität verloren gehen und Hemmung erzeugt werde (vgl. Zahn, 2007, S. 397). Das könnte zum Beispiel auch dazu führen, dass sich Reflexion negativ auf die Fähigkeit zum kreativen Problemlösen auswirkt (Rosseel & Anseel, 2021) oder dass es zu Performanzverschlechterungen durch Inferenzen zwischen Denken und Tun kommt (vgl. Neuweg 2007, S. 11).

Abschließend ist noch auf ein letztes Problem einzugehen, nämlich die Möglichkeit, Reflexion als Kontrollinstanz zu missbrauchen. Reflexionen, besonders wenn sie schriftlich festgehalten und einsehbar sind, können dazu dienen, den Lernfortschritt, die Haltungen und die Handlungen von Lehrkräften und Studierenden offenzulegen. Das bereits im Studium verordnete Reflektieren sollte sich im besten Fall im Lauf der Professionalisierung verautomatisieren, sodass die regelmäßige Reflexion zu einer selbstverständlichen und gewohnheitsmäßigen Tätigkeit wird, um eine beständige Selbstverbesserung ohne Eingriffe von außen zu gewährleisten. Eine anfängliche Kontrolle von außen wird also schrittweise ersetzt durch eine Selbstkontrolle in Form einer reflexiven Haltung der eigenen Praxis gegenüber. Reflexion wird so zu einer Art Rechenschaftsablegung, einem Akt der Selbstregulation und Selbstoptimierung, der zur Durchsichtigkeit und Verfügbarkeit des Subjekts beiträgt (Foucault, 2016). So kann Reflexion eine lebenslange Selbstprüfung werden und eine Selbstoptimierung (Münste-Goussar, 2017) nach sich ziehen, die ihren Ursprung eben nicht in der Freiwilligkeit, sondern in der Fremdkontrolle hat: „Reflection [becomes] a tool for control and discursive domination“ (Cushion, 2018, S. 3). Reflexion liegt zwischen professionellem und privatem Denken und vor allem, wenn so etwas wie ein reflexives Tagebuch eingesetzt wird, um Reflexion zu erheben und einsehbar zu machen, ist das möglicherweise eine Grenzüberschreitung. Auch hier zeigt sich aber wieder die Vielseitigkeit von Reflexion – kann sie doch gleichzeitig vor Kontrolle und Überwachung schützen, da es sich auch um eine Selbst-Technik handelt, die zum kritischen Hinterfragen von Kontrolle und Selbstoptimierung führen kann (Pongratz et al., 2004).

Zusammenfassend ist es wichtig, sich dieser möglicherweise problematischen Aspekte von Reflexion bewusst zu sein, damit, wenn nötig und möglich, Vorsichtsmaßnahmen getroffen werden können und der Einsatz kritisch abgewogen werden kann: „As teacher educators, we must be aware of the limitations and misinterpretations of the idea“ (Akbari 2007, S. 204).

2.1.3 Zusammenfassung und Bewertung

Im ersten Kapitel dieser Arbeit wurde zunächst gezeigt, dass der Begriff Reflexion auf verschiedene Entwicklungslinien zurückgeht und in verschiedenen Kontexten und Domänen eine Rolle spielt, sodass es nicht nur in der Wissenschaft, sondern auch im Alltagsverständnis zu unterschiedlichen Verständnissen und Definitionen des Begriffs kam. Zwar wird sich in der aktuellen Literatur häufig auf Dewey und Schön bezogen, deren theoretische Entwürfe zur Reflexion werden aber nicht systematisch berücksichtigt. Es wird deutlich, dass hinsichtlich verschiedener Aspekte von Reflexion nicht nur die Theorien dieser beiden Wissenschaftler voneinander abweichen, sondern es viele, teils widersprüchliche Überlegungen zu relevanten Aspekten des Konstrukts gibt. Es fehlt also nicht nur eine konsensuelle Definition, sondern auch eine Einigkeit darüber, welche Aspekte es bei Reflexion überhaupt zu berücksichtigen gilt und welchen Anteil diese jeweils ausmachen.

Deutlich geworden ist die Dialektik des Begriffs Reflexion: Einerseits kann sie als Kontrollmechanismus dienen, andererseits als Mittel, Kontrolle und Freiheitseinschränkungen zu enttarnen und sich zu emanzipieren. Einerseits kann sie verunsichern, erstarren lassen und ins Grübeln stürzen und andererseits zu echter persönlicher und beruflicher Weiterentwicklung führen. Sie kann uns vereinzeln, aber auch verbinden. Sie kann zurückblicken und in dieser Vergangenheitsorientierung hängen bleiben, sie kann aber auch zukunftsorientiert sein und Veränderung anbahnen. Sie wird oft als rein kognitiver Vorgang begriffen, ist aber gleichzeitig an vielen Stellen mit Emotionen verknüpft und steht unmittelbar mit dem Handeln in Verbindung. Sie kann zu Lernen und Entwicklung führen, aber gleichzeitig auch Flow und entlastende Routinen unterbrechen. Sie kann die Berufszufriedenheit steigern und Angst nehmen und gleichzeitig zu Stress und Unsicherheit führen. Reflexion ist ein

zweischneidiges Schwert: Sie kann entweder ein wertvolles Tool sein oder auch etwas, das, unhinterfragt eingesetzt, Probleme mit sich bringen kann. Während Reflexion viel Potential zugeschrieben wird, sich persönlich und professionell weiterentwickeln, gibt es kaum empirische Belege, die diese Wirkung auch tatsächlich zeigen. Es ist deshalb wichtig, Reflexion nicht uneingeschränkt positiv zu sehen und als Allheilmittel zu propagieren, sondern sie differenziert zu betrachten und in ihren Facetten zu analysieren und zu erforschen.

Zu viele relevante Fragen sind noch offen und empirische Belege fehlen praktisch in allen angerissenen Bereichen. Es ist deshalb ein offensichtliches Desiderat, mehr systematische Grundlagenforschung durchzuführen. Aktuell herrscht die Tendenz vor, Reflexion nach den eigenen Bedürfnissen zu definieren und zu operationalisieren und nur den Fokus des eigenen Forschungsvorhabens zu berücksichtigen. Natürlich ist es für die Messung und Erforschung von Reflexion nötig, die Komplexität des Konstrukts auf ein bewältigbares Maß zu reduzieren. Das kann aber auch dazu führen, dass es im Grunde nicht mehr als das erfasst wird, was es eigentlich ist, und dass entscheidende Einflüsse und Aspekte übersehen werden, die beispielsweise Messergebnisse beeinflussen oder erklären.

Im folgenden Kapitel soll ein Überblick über den aktuellen Stand der Forschung zur Messung von Reflexion gegeben werden. Welche Messmethoden und -instrumente wurden bisher eingesetzt und welche Erkenntnisse damit gewonnen? Welche Probleme gilt es bei der Messung von Reflexion zu berücksichtigen? Dabei wird geprüft, inwiefern die im ersten Kapitel erarbeiteten Aspekte des Konstrukts eine Rolle für die Messung spielen und inwiefern diese bereits berücksichtigt wurden.

2.2 Messung von Reflexion

„Reflection: can we assess it? Should we assess it?“ (Sumsion & Fleet, 1996)

Im ersten Teil dieses Kapitels werden sowohl quantitative als auch qualitative Messverfahren und -instrumente vorgestellt und kritisch beleuchtet und einige alternative Messmöglichkeiten zu den konventionellen Verfahren vorgestellt. Damit soll ein Überblick gegeben werden, welche Möglichkeiten zur Messung es gibt und inwiefern diese den Gütekriterien entsprechen.

Im zweiten Teil wird untersucht, zu welchen Zwecken bisher Messungen von Reflexionsfähigkeit vorgenommen wurden. Dabei wird primär drei Fragen nachgegangen, nämlich wie die Reflexionsfähigkeit von Studierenden insgesamt eingeschätzt wird, ob und inwiefern sich Reflexionsfähigkeit im Lauf der Professionalisierung entwickelt und ob, bzw. mit welchen Maßnahmen, sie förderbar ist. Im Fokus steht dabei auch die Frage, ob die gewonnenen Ergebnisse verlässlich sind.

Der dritte Teil rückt Probleme und Schwierigkeiten in den Fokus, die entweder bei bisherigen Messungen bereits auftraten oder bei Messungen der Reflexionsfähigkeit generell zu bedenken sind, aber bisher zu wenig Berücksichtigung gefunden haben. Hier werden insbesondere die Erkenntnisse aus dem vorangegangenen Kapitel noch einmal aufgegriffen und mit dem Stand der empirischen Forschung in Verbindung gebracht: Es geht unter anderem um die Frage, wie Reflexionen ausgelöst werden können. Thematisiert werden aber auch der Einfluss der Aufgabenstellung, eine mögliche Sichtbarmachung von Reflexion über ein Medium und abschließend die Frage nach Rahmenbedingungen der Messung.

2.2.1 Messinstrumente

2.2.1.1 Fragebögen

Tabelle 1 gibt einen Überblick über die Fragebögen, die entweder mit Selbstauskunft oder Fallvignetten arbeiten und im Folgenden genauer vorgestellt werden.

Tabelle 1. Überblick über die vorgestellten quantitativen Messinstrumente

Autor*innen/Jahr	Instrument	Skalen
Korthagen (1993)	Internal/External Orientation Questionnaire (IEO)	Self-internal Self-external Fellow students-internal Fellow students-external Mathematics-internal Mathematics-external
Trapnell & Cambell (1999)	Rumination-Reflection Questionnaire (RRQ)	Rumination Reflection
Grant, Franklin & Langford (2002)	Self-Reflection and Insight Scale (SRIS)	Reflection Insight
Kember, Leung, Jones, Loke, McKay, Sinclair, Tse, Webb, Yuet, Frances, Wong, Yeung (2000)	Reflection Questionnaire	Habitual action Understanding Reflection Critical Reflection
Sobral (2000)	Reflection-in-Learning Scale (RLS)	-
Aukes, Geertsma, Cohen-Schotanus, Zwierstra & Slaets (2007)	Groningen Reflection Ability Scale (GRAS)	Self-reflection Empathetic reflection Reflective communication
Akbari, Behzadpoor & Dadvand (2010)	Reflective Teaching Inventory	Practical Cognitive Meta-cognitive Affective Critical
Choy & Oo (2012)	Teacher's Reflective Thinking Questionnaire	Ability to self-assess Awareness of how one learns Developing lifelong learning skills Influence of belief about self and self-efficacy
Priddis & Rogers (2018)	Reflective Practice Questionnaire	Uncertainty Job satisfaction Desire for improvement Self-appraisal Confidence
Kayapinar & Erkus (2009)	Teacher Reflection Scale	Fallvignetten
Seyfried, Weinberger & Reitingner (2013)	Dispositions Inventory for Education (DInE)	Reflexionskompetenz Persönliche Einstellung Vertrauensorientierung Haltungen zu Grundbedürfnissen

Selbstauskunft. Als einer der ersten Fragebögen, der im Kontext der Messung von Reflexion eingesetzt wurde, kann der *Internal/External Orientation Questionnaire (IEO)* von Korthagen (1993a) gelten. Wenngleich er nicht direkt Reflexionshäufigkeit oder -qualität erhebt, wurde er zur Messung derselben verwendet. Gemessen wird die Orientierung des Lernens entweder aus inneren oder aus äußeren Impulsen, wobei die inneren Impulse mit einem Lernen durch Reflexion gleichgesetzt wurden. In drei Domänen, nämlich der Beziehung der Lehramtsstudierenden zu sich selbst, zu Kommiliton*innen und zum Fach (hier: Mathematik) prüft der Fragebogen, ob diese Beziehungen jeweils eher internal oder external orientiert sind, woraus sich sechs Subskalen ergeben. Internal orientiert bedeutet, dass durch eigene Reflexionen und Überlegungen gelernt wird, external bedeutet, dass Hilfe und Anweisungen von außen benötigt oder gewünscht werden. Das Instrument besteht aus insgesamt 61 Items, die auf einer fünfstufigen Likert-Skala zu beantworten sind (in Abhängigkeit von der Formulierung ein Teil von 1 = *not at all applicable to me* bis 5 = *totally applicable to me*, ein zweiter Teil von 1 = *never* bis 5 = *always*). Die Reliabilitäten der Subskalen variieren zwischen Cronbachs $\alpha = .77$ und $\alpha = .87$. Es ist nicht geeignet für individuelle Diagnostik, sondern wurde entwickelt, um Entwicklungstendenzen in größeren Gruppen, zum Beispiel in Seminaren, aufzuspüren. Ein Beispielitem für eine internale Orientierung lautet: „I think about my own development“, für eine externale Orientierung: „I want people to tell me what I am doing wrong“ (S. 229). Allerdings äußert Korthagen selbst Bedenken, ob es auch als valide gelten kann, da sich interne und externe Orientierung nicht immer klar unterscheiden ließen. Korthagen und Wubbels führten in einer weiteren Studie eine Validierung des Fragebogens durch, indem sie die Ergebnisse mit dem Ergebnis einer qualitativen Textanalyse abglich. Die Ergebnisse des Fragebogens und der Textanalyse wichen stark voneinander ab. Sie führten das darauf zurück, dass der Fragebogen noch nicht adäquat sei und an der schwachen Operationalisierung von Reflexion scheitere. Sie hielten aber auch eine Verzerrung durch sozial erwünschte Antworten für möglich (Korthagen & Wubbels, 2000).

Mit der Messung von Selbstreflexion befassen sich zwei aus der Psychologie stammende Fragebögen, nämlich der *Rumination-Reflection Questionnaire (RRQ)* von Trapnell und Campbell (1999) und die *Self-Reflection and Insight Scale (SRIS)* von Grant, Franklin und Langford (2002). Beide Instrumente entstanden ausgehend vom Konzept der *Private Self-Consciousness (PrSC)* (Fenigstein et al., 1975) und

differenzieren dieses weiter aus. Das Konstrukt der Selbstaufmerksamkeit (= „self-consciousness“) kann sich entweder auf Aspekte des Selbst richten, die nur für das Individuum zugänglich sind (= „private“) oder auf Aspekte, die auch von außen zugänglich und beobachtbar sind (= „public“), wobei nur der erste hier in den Fokus genommen wurde. Ein lange ungelöstes Paradoxon bestand darin, dass ein hoher *Private Self-Consciousness*-Wert, also eine hohe nach innen gerichtete Selbstaufmerksamkeit, einerseits im positiven Sinne mit einem genaueren und umfangreicheren Wissen über die eigene Person verbunden sein kann, aber im negativen Sinne auch mit einem höheren Stresslevel einhergehen kann. Der RRQ von Trapnell und Campbell (1999) wurde auf Basis der Annahme entwickelt, dass sich dieses Paradoxon durch zwei Faktoren erklären lässt, die dem Konzept der *Private Self-Consciousness* innewohnen. Das sind einerseits die intellektuelle Selbsterkundung oder Neugier darauf, sich selbst besser zu erkennen, bezeichnet als „reflection“, die für die positiven Effekte verantwortlich sein könnte und andererseits die neurotische Selbstaufmerksamkeit, bezeichnet als „rumination“, also Grübeln, welche möglicherweise Stress und negative Folgen nach sich zieht. Um diese Annahme zu überprüfen, wurde der *Rumination-Reflection-Questionnaire (RRQ)* entwickelt, der aus 24 Items besteht, von denen jeweils zwölf den zwei Faktoren „reflection“ (Beispielitem: „I love analyzing why I do things“, S. 293) und „rumination“ (Beispielitem: „Oftentimes I'm playing back over in my mind how I acted in a past situation“, ebd.) angehören. Beantwortet werden die Items auf einer fünfstufigen Likert-Skala von 1 = *strongly disagree* bis 5 = *strongly agree*. Die Reliabilität beider Skalen war hoch ($\alpha = .91$ für „reflection“, $\alpha = .90$ für „rumination“). Folgestudien, die ebenfalls das Instrument eingesetzt haben, konnten die vermuteten Effekte von Grübeln und Reflexion replizieren (Harrington & Loffredo, 2010). Reliabilität und Validität des RRQ konnten in diesen Studien ebenfalls bestätigt werden. Auch zeigten sich die erwarteten positiven Korrelationen zwischen der Subskala „reflection“ und Offenheit für neue Erfahrungen sowie der Skala „rumination“ und Neurotizismus (Teasdale & Green, 2004).

Das zweite Instrument, das sich mit der *Private Self-Consciousness-Scale* befasst, und diese weiter ausdifferenziert, ist die *Self-Reflection and Insight Scale (SRIS)*. Während Trapnell und Campbell versuchten, die *Private Self-Consciousness-Scale* in die zwei Faktoren „reflection“ und „rumination“ ausdifferenzieren, wurden

beim SRIS von Grant und Kollegen die zwei Faktoren „self-reflection“ und „insight“ gebildet. „Self-reflection“ misst dabei, wie hoch einerseits das Bedürfnis nach Selbstreflexion ist („need for self-reflection“, 6 Items) und andererseits in welchem Maß Selbstreflexion tatsächlich stattfindet („engagement in self-reflection“, 6 Items). Ein Beispielitem der Skala „self-reflection“ lautet: „I frequently take time to reflect on my thoughts“ (S. 825). „Insight“ misst den Grad der Bewusstheit von eigenen Gefühlen, Gedanken und Beweggründen (8 Items, Beispielitem: „I am usually aware of my thoughts“). Das Instrument umfasst also insgesamt 20 Items, die auf einer sechsstufigen Skala beantwortet werden (1 = *strongly disagree* bis 6 = *strongly agree*). Es wurde in mehreren Teilstudien von den Autoren validiert, in welchen sich gute Reliabilitäten zeigten („self-reflection“: $\alpha = .91$ bzw. $.71$; „insight“: $\alpha = .87$ bzw. $\alpha = .82$). Auch andere Studien bestätigten, dass es sich um zwei unabhängige Faktoren handelt und dass „insight“ positiv mit einem hohen Selbstwertgefühl und negativ mit depressiven Symptomen und Ängsten zusammenhängt, was bei „reflection“ aber nicht der Fall ist (Silvia & Phillips, 2011).

Der *Reflection Questionnaire* von Kember et. al (2000) ist ein Instrument, das auf den Einsatz in Seminaren ausgelegt ist und misst, inwiefern die Teilnehmenden in diesem Rahmen reflexives Denken einsetzen. Es wurde für den Einsatz in Medizin und Gesundheitswissenschaften entwickelt und getestet, ist aber den Autor*innen zufolge auf andere Einsatzbereiche übertragbar. Es basiert auf einer umfangreichen Literaturanalyse und stützt sich theoretisch hauptsächlich auf die Arbeit von Mezirow (1991). Aus dieser wurden die vier Skalen des Instruments abgeleitet: „habitual action“ (unreflektiertes, gewohnheitsmäßiges Handeln), „understanding“ (Verständnis akademischer Inhalte ohne Herstellen eines Bezugs zum eigenen Leben), „reflection“ (Nachdenken über eigenes Handeln und mögliche Alternativen) und „critical reflection“ (Hinterfragen von Gründen für das eigene Handeln und Denken). Diese vier Dimensionen steigern sich in der Qualität der Reflexion von Stufe 1 bis Stufe 4, womit das Instrument auch den Stufenmodellen der qualitativen Inhaltsanalyse (siehe folgendes Kapitel) sehr nahesteht. Für jede Dimension wurden je vier Items entwickelt, die auf einer fünfstufigen Skala von A = *definitely agree* bis E = *definitely disagree* zu bewerten sind. Ein Beispielitem aus der Subskala „reflection“ lautet: „I often reflect on my actions to see whether I could have improved on what I did“ (S. 395). Durch das Summieren der Scores in jeder Subskala kann für jede Versuchsperson ein Reflexionsprofil erstellt werden,

das jeweils die Ausprägungen in den vier Dimensionen zeigt. Die Reliabilitäten der Skalen reichten von Cronbachs $\alpha = .62$ bis $\alpha = .76$. Eine konfirmatorische Faktorenanalyse bestätigte die vierfaktorielle Struktur des Instruments. Eine Studie von Lethbridge, Andrusyszyn, Iwasiw, Laschinger und Fernando (2013) untersuchte die psychometrischen Eigenschaften des Tests und bestätigte dessen Validität und Reliabilität weitgehend.

Die *Reflection-in-Learning Scale (RLS)* von Sobral (2000) misst, in welchem Grad und auf welche Weise Reflexion während eines Lernprozesses genutzt wird. Sie existiert in zwei Fassungen. Eine besteht aus zehn Items, die auf einer siebenstufigen Skala von 1 = *never* bis 7 = *always* beantwortet werden. Der Originalfragebogen dazu wurde in portugiesischer Sprache verfasst, der Autor legt aber eine englischsprachige Übersetzung vor. Ein Beispielitem lautet: „To what extent have I: Been aware of what I was learning and for what purposes“ (S. 187). Eine zweite Fassung besteht aus 14 Items, bei der die vier zusätzlichen Items die empfundene Selbstwirksamkeit bezüglich „reflection-in-learning“ erheben sollen. Beide Versionen haben laut dem Autor sehr ähnliche psychometrische Eigenschaften. Für die Version mit 10 Items war dies ein Cronbachs $\alpha = .81$ (vgl. ebd.), in einer weiteren Studie für die Version mit 14 Items ein Wert von $\alpha = .84$ zum ersten Messzeitpunkt und $\alpha = .88$ zum zweiten Messzeitpunkt (Sobral, 2005). In dieser Validierungsstudie wurde auch eine Faktorenanalyse durchgeführt, die die einfaktorielle Struktur bestätigte und die Test-Retest-Stabilität ermittelt, welche vom Autor als gut bezeichnet wurde ($r = .709, p \leq .001, n = 72$ nach 6 Monaten; $r = .515, p \leq .001, n = 75$ nach 18 Monaten). Der Autor schließt aus den Daten, dass hier eher ein stabiler Charakterzug gemessen wird, als ein kontextspezifisches Konstrukt (vgl. ebd., S. 311).

Der *Reflection Questionnaire* von Kember und Kolleg*innen und die *Reflection-in-Learning Scale* wurden beide in einer Untersuchung von Kalk, Luik, Taimalu und Täht (2014) an einer Stichprobe aus Lehramtsstudierenden und jungen Lehrkräften aus verschiedenen Schularten eingesetzt und auf ihre Validität und Reliabilität geprüft. Beide Instrumente werden von den Autor*innen als grundsätzlich übertragbar auf den pädagogischen Bereich eingeschätzt und zur Anwendung empfohlen, obwohl sich die Ergebnisse der Originalstudien bei beiden Instrumenten nur zum Teil replizieren ließen. Die Untersuchung ergab außerdem, dass die *Reflection-in-*

Learning Scale 25 % der Subskala „reflection“ und 4 % der Subskala „critical reflection“ des *Reflection Questionnaire* von Kember et al. erklärt (vgl. ebd., S. 132). Sie folgern: „The three tested scales thus measure different aspects of reflection“ (ebd.).

Die *Groningen Reflection Ability Scale (GRAS)* von Aukes, Geertsma, Cohen-Schotanus, Zwierstra und Slaets (2007) wurde entwickelt, um die persönliche Reflexionsfähigkeit von Medizinstudierenden zu messen. Es kann den Autor*innen zufolge sowohl auf Gruppenebene als auch auf Individualebene eingesetzt werden und enthält insgesamt 23 Items. Diese werden auf einer fünfstufigen Likert-Skala (1 = *totally disagree* bis 5 = *totally agree*) beantwortet und gliedern sich in die drei Aspekte „self-reflection“ (10 Items), „empathetic reflection“ (6 Items) und „reflective communication“ (7 Items). Ein Beispielitem lautet: „I take a closer look at my own habits of thinking“ (S. 181). Die angedachte dreidimensionale Struktur konnte die Faktorenanalyse nicht bestätigen, die nur eine einfaktorielle Lösung zuließ. Das endgültige Instrument ist das Produkt einer umfassenden und professionellen Instrumentenentwicklung, bei der aus der Theorie abgeleitete Items in quantitativen und qualitativen Verfahrensschritten reduziert und optimiert wurden. Die interne Konsistenz kann mit einem Cronbachs $\alpha = .83$ (in einer ersten Messung) und $\alpha = .74$ (in einer zweiten Messung) als gut beurteilt werden. Trotzdem weisen die Autor*innen darauf hin, dass das Phänomen Reflexion in der Literatur als multidimensional und komplex beschrieben wird und das Instrument vermutlich nicht in der Lage ist, dies vollständig abzubilden. Sie heben aber die Einfachheit und Praktikabilität des Instruments hervor, mit dessen Hilfe auf zeitsparende und ökonomische Weise ein Gesamtscore der Reflexionsfähigkeit (von 23 Punkten bis max. 115 Punkten) gebildet werden könne (vgl. S. 182).

In einer Validationsstudie von Andersen, O’Neill, Gromsen, Hvidberg und Morcke (2014) wurden die psychometrischen Eigenschaften des Instruments untersucht, das zuvor ins Dänische übersetzt wurde (GRAS-DK). Die Studie wurde ebenfalls im medizinischen Bereich durchgeführt. Während beispielsweise Messfehler und Test-Retest-Score und Cronbachs Alpha ($\alpha = .87$) akzeptabel waren, konnte in einer Faktorenanalyse weder die einfaktorielle noch die ursprünglich angedachte dreidimensionale Struktur bestätigt werden. Außerdem gab es einige nicht erklärbare

extreme Ausreißer. Die Autor*innen führen die negativen Ergebnisse darauf zurück, dass hier das Konzept der Reflexion nicht angemessen operationalisiert wurde.

Das *Reflective Teaching Inventory* von Akbari, Behzadpoor und Dadvand (2010), wurde für die Messung von Reflexion bei Englischlehrkräften entwickelt, ist aber so allgemein formuliert, dass es fachunabhängig eingesetzt werden kann. Bei seiner Entwicklung wurde versucht, die theoretischen Schwächen und die mangelnde Operationalisierung von Reflexion bei vorherigen Fragebögen zu vermeiden. Aus einer umfassenden theoretischen Analyse wurden über sechshundert mögliche Items extrahiert, die in weiteren Schritten und mithilfe von Expert*innen gekürzt und immer wieder validiert wurden. In der finalen Fassung besteht es aus 29 Items, die auf einer fünfstufigen Likertskala von 1 = *nie* bis 5 = *immer* beantwortet werden. Bei diesem Instrument werden fünf Elemente von Reflexion über eine Selbstauskunft gemessen, nämlich „practical“, „cognitive“, „meta-cognitive“, „affective“ und „critical“. „Practical“ meint dabei den tatsächlichen Akt des Reflektierens und die Frage, wie dieser durchgeführt wird (Beispielitem: “I write about my teaching experiences in a diary or a notebook”, S. 216). „Cognitive“ meint Schritte der bewussten Professionalisierung, zum Beispiel durch Fortbildungen oder Lektüre. „Affective“ bezieht sich auf das Nachdenken über das Wissen, die Emotionen und Kognitionen der Lernenden. Unter „meta-cognitive“ ist das Wissen der Reflektierenden über ihre eigene Persönlichkeit und ihr Agieren als Lehrkraft zusammengefasst. „Critical“ bezieht sich auf die soziale und politische Dimension von Unterricht (vgl. S. 215). Im ursprünglichen Fragenpool war auch noch der Aspekt „moral“ vorhanden, der sich mit Fragen von Gerechtigkeit und Werten befasste, aber im Laufe der Optimierung des Fragebogens im Rahmen einer konfirmatorischen Faktorenanalyse ausgeschlossen wurde. Eine Validierungsstudie von Moghaddam, Davoudi, Adel und Amirian (2020) ermittelte einen Wert von Cronbachs $\alpha = .78$ für das gesamte Instrument und eine Faktorenanalyse bestätigte die Struktur des Fragebogens. Es wurde außerdem mit ähnlich guten Ergebnissen von Afshar und Farahni (2015) eingesetzt, scheint aber den iranischen Raum, in dem es auch entwickelt wurde, bisher nicht verlassen zu haben. Die letztgenannte Studie setzte außerdem noch einen zweiten Fragebogen ein und ermittelte eine Korrelation zwischen dem *Reflective Teaching Inventory* und dem *Teacher's Reflective Thinking Questionnaire* von Choy und Oo (2012) von $r = .56$ ($p \leq .001$) bei einer

Stichprobe von 233 Teilnehmer*innen. Der *Teacher's Reflective Thinking Questionnaire* besteht aus 33 Items, die auf einer fünfstufigen Skala von 1 = *strongly disagree* bis 5 = *strongly agree* beantwortet werden. Die Items sind gegliedert in die vier Abschnitte „ability to self-assess“ (12 Items, Beispiel: „I always think of what I had done during my lessons so that I can improve on it further“, S. 171), „awareness of how one learns“ (9 Items), „developing lifelong learning skills“ (9 Items) und „influence of belief about self and self-efficacy“ (3 Items). Die Autor*innen berichten nicht darüber, wie der Fragebogen genau entwickelt wurde und Daten zu dessen Reliabilität und Validität fehlen ebenfalls.

Der *Reflective Practice Questionnaire (RPQ)* (Priddis & Rogers, 2018) wurde entwickelt, um selbstberichtete Erfahrungen mit reflexiver Praxis in Gesundheitsberufen und pädagogischen Berufen und damit einhergehende mögliche Chancen und Risiken zu erheben. Es handelt sich also um ein disziplinenübergreifendes Instrument, dessen zehn Subskalen mit je vier Items auf einer sechsstufigen Skala beantwortet werden (1 = *not at all* bis 6 = *extremely*). Die Subskalen können laut den Autor*innen auch einzeln eingesetzt werden. Sie beziehen sich nur zum Teil direkt auf die Messung von Reflexion und zum Teil auf Konstrukte, die theoretisch oder empirisch mit Reflexion in Verbindung gebracht werden („uncertainty“, „job satisfaction“, „desire for improvement“, „self-appraisal“ und „confidence“). Die drei Skalen zur Reflexion wurden auch in dieser Studie aus der Literaturanalyse abgeleitet: „reflective-in-action“, „reflective-on-action“, „reflective with others“. Ein Beispielitem aus der Skala „reflective-on-action“ lautet „After interacting with clients I spend time thinking about what was said and done“ (S. 94). Die Reliabilitäten der einzelnen Skalen lagen in einer ersten Teilstudie der Autor*innen zwischen $\alpha = .82$ und $\alpha = .91$, in einer zweiten Teilstudie zwischen $\alpha = .64$ und $\alpha = .92$. Die Autor*innen erprobten das Instrument selbst in einer weiteren Studie im medizinischen Kontext mit einem Cronbachs $\alpha = .84$ für die drei Reflexionsskalen und Werten zwischen $\alpha = .75$ und $\alpha = .83$ für die weiteren Konstrukte (Rogers et al., 2019) und folgerten daraus, dass es sich um ein reliables Instrument handelt. In einer Studie von Fuertes-Camacho, Sulsat-Ortiz und Álvarez-Cánovas (2021) konnten die hohen Reliabilitäten nur teilweise repliziert werden und vor allem die theoretisch verwandten Konstrukte wie „self-appraisal“, „stress“ und „job satisfaction“ blieben unter einem Wert von $\alpha = .60$.

Zur Erhebung reflexionsbezogener Teilaspekte, die allerdings hier nur noch am Rande erwähnt werden sollen, existieren noch weitere Instrumente, zum Beispiel ein Fragebogen zur reflexionsbezogenen Selbstwirksamkeit (Lohse-Bossenz et al., 2019) oder das Instrument *Reflection on Success and Failures*, das, ähnlich wie bereits andere vorgestellte Instrumente, im medizinischen Bereich entwickelt wurde (Matsuo, 2020). Zur Messung der Einstellungen zur systematischen Reflexion, der Einstellungen zur kollegialen Reflexion und der Einstellungen zur individuellen Reflexion wurde von Neuber und Göbel (2016) ein Instrument entwickelt.

Neukombination und Teilverwendung bestehender Instrumente. Zum Teil wird keine formale Fragebogenkonstruktion unternommen, sondern bestehende Fragebögen werden abgewandelt und angepasst. Dies geschah nach eigenen Angaben beispielsweise bei Hosenfeld (2010) in deren Studie zum Einfluss von Unterrichtsrückmeldungen auf die Unterrichtsentwicklung unter anderem auch die einstellungs- und verhaltensbezogene Reflexionsbereitschaft gemessen werden sollte. Die Skala „Reflexionsbereitschaft verhaltensbezogen“ besteht aus sieben Items, die auf einer fünfstufigen Skala (1 = *nie* bis 5 = *sehr oft*) beantwortet werden und sich auf die Frage beziehen: „Wie oft tun Sie folgende Dinge?“. Ein Beispielitem lautet: „Ich denke darüber nach, wie ich meinen Erziehungsauftrag besser erfüllen kann“ (Hosenfeld, 2010, S. 97). Die Reliabilität ist mit $\alpha = .72$ zufriedenstellend. Die Skala „Reflexionsbereitschaft einstellungsbezogen“ besteht aus sieben Items, die ebenfalls auf einer fünfstufigen Skala beantwortet werden (1 = *stimme gar nicht zu*, 5 = *stimme voll und ganz zu*). Sie beziehen sich auf die Frage: „Wie sehr stimmen Sie den folgenden Aussagen zu?“ (ebd., S. 98) und ein Beispielitem lautet hier: „Ich nehme mir Zeit, um meinen Unterricht zu reflektieren.“ Cronbachs Alpha für diese zweite Skala beträgt $\alpha = .78$. Die beiden Skalen wurden von der Autorin selbst entwickelt und als Basis dafür der „Fragebogen zur Erfassung dispositionaler Selbstaufmerksamkeit“ (Filipp & Freudenberg, 1989) genannt. Da sich die Fragebögen aber nicht ersichtlich ähneln, stellt sich die Frage, nach welcher Logik die Items aus diesem Vorbild abgeleitet wurden. Auch stellt die theoretische Basis ein Problem dar, da das Verhältnis zwischen Selbstaufmerksamkeit und (Selbst-)Reflexion noch nicht vollständig geklärt ist, wie bereits zu Beginn des Kapitels deutlich wurde (siehe die Instrumente RRQ und SRIS).

Ebenfalls selbst entwickelt wurden die Skalen zur Reflexionsbereitschaft der Gießener Offensive Lehrerbildung (Fraij, 2018). Es handelt sich um drei reflexionsbezogene Skalen, die alle auf einer fünfstufigen Skala (abhängig von der Formulierung der Items von 1 = *nie* bis 5 = *immer* oder 1 = *trifft nicht zu* bis 5 = *trifft zu*) beantwortet werden. Diese lauten: „Reflexion im zukünftigen Lehreralltag“ (12 Items, $\alpha = .87$), „Reflexionsbezogene Selbstwirksamkeitserwartungen“ (8 Items, $\alpha = .87$) und „Reflexive Orientierungen“ (7 Items, davon drei der externen Orientierung zugeordnet, $\alpha = .70$, und vier der inneren Orientierung zugeordnet, $\alpha = .81$). Teilweise wurden für diese Skalen Items aus anderen Fragebögen übernommen, übersetzt und neu kombiniert, zum Beispiel aus der bereits vorgestellten GRAS von Aukes et al. (2007) oder dem IEO-Fragebogen von Korthagen und Wubbels (2001). Zum Teil wurden die Items in Anlehnung an andere Erhebungsinstrumente neu formuliert, beispielsweise einem Fragebogen zu Komponenten reflexiver beruflicher Entwicklung von Niggli (2004) und einem Fragebogen zu Elementen des reflexiven Praktikums von Felten (2005). Auch hier finden sich keine Angaben, nach welchen Prämissen die Vorlagen jeweils ausgewählt und verändert wurden oder wie sich die Auswahl der Items begründet.

Fallvignetten. Einen Sonderfall der Fragebogenerhebung stellen Erhebungen mit Fallvignetten dar. Zwei Beispiele für die Messung von Reflexion mit Fragebögen, die Fallvignetten einsetzen, sind die *Teacher Reflection Scale* von Kayapinar und Erkus (2009) und zum Teil auch das Instrument DInE von Seyfried, Weinberger und Reitinger (2009).

Die *Teacher Reflection Scale* von Kayapinar und Erkus (2009) wurde entwickelt, um die Reflexionsfähigkeit von Lehrkräften einzuschätzen und sie damit in ihrer professionellen Entwicklung zu unterstützen. Das einfaktorielle Instrument besteht aus 22 Items in der Form von Fallvignetten zu problematischen Situationen in Unterricht, Schule und Kollegium. Eine solche lautet beispielsweise: „One of your students prevents others from learning“ (ebd., S. 150). Zur Auswahl stehen bei jeder Fallvignette drei mögliche Antworten von a bis c, die potenzielle Reaktionen auf den Fall darstellen. Die Antwortmöglichkeiten zum genannten Beispiel lauten: „a. I look for the ways to end the situation./ b. I make him/her sit down in the front alone./

c. Everyone is responsible for himself/herself“ (ebd.). Je eine der Antwortmöglichkeiten signalisiert, dass von der Lehrkraft reflektiert wird (in diesem Fall a.), eine zweite Antwortmöglichkeit steht für das Fehlen von Reflexion mit Fokus auf eine externale Attribution (hier b.) und eine dritte Antwortmöglichkeit steht ebenfalls für das Fehlen von Reflexion mit Fokus auf Ablehnung der Verantwortungsübernahme (hier c.). Theoretische Voraussetzung des Instruments ist also, dass Reflexion mit einer internalen Attribution und einem hohen Maß an Verantwortungsübernahme einhergeht.

Hier zeigt sich auch bereits das Hauptproblem des Instruments. Obwohl das Instrument eine hohe Reliabilität aufweist ($\alpha = .87$), statistisch geprüft, mit Hilfe von Expert*innen entwickelt wurde und die Autoren die erwartbaren Korrelationen mit verwandten Konstrukten nachweisen konnten, ist fraglich, was hier eigentlich gemessen wird und ob es nicht eher der Grad der Verantwortung und die Art der Attribution sind als tatsächlich das Auftreten von Reflexion.

Das Instrument wurde in der Folge von Armutcu und Yaman (2010) bei Lehramtsstudierenden vor und nach dem Praktikum eingesetzt, um einen erwarteten Zuwachs an Reflexion zu messen. Die psychometrischen Eigenschaften des Instruments wurden dabei nicht mehr geprüft. Es fanden sich keine Unterschiede in den Tests zu Beginn und am Ende des Praktikums. Das Ergebnis deckt sich mit der sehr hohen Test-Retest-Reliabilität, die bereits in der Originalstudie beschrieben wurde. Die Autoren folgern daraus, dass hier, ähnlich wie bei dem Instrument von Sobral, eventuell hauptsächlich Reflexion als relativ stabiler Charakterzug gemessen wird. Das wäre wiederum plausibel, wenn es eigentlich Verantwortungsbewusstsein und Kontrollüberzeugen sind, die gemessen werden, da diese recht stabil sind.

Beim *Dispositions Inventory for Education (DInE)* von Seyfried, Weinberger und Reitinger (2013) handelt es sich um ein Inventar, das für Aufnahmeverfahren von Bewerber*innen an Pädagogischen Hochschulen entwickelt wurde. Es soll Informationen darüber liefern, welche Dispositionen die Kompetenzentwicklung beeinflussen und fokussiert insbesondere auch auf die Lehrer*innengesundheit. Das Inventar ist in vier Bereiche gegliedert, die unterschiedliche Aspekte erheben, und zwar „Reflexionskompetenz“, „Persönliche Einstellung“, „Vertrauensorientierung“ und „Haltungen zu Grundbedürfnisfragen“.

Relevant für diese Arbeit ist nur der Teil, der Reflexionskompetenz zu erfassen versucht. Dies geschieht über die Vorlage von sechs Bildvignetten, die Szenen aus dem beruflichen Alltag von Lehrkräften zeigen (z.B. eine Schülergruppe in der Pause). Zu jedem Bild werden zwei interpretierende Bildbeschreibungen präsentiert, die jeweils auf einer fünfstufigen Skala beantwortet werden (1 = *stimme gar nicht zu* bis 5 = *stimme voll zu*). Durch diesen Aufbau ergibt sich eine Gesamtskala von zwölf Items. Eine der beiden Aussagen steht dabei für eine niedrige und eine Aussage für eine hohe Reflexionskompetenz. Eine beispielhafte Aussage für eine hohe Reflexionskompetenz lautet:

Im Bild sehe ich einen Buben und fünf Mädchen. Vermutlich handelt es sich um eine Pause. Die Mädchen scheinen sich miteinander zu unterhalten, der Bub sitzt alleine und schaukelt auf dem Stuhl. Vielleicht hat er Langeweile, aber ich kann das nicht sicher sagen. Aufgrund des einzelnen Bildes lassen sich schwer Aussagen über die einzelnen Personen sagen [sic]. Ich könnte darüber nachdenken, wie es wohl den einzelnen Kindern in dieser Situation geht: Fühlen sich die Kinder wohl? (Ebd., S. 121)

Die zugehörige Aussage für eine niedrige Reflexionskompetenz lautet:

Das Bild zeigt eine Szene aus der Pause. Für das Alter der Kinder ist es eher eine typische Situation: Mädchen unterhalten sich miteinander, ein Bub wird da nicht dazu gezogen. Dieser sitzt alleine auf einem Stuhl und schaukelt aus Langeweile. Die Mädchen haben offensichtlich ein Thema von besonderem Interesse. Es geht um einen Nagellack. Ich würde gerne den Buben ansprechen, damit sich dieser nicht alleine fühlen muss. (Ebd.)

Was als niedrige und hohe Reflexionskompetenz gilt, wurde in einer früheren Veröffentlichung der Autoren erarbeitet und für die Formulierung der Antwortmöglichkeiten verwendet (Weinberger & Seyfried, 2009). Eine Aussage mit hoher Reflexionskompetenz zeichnet sich demzufolge durch deskriptive und interrogative Aussagen sowie Erklärungen und Interpretationen mit Beobachterbewusstsein aus. Auch die Bilder stammen aus dieser ursprünglichen Veröffentlichung zum qualitativen *Reflection Instrument for Education (RIfE)*, das sich mit der qualitativen, textanalytischen Messung von Reflexionsfähigkeit befasst hatte und im folgenden Kapitel genauer vorgestellt wird.

Das Instrument wurde in einem Test professionellen und narrativen Reflektierenden ($N = 38$) vorgelegt, wobei nicht erläutert wurde, nach welchen Kriterien diese beiden Gruppen gebildet wurden. Es konnte laut den Autoren signifikant zwischen

den Gruppen unterscheiden. Mit einem Cronbachs $\alpha = .83$ weist es eine gute Reliabilität auf, weshalb von den Autoren geschlussfolgert wird, dass es seinen Zweck, Reflexionskompetenz zu messen, erfüllt.

Ein Problem des Instruments ist, dass die präsentierten Bilder bereits in der ersten Studie zu RIfE als potenziell zu uneindeutig kritisiert wurden. Diese Uneindeutigkeit könnte Einfluss auf die Bewertung der präsentierten Texte zu den Bildern haben, die unabhängig von der Reflexionsfähigkeit sind. Da nur zwei unterschiedliche Aussagen präsentiert wurden, konnte nicht das ganze Spektrum der Reflexionsfähigkeit abgedeckt werden. Ferner kann kaum davon gesprochen werden, dass eine „Kompetenz“ durch das Vergeben von Zustimmung oder Ablehnung zu Aussagen gemessen werden kann. Die Messung bleibt unterkomplex. Außerdem kann immer nur einem aus ca. sechs Sätzen bestehenden Beschreibungsblock zugestimmt werden, sodass vor allem die Frage offenbleibt, welchem Teilaspekt hier vielleicht aus welchen Gründen zugestimmt oder welcher abgelehnt wurde. Womöglich wird also eine Aussage abgelehnt, weil einem Teilsatz nicht zugestimmt werden kann, der sich auf die Interpretation der Szene auf dem Bild bezieht und mit dem Beobachterbewusstsein in keinem Verhältnis steht. Diese Besonderheiten des Instruments gefährden eventuell die kognitive Validität.

Zwischenfazit. Als typische Messfehlerquelle von Fragebögen gilt die Tatsache, dass die Items von den Befragten anders verstanden werden, als dies von den Entwickler*innen intendiert war: „Interpreting items as they are intended by the instrument developer appears to be a difficult and demanding task for respondents“ (Faddar et al., 2017, S. 621). Das könnte besonders auf Fragebögen zur Reflexion zutreffen, denn wie bereits gezeigt wurde, wird der Begriff Reflexion sehr unterschiedlich verstanden. Viele der vorgestellten Fragebögen setzen aber in der Itemformulierung direkt das Wort „Reflexion“ oder „reflektieren“ ein, beispielsweise im Fragebogen von Korthagen (1993): „I reflect on myself“ (S. 229). Ob man ein solches Item als für sich zutreffend erachtet, hängt von der individuellen Definition von Reflexion ab.

Ein weiteres Problem, das sich beispielhaft an diesem Item zeigt, ist, dass diese häufig zu vage formuliert sind und eine Bezugsnorm fehlt, anhand derer eine Einordnung vorgenommen werden kann (N. B. Andersen et al., 2014). Wie oft sollte jemand sich selbst reflektieren, um diesem Item zustimmen zu können? Eine Folge aus der wenig konkreten Formulierung könnte auch sein, dass das Gefragte nicht erinnert werden kann. Reflexion ist schwer fassbar, weil es sich dabei um einen „nonfactual content“ handelt, zu dem beispielsweise „attitudes, beliefs“ und „self-perception“ (Alwin, 2010, S. 410) gehören. Über welche manifesten Eigenschaften man auf das latente Konstrukt Reflexion schließen kann, ist noch nicht abschließend geklärt.

Einen weiteren limitierenden Faktor bei der Erhebung von Reflexion über Selbstauskunft stellt die Tatsache dar, dass für eine Beurteilung der eigenen Reflexionsfähigkeit oder -bereitschaft bereits ein hohes Maß an Selbstreflexionsfähigkeit vorhanden sein muss: „Applicants are asked to judge their own reflection ability, which presupposes already a certain degree of self-reflection and self-observation (Aukes et al., 2007, S. 181). Ein verzerrtes Selbstbild könnte die Zuverlässigkeit der Messung beeinträchtigen. Aus der Forschung ist bekannt, dass „private Selbstaufmerksamkeit mit einer besseren Selbstkenntnis, zumindest aber mit einer erhöhten Veridikalität der Selbstberichte“ (Filipp & Freudenberg, 1989, S. 9) einhergeht. Es ist also davon auszugehen, dass Personen mit einer geringeren Selbstaufmerksamkeit und Reflexionsfähigkeit sich auch weniger zuverlässig selbst einschätzen und damit die Validität ihrer Selbstauskunft in Frage gestellt ist.

Auch das Problem sozial erwünschter Antworten, das bei der Selbstauskunft ohnehin immer besteht, könnte im Kontext der Reflexionsforschung dadurch verstärkt sein, dass das Thema in der Lehrkräftebildung sehr prominent vertreten und durchweg positiv konnotiert ist. Es ist deshalb anzunehmen, dass den Studierenden und Lehrkräften in besonderem Maße bewusst ist, dass eine reflexive Haltung eine wünschenswerte Zielkompetenz und damit auch eine soziale Norm darstellt, der sie möglicherweise durch angepasste Antworttendenzen zu entsprechen versuchen. Hierzu gibt es allerdings noch keine Forschung, sodass dieser Zusammenhang erst geprüft werden müsste.

Hinsichtlich vieler der untersuchten Instrumente fällt außerdem auf, dass die Skalen oft nicht ausbalanciert sind und sich wesentlich weniger oder sogar gar keine nega-

tiv formulierten Items finden, was die Verzerrung durch das Auftreten von Acquieszenz, also der Tendenz mit „stimme zu“ zu antworten, erhöhen könnte (Bogner & Landrock, 2015).

Wesentlich schwerwiegender als die beschriebenen möglichen Messfehler wiegt allerdings die Tatsache, dass begründete Zweifel bestehen, ob im Falle der beschriebenen Fragebögen das Kriterium der Inhaltsvalidität erfüllt ist. Zwar wurde bei fast allen Messinstrumenten eine mindestens zufriedenstellende Reliabilität erreicht. Reliabilität ist aber nur eine notwendige, keine hinreichende Voraussetzung für eine valide Messung. Kontrollen der Konstruktvalidität und der faktoriellen Validität offenbarten teilweise Probleme mit der Faktorenstruktur mancher Instrumente, fielen aber trotzdem meist im Sinne der Erwartungen der Entwickler*innen aus. Ob allerdings die Instrumente tatsächlich messen, was sie messen sollen, können diese eher statistisch orientierten Validierungsverfahren nicht bestätigen (Schaper, 2014). Das Hauptproblem liegt in der Definition und der Operationalisierung des Konstrukts Reflexion und damit in einem Theoriedefizit. Dieses verstärkt sich, wenn Fragebögen nicht mehr durch eine umfangreiche Theorieanalyse und eine Operationalisierung des Konstrukts entwickelt werden, sondern nur durch eine Anpassung bestehender Fragebögen an neue Kontexte, wobei sich zum Teil auch noch aus unterschiedlichen Instrumenten bedient wird, um eine neue Skala zu erstellen. Auch der Einsatz selbstformulierter Items, die sich zwar an der eigenen Fragestellung, jedoch nicht an der Theorie des Konstrukts orientieren, ist als problematisch zu betrachten. Doch auch sorgfältig entwickelte Fragebögen mit umfangreicher Theorieanalyse haben das gleiche Problem. Das ist daran erkennbar, dass die verschiedenen Fragebögen Unterschiedliches messen, was sich immer dann zeigt, wenn Korrelationen zwischen Fragebögen ermittelt wurden und diese klein ausfielen. Eben weil das Konstrukt Reflexion noch nicht konsensuell definiert und geklärt ist, führt dies zur Entwicklung von immer neuen Instrumenten: „Reflection is not well defined and there is lack of consensus about what reflection actually entails. Therefore, some authors have developed their own model of reflection to design an instrument“ (Kalk et al., 2014, S. 123).

Da ebenfalls nicht klar ist, wie sich ein hohes Level von Reflexion überhaupt definiert, kann dieses auch in den Items nicht umgesetzt werden. Es steht also inhaltlich gesehen nicht fest, was überhaupt eine wünschenswerte Antwort darstellen kann:

„We wonder, whether a high score indicates a high level of reflection. For example, is it a measure of high personal reflection to agree that one understands people of a different cultural background – or is it reflective to rather be self-critical and indicate that it is challenging to understand other people?“ (N. B. Andersen et al., 2014, 6f).

Diese Instrumente haben zusätzlich noch das Problem, dass sie Reflexion nicht eindeutig von ähnlichen psychologischen Phänomenen wie beispielsweise Selbstaufmerksamkeit abgrenzen können und am Ende vielleicht nicht (nur) Reflexion messen, sondern (zumindest zum Teil auch) Introspektion, Verantwortungsbewusstsein, Kontrollüberzeugungen oder Selbstwirksamkeitsüberzeugungen.

2.2.1.2 Inhaltsanalytische Verfahren

Während quantitative Messverfahren zwar gegenwärtig einen gewissen Aufschwung erleben, sind sie trotzdem noch wesentlich seltener vertreten als inhaltsanalytische Verfahren, die sich auf einen Textkorpus stützen. Dieser Korpus besteht meist aus schriftlichen Reflexionen, vereinzelt auch aus transkribierten Interviews. Bei der Analyse dieser Texte lassen sich verschiedene Herangehensweisen unterscheiden: Am häufigsten wird die Reflexionstiefe der Texte anhand von Levels oder Stufen gemessen und aus diesen ein Schluss auf die Reflexionsfähigkeit der Proband*innen gezogen. Manche Studien legen ihren Fokus aber auch auf die Reflexionsbreite und untersuchen, welche inhaltlichen Dimensionen von den Reflektierenden abgedeckt wurden. Meist wird die inhaltliche Breite der Reflexionen aber eher als zusätzliche Dimension in Kombination mit der Reflexionstiefe untersucht, sodass sich hier Mischmodelle ergeben. Im aktuellen Kompetenzdiskurs wird vermehrt davon gesprochen, dass die Reflexionsperformanz der Proband*innen erhoben werden müsse, um auf deren Reflexionskompetenz schließen zu können (Leonhard et al., 2010; Stender et al., 2021). Zurückgegriffen wird dabei aber immer noch auf die gleichen textanalytischen Verfahren wie bereits in den Jahrzehnten zuvor, sodass nicht von einer Veränderung der Messmethoden im Zuge der Kompetenzdebatte gesprochen werden kann.

Im Folgenden werden einige inhaltsanalytische Messverfahren vorgestellt und dabei zunächst auf Messungen der Reflexionstiefe eingegangen. Dabei wird beispielhaft das Stufenmodell von Hatton und Smith (1995) als typisches und häufig auch von anderen Forscher*innen eingesetztes Modell ausführlicher beleuchtet. Anschließend wird auf die Messung der Reflexionsbreite und auf Mischmodelle eingegangen. In Tabelle 2 findet sich ein Überblick über die vorgestellten Messinstrumente.

Tabelle 2. Überblick über die vorgestellten qualitativen Messinstrumente

Autor*innen/Jahr	Maßeinheiten	Reflexionstiefe	Reflexionsbreite
Van Manen (1977)	Drei Stufen	x	-
Sparks-Langer, Simmons, Pasch & Colton (1990)	Sieben Levels	x	-
Jay & Johnson (2002)	Drei Stufen	x	-
Kember, McKay, Sinclair & Wong (2008)	Vier Stufen	x	-
Weinberger & Seyfried (2009)	Vier Dimensionen	x	-
Hatton & Smith (1995)	Vier Stufen	x	-
Davis (2006)	Vier Kategorien	-	x
Luttenberg & Bergen (2008)	Je zwei Stufen in vier inhaltlichen Kategorien	x	x
Gödeke (2020)	Je vier Stufen in fünf inhaltlichen Kategorien	x	x
Bain, Ballentyne Packer & Mills (1999)	Je fünf Levels in vier inhaltlichen Kategorien	x	x
Artmann, Herzmann, Hoffmann & Proske (2013)	Vier Stufen und fünf Wissenstraditionen	x	x
Leonhard & Rihm (2011)	Vier Stufen und fünf Inhaltsdimensionen	x	x

Messung der Reflexionstiefe. Die ersten Versuche, Reflexion in verschiedene Arten einzuteilen, die in der Folge dann häufig auch hierarchisch geordnet wurden,

wurden deduktiv aus der Theorie abgeleitet und meist nicht zur Textanalyse eingesetzt. Als ein solcher Versuch, Reflexion in Levels einzuteilen, kann der Ansatz Van Manens (1977) gelten, der drei Ebenen des Umgangs mit Praxis unterscheidet. Auf der untersten Ebene stehen die Mittel und Wege, ein erwünschtes Ziel zu erreichen, im Zentrum und dafür wird theoretisches Wissen auf eine technologische Weise eingesetzt, ohne verschiedene Positionen zu hinterfragen. Auf der zweiten Ebene werden Wahrnehmungen, Vorurteile und Meinungen, die pädagogische Entscheidungen beeinflussen, in den Fokus gerückt und analysiert. Auf der dritten Ebene werden soziale, politische und ethische Faktoren in die Überlegungen mit einbezogen und das Wissen selbst hinterfragt. Als theoretische Basis für diese Stufeneinteilung bezieht sich der Autor in erster Linie auf Habermas.

Aus Van Manens drei Stufen, Gagnés „Hierarchy of Learning“ (1968) und Kolbs „Experiential Learning Theory“ (1984) leiten wiederum Sparks-Langer, Simmons, Pasch und Colton (1990) in ihrem „Framework for Reflective Thinking“ ebenfalls rein deduktiv sieben Levels der Reflexion ab („1: No descriptive language, 2: Simple layperson description, 3: Events labeled with appropriate terms, 4: Explanation with tradition or personal preferences given as the rationale, 5: Explanation with principle or theory given as the rationale, 6: Explanation with principle/theory and consideration of context factors, 7: Explanation with consideration of ethical, moral, political issues“, S. 27). Diese Stufen wurden von den Autor*innen zur Textanalyse eingesetzt. An der Benennung der Stufen lässt sich bereits ablesen, dass in diesen nicht nur die sprachliche Darstellung in Form eines sophistizierteren Schreibstils, sondern auch der Einbezug von Theorie eine wesentliche Rolle spielt und damit ein anderer Fokus gesetzt wurde, als es beispielsweise noch bei Van Manen der Fall war, der sich eher auf den Grad der Distanzierung im Sinne des Einnehmens einer Metaebene bezog.

Während die drei Stufen von Jay & Johnson (2002) auf den ersten Blick an Van Manen erinnern, zeigt sich jedoch, dass diese wiederum etwas anderes in den Fokus nehmen. Die drei Stufen werden beschrieben als „descriptive stage“ (Beschreibung), „comparative stage“ (Vergleich und Einbezug anderer Sichtweisen, z.B. auch aus der Literatur) und „critical stage“ (Abwägung unterschiedlicher Alterna-

tiven für zukünftiges Handeln, Integration von neuen Erkenntnissen in den bisherigen Wissensstand). Diese Stufen wurden auch von anderen Forscher*innen eingesetzt (Kaasila & Lauriala, 2012; Körkkö et al., 2016).

Kember, McKay, Sinclair und Wong erarbeiteten ebenfalls ein Instrument für schriftliche Reflexionen mit vier Stufen. Diese sind: „habitual action/non-reflection“, „understanding“, „reflection“ und „critical reflection“ (Kember et al., 2008). Hier wird die Stufe „reflection“ verstanden als Zusammenführung von Erfahrung bzw. Beobachtung und Theorie. „Critical reflection“ beinhaltet einen Perspektivenwechsel. Im Gegensatz zu den vorher dargestellten Stufenmodellen wurden mit diesem Schema keine einzelnen Kodiereinheiten unterschiedlicher Länge bewertet, sondern ein gesamter Text eingestuft. Auch die Stufen von Kember et al. wurden von anderen Wissenschaftler*innen für die Textanalyse übernommen (L. Andersen & Matkins, 2011).

Nicht nur ein Stufenschema allein, sondern ein gesamtes Instrument zur Erhebung der Reflexionstiefe einschließlich Fallvignetten und Aufgabenstellung wurde beispielsweise von Weinberger und Seyfried (2009) entwickelt. Solche Instrumente böten die Chance, eine höhere Vergleichbarkeit zwischen Studien herzustellen. Es handelt sich um das *Reflection Instrument for Education (RIfe)*, welches aus sechs Bildvignetten in Form von Fotos von Unterrichtssituationen besteht. Die Instruktion lautet, zu jedem Bild alles aufzuschreiben, was einem dazu in den Sinn kommt (S. 89), wobei die Schreibzeit zu jedem Bild in der Studie auf sieben Minuten begrenzt war. Die so entstandenen Texte werden inhaltsanalytisch mit Hilfe eines von den Autoren in einer früheren Studie entwickelten Reflexionsmodells ausgewertet. Es besteht aus vier Dimensionen des reflexiven Schreibens, in das die Textpassagen bei der Analyse eingeordnet werden. Diese vier sind erstens „descriptive statements“, also Situationsbeschreibungen, zweitens „interpretations“, also Aussagen zur Bedeutung oder Intention von Handlungen, drittens „interrogative statements“, also das Formulieren von Fragen an die Situation und viertens „explanative statements“, also das Aufstellen von Hypothesen und das Finden von Erklärungen. Für zwei der Dimensionen gibt es eine Untergliederung, und zwar für „interpretations“ und „explanative statements“. Bei diesen beiden wird noch unterschieden zwischen Aussagen mit und ohne „observer-awareness“, also einem Bewusstsein darüber,

dass man sich in einer Beobachterposition befindet, die keine Aussagen mit Abso-
luthet erlaubt, immer nur subjektiv sein kann und von einem eingeschränkten Wis-
sen über die Situation gekennzeichnet ist. Dieses Beobachterbewusstsein wird als
Zeichen von Reflexion gewertet. Es gelang, mit dem Instrument signifikant zwi-
schen zwei Interventions- und einer Kontrollgruppe zu unterscheiden, die Retest-
Reliabilität war aber nicht zufriedenstellend, woraus die Autoren schließen: „Re-
flexion is not a time stable trait“ (S. 93). Sie ziehen außerdem das Fazit, dass das
Instrument weiter überarbeitet werden müsse, es sich aber in Ansätzen bereits als
reliabel und valide herausgestellt habe.

Es wurden noch einige weitere Stufenmodelle entwickelt, die aber hier nicht mehr
im Detail vorgestellt werden sollen, weil sie im Wesentlichen dem gleichen Prinzip
folgen (Blomberg et al., 2014; El-Dib, 2007; Etscheidt et al., 2012; Larrivee, 2008).

Exemplarisch für die Messung der Reflexionstiefe wird zum Abschluss noch das
Stufenschema von Hatton und Smith (1995) genauer vorgestellt, denn auch wenn
sich bisher keines der Stufenmodelle durchsetzen konnte, so wurde das der beiden
Autoren am häufigsten von anderen Forscher*innen verwendet und überarbeitet.

Hatton und Smith entwickelten in ihrer Studie teils induktiv, teils theoriegestützt
aus $N = 94$ Reflexionstexten („written reports“ zur Planung, Implementierung und
Evaluation eines Unterrichtsversuchs im Praktikum) ein vierstufiges Schema zur
Einschätzung des Reflexionsniveaus einzelner Kodiereinheiten. Die unterste Kate-
gorie stellt dabei die Stufe „descriptive writing“ dar, die Beschreibungen der Situ-
ation oder Äußerungen der eigenen Meinung umfasst, ohne dass Begründungen,
Literaturbezüge oder Erläuterungen geliefert werden. Die zweite Stufe nennen die
Autoren „descriptive reflection“. Hierzu gehören die oben beschriebenen Äußerun-
gen, sofern eine Begründung gegeben wird oder Bezüge zu alternativen Sicht-
weisen in der Fachliteratur hergestellt werden. Die dritte Stufe ist die „dialogic re-
flexion“, die zeigt, dass sich eine Person bewusst von der Situation distanzieren
kann und in einen Diskurs mit sich selbst geht. Hierunter werden analytische, er-
klärende und integrative Überlegungen zu einem Aspekt zusammengefasst. Von
den Reflektierenden werden Hypothesen aufgestellt, evtl. auch Inkonsistenzen im
eigenen Denken und Handeln aufgedeckt oder verschiedene Perspektiven beleuch-
tet und zusammengeführt. Die höchste Stufe ist die „critical reflection“, die ein Be-
wusstsein dafür zeigt, dass die beschriebenen Situationen und Handlungen Teil von

historischen, sozio-politischen Kontexten und nur aus verschiedenen Perspektiven beschreib- und erfassbar sind. Die zweite und dritte Stufe sind jeweils noch in zwei Unterkategorien untergliedert, nämlich „reflection based on one perspective/factor as a rationale“ und „reflection based on the recognition of multiple factors and perspectives“. Zu jeder Kategorie wird außerdem ein Ankerbeispiel gegeben, außer zur ersten Kategorie, „descriptive writing“. Die Zuteilung zu den einzelnen Kategorien erfolgt nach eher formal orientierten Kriterien. Die Autoren geben beispielsweise als Kodierhilfe verschiedene Formulierungen (z.B. „On the one hand, ..., yet on the other...“ als typisch für „dialogic reflection“) vor. Hatton und Smith nennen als Limitation an diesem Vorgehen, dass die stark formalen Kriterien, die eine Stufe wie beispielsweise „dialogic reflection“ klassifizieren, dazu führen können, dass reflexive Aussagen, die dem Schema nicht entsprechen, übersehen werden. Gleichzeitig kann es passieren, dass Aussagen, die den formalen Kriterien entsprechen, als „dialogic reflection“ eingeordnet werden, aber inhaltlich betrachtet eigentlich nicht den Anforderungen an eine Reflexion entsprechen: „It becomes problematic only if what is being classified as reflective is not in fact reflection, and if what at present is being passed over is in fact reflective, but goes unrecognised“ (S. 42). Über die Festlegung von Kodiereinheiten sowie Interraterreliabilitäten wurde nicht berichtet. Eine Einigung über die Kodierung erfolgte, soweit es der Studie zu entnehmen ist, konsensuell. Ergebnisse der Studie werden in Kapitel 2.2.2.1 dargestellt.

Das Stufenschema von Hatton und Smith wurde verschiedenen späteren Studien zugrunde gelegt, die dieses entweder im Original übernommen oder angepasst haben. So benutzten Penso, Shoham und Shiloah (2001) die vier Hauptstufen zur Analyse und Kategorisierung von Reflexionstexten, ohne jedoch die Untergliederungen der Stufen zu übernehmen oder zusätzliche Angaben im Sinne eigener Ankerbeispiele oder individueller Beschreibung der Stufen zu machen. Als Länge der Analyseeinheiten, in die die Reflexionstexte zerlegt wurden, wurden einer bis mehrere Sätze angegeben, eine Einigung über die Zuordnung zu den vier Stufen erfolgte über Konsensvalidierung zwischen zwei Ratern (vgl. S. 328). Die Ergebnisse waren denen der Studie von Hatton und Smith sehr ähnlich. Der Einsatz der Stufen von Hatton und Smith wurde hier und auch in einigen weiteren Studien kombiniert mit

der Erhebung der Reflexionsbreite, auf die an späterer Stelle genauer eingegangen wird.

Auch in der Studie von Poom-Valickis und Mathews (2013) wurde eine Analyse mit den vier Hauptstufen nach Hatton und Smith durchgeführt und kombiniert mit einer induktiven Kategorisierung der gefundenen Sinneinheiten nach Themen. Interessant ist, dass hier auch eine Quantifizierung vorgenommen wurde, indem den Stufen Punkte von 1-4 zugewiesen wurden, wodurch es möglich war, einen Durchschnitt der gefundenen Reflexionsstufen anzugeben. Dies geschah allerdings nicht auf individueller, sondern auf Gruppenebene.

Leonhard und Rihm (2011) dagegen legten das Schema zwar dem ihren zugrunde, versuchten aber durch eine weitere Ausgliederung der Stufen ein noch differenzierteres Bild der Reflexionsfähigkeit zu entwerfen. Die zusätzliche Untergliederung zeigte sich allerdings hinsichtlich der Reliabilität als problematisch, da es den Ratern nicht gelang, eine zufriedenstellende Übereinstimmung zu erreichen.

Artmann et al. (2013) entschieden sich genau wie Leonhard und Rihm dafür, die Stufen weiter zu untergliedern und begründen, dass sich ansonsten aufgrund des häufigen Fehlens der höheren Reflexionsstufen kein auflösungsstarkes Bild der Reflexionsfähigkeit ermitteln lasse (vgl. S. 139). Sie kritisieren außerdem am Modell von Hatton und Smith die „fehlende Unterscheidung zwischen subjektiven und wissenschaftlichen Bezugsrahmen“, die sie für eine in diesem Kontext bedeutsame Kategorie halten (ebd.).

Lüsebrink und Grimminger (2014) basieren ihre „Stufenmodellierung unterrichtsbezogener Reflexionskompetenz“ nach eigenen Angaben ebenfalls auf den Stufen von Hatton und Smith (1995), differenzieren die Stufen allerdings noch wesentlich stärker aus. Die Benennungen und Beschreibungen der endgültigen Stufen lassen zum ursprünglichen Modell kaum mehr inhaltliche Ähnlichkeit erkennen. Auch sie berichten von gravierenden Problemen, bei der Kodierung zu eindeutigen Zuordnungen zu gelangen. Ein weiteres Problem bestand darin, dass häufig sehr niedrige neben sehr hohen Stufen in den Texten vorhanden waren, ohne dass sich die Zwischenstufen gezeigt hätten.

Ergänzung der Messung um die Reflexionsbreite. In den Anfangsjahren der Reflexionsforschung wurden primär Messungen der Reflexionstiefe versucht. Eine Fokussierung auf die Reflexionsbreite erfolgte erst in späterer Zeit und findet sich vereinzelt als alleinige Untersuchungsform, meistens aber in Verbindung mit der Messung der Reflexionstiefe. Ihren Ausgang findet sie bei LaBoskey (1993), die dafür plädiert, nicht auf eine hierarchische Ordnung von Levels zu fokussieren, weil es so zu einer nicht wünschenswerten Auf- bzw. Abwertung bestimmter Aspekte von Reflexion komme. Sie regt stattdessen dazu an, inhaltliche Kategorien so zu unterscheiden, „that the categories be referred to as potential foci or content of reflection rather than levels“ (S. 26).

Solche primär inhaltlich-thematisch arbeitenden Untersuchungen von Reflexionen gibt es wenige (Davis, 2006; Mauri et al., 2017; Moore-Russo & Wilsey, 2014). Davis (2006) untersuchte beispielsweise Tagebucheinträge von Studierenden nach den inhaltlichen Kategorien „learners and learning“, „subject matter knowledge“, „assessment“ und „instruction“. Das Schema von Davis wurde auch bei Moore-Russo und Wilsey (2014) eingesetzt und damit zum Beispiel den Forschungsfragen nachgegangen, welche Aspekte Studierende von sich aus thematisieren und ob sie in der Lage sind, „a complex view of teaching“ (S. 78) zu entwickeln. Das heißt, die Betrachtung mehrerer verschiedener Themen und Felder wurde interpretiert als die Fähigkeit, Unterricht in seiner Multidimensionalität und Komplexität wahrzunehmen. Beachtenswert ist die Bilanz dieser Studie, die hinsichtlich der Vergleichbarkeit mit der Originalstudie von Davis (2006) gezogen wird: „[I]t would be short sighted to compare the results of this study directly to [...] Davis“ (S. 87). Obwohl sogar das gleiche Messinstrument verwendet wurde, unterscheiden sich die Analyseeinheiten, die Länge und Art der Reflexionstexte und die Themen der Reflexionen so stark voneinander, dass die Ergebnisse nicht verglichen werden können.

Was sich insgesamt in diesen Ansätzen widerspiegelt, ist die Idee, dass man auch anhand der Reflexionsbreite eine Einschätzung der Reflexionsgüte vornehmen könne. Leonhard, Nagel, Rihm, Strittmeier-Haubold und Wengert-Richter (2010) definieren den Begriff wie folgt: „Reflexionsbreite beschreibt, welche Aspekte eine Lehrperson in die Vergegenwärtigung einer Situation einbezieht“ (S. 114). Die Autor*innen sehen in „einer[r] zunehmende[n] Vielfalt von (plausiblen) Aspekten“

(ebd.), die in einer Reflexion bedacht werden, eine Zunahme der Reflexionsbreite und damit auch eine Steigerung der Reflexionsgüte.

Luttenberg und Bergen (2008) konstatieren, dass Breite und Tiefe nicht getrennt werden können. Sie führen die beiden Dimensionen zusammen in einer Typologie von Reflexion, indem sie drei inhaltliche Kategorien definieren, nämlich „pragmatisch“, „ethisch“ und „moralisch“, und innerhalb dieser Kategorien jeweils unterscheiden, ob eine offene oder geschlossene Herangehensweise in der Reflexion vorliegt. Bei Gödecke (2020) wurden ebenfalls inhaltliche Kategorien festgesetzt („Auslöser von Reflexion“, „Auswahl des Schwerpunkts“, „Vorwissen“, „Vorannahmen und Überzeugungen“, „Motive und Ziele“) und innerhalb derer dann die Qualität mittels Stufen beurteilt (0 = Keine Reflexion, 1 = Geschlossene Reflexion, 2 = Halboffene Reflexion, 3 = Offene Reflexion). Eine sehr ähnliche Herangehensweise findet sich auch bei Risiko (2009) und bei Gutzwiller-Helfenfinger et al. (2017). Letztere analysierten Texte mit ihrem Instrument SERQU, das verschiedene Inhaltskategorien aus dem Reflexionsmodell EDAMA (Aeppli & Lötscher, 2017) ableitet. Eine etwas andere Herangehensweise wählten beispielsweise Bain, Ballentyne, Packer und Mills (1999), die zwar ebenfalls die Qualität der Reflexion innerhalb verschiedener thematischer Schwerpunkte beurteilten, aber dafür ein eigenes Stufenmodell entwickelten, das sich über fünf Levels erstreckt („reporting“, „responding“, „relating“, „reasoning“ und „reconstructing“). Diese Reflexionskala wurde auch von anderen Forscher*innen übernommen und weiterbearbeitet (P. Roberts, 2016; Turhan & Kirkgoz, 2018).

Die bisher vorgestellten Ansätze haben eine Verknüpfung von Reflexionsbreite und -tiefe gemeinsam, insofern bei ihnen innerhalb verschiedener inhaltlicher Aspekte die Qualität der Ausführungen mittels Stufen ermittelt wird. Ein zweiter Ansatz besteht darin, die Reflexionstiefe und -breite getrennt voneinander als zwei verschiedene Kategorien der Textqualität zu erfassen.

Das geschah beispielsweise bei Artmann et al. (2013), die sowohl eine inhaltliche Dimensionierung entlang von fünf Wissenstraditionen („academic tradition“, „developmentalist tradition“, „social efficiency tradition“, „social reconstructionist tradition“, „generic tradition“) als auch eine Graduierung mit dem Stufenmodell von Hatton und Smith (1995) durchführten und sogar noch um eine Perspektive erweiterten, die sich mit den Auslösemechanismen von Reflexion befasst (vgl. S. 142).

Von Leonhard und Rihm (2011) wurde ein vollständiges Instrument bestehend aus Aufgabenstellungen, Vignetten und Auswertungsschema entwickelt. Es besteht aus sechs Situationsvignetten in Text bzw. Bildform, je zwei davon bezogen auf die Kompetenzbereiche „Bewerten“, „Unterrichten“ und „Erziehen“. Die Texte schildern fiktive dilemmatische Situationen aus dem Schulkontext, zu denen jeweils Leitfragen gestellt wurden. Eine Auswertung der Texte erfolgte mittels qualitativer Inhaltsanalyse sowohl hinsichtlich Reflexionstiefe als auch -breite. Das Instrument wurde oben bereits erwähnt, da es hinsichtlich der Reflexionstiefe mit den Stufen von Hatton und Smith (1995) arbeitet. Zusätzlich wurden fünf verschiedene Inhaltsdimensionen untersucht, die sich auf einer Skala von einem engen Individualfokus über den Klassenfokus und Schulfokus bis zur Betrachtung der Gesellschaft als Ganzes immer stärker weiten. Die Autoren bilanzieren, dass das Instrument zwar grundsätzlich in der Lage war, zwischen den beiden Messzeitpunkten und unterschiedlichen Reflexionskompetenzlevels zu unterscheiden, sich allerdings im Entwicklungs- und Auswertungsprozess erhebliche Schwierigkeiten ergeben hätten, die deutlich machten, dass eine Weiterentwicklung des Instruments vor einem weiteren Einsatz unbedingt nötig sei.

Auch bei den Messungen mit Mischformen aus Reflexionstiefe und -breite gäbe es noch eine Vielzahl verschiedener Studien zu berichten, die alle nach einem ähnlichen Prinzip, aber mit unterschiedlichen Stufungen und inhaltlichen Schwerpunkten arbeiten (Fund et al., 2002; Hayden & Chiu, 2015; Lee, 2005; Leijen et al., 2012; Ward & McCotter, 2004). Es wurde aber bereits anhand der vorgestellten Modelle deutlich, wie unterschiedlich und wenig vergleichbar diese sind und wie viele verschiedene Ansätze in Abhängigkeit von den Forschungsfragen, Untersuchungsmethoden, der Stichprobe und dem Datenmaterial verfolgt wurden.

Quantifizierung. Die meisten qualitativen Instrumente bzw. Analysemethoden ermitteln keine Reflexionsfähigkeit oder -leistung auf individueller Ebene, sondern auf der Ebene von Teilgruppen oder sie beschreiben die gesamte Stichprobe. Dabei beschränken sie sich größtenteils auf deskriptive Auswertungen. Meist werden Auftretenshäufigkeiten, prozentuale Verteilungen oder Mittelwerte der verschiedenen Stufen oder inhaltlichen Felder berichtet. Die Länge der Kodiereinheiten unterscheidet sich von Studie zu Studie stark von der Untersuchung auf Satzebene über

verschieden lange Sinneinheiten bis hin zur Kategorisierung des ganzen Textes ohne Verwendung kleinerer Analyseeinheiten. Auch die Länge und Anzahl der untersuchten Texte unterscheidet sich naturgemäß stark zwischen den einzelnen Studien und die Spannweite ist sogar innerhalb einzelner Untersuchungen zum Teil sehr hoch.

Gerade in neueren Studien gibt es aber durchaus auch Bemühungen, qualitative Analysen so anzulegen, dass sich Reflexionscores auf individueller Ebene ermitteln lassen. Einen sehr interessanten Ansatz entwickelten beispielsweise Kempin, Kulgemeyer und Schecker (2017, 2019), die einen Online-Reflexionsperformanztest einsetzten, der auf Video-Vignetten basiert. Der Test hat als Rahmenthema die fiktive Situation, dass ein Mitpraktikant Feedback zu seinem Unterricht erbittet. Einzelne Videoausschnitte seines Unterrichts mit Impulsfragen (z.B. „Was sagst du zu meinem Unterrichtseinstieg?“ 2017, S. 750) werden präsentiert und die Antworten der Proband*innen über Audioaufnahmen erfasst, um, wie die Autor*innen erläutern, Schreibhemmungen durch diese Variante des fingierten Gesprächs zu vermeiden. Zu kommentieren sind auf diese Weise 54 Reflexionsanlässe zu einzelnen Ausschnitten des Unterrichts und 12 Reflexionsanlässe zur Stunde als Ganzes. Die Aufnahmen werden anschließend ausgewertet, indem sie einer von sieben Stufen der Reflexionskompetenz zugewiesen und entsprechend bepunktet werden. Diese sind: 0 = Nicht angesprochen, 1 = Beschreibung, 2 = Bewertung, 3 = Begründete Bewertung, 4 = Alternative, 5 = Begründete Alternative, 6 = Konsequenz, 7 = Begründete Konsequenz (2019, S. 440). Durch das Bilden einer Summe aus den einzelnen Bewertungen kann ein Gesamtscore ermittelt werden. Solche Bemühungen, einen Gesamtscore der individuellen Reflexionsfähigkeit zu ermitteln, sind aber noch sehr selten vorzufinden.

2.2.1.3 Weitere Messversuche

In diesem Abschnitt werden noch einige weitere Methoden beleuchtet, anhand derer Versuche unternommen wurden, Reflexion zu messen.

Für die Analyse von Texten wurden vereinzelt auch andere Methoden als die qualitative Inhaltsanalyse eingesetzt. So fanden sich beispielsweise die Verwendung

der Dokumentarischen Methode nach Bohnsack (z.B. bei Hinzke, 2020), die Methode des permanenten Vergleichs, einer Arbeitsform der Grounded Theory (z.B. bei Braun & Crumpler, 2004) oder automatisierte digital gestützte Auswertungsverfahren. Bei letzteren wurden Methoden eingesetzt wie die Semantic Networks Technique (Lew & Schmidt, 2011), die Latent Semantic Analysis (Cheng & Chau, 2013) oder es wurde mit Machine-Learning gearbeitet, wie in der Studie von Ullmann (2019). Bei dieser wurden mittels eines selbstlernenden Programms Textpassagen in fünf Inhaltskategorien einsortiert und in zwei Qualitätsstufen eingeordnet (reflexiv vs. nicht-reflexiv). Es arbeitete mit einer Genauigkeit, die durchschnittlich um zehn Prozent geringer war als bei einer manuellen menschlichen Auswertung. Die Analyse orientierte sich an vorher definierten Schlagworten. Noch ist das Verfahren nicht ausgereift genug, aber es ist ein vielversprechender Forschungsansatz, der vor allem dann bessere Ergebnisse liefern könnte, wenn größere Datensätze zur Verfügung stehen, um die Algorithmen zu trainieren.

Abgesehen von schriftlichen Reflexionsdokumenten gibt es aber auch noch andere Ansätze. Behr und Welzel (2009) beispielsweise führten eine Studie durch, in der sie Reflexionskompetenz durch Aufzeichnung mittels Video messen wollten. Sie zeichneten Reflexionsphasen von Proband*innen mithilfe von Video auf und analysierten das Bildmaterial kategoriengeleitet. Vier Kategorien wurden herausgearbeitet, und zwar Sprechakte, Reflexionsinhalte, Reflexionsstufen und Sensibilisierung. Die Reflexionsstufen wurden dabei in Anlehnung an das Kompetenzstufenmodell von PISA entwickelt, eine Beschreibung derselben fehlt aber, sodass diese nicht nachvollzogen werden können. Eine Auswertung erfolgte nicht individuell für die Teilnehmer*innen, sondern als Summenparameter (vgl. S. 120). Zwar gelang es den Autor*innen zufolge, mit dieser Methode qualitative Unterschiede der Reflexionen zwischen vier Seminaren einer Fortbildungsreihe herauszuarbeiten, wie diese Ergebnisse genau gewonnen wurden, ist aber kaum nachvollziehbar.

Es gibt auch Versuche, Reflexion mithilfe physiologischer Methoden direkt dort zu messen, wo sie passiert: im Gehirn. In einer Studie von Johnson, Baxter, Wilder Heiserman und Prigatano (2002) wurden Proband*innen beispielsweise mithilfe von Magnetresonanztomographie überwacht, während sie reflektierten bzw. in der Kontrollgruppe einfache Wissensfragen beantworteten. Die Analyse ergab, dass an Reflexionsprozessen wahrscheinlich verschiedene Hirnareale beteiligt sind.

Schwerpunkte fanden sie aber im medialen präfrontalen und posterioren cingulären Cortex, was sich den Autor*innen zufolge auch mit Erkenntnissen aus der Läsionsforschung deckt, sodass sie davon ausgehen, damit die Einheiten identifiziert zu haben, die für Reflexion primär verantwortlich sind. Moran (2016) hingegen beschreibt, Selbstreflexion finde in teilweise anderen Strukturen statt, in denen das Selbstbild verortet sei.

Ein etwas anderer Versuch, Reflexion auf physiologischer Ebene sichtbar zu machen, wurde von Rominger, Reitinger, Seyfried, Schneckenleitner und Fink (2017) unternommen, die eine elektroenzephalographische Überwachung (EEG) dafür nutzten, Reflexionen bei Lehrkräften nachzuweisen. Sie fanden eine höhere Aktivität der Alpha-Wellen beim Reflektieren, schränken aber ihr Ergebnis dahingehend ein, dass derartige Veränderungen auch bei anderen kognitiven Prozessen denkbar sind, also nicht unbedingt reflexionsspezifisch auftreten. Außerdem ließen sich die Ergebnisse nicht durch eine Intervention zur Steigerung der Reflexionsfähigkeit verändern.

2.2.1.4 Zusammenfassung und Bewertung

Es existieren zahlreiche Fragebögen, die entweder Reflexion insgesamt, Reflexion in unterschiedlichen Kontexten oder unterschiedliche Teilaspekte derselben zu messen versuchen. Sie wurden mit unterschiedlichen Verfahren und vor allem auch unterschiedlicher Qualität entwickelt und getestet. Gemeinsam ist ihnen die Kritik, dass sie Reflexion, ja teilweise sogar Reflexionskompetenz, meist über Selbsteinschätzungen zu erheben versuchen und sich damit auf einer sehr niedrigen Ebene der Zuverlässigkeit bewegen. Es wurde herausgearbeitet, dass speziell bei einem Thema wie Reflexion die Zuverlässigkeit von Selbstberichten noch mehr als sonst beeinträchtigt sein könnte (Filipp & Freudenberg, 1989, S. 7). Während sich mit diesen Verfahren weitgehend reliable Ergebnisse erzielen ließen, gibt es erheblichen Zweifel, ob die vorgestellten Fragebögen das wichtigere Kriterium der Validität erfüllen können. Die Operationalisierung von Reflexion zeigt viele Überschneidungen zu anderen Konstrukten, sodass oft nicht klar ist, was mit den Fragebögen überhaupt genau gemessen wird. Für Vergleiche großer Gruppen oder der

Einschätzung der Entwicklung von Kohorten gibt es allerdings kaum Alternativen, weil hier qualitative Instrumente zu aufwändig sind und an ihre Grenzen stoßen.

Alternativ zu diesen rein quantitativen Messmethoden wurde als zweites großes Feld in der Reflexionsmessung die primär qualitative inhaltsanalytische Vorgehensweise vorgestellt, die sich in erster Linie auf Reflexionstexte als Korpus stützt. Es gibt zwei Dimensionen, die hier betrachtet wurden, und zwar einerseits die Reflexionstiefe und andererseits die Reflexionsbreite.

Die Messung der Reflexionstiefe basiert auf der Vorstellung einer Abstufung von Reflexionsqualität, die an Kompetenzniveau Modelle erinnert (vgl. Hartig & Klieme, 2006). Solche Modelle von Stufen oder Levels wurden teils deduktiv aus der Theorie, teils induktiv aus den Reflexionstexten oder in Mischformen entwickelt und zur Textanalyse herangezogen. Versuche, vorhandene Stufenmodelle weiter auszudifferenzieren, führen meist zu Problemen, vor allem bezüglich der Interraterreliabilität. Über diese Probleme berichten zahlreiche Studien (Leonhard & Rihm, 2011; Lüsebrink & Grimminger, 2014; Sumsion & Fleet, 1996). In den meisten Studien wurde deshalb konsensuell kodiert und auf die Berechnung eines Reliabilitätsmaßes verzichtet. Zudem bilden die vorgenommenen Ausdifferenzierungen meist die eigenen Forschungsfragen und das individuelle Material ab und sind nicht immer theoretisch gestützt. Auch wenn die Stufen nicht stark ausdifferenziert werden, sind sie häufig nicht völlig disjunkt und die Zuordnung hängt von der Interpretation der Aussagen ab, sodass es sich nicht um exakte Messinstrumente handeln kann: „Such a scheme will obviously not provide a precise measurement, but will provide guidance in making judgements“ (Kember et al., 2008, S. 371). Es handelt sich dabei also eher um „einigermaßen plausible, analytisch brauchbare Ordnungen kognitiver Teiltätigkeiten“ (Berndt & Häcker, 2017, S. 250). Der Einsatz solcher Stufenmodelle ist über die Reliabilitätsprobleme hinaus dem Vorwurf ausgesetzt, dass noch gar nicht abschließend geklärt ist, ob Reflexionsfähigkeit überhaupt hierarchisch unterscheidbaren Schritten folgt (van Beveren et al., 2018, S. 2). Selbst wenn man davon ausgeht, dass Reflexionskompetenz, wie andere Kompetenzen auch, über Entwicklungsstufen beschrieben werden könnte, so ist noch nicht klar, wie sich diese „höhere“ Reflexionsleistung jeweils definiert und an welchen Kriterien sie festzumachen ist: „Multiple expressions indicate that the goal

for teachers is to reflect at a ‚higher‘ or ‚deeper‘ level, but there is a lack of description as to what this means or how to determine if it has occurred“ (Moore-Russo & Wilsey, 2014, S. 77).

Bei der Erhebung der Reflexionsbreite wird davon ausgegangen, dass Reflexionen hochwertiger sind, wenn möglichst viele thematische Ebenen bedacht wurden. Auch hier zeigt sich, ähnlich wie bei der Reflexionstiefe, eine große Varianz bei der Wahl der inhaltlichen Schwerpunkte, die eine Vergleichbarkeit der Studien kaum ermöglicht. Die Identifikation inhaltlicher Themen und die dadurch ermöglichte genauere Beschreibung von Reflexionen kann sicherlich einige interessante Forschungsfragen beantworten. Die Ableitung einer Graduierung der Reflexionsqualität aus der thematischen Breite fußt allerdings auf keiner theoretischen Basis. Ob Reflexionen erfolgreicher oder gewinnbringender sind, wenn sie bestimmte oder möglichst viele inhaltliche Bereiche abdecken, konnte noch nicht gezeigt werden. Der Ansatz steht in keiner Verbindung zu den Grundlagen – beispielsweise Dewey oder Schön – aus dem ersten Teil der Arbeit.

Außerdem ist zu beobachten, dass in vielen Studien „zu wenig genau beschrieben [wird], wie Reflexion gemessen wurde, sodass schwer nachvollzogen werden kann, was deren Ergebnisse bedeuten“ (Gutzwiller-Helfenfinger et al., 2017, S. 134). In den seltensten Fällen sind die Auswertungsverfahren intersubjektiv so nachvollziehbar darstellbar, sodass ein Versuch unternommen werden könnte, die Ergebnisse zu reproduzieren.

Es existieren also sowohl hinsichtlich der Reflexionstiefe als auch der Reflexionsbreite viele verschiedene Modelle gleichwertig nebeneinander und viele Mischmodelle, in denen Verfahren aus beiden Kategorien kombiniert wurden. Keines konnte sich bisher als eines herauskristalisieren, das Reflexion in den Augen der meisten Wissenschaftler*innen valide und reliabel abbildet, wobei sich bei den textanalytischen Verfahren primär die Reliabilität als herausfordernd herausstellt. Die Existenz dieser vielen verschiedenen Stufenmodelle zeigt, dass das Konstrukt Reflexion noch nicht ausreichend operationalisiert werden konnte. Leonhard und Rihm (2011) stellen fest, „dass die empirisch vorfindbare Komplexität reflexiver Aussagen sich in den Modellen bisher nicht hinreichend abbildet“ (S. 245). Boud (2010) sieht die qualitativen Verfahren ebenfalls als problematisch: „They took the language of reflection – elements, stages, whatever – and turned these into procedures which they

could easily identify as either being completed or not, as if we can ever tell when another person has reflected enough“ (S. 27f). Trotz dieser Kritik ist festzuhalten, dass inhaltsanalytische Verfahren im Gegensatz zu Fragebogenerhebungen immerhin auf der Ebene der Performanz ansetzen und damit möglicherweise ein genaueres Bild einer Kompetenz erzeugen könnten. Ähnlich wie Fragebögen haben aber auch sie das Problem, dass eine nicht genau identifizierte Nähe zu anderen Konstrukten, zum Beispiel „problem-solving and decision-making“ (Hatton & Smith, 1995, S. 43), besteht.

Die über die herkömmlichen Verfahren hinausgehenden weiteren Messversuche, wie Hirnstrom- oder EEG-Messung und automatisierte Textanalysen, die vor allem mithilfe moderner technischer Unterstützung durchgeführt werden, sind momentan noch nichts weiter als Vorstöße in verschiedene Richtungen, die erst noch Belege für ihre Zuverlässigkeit erbringen müssen und – wenngleich sie interessante Ansätze darstellen – aktuell noch nicht in der Breite einsetzbar sind.

Es gibt also verschiedene Arten von Messungen von eher traditionell bis modern über alle Forschungsparadigmen hinweg und keine von ihnen konnte sich bisher als „state-of-the-art“ herausbilden. Die im ersten Kapitel herausgearbeitete Komplexität des Konstrukts findet sich in den Messungen nicht wieder, die immer nur einzelne Teilaspekte in den Fokus nehmen. Die vorliegenden Instrumente sind einzeln nicht geeignet, um sich ein umfassendes und differenziertes Bild über die Reflexionskompetenz von Personen zu machen, und es gibt viele Punkte, die die Validität der einzelnen Messungen zweifelhaft machen. Die zentrale Frage ist nach wie vor unbeantwortet: „What, exactly, are we looking for as evidence of reflection?“ (Rodgers, 2002, S. 843).

2.2.2 Messergebnisse

2.2.2.1 Befunde zu Qualität und Häufigkeit von Reflexion

Bevor Befunde zu Qualität und Häufigkeit von Reflexion bei Lehramtsstudierenden dargestellt werden, wird zunächst noch auf Einstellungen zu Reflexion eingegangen.

Unabhängig von der tatsächlichen Qualität der Reflexionstexte ergeben Befragungen zur Einstellung von Studierenden und Lehrkräften größtenteils, dass sie Reflexion eher positiv und als relevant bewerten (Göbel et al., 2021; Raaflaub et al., 2019; Slepcevic-Zach & Stock, 2018; Turhan & Kirkgoz, 2018). Slepcevic-Zach und Stock (2018) fanden dabei auch noch einen kleinen Geschlechterunterschied, nämlich dass weibliche Teilnehmerinnen das Lernen durch (Selbst-)Reflexion wichtiger fanden und diesem offener gegenüberstanden als männliche Teilnehmer. Dem entgegen steht eine Untersuchung von Williams und Grudnoff, in der sowohl Studienanfänger*innen als auch erfahrene Lehrkräfte sich eher skeptisch gegenüber Reflexion äußerten. Nachdem sie allerdings in einer Intervention ein Reflexionsmodell kennengelernt und angewandt hatten, wurde auch in dieser Studie Reflexion wertschätzend als Mittel zu Analyse und Verbesserung von Praxis eingeschätzt (Williams & Grudnoff, 2011). Studierende scheinen sich eher kollegial durchgeführte Reflexionen zu wünschen, während Lehrkräfte in der Praxis diese als weniger wichtig beurteilen (Göbel et al., 2021).

Während Reflexion zwar überwiegend positiv und hilfreich eingeschätzt wird, gibt es hinsichtlich der tatsächlichen Qualität und Menge der Reflexionen Hinweise, dass sie eher selten praktiziert wird und in den meisten Fällen oberflächlich. Eine Studie von Wyss (2011) ergab, dass die befragten Lehrkräfte zwar laut eigener Selbsteinschätzung „ihren Unterricht bewusst und regelmäßig reflektieren“ (S. 253), dies aber nur alleine und in Gedanken, also eher unstrukturiert, täten und nur selten in kollegialen Settings. Es zeigte sich außerdem, dass keine theoretischen, moralischen oder gesellschaftlichen Aspekte thematisiert wurden und sich die Reflexionen auf der Mikroebene bewegten (vgl. S. 263). Solche kritischen Reflexionen erfolgen auch nach Hall (2018) extrem selten spontan, also ohne explizite Anregung. Dass bei der Auswertung von Reflexionen kritische Anteile oder moralische Erwägungen kaum oder sogar gar nicht identifiziert werden konnten, ergaben auch einige weitere Untersuchungen (Choy & Oo, 2012; El-Dib, 2007; Hatton & Smith, 1995). Dem gegenüber stehen nur sehr vereinzelte optimistischere Befunde wie beispielsweise die Studie von Andersen und Matkins (2011), laut der sogar 30% der Lehramtsstudierenden kritisch reflektierten. Eine Untersuchung von Mönig (2013) zeigte, dass “wissenschaftsorientierte Strategien vergleichsweise weniger genutzt werden, als Strategien, die sich auf den Rückgriff auf eigene Erfah-

rungen oder die Kooperation mit Kollegen beziehen” (S. 285). Die Einbindung wissenschaftlicher Bezüge erfolgte signifikant häufiger bei fachlichen Fragestellungen als bei allgemein-pädagogischen Problemen (vgl. ebd.).

Gutzwiller-Helfenfinger, Aepli und Lötscher (2017) kamen zu dem Ergebnis, dass sich in den ausgewerteten Reflexionstexten außerdem kaum multiperspektivische Betrachtungen fanden. Die Darstellungen waren „mehrheitlich wenig dicht und reichhaltig“ (S. 144). Aus einem Abgleich mit anderen Studien ziehen sie die Bilanz, dass die Fähigkeit des multiperspektivischen Denkens zwar vielleicht grundsätzlich vorhanden wäre, diese Fähigkeit aber nicht spontan zur Anwendung komme, sondern nur bei gezielter Aufforderung (vgl. S. 138). Artmann et al. (2013) kamen in ihrer Studie zu dem Ergebnis, dass sich „in allen Fällen das Bemühen [zeigte], der anfänglichen Frage mit einer klaren Antwort zu begegnen bzw. mit einer eindeutigen Gewissheit die eigene Reflexion zu Unterricht abzuschließen“, vorrangig mit dem Ziel „Handlungssicherheit zu erlangen“ (S. 144f). Göbel, Wyss, Neuber und Raaflaub (2021) stellten fest, dass die Reflexionen Studierender zu eigenen Unterrichtsstunden auf einer sehr oberflächlichen, offensichtlichen Ebene blieben und nicht in die Tiefe gingen. Auch wurden die Unterrichtsstunden sehr positiv wahrgenommen und kaum kritisch gesehen.

Insgesamt stellt sich das Reflexionsniveau von Studierenden also als eher niedrig dar, wie von den meisten Forscher*innen mit einer gewissen Enttäuschung festgestellt wird. Es wurden überwiegend niedrige Stufen bzw. Levels in den Texten identifiziert. Das war auch in der ausführlicher vorgestellten Studie von Hatton und Smith (1995) der Fall. In dieser wurde leider nicht berichtet, wie hoch der Anteil der nicht-reflexiven Textanteile („descriptive writing“) war, sondern nur die reflexiven Anteile aufgeschlüsselt. 60-70 % davon wurden als „descriptive reflection“ klassifiziert (die uneindeutigen Angaben ergeben sich aus verschiedenen Erhebungszeitpunkten), etwa 30 % als „dialogic reflection“ und nur acht Belege insgesamt als „critical reflection“ (vgl. S. 41). Auch in den ebenfalls bereits erwähnten Studien, die die Stufen von Hatton und Smith eingesetzt hatten, zeigten sich ähnliche Befunde. Diese berichten aber auch die Trefferzahl auf der untersten Stufe des „descriptive writing“, so beispielsweise die Studie von Penso, Shoham und Shiloah (2001): 29 % der Texte (in einer ersten Erhebung) bzw. 32 % (in einer zweiten Erhebung) entfielen auf nicht reflexive Anteile („descriptive writing“), 63 % bzw. 50 % auf „descriptive reflection“, 3 % bzw. 12 % auf „dialogic reflection“ und 5 %

bzw. 6 % auf „critical reflection“. Poom-Valickis und Mathews (2013) fanden sogar, dass die nicht reflexiven Anteile über die Hälfte (52 %) der Texte ausmachten, auch hier gefolgt von „descriptive reflection“ (41 %) und „dialogic reflection“ (7 %). In dieser Studie fanden sich keine Hinweise auf Reflexionen auf der Ebene „critical reflection“. Auch Fund (2002) fand hauptsächlich beschreibende Anteile, sowie Artmann et. al (2013) nur die beiden untersten Stufen. Es zeigt sich also zusammenfassend, dass große Teile der Texte keine Reflexionen enthalten und die reflexiven Anteile sich auf einem niedrigen, das heißt, unkritischen und wenig differenzierten Niveau bewegen.

Weinberger und Seyfried (2009) kamen mit ihrem eigenen Instrument (RIfE - *Reflection Instrument for Education*) zu einem ähnlichen Ergebnis wie auch die Autor*innen, die andere Stufenmodelle verwendeten: Zum größten Teil fanden sich die beiden untersten Stufen in den Texten, nämlich Beschreibung und Interpretation.

Verschiedene Studien konnten zeigen, dass Reflexionsprozesse vor allem dann weniger produktiv waren, wenn Studierende beim Reflektieren auf sich allein gestellt waren (Davis 2006, Michalsky und Kramarski 2015, Rodgers 2002).

Abgesehen von einer Beurteilung der Qualität der Reflexionen stellt Wyss (2011) auf einer inhaltlichen Ebene fest, dass sich die Reflexionen in 80 % der Fälle auf die Lehrkraft bezogen, wohingegen die Schüler*innen und deren Lernen kaum thematisiert wurden (vgl. S. 260). 60 % der Reflexionen befassten sich mit didaktischen Fragen, inhaltliche Fragen wurden dagegen selten in den Fokus genommen (vgl. S. 257).

2.2.2.2 Entwicklung von Reflexionsfähigkeit

In Forschung und Lehre herrscht weitestgehend Einigkeit darüber, dass Reflexionskompetenz grundsätzlich gefördert werden kann und sich auch im Lauf des Lebens und des Professionalisierungsprozesses entwickeln kann.

Einige Studien zeigen, dass es einen deutlichen qualitativen Unterschied in den Reflexionen von Noviz*innen bzw. Lehramtsstudierenden und erfahrenen Lehrkräften bzw. Expert*innen gibt (Gutzwiller-Helfenfinger et al., 2017, S. 135). Höhersemestrige Studierende erreichten häufig höhere Reflexionslevels als jüngere (Bain

et al., 1999; Kember et al., 2000). Eine Studie von Afshar und Farahni (2015) zeigte, dass erfahrene Lehrkräfte signifikant mehr und besser reflektierten als unerfahrene Lehrkräfte.

Dass Junglehrer*innen und Studienanfänger*innen weniger hochwertig reflektieren können, wird beispielsweise darauf zurückgeführt, dass sie auf weniger Erfahrungen und Wissen zurückgreifen können (Gruber, 2021; Moore-Russo & Wilsey, 2014; Penso et al., 2001). Probleme überhaupt zu erkennen, stellt für sie schon eine Herausforderung dar, wie auch bereits in Kapitel 2.1.2.6 zu den Voraussetzungen von Reflexion herausgestellt wurde. Expert*innen können über ihre Wahrnehmungen nachdenken, sie analysieren und daraus Schlüsse ziehen, wohingegen Noviz*innen häufig primär ihre Wahrnehmungen beschreiben. Auch Herzmann und Proske konnten zeigen, dass Unterrichtsanalysen zu Beginn meist eher flach und wenig differenziert ausfielen und sich erst mit zunehmendem Wissen ausdifferenzierten und präziser wurden (Herzmann & Proske, 2014, S. 36). Eine Studie von Lee (2005) ergab hinsichtlich thematischer Schwerpunkte der Reflexionen, dass die Studierenden sich anfangs eher mit sich selbst beschäftigten und erst mit der Zeit schülerorientierter wurden (Lee, 2005).

Diesen Befunden stehen wiederum Studien gegenüber, die mit zunehmender Erfahrung keine Zunahme der Reflexionsfähigkeit feststellen konnten (Turhan & Kirkgoz, 2018) oder es sogar in einzelnen Teilaspekten zu einer Verschlechterung der Reflexionen kam (Wyss, 2011). Poom-Valickis und Mathews (2013) konnten im Rahmen einer thematischen Textanalyse zeigen, dass die Reflexionen zu Disziplinproblemen bei noch unerfahrenen Lehramtsstudierenden zunächst eher unsicher ausfielen und sich dann mit zunehmender Erfahrung nach einem Jahr mit Praxisphasen eher ins Autoritäre und Bevormundende wandelten, meist ohne dass diese Haltung hinterfragt worden wäre oder Alternativen dafür gesehen worden wären. Diese Ergebnisse decken sich mit dem allgemeinen Forschungsstand zur Entwicklung von Professionalität im Lehrberuf, nämlich dass Lehrpersonen mit zunehmender Erfahrung nicht automatisch kompetenter werden (Hauser, 2021, S. 27).

Es darf also nicht davon ausgegangen werden, dass sich die Reflexionsfähigkeit mit zunehmender Erfahrung von selbst entwickelt und verbessert und sich thematisch in die gewünschte Richtung bewegt. Die Entwicklung von Reflexionsfähigkeit ist nicht zwingend ein linearer Prozess (Fox et al., 2019; Larrivee, 2008; Sydnor et al.,

2020). Behr und Welzel beispielsweise (2009) legen ebenfalls die Interpretation nahe, dass „eine Entwicklung von Reflexionskompetenz nur über einen längeren Zeitraum möglich ist und diese nicht linear verläuft“ (S. 121). Das könnte möglicherweise auch die vielen nicht signifikanten Ergebnisse von Untersuchungen zur Förderung von Reflexion (siehe folgendes Kapitel) erklären, da diese häufig nur kürzere Zeiträume in den Fokus nahmen.

Während qualitative Untersuchungen also meist zu dem Schluss gelangen, dass es sich bei Reflexion nicht um ein über längere Zeit stabiles Konstrukt handelt und dies vor allem bei Untersuchungen von Studierenden in der Anfangsphase der Lehrerbildung zu bedenken geben (Weinberger & Seyfried, 2009, S. 93), konnten die meisten quantitativen Instrumente kaum Veränderungen bei erneuten Messungen detektieren und kamen häufig zu dem Schluss, bei Reflexion handele es sich um einen relativ stabilen Charakterzug, wie in Kapitel 2.2.1.1. gezeigt wurde.

Die Antwort auf die Frage, ob und wie sich Reflexionsfähigkeit entwickelt, hängt also auch vom gewählten Messinstrument ab.

2.2.2.3 Förderbarkeit von Reflexionsfähigkeit

Im Folgenden sollen einige beispielhafte Einblicke in die Forschung gegeben werden, ob und wie Reflexionsfähigkeit gefördert werden kann. Dabei soll vor allem auf methodische Probleme hingewiesen werden, die für die vorliegende Studie relevant sind.

Zunächst fällt auf, dass einigen Maßnahmen aufgrund von Plausibilitätsannahmen eine Förderung der Reflexionsfähigkeit zugesprochen wird, ohne dass eine solche Wirkung bisher empirisch schlüssig belegt wäre. Das gilt zum Beispiel für die Biografiearbeit und die Arbeit mit sogenannten Life-History-Accounts (Knowles, 1993) sowie die pädagogische Beratung (Gröning, 2011, S. 44).

Ein ähnlich normativ aufgeladenes Konstrukt ist das Portfolio oder auch E-Portfolio, von dem in höchstem Maße eine reflexionsanregende und -förderliche Wirkung erwartet wird. Auch hier zeigt sich bei genauerem Hinsehen, dass es in der immensen Menge an Literatur zum Thema Reflexionsförderung durch (E-)Portfolio (Ayan & Seferoğlu, 2011; Chittum, 2018; Czerwionka et al., 2010; Gödecke, 2020; Jenson, 2011; Pineker & Störtländer, 2013; P. Roberts, 2018; C. Roberts & Stark,

2008) nur wenige verlässliche Studien gibt, die tatsächlich eine Steigerung der Reflexionsfähigkeit zeigen konnten (Ayan & Seferoğlu, 2011; Sung et al., 2009). Ein weiteres Problem besteht unter anderem darin, dass verschiedene Arten von Portfolios zum Einsatz kamen und in einem Portfolio verschiedene Aufgaben eingeschlossen sein können. Auch wurden unterschiedliche Erfolgskriterien festgelegt. Dass die genauen Aufgaben und Reflexionsanlässe häufig gar nicht in den Studien berichtet werden, zeigt, dass für diesen ganz wesentlichen Punkt noch zu wenig Sensibilität herrscht.

Ähnlich wie dem Portfolio wird auch der Action Research als primärem Effekt die reflexive Verbindung von Theorie und Praxis zugeschrieben (El-Dib, 2007).

Im Rahmen eines Action-Research-Projekts von Roberts (2016) zeigten sich diesbezüglich positive Tendenzen, die aber wegen einer zu geringen Fallzahl statistisch nicht nachweisbar waren. Hier wird ein weiteres Problem dieses Forschungsstrangs deutlich: Sehr häufig finden sich über alle möglichen Fördermaßnahmen hinweg lediglich Fallstudien mit sehr wenigen Untersuchungsobjekten (Lambert et al., 2014; Lee, 2005; Spalding & Wilson, 2002; Tiainen et al., 2018), anhand derer keine statistisch abgesicherten Aussagen getroffen werden können.

Weitere häufig eingesetzte Maßnahmen zur Förderung von Reflexion sind der professionelle Dialog mit Mentor*innen, Kolleg*innen oder Peers, das Schreiben, der Einsatz von Videoaufzeichnungen und Seminare bzw. Begleitveranstaltungen. Diese vier großen Felder werden im Folgenden noch genauer beleuchtet.

Als förderlich für die Reflexion zeigte sich im Kontext professioneller Gespräche, wenn die Gesprächsebene bzw. der Austausch zwischen den Beteiligten eher konstruktivistisch geprägt war als transmissiv (Linninger, 2016) und aktiv verschiedene Positionen diskutiert wurden (Decker et al., 2015). Als hilfreich wurden zum Teil auch „critical friend“-Gespräche (Hatton & Smith, 1995) oder „reflective teams“ eingeschätzt (Göbel & Gösch, 2019). Allerdings zeigte sich in anderen Studien, dass Zusammenarbeit und Feedback unter Peers die Reflexionsfähigkeit entweder gar nicht oder sogar negativ beeinflusste (Risko, 2009, S. 54) und zu eher technischen als kritischen Reflexionen führte (King, 2008). Eine qualitative Untersuchung von Raaflaub, Wyss und Hüsler (2019) zeigte, dass in Peer-Reflexionsgesprächen nur selten auf kritische Aspekte eingegangen und unterschiedliche Ansichten kaum als Anregungen wahrgenommen wurden. Problematische Aspekte wurden eher relativiert und abgeschwächt, als als Lernausgangspunkt genutzt (vgl.

S. 54). Ein Teilen von Wissen zwischen Lehrkräften via WhatsApp förderte einer Studie von Farahian und Parhamnia (2021) zufolge das Reflektieren. Eine Studie von Mauri, Clarà, Colomina und Onrubia (2017) verglich die Reflexionen einer Gruppe, in der Studierende und Tutor*innen gemeinsam reflektiert hatten, mit einer Kontrollgruppe ohne gemeinsame Reflexionen. Sie kommt zu dem Ergebnis, dass die Intervention bei manchen Studierenden zu Verbesserungen der Reflexionen führte, bei manchen allerdings nicht: „While the students in one case appear to have used the joint reflection process as a learning process, the students in the other did not“ (S. 439). Erklärungen zu diesem Befund liegen nicht vor. Auch Schülerrückmeldungen wurden eingesetzt, um Reflexionen von Lehramtsstudierenden anzuregen (Göbel et al., 2021; Neuber & Göbel, 2016). Diese waren nur teilweise geeignet, Reflexionen und Veränderungsprozesse anzustoßen und hatten eher den Effekt, dass sie als Rückversicherung und Bestätigung bestehenden Verhaltens dienten. Die Betreuung durch Mentor*innen wurde laut einer Studie von Freisler-Mühlemann, Winkler, Schafer und Böhlen (2021) von Lehramtsstudierenden zwar subjektiv als hilfreich und unterstützend erlebt, hatte aber auf die tatsächliche Zunahme von Reflexionskompetenz keinen signifikanten Einfluss.

Auch dem Verfassen reflexiver Texte wird eine Steigerung der Reflexionsfähigkeit zugeschrieben. Erforscht wurde beispielsweise das Tagebuchschreiben (Hartung-Beck & Schlag, 2020; Santagata & Angelici, 2010; Spalding & Wilson, 2002), aktuell auch oft verlagert in den digitalen Raum, auf das Verfassen von Blogs (L. Andersen & Matkins, 2011; Garza & Smith, 2015; Killeavy & Moloney, 2010) oder Wikis (Wegner et al., 2017). Häufig finden sich aber auch in diesem Bereich neben einigen optimistisch stimmenden Befunden bei anderen Autor*innen keine signifikanten Ergebnisse bezüglich einer Steigerung der Reflexionsfähigkeit (Bohndick, 2016; Killeavy & Moloney, 2010; Totter et al., 2016).

Auch Videoaufzeichnungen eigenen oder fremden Unterrichts werden häufig eingesetzt. Tripp und Rich (2012) untersuchten in ihrem Review 63 Studien zur Förderung von Reflexion durch Videoanalyse und stellen heraus, dass selbst innerhalb dieses Oberbegriffs „Videoanalyse“ die Studien so unterschiedlich sind, dass sie schwer vergleichbar sind. Die Studien unterscheiden sich beispielsweise durch unterschiedliche Aufgabenstellungen und Anleitungen zu Reflexion und Analyse, und außerdem dadurch, ob individuell oder kollektiv reflektiert wurde, hinsichtlich der Länge der Videos, der Anzahl an Reflexionen und der Messung von Reflexion

(vgl. S. 680). Außerdem gibt es sowohl Studien zum Einsatz von Videoaufzeichnungen des eigenen Unterrichts der Reflektierenden (Allas et al., 2017; Bergman, 2015; Husu et al., 2008; Navin, 2018) als auch des Unterrichts von anderen Lehrkräften oder Studierenden. Die Frage, welche dieser beiden Möglichkeiten gewinnbringender ist, wird in Kapitel 2.2.3.2 gesondert betrachtet. Häufig wird der Einsatz von Videos mit anderen Maßnahmen wie schriftlichen Reflexionen oder Diskussionen kombiniert (Cherrington & Loveridge, 2014; Epler et al., 2013; Husu et al., 2008).

Auch Seminare, Kurse, Begleitveranstaltungen (Fund et al., 2002; Linninger, 2016; Volmer et al., 2021) und Workshops (Bohdick, 2016; Penso et al., 2001) sind häufig untersuchte Formate zur Steigerung der Reflexionsfähigkeit. Bei diesen Studien handelt es sich größtenteils um Evaluationen von eigens entwickelten Formaten, die ebenfalls unter dem Problem leiden, dass eine Vielzahl unterschiedlicher Maßnahmen, Schwerpunkte und Inhalte in diese eingeflossen ist, die sich im Ergebnis nicht mehr trennen lässt. Bei einer solchen Menge unkontrollierter Variablen ist am Ende nicht mehr festzustellen, worauf sich eine mögliche Steigerung der Reflexionsfähigkeit zurückführen lässt. Seminarkonzepte werden beispielsweise mit dem Einsatz von Unterrichtsvideos (Herzmann & Proske, 2014) oder theoriebasierten Fallreflexionen (Bonanati et al., 2020), mit schriftlichen Reflexionsaufgaben (Allen et al., 2018; Subramaniam, 2013) oder auch verschiedenen Methoden gleichzeitig (Snyder, 2011) kombiniert. Auch der Einsatz von Feedback wird natürlicherweise häufig nicht als einziges Mittel genutzt, sondern mit anderen Maßnahmen zusammen eingesetzt (Hayden & Chiu, 2015; Quinton & Smallbone, 2010; Sydnor et al., 2020). Sehr häufig sind es komplette (Kurs-)Programme bestehend aus verschiedenen Maßnahmen, die evaluiert werden (Martin, 2005) und bei denen am Ende nicht mehr feststellbar ist, worauf genau der eventuelle positive Effekt zurückzuführen ist (Mulryan-Kyne, 2020). Derartige Evaluationsstudien stellen den größten Teil von Untersuchungen zur Förderung von Reflexionsfähigkeit dar und weisen häufig auch methodische Mängel wie fehlende Kontrollgruppen und fehlende Prätests auf. Für die Effektivität eines solchen kombinierten Einsatzes von verschiedenen Maßnahmen argumentieren Kaasila und Lauriala (2012). Das Ergebnis ihrer Studie ist, dass die Reflexionen der Teilnehmenden sich sowohl in der Breite als auch in der Tiefe verbesserten, wenn verschiedene Anreize und Kontexte zur Reflexion wie

beispielsweise Forschungstexte, autobiographische Überlegungen und andere gegeben wurden. Das mag für die Förderung der Reflexionsfähigkeit möglicherweise zutreffen, eine systematische Erforschung der Wirksamkeit einzelner Maßnahmen kann so aber nicht erfolgen.

Zusammenfassend kann man festhalten, dass sich mit verschiedenen Maßnahmen Erfolge bei der Steigerung der Reflexionsleistung erzielen ließen. Allen et al. (2018) ziehen deshalb das Fazit, „there may not be a singular ‚best‘ way to prepare reflective practitioners“ (S. 92). Auch Spalding und Wilson (2002) bilanzieren, dass keine bestimmte pädagogische Strategie deutlich besser abschneide als andere und begründen das damit, dass Personen individuell unterschiedlich auf Maßnahmen reagieren. Leonhard und Rihm (2011) kommen zu dem Schluss, dass eine untersuchte Begleitveranstaltung zwar im Mittel die Reflexionskompetenz nicht gesteigert hatte, sich aber bei genauerer Betrachtung an der „erheblichen Zunahme der Standardabweichung [zeige], dass die Entwicklung innerhalb der Gruppen ausgesprochen heterogen ist.“ (S. 260). Dass Heterogenität in diesem Kontext eine zu wenig beachtete Rolle spielt und nicht alle Studierenden gleich auf Förderung und Angebote reagieren, zeigt sich ebenfalls sehr deutlich in qualitativen Fallanalysen (Hayden & Chiu, 2015), aber auch in Studien, die methodisch sauber verschiedene Fördermaßnahmen systematisch vergleichen (Hager, 2020). Worin diese Unterschiede genau bestehen, wie sie sich auswirken und wovon sie beeinflusst werden, ist bisher ungeklärt.

Eine zweite Folgerung, die man aus der Studienlage ziehen kann, ist, dass es womöglich weniger relevant ist, welche Tools und Fördermaßnahmen konkret eingesetzt werden, sondern eher, *wie* sie eingesetzt werden. Anleitung, Begleitung und Unterstützung scheinen wesentliche Faktoren zu sein, die die Wirksamkeit der Methoden beeinflussen (Benko et al., 2016; Hall, 2018). Das trifft vor allem für die Anfangsphase der Lehrkräftebildung zu, in der noch mehr Struktur und Lenkung von den Studierenden gewünscht und gebraucht werden: „It may be advisable in the initial stages to offer these students the external structure which they prefer“ (Korthagen, 1993a, S. 234).

Daran anschließend gibt es auch Wissenschaftler*innen, die eine explizite Förderung von Reflexion vertreten, also im Sinne eines Einübens und einer direkten Vermittlung dessen, was Reflektieren eigentlich bedeutet und wie man beim Reflektieren vorgehen sollte (LaBoskey, 1993; Ryan, 2011). Einige Studien ermittelten dafür

auch positive Effekte (Stenberg et al., 2016; Weinberger & Seyfried, 2009). Reflektieren wird hier als eine professionelle Technik gesehen, die als solche erlernbar ist und deren Beherrschung nicht selbstverständlich vorausgesetzt werden kann.

Leonhard und Abels dagegen postulieren, dass es wichtig sei, immer wieder sinnvolle Anlässe zur Reflexion anzubieten, denn Reflektieren „lern[e] man durch Reflektieren“ (2017, S. 52). In eine ähnliche Richtung argumentieren Prieto, Magnuson, Dillenbourg und Saar (2020) die in ihrer Studie mit technologischen Mitteln Reflexionsanregungen lieferten und direkt ans Reflektieren erinnerten. Sie gingen von dem Standpunkt aus, dass das Reflektieren prinzipiell beherrscht wird, aber im beruflichen Alltag weder institutionalisiert noch selbstverständlich ist und immer wieder in Erinnerung gerufen werden müsse. Zusätzlich zu den strukturellen Bedingungen spielt sicherlich die Vorbildfunktion der Dozierenden und der Verantwortlichen in allen Phasen der Lehrkräftebildung eine Rolle: „Before we can have reflective teachers, we need reflective schools and reflective teacher educators“ (Valli, 1992, S. 146).

2.2.2.4 Zusammenfassung und Bewertung

Während sich also insgesamt eher positive Einstellungen von Studierenden der Reflexion gegenüber finden, wird die tatsächliche Reflexionsfähigkeit fast durchgehend als enttäuschend gering ermittelt und Reflexionen als überwiegend oberflächlich und eindimensional eingestuft. Während weitgehend übereinstimmend über schlechte Reflexionen von Studierenden und Lehrkräften geklagt wird, wird zu wenig darüber gesprochen, was die Ursachen für diese Ergebnisse sein könnten. Es ist möglicherweise irreführend, wenn die Messergebnisse der Studien allein auf die Fähigkeit der Studierenden zurückgeführt werden, da sie, wie bereits im vorherigen Kapitel gezeigt, mit größeren Messfehlern behaftet sein können, die die Studienergebnisse in Frage stellen.

Als Erklärung für solche Tendenzen würde sich möglicherweise auch anbieten, dass Studierende zum Teil davon ausgehen, dass eine klare Antwort oder ein eindeutiges Fazit, eine Positionierung von ihnen erwartet wird und sie eine multiperspektivische Betrachtung und ein Aufzeigen von Handlungsalternativen als ein Zeichen des Eingestehens von Unsicherheit oder einer Wissenslücke betrachten, die es zu

vermeiden gilt. Auch die universitäre Prüfungspraxis mit eindeutig richtigen oder falschen Antworten könnte einem solchen Verhalten Vorschub leisten. Chye, Zhou, Koh und Liu (2021) legen die Interpretation nahe, dass Studierende möglicherweise deshalb keine hohen Levels von Reflexion zeigen, weil dies im Kontext der Universität, des Seminars oder Praktikums nicht legitimiert oder belohnt wird. Das Reflexionslevel wäre demzufolge Ausdruck des soziokulturellen Kontextes, den die Autor*innen empfehlen, bei der Messung von Reflexion miteinzubeziehen.

Kritisch zu sehen ist auch die Abwertung des Beschreibens als niedrige Ebene der Reflexion oder die Wertung als gar kein Teil von Reflexion, wie es in einigen Studien und Stufenmodellen geschieht: „Bisherige Studien [...] vernachlässigen die Erfahrungsdarstellung“ (Gutzwiller-Helfenfinger et al., 2017, S. 133). Im Kapitel 2.1.2.4 zum Ablauf von Reflexion wurde gezeigt, dass die Beschreibung in allen Modellen von Reflexion eine Rolle spielt. Werden diese zyklisch gedacht, kann nicht einmal argumentiert werden, dass die Beschreibung dabei eine niedrige Stufe darstellt. Rosenberger beispielsweise betont, dass es für die Qualität der Reflexionen von großer Bedeutung war, die „erlebten Situationen möglichst genau zu beschreiben“ (Rosenberger, 2017, S. 196) und sich facettenreich noch einmal zu vergegenwärtigen. Das stelle auch die Ausgangsbasis für eine spätere multiperspektivische Betrachtung der Situation dar. Auch Dewey bezeichnet eine genaue Beobachtung und deren Beschreibung als unverzichtbaren Anteil von Reflexionen: „Reflection includes observation“ (Dewey, 1933, S. 102). An dieser Stelle zeigt sich exemplarisch, wie sich die Messungen zum Teil von der theoretischen Basis entfernen, ohne dass es dafür eine Notwendigkeit gibt. Auch zeichnet sich als Entwicklungstendenz ab, dass dem kritischen Reflektieren und moralischen Aspekten weniger Beachtung geschenkt wird und es sich häufig nicht mehr in den Messverfahren abbildet (vgl. Akbari, 2007, S. 197).

Hinsichtlich der Frage, ob Reflexionsfähigkeit sich entwickeln kann, kommt die Forschung überwiegend zu einem optimistischen Ergebnis mit der Einschränkung, dass es nicht unbedingt von selbst zu einer positiven Entwicklung kommt. Die Beurteilung dieser Frage hängt allerdings vom gewählten Messinstrument ab: Während zu den Fragebögen meist bilanziert wurde, dass sie Reflexion als eher stabilen Charakterzug erheben, erzeugen die Textanalysen ein anderes, sehr wandelbares

und situationsabhängiges Bild von Reflexionsfähigkeit, was wiederum zeigt, dass die Wahl des Messinstruments die Ergebnisse beeinflussen kann.

Hinsichtlich der Förderbarkeit von Reflexionskompetenz ist die Forschungslage zum Teil widersprüchlich und wenig belastbar. Ob es nun förderlicher ist, direkt oder indirekt auf die Reflexionsqualität und -häufigkeit Einfluss zu nehmen und ob bestimmte vorgestellte Maßnahmen wirksamer sind als andere, kann aus den Studien derzeit noch nicht abgeleitet werden. Das liegt primär an der mangelnden Vergleichbarkeit der Einzelstudien, bei denen Reflexion auf unterschiedliche Arten gemessen und operationalisiert wurde: „It seems that a standard comparison is not possible, because each study is held in a different context with different data gathering and analysis instruments“ (Kayapinar & Erkus, 2009, S. 148). Collin, Karsenti und Komis (2013) kommen ebenfalls zu dem Fazit, dass Methodik und Operationalisierung der Studien zu unterschiedlich sind, um einen Vergleich anstellen zu können oder Ergebnisse zu generalisieren. Bemerkenswert ist auch das Ergebnis, dass die Studierenden sehr heterogen auf die unterschiedlichen Maßnahmen, Erhebungssituationen und -instrumente reagieren.

Verbesserungen der Reflexionen im Rahmen verschiedener Interventionen lassen sich aber auch durch andere Ursachen als die Intervention selbst erklären. Berndt und Häcker legen die Interpretation nahe, dass sich die Reflexionen nicht (nur) aufgrund der Interventionen verbessert haben könnten, sondern die Studierenden schlicht – z.B. im Verlauf eines Seminars – erkannt haben, was von ihnen verlangt wird und das gewünschte Verhalten zeigen: „Werden die Studierenden tatsächlich ‚reflektierter‘ oder stellen sie sich lediglich besser darauf ein, zu ‚liefern‘, was von ihnen erwartet wird?“ (2017, S. 243). Dieser Verdacht könnte sich durch methodisch entsprechend angelegte Studien mit Kontrollgruppen, größeren Teilnehmerzahlen und Follow-up-Untersuchungen ausräumen oder bestätigen lassen, diese werden allerdings nur sehr selten oder gar nicht durchgeführt. Wegen des Fehlens solcher Untersuchungen kann keine belastbare Aussage dazu getroffen werden, ob durch die Maßnahmen eine dauerhafte, womöglich domänenübergreifende Kompetenz erworben wurde, oder ob diese nur spezifisch, kontextbezogen und zeitlich begrenzt in Erscheinung trat.

Collin (2013) sowie Berndt und Häcker (2017) kritisieren, dass es sich beim Großteil der Studien um Evaluationen eigener Projekte handelt, die mit einer „legitimatorischen Absicht“ (Berndt & Häcker, 2017, S. 250) verfasst werden und deren Ergebnisse häufig wenig verallgemeinerbar sind.

In Referenz auf das Kapitel 2.1.2.9, in dem mögliche negative Wirkungen von Reflexion herausgearbeitet wurden, soll außerdem noch darauf hingewiesen werden, dass bei den vorliegenden Studien und Fördermaßnahmen auf mögliche unerwünschte Effekte keine Rücksicht genommen wurde. Mögliche Nebeneffekte wie Verunsicherung oder Grübeln wurden, soweit es in den Studien berichtet wurde, nicht überprüft oder aufgefangen. Ferner wurde nirgends davon berichtet, dass mit den Teilnehmenden thematisiert wurde, in welchen Kontexten Reflexion überhaupt sinnvoll angewendet werden kann und wo ihre Grenzen liegen.

Abschließend ist noch kritisch zu erwähnen, dass beim Feld „Schreiben“ sowie auch beim Portfolio die Grenze zwischen Intervention und Mess- bzw. Bewertungsinstrument verschwimmt (Leonhard, 2013, S. 180). Es wurde gezeigt, dass beide sowohl als Maßnahme zur Steigerung von Reflexionsfähigkeit als auch als Instrument, diese zu erheben, eingesetzt werden. Das gleiche trifft im Grunde auch auf Interviews zu, die ebenfalls als Messinstrument eingesetzt werden, aber möglicherweise auch die reflexionsfördernde Wirkung eines Gesprächs mit einem kritischen Gegenüber haben könnten (Li & Peng, 2018). Dass es durch die Messmethode selbst zu einer Leistungssteigerung kommen könnte, würde wiederum bedeuten, dass die Validität der Messungen in Frage gestellt ist.

Zusätzlich zu den Messinstrumenten und deren Ergebnissen spielen auch noch weitere Faktoren im Kontext der Messung von Reflexionsfähigkeit eine wichtige Rolle, die zum Teil bereits angeschnitten wurden und im folgenden Kapitel genauer betrachtet werden sollen.

2.2.3 Schwierigkeiten bei der Messung

2.2.3.1 Reflexionsanlass

Wie bereits gezeigt wurde, steht Reflexion immer unter einem Möglichkeitsvorbehalt. Deshalb ist die Frage, wie eine Reflexion erfolgreich angestoßen werden kann,

von besonderer Bedeutung. Dabei soll es zunächst darum gehen, was den Anlass zur Reflexion geben soll, also worüber eigentlich reflektiert werden sollte. In Kapitel 2.1.2.1 wurde bereits dargestellt, dass theoretisch nicht ganz geklärt ist, ob Reflexion nur nach einem natürlich entstandenen Anlass auftritt oder didaktisch induziert werden kann. Als Fazit wurde gezogen, dass Reflexion im Sinne Deweys und Schöns nur anlassbezogen in Folge einer erlebten Überraschung oder eines Problems erfolgt, dass sie aber eventuell auch bewusst ausgelöst werden kann, sofern es gelingt, eine Irritation oder überraschende Situation künstlich herzustellen. Gelingt dies nicht, könnte es sein, dass die künstlich hergestellte Reflexionssituation für die Proband*innen nicht als problematisch, uneindeutig und damit „reflexionsbedürftig“ eingeschätzt wird. Daraus könnte folgen, dass Proband*innen keine oder kaum Reflexionshandlungen zeigen, obwohl sie prinzipiell über eine hohe Reflexionsfähigkeit verfügen, was ihnen dann bei der Messung als mangelnde Reflexionsfähigkeit ausgelegt werden könnte.

In diesem Kapitel soll diese Problematik noch einmal aufgegriffen und die Frage beleuchtet werden, ob es zur Messung von Reflexion sinnvoller ist, die Studierenden zu selbst erlebten Ereignissen reflektieren zu lassen oder zu fiktiven Erlebnissen und Ereignissen bzw. theoriebezogenen Themen.

Einzelne Studien kommen zu dem Ergebnis, dass die Reflexion über fremde Erlebnisse leichter fällt. Dies könnte, wie Poom-Valickis und Mathews (2013) vermuten, daran liegen, dass diese mit weniger starken Emotionen verbunden sind als persönliche Erlebnisse. Diese emotionale Involviertheit beeinträchtigt die Reflexionsleistung: „Reflecting on others’ practices does not carry the same emotional biases as reflecting on one’s own practice“ (S. 422). Auch Leijen, Valtna, Leijen und Pedaste (2012) beschreiben, dass es ihren Studienteilnehmer*innen schwerer fiel, zu eigenen Erlebnissen tiefergehende Reflexionen anzustellen, als wenn die Handlungen anderer thematisiert wurden. Risiko (2009) verglich drei verschiedene Reflexionsanlässe, nämlich Reflexion zu eigenen Praxiserfahrungen, zu einem wissenschaftlichen Artikel und zu einer Fallstudie. Reflexion konnte laut der Studie von allen dreien ausgelöst werden, wenngleich die Fallstudie am besten abschnitt. Risiko zog außerdem das Fazit, dass es Studierenden schwerer falle, Probleme zu erkennen und Lösungen zu finden, wenn es um ihre eigenen Praxiserfahrungen gehe (S. 54). Herzmann, Artmann und Wichelmann führen als Vorteil der Arbeit mit

fremden Beispielen „ein handlungsentlastetes und rechtfertigungsfreies Einüben“ an (Herzmann et al., 2017).

Göbel und Gösch stellen fest, dass „fremde Videos eher zu einer distanzierteren, kritischeren Haltung gegenüber den Handlungen [anregen,] während Videos von eigenem Unterricht eine stärkere persönliche Betroffenheit und damit einhergehend eine höhere Motivation zur Auseinandersetzung mit dem Material fördern“ (2019, S. 11). Dies wird bestätigt von Seidel, Stürmer, Blomberg, Kobark und Schwindt (2011) sowie Kosinár (2017). Letztere geht ebenfalls davon aus, dass Motivation für eine Reflexion nur gegeben ist, wenn eine subjektive Bedeutung der Situation für das Individuum vorhanden ist, weshalb sie dafür plädiert, dass der Gegenstand der Reflexion von den Akteur*innen selbst gewählt werden sollte (S. 157). Dafür spräche auch, dass es mit keinem künstlichen Reflexionsanlass gelingen kann, dass wirklich alle Studierenden diesen übereinstimmend als problematisch empfinden. Rosenberger (2017) weist darauf hin, dass die Studierenden, die eben keinen Zweifel und keine Notwendigkeit einer Problemlösung bei sich spüren, eine „Dissonanz beim Reflektieren-Müssen“ (Rosenberger, 2017, S. 197) empfinden. Es besteht die Gefahr, dass „sie den Sinn nicht einsehen“ (Wyss & Mahler, 2021, S. 22). Andersen und Matkins (2011) konnten in ihrer Studie zeigen, dass höhere Reflexionsstufen erreicht wurden, wenn zu eigenen Erlebnissen reflektiert wurde, als wenn andere Lehrkräfte beobachtet wurden. Martin (2005) stellt heraus, dass in der Literatur vor allem angezweifelt wird, ob junge Lehrkräfte und Studierende in der Lage sind, die eigene Praxis zu reflektieren, weil sie sich anfangs primär auf technische Aspekte konzentrieren und nur wenig andere Aspekte wahrnehmen. Sie selbst stellt in ihrer Untersuchung aber fest, dass die unerfahrenen Proband*innen sich durchaus in ihrer Reflexionsfähigkeit gesteigert hätten (S. 534).

Persönliche Erlebnisse als Reflexionsanlässe könnten als mögliche Schwierigkeit beinhalten, dass mit zunehmendem zeitlichem Abstand zur Situation die Erinnerungen ungenauer werden und sich erlebte Situationen ohnehin einer objektiven Rekonstruktion entziehen: „Jedes soziale Handeln enthält inkorporierte Wissensanteile, derer sich die Akteurinnen und Akteure teilweise nicht bewusst sind bzw. die von ihnen auch im Nachhinein nicht vollständig rekonstruiert werden können“ (Rosenberger, 2017, S. 195). Wenn mit Videos von eigenen Stunden gearbeitet

wird, könnte das den Vorteil haben, dass Erinnerungsunschärfen durch das nochmalige Anschauen ausgeglichen werden könnten (vgl. Tripp & Rich, 2012, S. 681). Außerdem könnte bei einer Messung die Vergleichbarkeit der persönlichen Reflexionsanlässe ein Problem darstellen, weil aus einer unendlichen Fülle von möglichen Erfahrungen gewählt werden kann.

Als Fazit lässt sich festhalten, dass es sowohl Studien gibt, die die Reflexion zu Erlebnissen anderer Personen oder fiktiver Erlebnisse für gewinnbringender halten und andere, die betonen, es müsse sich um eigene Erlebnisse und authentische Probleme handeln. Es ist in dieser Frage noch zu keiner abschließenden Klärung gekommen.

2.2.3.2 Aufgabenformulierung und Prompts

Neben dem konkreten Reflexionsanlass ist eine zweite wichtige Frage, wie die Aufgabenstellung, bzw. der Prompt, formuliert sein sollte, um die gewünschte Reflexion zu erzeugen.

Dabei ist unter anderem von Bedeutung, welcher Operator bzw. welche Operatoren verwendet werden, geben diese doch vor, was die Studierenden konkret tun sollen. Sparks-Langer und Colton bilanzieren beispielsweise am Ende ihrer Auswertungen: „The questions did not elicit [...] the kind of thinking we were coding“ (Sparks-Langer & Colton, 1991, S. 27). Die Fragen und Aufgaben selbst erzeugen also bereits eine bestimmte Art des Denkens. Der Einsatz des Operators „Reflektieren“ könnte problematisch sein, da unter diesem Wort verschiedene Dinge verstanden werden, wie in Kapitel 2.1.1.2 bereits erläutert wurde. Bei jedem anderen Operator legt man allerdings bereits fest, welche „Form“ von Reflexion erzeugt wird. Formuliert man beispielsweise als Aufgabenstellung „Nehmen Sie kritisch Stellung zu...“ erhält man mit größerer Wahrscheinlichkeit Ausführungen auf einer kritischen und vielleicht auch moralischen Ebene, als wenn dies nicht explizit gefordert wird. Hieraus folgt also wieder eine Beeinflussung der Messergebnisse durch Rahmenbedingungen der Messung.

Eine weitere Frage, die sich in diesem Zusammenhang stellt, ist die nach der Ausführlichkeit und Konkretheit der Aufgabenstellung. Sollte man möglichst konkrete

Aufgabenstellungen wählen (im Sinne auch einer „Anleitung“ zum Reflektieren) oder sollte die Anweisung so allgemein und kurz wie möglich gehalten werden? Weinberger und Seyfried nennen als mögliche Ursache für die große Menge an oberflächlichen und auf den beiden untersten Ebenen stattfindenden Studierendenreflexionen beispielsweise die Aufgabenstellung ihres Instruments: „The instruction of the instrument [may be] too vague“ (Weinberger & Seyfried, 2009, S. 92). Sollen die Aufgaben also vorzugsweise konkret formuliert sein? Forschungsergebnisse zu dieser Frage sind nicht eindeutig. So testete Davis (allerdings mit Schüler*innen) in ihrer Studie „generic prompts“ und „directed prompts“, also eher offen gehaltene Anweisungen und direkte Anweisungen im Vergleich. Die genaueren Anweisungen führten dabei zu unproduktiveren Reflexionen. Die allgemein gehaltenen Anweisungen führten dagegen nach Meinung der Autorin dazu, dass die Schüler*innen wirklich innehielten und zu komplexeren Reflexionen und einem kohärenterem Verständnis gelangten (Davis, 2003). Im Gegensatz dazu stellten Kaufman und Kollegen fest, dass ihre „reflection prompts“ bessere Ergebnisse lieferten, wenn sie mit konkreten Problemlösehinweisen verknüpft waren. Ein ähnliches Bild ergab sich auch in einer Studie von Michalsky und Kramarski, bei denen sich höhere Reflexionsebenen beim Einsatz von „directed prompts“ im Gegensatz zu generischen Aufforderungen zum Reflektieren zeigten (Michalsky & Kramarski, 2015). Auch bei Husu und Kollegen ergab sich, dass konkretere Instruktionen und Scaffolding die Reflexionsqualität positiv beeinflussten (Husu et al., 2008). Gutzwiller-Helfenfinger et al. (2017) nutzen in ihrer Studie bewusst offene Prompts und erhielten ihrer Meinung nach wenig reichhaltige Ergebnisse und die Kriterien, nach denen ausgewertet wurde, wurden von den Studierenden kaum erfüllt. So wurden kaum Fragen aufgeworfen, Perspektiven gewechselt oder Bezüge zur eigenen Rolle hergestellt, was die Autor*innen darauf zurückführten, dass zu wenig konkret angeleitet wurde. Sie folgern, dass Studierende wohl in der Lage wären, diese Dimensionen von Reflexion zu erfüllen, „dies jedoch nicht spontan tun, sondern nur dann, wenn sie dazu aufgefordert werden“ (S. 138). Ein ähnliches Ergebnis zeigte sich in der Studie von Kauffman, Ge, Xie und Chen (2008), nämlich dass die bloße Aufforderung, ein Arbeitsergebnis oder Erlebnis zu reflektieren, die Problemlösefähigkeit und die Arbeitsergebnisse nicht verbesserte, sondern nur in Kombination mit weiteren, konkreteren Problemlöseprompts erfolgreich war. Auch

in der Studie von Schäfer, Blomberg und Seidel (2012) löste eine höhere Strukturierung von Lerntagebüchern mithilfe von Prompts tiefenorientiertere Elaborationsstrategien aus als eine geringere Strukturierung, was als expertenhaftere Reflexion interpretiert wurde (vgl. S. 271).

Es ergibt sich also fast durchgehend das Ergebnis komplexerer Reflexionen auf höheren Ebenen bei genaueren Anweisungen, woraus im Grunde die Empfehlung für zukünftige Forschung folgen müsste, Aufgaben mit höherer Strukturierung und Detailliertheit einzusetzen. Zu bedenken ist allerdings ein gravierendes Validitätsproblem: Solche Anweisungen, die beispielsweise die für Reflexion erforderlichen Stufen abdecken, lösen zwar vielleicht aus, dass Studierende analysieren, beschreiben oder diskutieren, zeigen aber – im Sinne einer Reflexionskompetenz – nicht, ob eine Person Reflexion und die dafür erforderlichen Fähigkeiten und Fertigkeiten selbstständig einsetzen kann und will, um ein Problem zu lösen (vgl. Weinert, 2014, S. 27). Der „Lösungsweg“ wird also im Grunde mit der Aufgabenstellung vorweggenommen. Die Studierenden nehmen dadurch vielleicht Aspekte, die sie allein nicht bedacht hätten (z.B. den Einbezug verschiedener Perspektiven), in ihre Reflexionen auf, wodurch diese dann natürlich, nach Stufen untersucht, hochwertiger erscheinen. Es ist aber ein unzulässiger Schluss, beispielsweise in Pro- und Kontraabwägungen ein Zeichen für eine höherwertige Reflexion zu sehen, wenn sie explizit mit der Aufgabenstellung eingefordert wurden. Während es im Kontext der Förderung von Reflexion durchaus von Bedeutung ist, reflexionsförderliche Prompts und unterstützende Hinweise zu geben (Göbel & Gösch, 2019, S. 284f), so kann es kaum sinnvoll sein, diese während der Messung zu geben, da diese ansonsten verfälscht wird. Auch Moore-Russo und Wilsey (2014) raten zu möglichst offen und wenig lenkenden bzw. vorstrukturierten Aufgaben und Reflexionsanlässen, da diese Aufgaben am ehesten dem entsprechen, was die Studierenden als zukünftige Lehrkräfte in ihrem Arbeitsfeld als authentische Reflexionsanlässe vorfinden werden (S. 86).

2.2.3.3 Sichtbarmachung über ein Medium der Reflexion

Eine bedeutende Frage in der Beforschung von Reflexion ist bisher nur scheinbar beantwortet, nämlich die Frage, in welcher Form Reflexion sichtbar gemacht werden kann, sodass eine Messung überhaupt möglich wird. Wir verfügen bisher nicht über Möglichkeiten, Reflexion in ihrer ursprünglichen Form, nämlich vorliegend in Gedanken und Gefühlen, unmittelbar zu messen oder auch nur zu beobachten. Den direktesten Zugang dazu hätten eventuell die neurobiologischen und hirnhysiologischen Ansätze der Messung, die in Kapitel 2.2.1.3 vorgestellt wurden, weshalb es interessant sein wird, wie sich diese Forschungsansätze in Zukunft entwickeln werden. Momentan sind solche Methoden noch viel zu ungenau und es liegt nicht genügend Wissen vor. Die empirische Sozialforschung jedenfalls ist darauf angewiesen, dass diese reflexiven Gedanken und Gefühle in einer Form sichtbar gemacht werden, die mit ihren Mitteln untersucht werden kann. Im Folgenden sollen die bisherigen Möglichkeiten der Sichtbarmachung kritisch beleuchtet werden.

Einerseits ist es möglich, etwas über die Reflexionsfähigkeiten, -gewohnheiten und -vorlieben herauszufinden, indem mittels Items in Fragebögen direkt nach ihnen gefragt wird und in Form einer Selbstauskunft der Proband*innen die Ergebnisse gewonnen werden. Diese können entweder geschlossene Antwortformate beinhalten, wie es meist der Fall ist, oder offene. Über die Probleme dieser Erhebungsform wurde bereits ausführlich in den Kapiteln 2.2.1.1 und 2.2.1.4 berichtet.

Der Weg, der am häufigsten beschritten wird, ist die Auswertung schriftlich vorliegender Dokumente in Form von Tagebucheinträgen, schriftlichen Reflexionsaufgaben oder beispielsweise sogenannten „critical incident reports“ (Kember et al., 2008). Die Art des Textes, die sich aus der Aufgabenstellung ergibt, beeinflusst allerdings den Schreibstil, die Wortwahl und die Art des Argumentierens (vgl. Artmann et al., 2013, S. 141f), die wiederum Einfluss auf die Auswertung nehmen. Als Begründung für die Verwendung schriftlicher Reflexionsdokumente wird immer wieder ein einziges Studienergebnis zitiert. Es handelt sich um das Fazit von Hatton und Smith (1995), die in ihrer Studie mit Interviews, schriftlichen Aufzeichnungen, Videoaufzeichnungen und Selbsteinschätzungen arbeiteten und bilanzieren: „While all data were analysed, the written reports provided most evidence of reflection“ (Hatton & Smith, 1995, S. 92). Das mag zwar für diese Studie zutreffend

sein, es ließen sich aber kaum weitere Studien finden, die sich ebenfalls gezielt mit der Fragestellung auseinandergesetzt hätten, welche Dokumente die besten Belege für Reflexion liefern. Eine sehr kleine qualitative Studie von Lambert, Sorensen und Elliot (2014) kam beispielsweise im Gegenteil zu dem Ergebnis, dass die schriftlichen Reflexionen in ihrer Studie eher zusammenfassend und detailärmer ausfielen als reflektierende mündliche Interviews und auch Lüsebrink und Grimminger (2014) berichten, dass die schriftlichen Reflexionen im Vergleich zu den engagierten Diskussionen im Seminarkontext eher „lustlos“ (S. 209) ausfielen. Sie stellen außerdem fest, dass die schriftliche Erhebung den weiblichen Teilnehmerinnen eher entgegenkam, was sich in längeren Texten zeigte (vgl. S. 208). Fragebogenerhebungen fanden im Gegensatz dazu keine geschlechterspezifischen Unterschiede hinsichtlich Qualität oder Quantität von Reflexion (Sobral, 2005) oder sogar einen signifikanten Vorsprung der männlichen Reflektierenden gegenüber den weiblichen (N. B. Andersen et al., 2014). Guardiera und Kolleg*innen halten als Bilanz ihrer Studie ganz allgemein fest, dass bei einigen Studierenden eine geringere Bereitschaft zur Verschriftlichung von Reflexionsprozessen festzustellen war, was die inhaltliche Zuverlässigkeit der Auswertung in Frage stellt (Guardiera et al., 2018). Eine Gefahr liegt auch darin, dass schlicht nicht alles geschrieben wird, was prinzipiell gedacht oder gewusst wird. Eine Studienteilnehmerin von Martin (2005) drückt es so aus: „You [die Studienteilnehmerin über sich selbst; M.R.] wrote whatever you wrote and then the comment would be – Why is this? And you’re like, if I’m talking to you I could tell you why this is. I know why it is but I couldn’t write another ten pages“ (S. 533). Die Probandin gibt also an, nicht alles verschriftlichen zu können, was sie denke und wisse und dass das in einem Gespräch bzw. Interview vielleicht leichter möglich gewesen wäre. Zu einem ähnlichen Schluss kommt auch Griffin (2003): „A format other than writing, such as an interview, may have elicited more information“ (S. 218). Gödecke schreibt, dass über die Gründe, warum bestimmte Reflexionsinhalte von den Proband*innen verschriftlich werden und andere weggelassen werden, nur Vermutungen angestellt werden könnten, aber dadurch schwer interpretierbare Leerstellen entstehen (Gödecke, 2020, S. 325) und auch Lüsebrink und Grimminger folgern aus dem gleichen Befund bei ihrer Textanalyse, es komme durch die Erhebung über das Schreiben zu einem „Bias, welcher sich in einer anderen, evtl. motivierenderen Erhebungsform nicht so zeigen würde“ (Lüsebrink & Grimminger, 2014, S. 209). Sogar

Hatton und Smith (1995) selbst erwähnen, dass die Analyse schriftlicher Texte aus sprachlichen Gründen eventuell Menschen mit schwächerem sozio-ökonomischen Hintergrund (vgl. ebd., S. 42) und niedrigerem Bildungsniveau unterschätzen könnte. Der Grund dafür könnte die Abhängigkeit der Beurteilung der Reflexionsqualität von der Sprachbeherrschung sein, was die Testfairness in Frage stellt. Durch Schwächen im sprachlichen Ausdruck kann die tatsächliche Reflexionsqualität als geringer eingestuft werden, als diese tatsächlich ist. In Kapitel 2.2.1.2 wurde bereits dargestellt, dass die vorgestellten Stufenmodelle zur Textanalyse zum Teil auch die sprachliche Gestaltung in die Stufenbeschreibungen miteinbeziehen und diese einen deutlichen Einfluss nimmt. Durch eine versiertere sprachliche Darstellung könnte also eventuell eine geringe Reflexionsleistung überschätzt und durch sprachliche Schwächen eine höhere Reflexionsleistung unterschätzt werden.

Eine Alternative könnten Interviews oder auch Gruppendiskussionen (Walz & Roth, 2017) darstellen. Positiv an diesen Erhebungsformen ist die Tatsache, dass ein besseres Vertrauensverhältnis und größere Nähe hergestellt werden können, was im Kontext von Reflexion eine wichtige Rolle spielt und im folgenden Kapitel noch genauer beleuchtet wird. Gerade diese Nähe kann jedoch auch die Tendenz zu sozial erwünschtem Antwortverhalten, das bereits in Kapitel 2.2.1.1. angesprochen wurde, erhöhen (Bogner & Landrock, 2015). Die Auswertung der Interviews bzw. Diskussionen erfolgt über deren Transkription und Inhaltsanalysen, ähnlich der Auswertung der schriftlichen Texte. Das macht das Verfahren gleichzeitig wesentlich aufwändiger und zeitintensiver als die Analyse der von Studierenden verfassten Texte, sodass auf diese Weise nur kleinere Stichproben bewältigt werden können.

Als weitere mögliche Alternative, die die bisher genannten Probleme eventuell umgehen könnte, böte sich eine Verhaltensbeobachtung, beispielsweise auch mithilfe von Videoaufzeichnungen, an. Diese Art des Messens ginge von der Prämisse aus, dass sich ein erfolgreicher Reflexionsprozess in einer Verhaltensänderung oder in mündlichen sprachlichen Äußerungen niederschlagen müsste. Eine solche Beobachtung wäre freier von sozialer Erwünschtheit und Verzerrungen durch Selbsteinschätzung oder sprachliche Probleme, hätte aber die Schwierigkeit, dass ein durch Reflexion verändertes Verhalten möglicherweise nicht genau im Beobach-

tungszeitraum auftreten würde. Außerdem wäre schwer zu prüfen, ob die Verhaltensänderungen tatsächlich auf Reflexionsprozesse oder evtl. auf andere Ursachen zurückzuführen wären, da nicht alle potenziellen Einflüsse kontrolliert werden können. Die Validität einer solchen Messung könnte außerdem durch nicht-natürliches Verhalten der Versuchspersonen beeinträchtigt werden, das der Beobachtungssituation geschuldet sein könnte (Paradis & Sutkin, 2017).

Eine höhere Zuverlässigkeit der Ergebnisse könnte durch die Kombination und Triangulation verschiedener Verfahren erreicht werden (z.B. Allas et al., 2017; Wyss, 2011).

2.2.3.4 Rahmenbedingungen

Hinsichtlich der Rahmenbedingungen einer Reflexionssituation scheinen vor allem drei Aspekte von Bedeutung zu sein: Vertrauen, bewertungsfreier Raum und Freiwilligkeit.

Wenn Reflexionen in sozialen Kontexten stattfinden sollen – seien dies Peer-Gespräche oder Interviews – aber auch, wenn die Reflexionen verschriftlicht werden und später von Dozent*innen und Forscher*innen eingesehen werden, ist eine wichtige Voraussetzung dafür eine entsprechende Vertrauensbasis. Das „Öffentlichmachen“ von Reflexionen stellt sich als besonders schwierig dar, wenn die Personen, die Zugang zur Verschriftlichung haben, als Bewertende wahrgenommen werden. In einem Peer-Interview äußerte sich ein*e Studierende*r zur Frage, ob er oder sie das Gefühl habe, bei den schriftlichen Reflexionen ehrlich sein zu können, so: „Also das Herz ausschütten tue ich einer Freundin oder eventuell einer netten Kollegin. Bei meinen Betreuern mache ich das eher nicht [...] Für mich ist sie [die Betreuerin; M.R.] eine Vorgesetzte, von der man abhängig ist“ (Rosenberger, 2017, 197f). Auch auf Peer-Ebene scheinen Reflexionsgespräche gewinnbringender zu sein, wenn die Zusammensetzung der Reflektierenden auf Sympathie und einer gemeinsamen Grundhaltung basiert (Raaflaub et al., 2019, S. 55). Wenn also analysierte Reflexionen die nötige Tiefe und persönliche Involviertheit vermissen lassen, könnte dies den Ursprung in mangelndem Vertrauen und geringer Nähe zu den Beurteilenden oder Mitreflektierenden haben und nicht (nur) auf mangelnde

Fähigkeit zurückzuführen sein. Slepcevic-Zach und Stock stellen fest, dass die Preisgabe persönlicher Informationen in akademischen Settings problematisch ist und nicht vorausgesetzt werden kann: „The core of the problem was the unusual focus on personal information in an academic setting since the ePortfolio-work includes addressing personal aspects“ (2018, S. 296). Das Problem fehlender Vertraulichkeit und Privatheit verstärkt sich außerdem im digitalen Raum. Wie bereits dargestellt wurde, werden aktuell viele Reflexionsaufgaben im Rahmen von E-Portfolios, Blogs oder Social-Media-Kanälen gestellt, in denen noch weniger klar ist, wer die Inhalte später einsehen kann und wie diese weiterverwertet werden (Ghaye, 2007).

Neben der fehlenden Vertrauensbasis besteht ein weiterer problematischer Punkt bei Reflexionsaufgaben darin, dass diese häufig Gegenstand von Bewertung sind. Brosziewski (2013) beschreibt ausführlich die aus einer Bewertung von Reflexionsportfolios resultierenden Probleme. Pineker und Störtländer (2013) weisen darauf hin, dass für Studierende selbst dann das Gefühl des Beurteiltwerdens noch einen wesentlichen Einfluss hat, wenn rein formal gar keine Bewertung vorgenommen wird (vgl. S. 202). Das Bewerten, so beispielsweise Boud (2010), unterdrücke echte und aufrichtige Reflexion, wie man sie eigentlich durch Portfolios oder Reflexionstagebücher erzeugen will. Er kritisiert deshalb „inappropriate assessment procedures such as the marking of raw reflective journals that has the effect of inhibiting the expression of any reflection by learners that might have been fostered through keeping a journal“ (S. 27). Hays und Gay berichten von strebsamen Studierenden, die mit ihren Reflexionen gute Bewertungen erzielen wollen und deshalb besonders lange Texte verfassen oder Webseiten nutzen, die sie – zumindest scheinbar – zum reflexiven Schreiben anleiten: „Unfortunately, the latter strategy often includes using websites that offer sentence structures and keywords that suggest reflection and potentially attract good marks“ (2011, S. 117). Artmann et. al (2013) weisen darauf hin, dass das Wissen der Studierenden um die Einsicht in die Reflexionen und Besprechungen mit den Mentor*innen zu „einem Schreiben im Sinne der sozialen Erwünschtheit führen“ (S. 142) könnte. Hauser beschreibt, dass sich folglich daraus, dass beispielsweise „gelegentlich in einer Reflexion dem Dozenten zuliebe auch noch eine wissenschaftliche Quelle“ angegeben werde, nicht ableiten

lasse, dass jemand sich tatsächlich am wissenschaftlichen Wissensbestand orientiere (Hauser, 2021, S. 29). Dass dadurch selbstverständlich auch Messergebnisse verzerrt werden können, wurde bereits erläutert.

Abschließend ist noch zu erwähnen, dass Freiwilligkeit einen wesentlichen Faktor für authentische Reflexionen darstellt, die zu häufig als Pflichtaufgaben erzwungen werden. Durch die Popularität, die Reflexionen im universitären Kontext zunehmend erfahren, und dem gleichzeitigen Druck, der unter anderem durch ein modularisiertes Studium entsteht, ist auch oft nicht genügend Zeit vorhanden, sich auf die vielfältigen Reflexionsangebote und -verpflichtungen einzulassen. Dozierende und Forscher*innen stellen deshalb in zunehmendem Maße fest, dass beispielsweise „Reflexionen im Seminar [...] in unterschiedlichem Maße immer wieder auch im Modus des Abarbeitens bewältigt“ werden. (Berndt & Häcker, 2017, S. 248). Dass sich der Druck im Studium, beispielsweise während der Prüfungszeit, negativ auf die Qualität und den Umfang von Reflexionen auswirkt, stellte auch Leonhard in seiner Untersuchung fest: „[Es wird] deutlich, dass die Messung von Reflexionskompetenz konstitutiv an die *Bereitschaft* der Studierenden zur Reflexion gebunden ist“ (Leonhard, 2013, S. 190; Hervorhebungen im Original). Wenn Reflexionsaufgaben zu erledigen sind und der eigene Antrieb dazu fehlt, kann das auch dazu führen, dass etwas erfunden wird, wie ein*e Studienteilnehmer*in in der Untersuchung von Rosenberger schildert: „Es ist halt oft so der Fall, mir fällt nichts ein, und dann dichtet halt etwas hin, dass zufrieden sind, die Lehrer“ (2017, S. 197). Die Gefahr der Reflexionsermüdung, wie sie auch in Kapitel 2.1.2.9 schon angesprochen wurde, der Rückgang von Motivation beispielsweise in Prüfungszeiten und erdichtete Reflexionen, um die Aufgabe zu erfüllen, sind also Einflussfaktoren, die nicht außer Acht gelassen werden dürfen.

Das Fehlen von Vertrauen und Privatheit sowie die Bewertung und das Forcieren von Reflexionen sind also Einflüsse, die die Validität von Messungen beeinträchtigen können. Beispielsweise im Fall der „hingedichteten“ Reflexion würden die Ergebnisse wohl erheblich verfälscht. Speziell wenn Reflexionen Forschungsgegenstand sein sollen, gilt es dies zu berücksichtigen, aber auch unabhängig davon sollten diese Überlegungen im didaktischen Kontext eine größere Rolle spielen.

2.2.3.5 Zusammenfassung und Bewertung

Im Kontext der Messung von Reflexion gibt es also einige relevante Punkte, die teilweise noch ungeklärt sind, in jedem Fall aber bisher zu wenig Beachtung gefunden haben. Sie sind deshalb so wichtig, weil sie die Voraussetzung dafür schaffen, dass Reflexion überhaupt sichtbar und damit messbar wird. Da es sich um einen Vorgang handelt, der sich im Individuum abspielt, kann ein Zugriff darauf nur erfolgen, wenn die richtigen Parameter dafür vorliegen und eine Umgebung geschaffen werden kann, die dazu anregt, Gedanken und Gefühle zu teilen: „[We] only have access to individuals' thoughts and feelings through what individuals choose to reveal about themselves“ (Boud et al., 1985b, S. 11).

Es wurde gezeigt, dass die Wahl des Auslösers und Anlasses einen entscheidenden Einfluss auf die Reflexion haben kann, es aber nicht abschließend geklärt ist, ob sich dafür eigene Erlebnisse oder fremde Erfahrungen besser eignen. Das gleiche Problem besteht bei der Formulierung von Aufgabenstellungen und Prompts, bei denen vor allem noch unklar ist, wie detailliert und genau diese formuliert sein sollten. In Studien zeigen sich zwar häufig höherwertige Reflexionen bei ausführlicheren und strukturierteren Aufgabenstellungen. Diese könnten aber gleichzeitig die Validität der Messung beeinträchtigen, weil sie schon im Vorhinein das Ergebnis in die gewünschte Richtung lenken. Während einige Forscher*innen die eigenen Reflexionsaufgaben und -anlässe kritisch in die Interpretation der Ergebnisse einbeziehen (Cherrington & Loveridge, 2014; Weinberger & Seyfried, 2009), gibt es viele Studien, die diese nicht einmal nachvollziehbar berichten, woraus sich schließen lässt, dass insgesamt zu wenig Bewusstsein für den Einfluss dieser Aspekte vorhanden ist.

Da Reflexion nicht direkt beobachtbar ist, stellt sich ferner die Frage, wie sie sichtbar gemacht werden kann. Unterschiedliche Möglichkeiten, beispielsweise Selbstauskünfte, schriftliche Darstellung oder Interviews, wurden aufgezeigt und deutlich gemacht, dass alle gewisse Vor- und Nachteile mit sich bringen, die bedacht und abgewogen werden müssen. Das trifft selbstverständlich auch für die am häufigsten eingesetzte Variante der Verschriftlichung von Reflexionen zu.

Abschließend wurde auf Rahmenbedingungen wie Vertrauen, Freiwilligkeit und Bewertungsfreiheit eingegangen, die für ein sensibles Thema wie Reflexionen eine

besonders wichtige Rolle spielen und die auch bereits im ersten Teil der Arbeit, in Kapitel 2.1.2.6, als bedeutsam hervorgehoben wurden.

Dass alle hier untersuchten Bedingungen möglicherweise Einfluss auf das Reflexionsergebnis nehmen, wurde an mehreren Stellen gezeigt. Wenn sie nicht bedacht, kontrolliert und berichtet werden, könnten die Messergebnisse falsch interpretiert werden. Diese Faktoren gilt es stärker zu beforschen und in den Fokus zu rücken, was in der vorliegenden Untersuchung geschehen soll.

3 Fragestellung und Hypothesen

3.1 Fragestellung

Die bisherigen Ausführungen zeigen, dass es im Kontext von Reflexion, sei es bezüglich der Wirkungen, der Förderbarkeit, der Entwicklung und anderer Bereiche, noch sehr viele offene Forschungsfragen gibt, die die bisherigen Studien nicht konklusiv beantworten konnten. Dafür gibt es drei wesentliche Ursachen, die unmittelbar miteinander zusammenhängen: Erstens ist die Operationalisierung des komplexen Konstrukts Reflexion schwierig, zweitens sind die bisherigen Messversuche uneinheitlich und nicht vergleichbar und drittens bestehen erhebliche Zweifel an ihrer Validität. Eine Möglichkeit, dieses Problem zu lösen, bestünde darin, ein neues Messverfahren zu entwickeln, das das Konstrukt zuverlässiger erfassen kann. Dass es diesbezüglich einige möglicherweise vielversprechende Ansätze gibt, wurde bereits gezeigt, aber zum aktuellen Zeitpunkt erscheint kein alternatives Verfahren gewinnbringender als die bestehenden. Eine zweite Möglichkeit, mit diesem Problem umzugehen, ist, die bereits bestehenden Verfahren systematischer zu untersuchen und zu überprüfen. Messverfahren sollten nicht nur einmalig für punktuelle Untersuchungen eingesetzt werden, sondern selbst zum Untersuchungsgegenstand werden. Sie müssen einer gründlichen Überprüfung hinsichtlich der Gütekriterien unterzogen werden. Dabei gilt es vor allem auch die in Kapitel 2.2.3 herausgearbeiteten Rahmenbedingungen der Messung mehr zu berücksichtigen. Mit diesem Vorgehen könnten sich möglicherweise eines oder mehrere bestehende Verfahren, evtl. mit Anpassungen, herauskristallisieren, die das Konstrukt Reflexion so zuverlässig erfassen können, dass das Verfahren auch von anderen eingesetzt werden kann und die Forschungsarbeiten dadurch vergleichbarer werden.

Die vorliegende Arbeit will einen Schritt in diese Richtung unternehmen und eines der bisher am häufigsten angewandten Messverfahren auf Validität überprüfen. Ausgewählt wurde dafür die inhaltsanalytische Auswertung von Reflexionstexten mithilfe des Stufenschemas von Hatton und Smith (1995), das die Reflexionstiefe erfasst. Für diese Messmethode soll untersucht werden, ob sie ein valides Ergebnis

liefern kann oder ob die Messung möglicherweise durch bisher nicht beachtete Einflussfaktoren verfälscht wird. Häcker (2017) formuliert als eine der zentralen Leitfragen für die reflexive Lehrkräftebildung: „Werden bei der Einschätzung der Qualität von Reflexionsprozessen und -ergebnissen die jeweils gegebenen Rahmenbedingungen der Reflexionsanforderungen (etwa Anlass bzw. Veranlassung) angemessen berücksichtigt?“ (S. 31). Im Anschluss an dieses Desiderat werden solche Rahmenbedingungen in den Blick genommen. Dabei wird besonders auch auf den Reflexionsanlass fokussiert. Es ergibt sich als erste Forschungsfrage:

1. Welche Rolle spielt der reflexionsauslösende Anlass als Rahmenbedingung für die Validität der Messung?

Um die Vergleichbarkeit der Messungen zu erhöhen, ist es ein zentrales Anliegen, nicht nur auf der Ebene der Messverfahren eine Vereinheitlichung zu erzielen, sondern auch die Rahmenbedingungen der Messung zu berücksichtigen. Im Fokus dieser Arbeit steht insofern die Aufgabenstellung mit dem Reflexionsanlass, der zur Reflexion anregen soll. Neben der Frage der Vergleichbarkeit liegt hier auch ein Erkenntnisinteresse hinsichtlich der Validität der Messung: Es wäre denkbar, dass in bisherigen Studien den Teilnehmenden eine geringe Reflexionsfähigkeit attestiert wurde, obwohl es nur der Aufgabe oder Thematik nicht gelungen ist, eine ernsthafte, tiefergehende Reflexion auszulösen. Im Speziellen soll in dieser Studie untersucht werden, wie sich der Grad der persönlichen Involviertheit bei der Aufgabenstellung auf verschiedene Faktoren auswirkt. Ist es wichtig, dass man sich bei einer Reflexion mit eigenen Erlebnissen auseinandersetzt oder gelingt Reflexion auch mit Erfahrungen aus zweiter Hand? Dieser Frage wird in den **aufgabenbezogenen Hypothesen** nachgegangen.

Zweitens wird der Einfluss einiger weiterer Faktoren auf das Messergebnis untersucht, die sich aus den vorgestellten theoretischen Ergebnissen aus dem ersten Teil der Arbeit ableiten, woraus sich als zweite Forschungsfrage ergibt:

2. Prädizieren bisher nicht kontrollierte weitere Faktoren das Messergebnis?
Ziel dieser Forschungsfrage ist es, die Validität aus einer zweiten Richtung zu untersuchen. Auch hier geht es wieder um das Aufdecken bisher nicht

identifizierter Einflüsse auf das Messergebnis. Dabei werden sowohl situationalen Faktoren als auch Persönlichkeitsvoraussetzungen in den Blick genommen. Auch das uneinheitliche Verständnis des Begriffs Reflexion und der empfundene Problemgehalt könnten eine Rolle spielen.

Dieser Frage wird in den **prädiktorbezogenen Hypothesen** nachgegangen.

3.2 Hypothesen

3.2.1 Aufgabenbezogene Hypothesen

Das Kapitel 2.2.3.1, in dem der Forschungsstand zur Frage nach dem geeignetsten Reflexionsanlass dargestellt wurde, wurde mit einem uneindeutigen Fazit abgeschlossen. Dewey geht davon aus, dass Reflexion nur von Situationen und Problemen mit persönlichem Bezug ausgelöst werden kann und unmittelbar mit dem Erleben des Individuums verknüpft ist. Einige Forscher*innen geben ebenfalls solchen Reflexionsanlässen den Vorzug. Andere wiederum kommen zu dem Ergebnis, dass bessere Reflexionen ausgelöst werden, wenn es sich um unpersönliche Situationen handelt, zu denen eine gewisse Distanz besteht. Diese Frage soll hier noch einmal systematisch in den Blick genommen werden. Dabei wird von der Theorie Deweys ausgegangen: Wenn Reflexion persönliche Erfahrungen und eine individuelle Involviertheit voraussetzt, sollten Erfahrungen aus zweiter Hand und Situationen ohne persönlichen Bezug weniger oder gar nicht zur Reflexion anregen. Um diese Annahme zu testen, soll anhand dreier unterschiedlich persönlicher Reflexionsaufgaben geprüft werden, ob die eigenen Erfahrungen stärker zur Reflexion anregen, als es Erfahrungen aus zweiter Hand leisten können. Daraus ergibt sich die folgende Hypothese:

H 1.1: Je größer der über die Aufgabe induzierte persönliche Bezug zum Reflexionsproblem ist, desto höher ist die Reflexionsleistung.

In einzelnen Studien wurde außerdem vermutet, dass ein persönlicherer Bezug zur Aufgabe in Form eigener Erlebnisse und Praxiserfahrungen eine höhere Motivation auslösen könnte (Göbel & Gösch, 2019; Kosinár, 2017; Seidel et al., 2011). Diese

höhere Motivation könnte ursächlich sein für eine höhere Reflexionsleistung (siehe Hypothese 2.1). Zunächst muss aber geprüft werden, ob Aufgaben mit persönlichem Bezug tatsächlich eine höhere Motivation zur Folge haben und ob sie als sinnvoller und nützlicher eingeschätzt werden als unpersönlichere Reflexionsanlässe aus zweiter Hand. Daraus ergibt sich als zweite Hypothese:

H 1.2 Je größer der über die Aufgabe induzierte persönliche Bezug zum Reflexionsproblem ist, desto mehr Motivation löst sie aus und desto höher wird ihr Nutzen eingeschätzt.

In Kapitel 2.2.1.2 wurde das Stufenschema von Hatton und Smith (1995) bereits vorgestellt und gezeigt, dass es unter anderem dafür kritisiert wurde, dass es bestimmte für Reflexion wichtige Aspekte nicht abdeckt. Das sind einerseits Bezüge zur fachwissenschaftlichen Literatur, aber auch die in Kapitel 2.1.2.7 beschriebenen Emotionen, die beim Reflektieren eine Rolle spielen können. Diese beiden Aspekte werden zusätzlich zum Stufenschema aus den Texten herausgearbeitet. Einerseits, um ein noch komplexeres Bild der Reflexionen zu erhalten und zu prüfen, ob die drei Reflexionsanlässe noch weitere systematische Unterschiede in den Texten bewirken. Andererseits, um einer weiteren Frage nachzugehen, die sich im Kontext des Reflexionsanlasses aufgetan hat: Manche Forscher*innen vermuten, dass die Reflexion zu fremden Erlebnissen deshalb leichter falle, weil weniger persönliche Involviertheit vorhanden sei (Poom-Valickis & Mathews, 2013). Während fremde Videos eher zu einer distanzierten, kritischeren Haltung führten, lösten Reflexionen eigenen Unterrichts eine stärkere Betroffenheit aus (Göbel & Gösch, 2019). Die Evidenz zu dieser Frage ist aber noch gering. Es wird erwartet, dass die Aufgaben ohne persönlichen Bezug eher auf der distanziert-kognitiven Ebene bearbeitet werden und sich demzufolge mehr Theoriebezüge und weniger Referenzen zu eigenen Gefühlen finden. Für die Aufgaben mit persönlichem Bezug und insbesondere für die Aufgabe, die auf ein persönliches Erlebnis referiert, wird der entgegengesetzte Effekt erwartet:

H 1.3 Je höher der durch die Aufgabe induzierte persönliche Bezug zum Reflexionsproblem ist, desto mehr emotionale und desto weniger kognitive Anteile finden sich in den Reflexionstexten.

3.2.2 Prädiktorbezogene Hypothesen

War bei den aufgabenbezogenen Hypothesen von Interesse, ob sich empfundene Motivation und wahrgenommener Nutzen der Aufgabe zwischen den Aufgabentypen unterscheiden, so soll in einem zweiten Schritt noch geprüft werden, ob Motivation und Nutzen der Aufgabe einen Einfluss auf die gemessene Reflexionsleistung haben. In vorangegangenen Studien wurde bereits ein solcher Effekt der Motivation vermutet (Leonhard & Rihm, 2011), die ja auch von Dewey als Einflussfaktor genannt wird. Es wird angenommen, dass ein höheres Maß an Motivation und empfundenem Nutzen der Aufgabe dazu führen, dass höherwertigere Reflexionen erzeugt werden und damit eine höhere Reflexionsleistung gemessen wird. Werden diese Faktoren nicht kontrolliert, kann es zu einer Fehlinterpretation der Reflexionsergebnisse kommen. Daraus folgt als zu prüfende Hypothese:

H 2.1 Motivation und wahrgenommener Nutzen der Aufgabe präzisieren die Reflexionsleistung.

In Kapitel 2.1.2.6 wurden verschiedene Voraussetzungen für Reflexionsfähigkeit dargestellt, die auch die Qualität der Reflexionen beeinflussen können. Dabei wurde gezeigt, dass einzelne Aspekte, z.B. der Einfluss von Wissen, auch bereits empirisch belegt werden konnten. Andere Punkte wurden bisher nur theoretisch angenommen, ohne dass sie empirisch geprüft worden wären. Die Vermutung liegt also nahe, dass bisher unentdeckte Einflüsse auf das Messergebnis der Reflexionsleistung bestehen könnten, für die bis dato empirische Belege fehlen. Die vorliegende Arbeit will einen Vorstoß in diese Richtung unternehmen und untersucht ausschnittsweise drei theoretisch genannte Voraussetzungen für Reflexionsfähigkeit (ausführlich hergeleitet in Kapitel 2.1.2.6): Offenheit, Ungewissheitstoleranz und Kreativität.

Offenheit für neue Erfahrungen und alternative Sichtweisen sollte dazu führen, dass ein Interesse daran besteht, sich mit einer herausfordernden problematischen Situation auseinanderzusetzen. Kreativität sollte im Rahmen dieses Lösungsprozesses ermöglichen, dass unterschiedliche Ideen und Erklärungsansätze gefunden werden können. Ungewissheitstoleranz sollte dazu führen, dass der Aufschub einer abschließenden Gewissheit so lange hinausgezögert werden kann, wie es nötig ist,

sich intensiv Gedanken zur Situation zu machen und zu kreativen, multiperspektivischen Lösungsansätzen zu gelangen.

Für diese drei Persönlichkeitsmerkmale wird erwartet, dass eine höhere Ausprägung von Offenheit, Ungewissheitstoleranz und Kreativität eine höhere Reflexionsleistung zur Folge haben. Daraus ergibt sich die Hypothese:

H 2.2: Die Persönlichkeitsmerkmale Ungewissheitstoleranz, Offenheit und Kreativität präzisieren die Reflexionsleistung.

Reflexion ist nach Dewey und Schön in erster Linie ein Mittel zum Klären und Lösen von auftretenden Problemen. Eine ihrer zentralen Prämissen ist, dass Reflexionen nur ausgelöst werden, wenn das Individuum mit einer problematischen Situation konfrontiert wird. Es ist also davon auszugehen, dass ein höheres Problemempfinden hinsichtlich des potentiellen Reflexionsauslösers eine intensivere Auseinandersetzung mit der Situation und damit ein höheres Niveau der Reflexionsleistung nach sich zieht. In Kapitel 2.1.2.1 wurde gezeigt, dass manche Forscher*innen davon ausgehen, dass Reflexionen auch unabhängig von einem individuell auftretenden Problem von außen didaktisch induziert werden können (Boud et al., 2013; Jenert, 2008), dass es aber dann zu einer „aktiven Übernahme“ (Häcker, 2017, S. 26) durch das Individuum kommen muss. Geringe Reflexionsleistungen könnten demzufolge Ausdruck eines fehlenden Problemempfindens sein. Der Rückschluss auf eine mangelnde Reflexionskompetenz wäre dann fehlerhaft. In bisherigen Studien wurde nicht kontrolliert, ob die zu reflektierenden Situationen von den Studierenden überhaupt als problematisch wahrgenommen wurden und ob eine solche aktive Übernahme erfolgt ist. Deshalb soll als Hypothese formuliert und untersucht werden:

H 2.3 Der empfundene Problemgehalt des Reflexionsthemas präzisiert die Reflexionsleistung.

Ferner soll geprüft werden, inwiefern das Bild von Reflexion, das die Proband*innen mitbringen, die Reflexionsergebnisse beeinflusst. Hierbei wird kein

Einfluss auf die Reflexionsleistung insgesamt, sondern auf das Vorkommen bestimmter Anteile in den Texten erwartet. In bisherigen Studien konnte gezeigt werden, dass unter dem Begriff Reflexion teilweise sehr unterschiedliche Konzepte verstanden werden, wie in Kapitel 2.1.1.2 dargestellt wurde. Im Alltagsverständnis spiegelt sich die definatorische Vielfalt des Begriffs wider. Da Reflexionsaufgaben, wie auch in dieser Studie, in vielen Fällen den Begriff Reflexion nennen oder sogar als Operator verwenden, besteht Grund zu der Annahme, dass die daraus entstehenden Reflexionstexte dem individuellen Begriff von Reflexion entsprechend verfasst werden. Wird also Reflexion beispielsweise als Begriff verstanden, der eine multiperspektivische Herangehensweise erfordert, so würden sich folglich, wenn die Hypothese zutrifft, in den Texten mehr multiperspektivische Abwägungen und Elemente finden. Diese könnten dann als Qualitätsmerkmal von Reflexion gewertet werden. Wenn ihr Auftreten aber vom individuellen Verständnis des Reflexionsbegriffs abhängt und keinen Ausdruck einer grundsätzlichen Kompetenz darstellt, so läge hier eine Validitätseinschränkung der Messung vor. Um zu prüfen, ob eine solche Einschränkung vorliegen könnte, wird als Hypothese formuliert:

H 2.4 Die Erwartungen der Proband*innen zu Reflexionsaufgaben präzisieren die in den Texten gefundene Menge an reflexionsspezifischen Elementen.

4 Methode

4.1 Stichprobe

Untersucht wurden insgesamt $N = 95$ Lehramtsstudierende, die sich zum Erhebungszeitpunkt durchschnittlich zwischen dem vierten und fünften Semester befanden ($M = 4.59$, $SD = 2.00$). 29 Versuchspersonen waren männlich (entspricht 31 %), 62 Personen weiblich (entspricht 65 %), vier Personen gaben keine Angabe zum Geschlecht, die Kategorie „divers“ wurde nicht gewählt.

Die Erhebung wurde an zwei bayerischen Universitäten durchgeführt; 41 Praktikant*innen der Friedrich-Alexander-Universität Erlangen/Nürnberg und 52 Praktikant*innen der Universität Passau nahmen teil. Zwei Personen gaben nicht an, welcher Universität sie angehören.

An den beiden Universitäten werden standardmäßig das pädagogisch-didaktische Praktikum angeboten, das 26 Studierende zum Zeitpunkt der Untersuchung absolvierten (Studiensemester: $M = 3.50$, $SD = 1.11$) sowie die beiden Alternativpraktika „Exercitium Paedagogicum“ (19 Teilnehmer*innen, Studiensemester: $M = 3.95$, $SD = 2.51$) und „Lehr:werkstatt“ (50 Teilnehmer*innen, Studiensemester: $M = 5.42$, $SD = 1.80$).

Reflexionsaufgabe 1 wurde von 32 Studierenden (davon 10 männlich) bearbeitet, Reflexionsaufgabe 2 von 31 (davon 10 männlich) und Reflexionsaufgabe 3 von 29 (davon 9 männlich).

Die Bearbeitung der Reflexionsaufgabe und das Ausfüllen des Fragebogens waren zwar im Rahmen des verpflichtenden praktikumsbegleitenden Portfolios obligatorisch, die darüber hinausgehende Teilnahme an der Studie jedoch explizit freiwillig. Die Zustimmung zur Teilnahme wurde von allen Studierenden gegeben. Von zwei Proband*innen liegt allerdings kein Fragebogen vor und drei Proband*innen reichten keinen Text zur Reflexionsaufgabe ein. Sie wurden nicht aus der Studie ausgeschlossen, da jeweils entweder die Fragebogendaten oder die Reflexionsaufgabe

vorhanden waren und als zusätzliches Datenmaterial zur Vergrößerung der Stichprobe in die Auswertung miteinfließen.

4.2 Instrumente

4.2.1 Aufgaben

Für die Untersuchung wurden drei Aufgaben entworfen, die bezüglich des Grades der persönlichen Involviertheit variieren, aber hinsichtlich Thematik und Aufgabenstellung identisch oder sehr ähnlich sind. Der Grad des persönlichen Bezugs zum Reflexionsproblem steigt von Aufgabe 1 bis Aufgabe 3. Die Formulierungen der eingesetzten Aufgaben finden sich in Abbildung 1.

Aufgabe 1:

Eine Lehrkraft unterrichtet eine neue Klasse, in der auch Paul, ein sehr unruhiger und impulsiver Schüler, ist. Er ist bei anderen Lehrkräften schon durch freche Bemerkungen aufgefallen. Eine Kollegin hat angedeutet, dass er in einem sehr schwierigen familiären Umfeld lebt.

Sie plant gerade eine Unterrichtsstunde und denkt mit Sorge daran, dass er auch ihren Unterricht stören könnte.

Wie könnte die Lehrkraft dieses Problem lösen?

Notieren Sie auf nachvollziehbare Weise alle Ihre Überlegungen diesbezüglich!

Aufgabe 2:

Sie dürfen im Praktikum einen Unterrichtsversuch halten.

In der Klasse, in der Sie unterrichten werden, ist auch Paul, ein sehr unruhiger und impulsiver Schüler, der Ihnen schon häufiger durch freche Bemerkungen aufgefallen ist.

Ihre Praktikumslehrkraft hat kürzlich angedeutet, dass er in einem sehr schwierigen familiären Umfeld lebt.

Sie planen gerade Ihre Unterrichtsstunde und denken dabei mit Sorge daran, dass er auch Ihren Unterricht stören könnte.

Wie könnten Sie dieses Problem lösen?

Notieren Sie auf nachvollziehbare Weise alle Ihre Überlegungen diesbezüglich!

Aufgabe 3:

Denken Sie an Ihr Praktikum zurück und rufen Sie sich eine Situation aus dem Kontext Klassenführung/Unterrichtsstörungen ins Gedächtnis, die für Sie entweder überraschend kam oder die Sie als irritierend oder problematisch empfunden haben.

Wenn Sie noch einmal in dieser Lage wären:

Wie könnten Sie dieses Problem lösen?

Notieren Sie auf nachvollziehbare Weise alle Ihre Überlegungen diesbezüglich!

Abbildung 1. Formulierung der drei eingesetzten Reflexionsaufgaben.

Die Formulierung der Aufgaben begründet sich wie folgt:

Persönliche Involviertheit. Mit den drei Aufgaben sollte ein Spektrum von sehr unpersönlicher Reflexionssituation bis sehr persönlicher Reflexionssituation abgedeckt werden. Auch wurde der Hinweis von Weinberger und Seyfried (2009) berücksichtigt, deren Instrument ausschließlich Situationen ohne persönlichen Bezug beinhaltete und aus dessen Erprobung sie aufgrund sehr flacher Reflexionen folgerten, dass die Reflexionsanlässe persönlicher gestaltet werden sollten, beispielsweise, indem die Studierenden den Auftrag erhalten, sich vorzustellen, sie wären selbst in dieser Situation (vgl. S. 93). Dieser Vorschlag wurde für Aufgabe 2 übernommen. Aufgabe 1 wurde als Problem einer fremden Lehrkraft geschildert und ist damit sehr unpersönlich. Aufgabe 3 bezieht sich auf ein eigenes Erlebnis

aus dem Praktikum und basiert damit auf der individuellen und persönlichen Erfahrung der Teilnehmenden. Aufgabe 2 wurde als Mischform formuliert. In ihr wird zwar das gleiche Problem wie in Aufgabe 1 vorgegeben, die Studierenden sollen sich aber in die Situation hineinversetzen und sich vorstellen, dass ihnen dieses Problem im Praktikum selbst begegnet.

Vergleichbarkeit. Um die Vergleichbarkeit der Aufgaben zu gewährleisten, wurde ein gemeinsames Rahmenthema festgelegt. Die Schilderung eines eigenen Erlebnisses aus dem Praktikum hätte zu viele verschiedene Aspekte zugelassen. Deshalb wurde ein Thema gewählt, mit dem jeder Praktikant und jede Praktikantin im Praktikum höchstwahrscheinlich in Berührung kommt, damit gewährleistet werden konnte, dass entsprechende Szenarien auch erlebt wurden und erinnert werden können. Zudem sollte das Thema ein gewisses Problempotential haben, das heißt, es sollte sich um einen Aufgabenbereich von Lehrkräften handeln, der störanfällig ist. Befragungen von Lehrkräften haben wiederholt ergeben, dass die Themen Disziplin im Klassenzimmer und Verhaltensauffälligkeiten von Schüler*innen als besonders problematisch und potenziell belastend erlebt werden (Schaarschmidt & Kieschke, 2013). Eine Studie von Liu (2022) zeigte außerdem, dass Lehramtsstudierende sich in ihren Reflexionen mehrheitlich auf Klassenführungsaspekte bezogen, wenn ihnen das Thema freigestellt wurde. Auch Poom-Valickis und Mathews (2013) halten Disziplinprobleme für das Thema, das Lehramtsanfänger*innen am meisten beschäftigt und erhielten mit diesem Rahmenthema reichhaltige Ausführungen von Studierenden im Praktikum. Deshalb wurde es als thematischer Rahmen in der vorliegenden Studie gesetzt.

Operator und Formulierung der Aufgaben. Die konkrete Formulierung der Aufgaben ergibt sich aus den theoretischen Überlegungen in Kapitel 2.2.3.2 zu Aufgabenformulierung und Prompts. Dort wurde zwar festgestellt, dass eine detailliertere Aufgabenstellung mit Lösungshinweisen und Operatoren bessere Reflexionstexte erzeugte. Zur Reflexionskompetenz gehört aber auch, dass Phasen und Ebenen eines Reflexionsprozesses spontan und ohne Hilfestellung durchlaufen werden können und diese nicht bereits durch untergliederte Aufgabenstellungen in der Art eines Leitfadens oder Advance Organizers vorweggenommen werden. Eine Formulierung der Aufgabe gemäß der erwarteten Tätigkeiten würde, wie beispielsweise auch

von Leonhard und Rihm kritisch angemerkt wird, zu einer „Prädisposition der Reflexionen“ führen (Leonhard & Rihm, 2011, S. 258). Deshalb werden hier die gewünschten Aspekte nicht durch Operatoren wie „Beschreiben Sie“, „Erklären Sie“ oder „Wägen Sie ab“, vorweggenommen, sondern die Reflexionsaufgabe abstrakter formuliert.

Zentrale Aufgabe der Reflexion ist nach Dewey das Lösen eines Problems (ausführlich siehe Kapitel 2.1.2.2 zum Ziel der Reflexion), sodass letztendlich die Formulierung als Problemlöseimpuls gewählt wurde.

Beschreibung des Fallbeispiels. Die Schilderung der problematischen Klassenführungssituation in den beiden ersten Aufgaben wurde so detailreich formuliert, dass sie einen Anlass für konkrete Überlegungen darstellt, andererseits nicht so ausführlich, dass die Fülle an möglichen Lösungen zu sehr eingeschränkt wird. In Kapitel 2.2.2.1 wurde dargestellt, dass die meisten Studien zu dem Schluss kommen, dass sich wenig moralische, ethische und systemkritische Reflexionen in den Texten finden und die Autor*innen vereinzelt anmerken, dass dies auch den eigenen Aufgabenstellungen und Reflexionsanlässen geschuldet sein könnte. Deshalb sollte die Beschreibung der Situation mit dem Schüler Paul auch die höchste Stufe der Reflexion, die moralischen Erwägungen, ermöglichen. Insofern wurde er nicht einseitig als mutwilliger Störer beschrieben, sondern auch die schwierige familiäre Situation erwähnt, die möglicherweise ein komplexeres Herangehen auf verschiedenen Ebenen nötig macht, da auch er ein Opfer der sozialen Umstände sein könnte. Es sollte damit ermöglicht werden, dass vielfältige Aspekte und Rahmenbedingungen erwogen werden können, wenn über das Problem nachgedacht wird und damit auch zum Nachdenken über unterschiedliche Ursachen von Störungen oder Problemen und mögliche alternative Handlungsweisen angeregt werden.

Sichtbarmachung der Reflexionsprozesse über Verschriftlichung. In Kapitel 2.2.3.3 zur Sichtbarmachung der Reflexion über ein Medium wurde herausgearbeitet, dass in der bisherigen Forschung am häufigsten schriftliche Texte zur Erhebung von Reflexion herangezogen wurden. Gleichzeitig wurde gezeigt, dass es begründete Kritik an diesem Vorgehen gibt. In der vorliegenden Untersuchung wurde trotzdem auf schriftliche Reflexionen zurückgegriffen, weil für die Triangulation mit den quantitativen Daten eine möglichst große Stichprobe erzeugt werden

musste. Dies wäre mit Interviews oder Verhaltensbeobachtungen nicht zu bewältigen gewesen. Einige erwähnte Probleme mit schriftlichen Reflexionsdokumenten, wie beispielsweise mögliche Geschlechterunterschiede, wurden aber kontrolliert.

4.2.2 Fragebogen

Der Online-Fragebogen, der in der Studie zum Einsatz kam, bestand aus vier Teilen. Im ersten Teil wurden soziodemographische Daten erhoben, im zweiten Teil allgemeine Fragen zum Verständnis des Reflexionsbegriffs gestellt, im dritten Teil wurden spezifische Aspekte zur vorliegenden Reflexionsaufgabe (Problemgehalt, Motivation, Nutzen) erfragt und im vierten Teil wurden einige Persönlichkeitsmerkmale (Ungewissheitstoleranz, Offenheit, Kreativität) erhoben (Fragebogen siehe Anhang 1).

Der aus den angegebenen Elementen zusammengesetzte Fragebogen wurde in einem Pretest in einer Stichprobe von $N = 112$ Lehramtsstudierenden der Universität Würzburg eingesetzt. Relevante Ergebnisse des Pretests werden im Folgenden an den jeweils betreffenden Stellen berichtet.

Teil 1: Soziodemographische Daten. Zur Kontrolle möglicher Einflüsse wurden das Geschlecht, das Fachsemester, die besuchte Universität und die gewählte Praktikumsform erhoben.

Teil 2: Erhebung des Reflexionsverständnisses. Zur Erhebung des Reflexionsverständnisses wurde in acht Items erfragt, wie bedeutsam die jeweils aufgeführten Aspekte im Rahmen einer Reflexion eingeschätzt werden (Frage: „Für wie wichtig halten Sie die folgenden Aspekte bei der Bearbeitung einer Reflexionsaufgabe?“ Beispielitem 1: „Sich auf eigene Erlebnisse und Beobachtungen zu beziehen“). Für jedes Item konnte ein Schieberegler zwischen 0 (*gar nicht wichtig*) und 100 (*sehr wichtig*) platziert werden. Sieben dieser Items wurden aus den beiden Kategoriensystemen der qualitativen Inhaltsanalyse abgeleitet (siehe Kapitel 4.4.1.1) und spiegeln die dort herausgearbeiteten Aspekte wider. Das heißt, eine Kategorie der Inhaltsanalyse war beispielsweise „Bezug zu fachwissenschaftlicher Theorie“ und

als zugehöriges Item wurde formuliert: „Für wie wichtig halten Sie bei einer Reflexion: Fachwissenschaftliche Theorien in die Überlegungen miteinzubeziehen“. Damit wurde geprüft, ob beispielsweise das Fehlen fachwissenschaftlicher Theoriebezüge auf fehlendes Wissen zurückzuführen ist oder ob im Rahmen einer Reflexionsaufgabe eventuell schlicht keine Notwendigkeit dafür gesehen wurde.

Zusätzlich wurde in einem Item kontrolliert, ob die gewählte Aufgabenstellung „Wie würden Sie das Problem lösen?“ von den Proband*innen tatsächlich als Operator für Reflexionsprozesse wahrgenommen wurde (Item: „Für wie wichtig halten Sie: Lösungen für Probleme zu finden?“). Diese Frage ist vor dem Hintergrund relevant, dass die gewählte Aufgabenstellung aus der Theorie abgeleitet wurde, aber in der Praxis nicht die üblicherweise gewählte Formulierung bei einer Reflexionsaufgabe darstellt und dient der Validitätsprüfung der Aufgabenformulierung.

Teil 3: Fragen zur vorliegenden Reflexionsaufgabe

Problemgehalt der Aufgabe. Das Item „Wie schwerwiegend erschien Ihnen das Problem, zu dem Sie reflektiert haben?“ wurde selbst formuliert und konnte auf einer fünfstufigen Skala von 1 (*überhaupt nicht problematisch*) bis 5 (*sehr problematisch*) beantwortet werden.

Motivation und Nutzen der Aufgabe. Zur Erfassung der Motivation wurde das *Intrinsic Motivation Inventory (IMI)* (Deci & Ryan, 1985) in der deutschen Version von Schaal (2006) eingesetzt. Das mehrdimensionale Instrument wird häufig und auch in unterschiedlichen Forschungsfeldern verwendet und hat sich zur Erhebung von tätigkeitsbezogener Motivation und vor allem im Zusammenhang mit Lern- und Leistungssituationen etabliert (Leng et al., 2010; McAuley et al., 1989; Monteiro et al., 2015). Dieses Inventar umfasst sechs Skalen mit insgesamt 45 Items. Die Skalen sind: „Wahrgenommene Kompetenz“, „Anstrengung/Wichtigkeit“, „Wahrgenommene Wahlfreiheit“, „Druck/Anspannung“ (als negativer Prädiktor für Motivation), „Interesse/Vergnügen“ und „Nutzen/Brauchbarkeit“. Die beiden letzteren wurden in der vorliegenden Studie eingesetzt. „Interesse/Vergnügen“ wurde ausgewählt, weil mit dieser Skala die selbstberichtete intrinsische Motivation erhoben wird und „Nutzen/Brauchbarkeit“ gibt einen Hinweis

auf den Grad der Internalisierung der Aufgabe; mit höherer Einschätzung der Nützlichkeit und damit Wichtigkeit einer Aufgabe steigt der Grad der Selbstregulation bei der Ausführung, auch wenn wenig oder kein Interesse an der Aufgabe vorliegt (Deci et al., 1994). Die Items konnten auf einer fünfstufigen Likert-Skala von 1 (*trifft überhaupt nicht zu*) bis 5 (*trifft völlig zu*) beantwortet werden.

Für die Skala „Interesse/Vergnügen“ ergab sich im Pretest eine Reliabilität von Cronbachs $\alpha = .82$ und für „Nutzen/Brauchbarkeit“ von $\alpha = .78$. Diese Werte sind nach Bortz und Döring (2006) als durchschnittlich bzw. akzeptabel zu betrachten. In jeder Skala wurde in der Haupterhebung je ein Item aufgrund geringer Trennschärfen, einer deutlichen Verbesserung der Reliabilität und großer sprachlicher Ähnlichkeit zu anderen Items ausgeschlossen („Während ich die Reflexionsaufgabe bearbeitet habe, habe ich darüber nachgedacht, wie sehr mir das gefällt“, aus der Skala Interesse/Vergnügen und „Ich denke, Reflektieren ist eine wichtige Tätigkeit“ aus der Skala Nutzen/Brauchbarkeit). In der Haupterhebung wurden folglich nur noch je sechs Items pro Skala eingesetzt. In dieser erzielte die Skala „Interesse/Motivation“ einen Wert für Cronbachs $\alpha = .83$ und „Nutzen/Brauchbarkeit“ den Wert $\alpha = .85$.

Teil 4: Erhebung der Persönlichkeitsmerkmale

Ungewissheitstoleranz. Zur Erhebung der Ungewissheitstoleranz wurde die Ungewissheitstoleranzskala (UGT) von Dalbert (1999) eingesetzt. Die acht Items der Skala werden auf einer sechsstufigen Skala von 1 (*stimmt überhaupt nicht*) bis 6 (*stimmt genau*) beantwortet. Die UGT wurde von der Autorin nach mehreren Prüfdurchgängen als homogen, valide und reliabel beurteilt. Die Skala zeigte sich dabei unabhängig von sozial erwünschtem Antwortverhalten. Abhängig von Alter und Geschlecht der Versuchsteilnehmer*innen bewegte sich Cronbachs Alpha um den Wert von ca. $\alpha = .72$. Im Pretest zu dieser Studie konnte dieses Ergebnis mit $\alpha = .70$ nahezu repliziert werden. Auch in der Haupterhebung erwies sich die Reliabilität der Skala als zufriedenstellend ($\alpha = .71$).

Offenheit. Zur Erhebung des Persönlichkeitsmerkmals „Offenheit“ wurde der deutsche Big-Five-Fragebogen von Hartig, Jude und Rauch (2003) eingesetzt, der auf dem *International Personality Item Pool (IPIP)* (Goldberg, 1999) basiert. Es

handelt sich dabei um ein Instrument, das speziell für die Online-Nutzung und als frei verfügbarer und nicht urheberrechtlich geschützter Persönlichkeitstest entwickelt wurde. Eine Auswahl von Items des IPIP wurde von den Autor*innen übersetzt und daraus eine gekürzte Version des ursprünglichen Fragebogens mit jeweils acht Items für jeden der fünf Persönlichkeitsaspekte erarbeitet. Für den hier eingesetzten Fragebogen wurde die Skala „Offenheit“ mit acht Items extrahiert. Diese sind auf einer fünfstufigen Likert-Skala von 1 (*sehr unzutreffend*) bis 5 (*sehr zutreffend*) zu bewerten.

Cronbachs Alpha betrug in der Studie von Hartig et al. für den Faktor Offenheit $\alpha = .76$. Im Pretest zur vorliegenden Studie ($\alpha = .73$) und in der Haupterhebung ($\alpha = .74$) waren die Reliabilitäten etwas niedriger, aber ebenfalls noch akzeptabel.

Bei der Skala Offenheit traten fünf einzelne fehlende Werte auf. Diese konnten jedoch mithilfe des Little's Test als MCAR eingestuft werden, $\chi^2(180) = 193.36$, $p = .24$. Für die fünf Personen mit fehlenden Werten wurde der Skalendurchschnitt aus den sieben verbleibenden Itemwerten gebildet.

Kreativität. Für die Messung der Kreativität wurde der *Alternative Uses Test* von Guilford (1967) eingesetzt. Mehrere einschlägige Tests zur Messung von Kreativität wurden in den 60er Jahren in der psychologischen Forschung entwickelt, beispielsweise von Torrance (1966) oder Mednick (1962). Diese werden bis heute eingesetzt, validiert und weiterentwickelt und gelten als anerkannte Erhebungsverfahren. Als „most elaborate battery of tests to assess creativity“ wurde allerdings der Test von Guilford und Mitarbeitern betrachtet (Taylor & Getzels, 1975, S. 23). Der Test besteht darin, dass die Versuchspersonen für einen einfachen Alltagsgegenstand (im Original ein Ziegelstein, der Gegenstand kann aber variiert werden) möglichst viele alternative Verwendungsmöglichkeiten nennen.

Der *Alternative Uses Test* erfasst nicht alle diskutierten Komponenten von Kreativität, deckt aber drei Unterkategorien divergenten Denkens ab, die hier erhoben werden sollten. Diese sind Fluidität (die Gesamtzahl der genannten Einsatzmöglichkeiten), Flexibilität (Nennungen aus unterschiedlichen Kategorien) und Originalität (Seltenheit der Nennungen). Wie diese im Einzelnen berechnet wurden, wird in Kapitel 4.4.2 erläutert.

4.3 Untersuchungsdesign

Um die vorliegende Fragestellung zu beantworten, wurde eine Querschnittsuntersuchung an Lehramtsstudierenden im ersten studienbegleitenden Praktikum durchgeführt. Dieses wird nach dem erziehungswissenschaftlichen Basismodul ca. im dritten Semester begonnen und ist in Bayern entweder das sechswöchige „pädagogisch-didaktische Praktikum“ oder ein universitätsspezifisches Alternativpraktikum¹.

Für die Studie wurden in einem Between-Subjects-Design (vgl. Charness et al., 2012) drei Treatmentgruppen gebildet und in jeder Gruppe eine der drei Reflexionsaufgaben eingesetzt. Die Zuteilung der Proband*innen auf die Versuchsbedingungen erfolgte randomisiert unter der Restriktion der Bildung etwa gleich großer Gruppen und parallelisiert hinsichtlich der gewählten Praktikumsform und des Geschlechts.

Die Aufgaben wurden in das praktikumsbegleitende E-Portfolio eingebunden, das an den beteiligten Universitäten grundsätzlich verpflichtend anzufertigen ist. Jedem Praktikanten bzw. jeder Praktikantin wurde zufällig eine Aufgabe zugewiesen und im Rahmen des Portfolios zugewiesen. Die jeweils anderen beiden Reflexionsaufgaben waren für die Proband*innen nicht sichtbar, um eine Beeinflussung durch diese auszuschließen. Unmittelbar im Anschluss an die Bearbeitung der Reflexionsaufgabe füllten die Proband*innen einen Online-Fragebogen aus, zu dem sie über einen Link am Ende der Reflexionsaufgabe geführt wurden. Der Fragebogen erfasste neben soziodemographischen Daten auch Fragen zu den Reflexionsaufgaben und erhob die als Moderatoren vermuteten Persönlichkeitsmerkmale.

Die Datenerhebung wurde zu zwei Zeitpunkten durchgeführt. Im Juli 2019 nahmen Lehramtsstudierende des Praktikumsformats „Lehrwerkstatt“ an den Universitäten Passau und Erlangen/Nürnberg am Ende ihres ganzjährigen Praktikums an der

¹ Die für die Studie relevanten Alternativpraktika sind die Lehrwerkstatt (225 Stunden) und das Exercitium Paedagogicum (240 Stunden), beides einjährige, studienbegleitende Praktika, die das Ziel einer intensiveren Praxisbegegnung und einer hochwertigeren Betreuung (sowohl an der Schule durch eine verantwortliche Tandemlehrkraft als auch wissenschaftlich im universitären Begleitprogramm) verfolgen. Sie können freiwillig gewählt werden und ersetzen dann die Regelpraktika „Orientierungspraktikum“ (60-80 Stunden) und „pädagogisch-didaktisches Praktikum“ (150-160 Stunden).

Untersuchung teil. Im Oktober 2019 wurde die Stichprobe um Lehramtsstudierende des pädagogisch-didaktischen Praktikums und des Exerzitium Paedagogicum der Universität Passau vergrößert. Abbildung 2 veranschaulicht das Untersuchungsdesign.

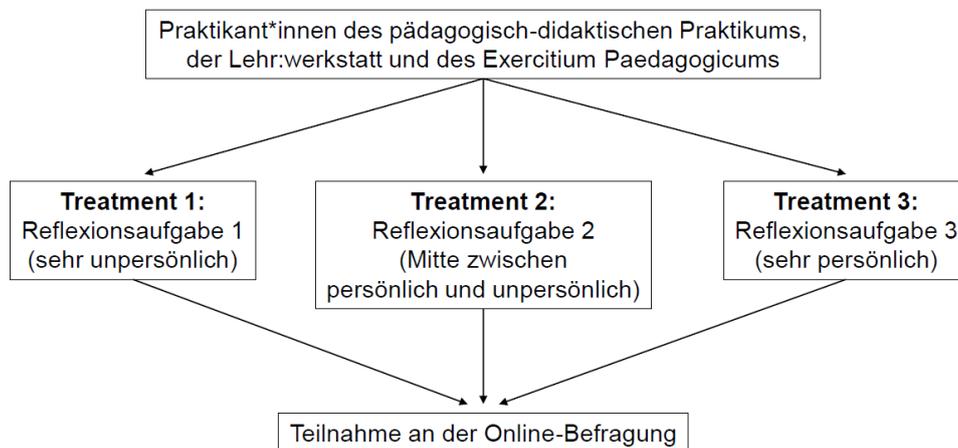


Abbildung 2. Untersuchungsdesign.

Die Reflexionstexte wurden inhaltsanalytisch ausgewertet, die Ergebnisse dieser Auswertung quantifiziert (Erläuterung zu diesem Verfahren siehe Kapitel 4.4.1.2) und mit den Daten des Fragebogens in einer Mixed-Methods-Auswertung zusammengeführt (vgl. Teddlie & Tashakkori, 2010).

4.4 Vorbereitende Auswertung und Datenanalyse

4.4.1 Auswertung der Reflexionstexte

4.4.1.1 Qualitative Inhaltsanalyse

Zur Auswertung der studentischen Reflexionstexte wurde eine Qualitative Inhaltsanalyse nach Mayring (2015) durchgeführt und dabei das Programm MAXQDA Plus 2018 bzw. 2020 (VERBI Software) eingesetzt.

Kodiereinheit. Kodiert wurde auf Satzebene. Das Kodieren nach Sinneinheiten wurde während des Pretests in Absprache mit mehreren Probekodier*innen abgeschlossen, da sich eine Einigung auf Sinnabschnitte sehr schwierig gestaltete und

die Kategorien auf diese Weise nicht trennscharf vergeben werden konnten. Das gewählte Vorgehen auf Satzebene hat zwar den Nachteil, dass es nicht zwischen unterschiedlichen Schreibstilen mit verschachtelten Sätzen und sehr kurzen Sätzen unterscheidet, es bietet aber den klaren Vorteil eindeutiger Kodiereinheiten und damit einer höheren Reliabilität, was für die spätere Quantifizierung der Ergebnisse eine unabdingbare Voraussetzung darstellt. Die Zergliederung in relativ kleine Einheiten lässt kaum noch latente Strukturen und Zusammenhänge erkennen – eine Entscheidung, die hinsichtlich Mayrings Empfehlung, die Validität der Reliabilität gegenüber bevorzugt zu verfolgen, gründlich bedacht werden muss (Mayring, 2015, S. 53). Es ist aber in diesem Fall gerechtfertigt, denn es passt zu den Eigenschaften des gewählten Kategoriensystems: Dieses erfordert primär eine Zuordnung der Aussagen über manifeste, teilweise auch sprachlich festgemachte Charakteristika. Es liegt das Erkenntnisinteresse also eher auf der manifesten Ebene und die Vergabe der Kategorien erfolgt primär auf der Basis von sprachlichen Mustern und weniger auf der Basis einer inhaltlichen Interpretation. Außerdem kann mit einer eher formalen und weniger inhaltlich orientierten Auswertung das Problem der schwierig herzustellenden Interraterreliabilität, die von anderen Studien bereits berichtet wurde (Kapitel 2.2.1.4), reduziert werden.

In die Auswertung gingen insgesamt 1652 Kodiereinheiten (= Sätze) ein. Im Durchschnitt umfasste jeder Reflexionstext $M = 17.76$ Sätze ($SD = 12.66$, $Min = 1$, $Max = 68$).

Kategoriensystem. Es wurden zwei deduktive Kategoriensysteme nacheinander eingesetzt. In einem ersten Durchgang wurde jedem Satz eine Reflexionsstufe nach Hatton und Smith (1995) zugewiesen, im zweiten Durchgang wurden zwei weitere Aspekte (Gefühle und Theoriebezüge) kodiert.

1. Teil: Reflexionsstufen nach Hatton und Smith (1995)

Mit einem ersten Kategoriensystem (Abb. 3) wurde eine skalierende Strukturierung der Texte vorgenommen. Dafür wurde das Kategoriensystem von Hatton und Smith (1995) eingesetzt (vgl. Kapitel 2.2.1.2). Das Kategoriensystem wurde sowohl im Original als auch in Abwandlungen in verschiedenen Nachfolgestudien eingesetzt

und repräsentiert gut die üblicherweise verwendeten Stufenschemata zur Hierarchisierung von Reflexionsfähigkeit. In mehreren Probedurchläufen wurde das Kategoriensystem von Hatton und Smith überarbeitet und verfeinert, da es im Original nur ungenaue Kodierregeln und wenige Beispiele enthält (Codebuch siehe Anhang 2). Wie bereits erörtert, sind es unter anderem die individuelle Neugestaltung bestehender Schemata oder der Entwurf neuer Stufenschemata, die die Vergleichbarkeit der bisherigen Studien erschweren. Deshalb wurde in der vorliegenden Studie versucht, weitgehend nach dem Originalschema von Hatton und Smith vorzugehen.

Dem ursprünglichen Kategoriensystem wurde als einzige inhaltliche Veränderung eine Differenzierung der Stufe „descriptive writing“ hinzugefügt, und zwar eine Unterscheidung von persönlicher Meinung und der Beschreibung erlebter Situationen. Damit soll dem Stellenwert, den Dewey und andere der genauen Beschreibung der Situation beimessen (dazu ausführlich Kapitel 2.2.1.4), Rechnung getragen werden, weil es sich Dewey zufolge bereits um die erste Stufe der Reflexion handelt. Bei Hatton und Smith hätte es sich dabei lediglich um nicht-reflexives Schreiben gehandelt, was einer nicht theoriekonformen Abwertung dieser Textteile gleichkommt, sobald die Stufen mit einer Bepunktung versehen werden. Hatton und Smith (1995) selbst betonen in ihrer Studie, dass es beim reflektierenden Schreiben unvermeidbar ist, dass Situationen zunächst beschrieben werden, um einen gewissen Kontext herzustellen (vgl. S. 42). Um an dieser Stelle also eine differenziertere Betrachtung zu ermöglichen, wurden diese beiden Aspekte separat erfasst.

Als „Persönliche Meinung“ wurden alle Aussagen kodiert, die nicht die Kriterien einer der höheren Reflexionsstufen erfüllten und damit als nicht-reflexives Schreiben gewertet wurden. Sie gehören nach dem Schema von Hatton und Smith (1995) zur untersten Kategorie „descriptive writing“ (Beispiel: „Je nach Fach gibt es unterschiedliche Möglichkeiten, aber als Englisch- und Geschichtslehrer würde ich raten viel vorlesen zu lassen.“ 1_SANROB09). In diese Kategorie fielen insgesamt 833 Sätze, was einem Anteil von 50.43 % am gesamten Textkorpus entspricht.

Als „Situationsbeschreibung“ wurden alle Aussagen erfasst, die nicht begründend, sondern beschreibend auf die erlebte Situation, die reflektiert werden sollte, eingingen (Beispiel: „Die Vertretungsstunde verlief die ganze Zeit so, ohne irgendwelche Veränderungen.“ BIARIC10_3). Auch die als Situationsbeschreibungen erfassten Sätze gehören zur Kategorie „descriptive writing“ und stellen noch unreflektiertes

Schreiben dar. Auf diese Weise wurden 259 Sätze kodiert, was einem Anteil von 15.68 % entspricht. Wenn man die beiden in der Studie untergliederten Aspekte „Persönliche Meinung“ und „Situationsbeschreibung“ wieder zur ursprünglichen, von Hatton und Smith vorgeschlagenen Kategorie „descriptive writing“ (Wertung: 0) zusammenfasst, ergibt sich eine Gesamtsatzzahl von 1092 Sätzen mit einem Anteil von 66.10 % am gesamten Textkorpus.

Unter „descriptive reflection“ wurden alle Aussagen erfasst, die eine Begründung für eine Handlung oder Aussage liefern und damit die erste Stufe des Reflektierens darstellen (Beispiel: „Am besten sollte er allein sitzen, damit er niemanden mit seiner Unruhe anstecken oder stören könnte.“ JANGUE06_1). Dabei wurde genau wie in der Studie von Hatton und Smith (1995) untergliedert in Aussagen mit nur einer Begründung („a. one perspective/factor“; Wertung: 1) und Aussagen mit mehreren verschiedenen Aspekten bzw. Begründungen („b. multiple perspectives/factors“; Wertung: 2). Es wurden 386 Sätze mit einer Begründung identifiziert (23.37 %) und 93 Sätze mit mehreren Begründungen (5.63 %). Insgesamt fallen in die Kategorie „descriptive reflection“ also 479 Sätze, was einem Anteil von 29 % am Textkorpus entspricht.

Diese Untergliederung in a) und b) findet sich auch in der nächsthöheren Stufe wieder, der „dialogic reflection“. Auf dieser Stufe werden Aussagen erfasst, die ein Für und Wider beleuchten oder sich mit Vor- und Nachteilen auseinandersetzen, oder etwas von zwei Seiten im Sinne eines „einerseits – andererseits“ betrachten, sich also „dialogisch“ nähern (Beispiel: „So wird er zwar ruhig gestellt, nimmt aber voraussichtlich kaum am Unterricht teil und lernt folglich auch nicht so viel wie die anderen Schüler.“ 1_ANWOL11). Bei Teilaspekt „a) one perspective/factor“ (Wertung: 3) gab es 59 Treffer (3.57 %) und bei Teilaspekt „b) multiple perspectives/factors“ (Wertung: 4) zehn Treffer (0.61 %). Insgesamt fielen in die Kategorie „dialogic reflection“ also 69 Sätze (4.18 % des Textkorpus).

Die höchste Stufe, die „critical reflection“ (Wertung: 5), erfasst Aussagen, die erkennen lassen, dass die konkrete Ebene des Geschehens verlassen wird und über grundsätzliche Konzepte, ethische, moralische oder gesellschaftliche Probleme nachgedacht wird (Beispiel: „Das wirklich allerletzte was ich tun würde, wären Bestrafungen zu geben, vor allem durch das Wissen über sein familiäres Umfeld, da ich Schülern mit Respekt und Anerkennung begegnen möchte.“ 2_GABRIC10).

Insgesamt wurden zwei Sätze in diese Kategorie eingeordnet, was einem Anteil von 0.12 % am Textkorpus entspricht.

1 Descriptive writing	Wert
1.1. Persönliche Meinung	0
1..2 Situationsbeschreibung (erlebt)	0
2 Descriptive reflection	
2.1 a) One perspective/factor	1
2.2 b) Multiple perspectives/factors	2
3 Dialogic reflection	
3.1 One perspective/factor	3
3.2 Multiple perspectives/factors	4
4 Critical reflection	5

Abbildung 3. Kategoriensystem 1 (nach Hatton und Smith, 1995) mit Punktwertungen der Kategorien.

2. Teil: Weitere reflexionsrelevante Aspekte

Es wurde ebenfalls bereits erörtert, dass das Kategorienschema von Hatton und Smith einige Aspekte von Reflexionsfähigkeit nicht abbildet und unterkomplex ist. Dies betrifft mehrere Aspekte; einen Überblick über möglicherweise fehlende Merkmale gibt beispielsweise die Kompetenzmodellierung von Hilzensauer (2017). Zwei als besonders wichtig erachtete Merkmale von Reflexion wurden exemplarisch in dieser Arbeit ausgewählt und zusätzlich zu den vorgestellten Kategorien aus den Texten herausgearbeitet, nämlich Bezüge zu Gefühlen und zur fachwissenschaftlichen Theorie.

Wie in Kapitel 2.1.2.7 gezeigt wurde, gibt es zahlreiche Stimmen, die Reflexion als nicht nur kognitiven Prozess sehen, obwohl er bisher in der Forschung primär so behandelt wurde. Während Leonhard und Rihm zwar beispielsweise explizit einräumen, dass Reflexion als Kompetenz mehr umfassen muss, als den bloß kognitiven Anteil, beschränken sie die Messung trotzdem bewusst auf die kognitiven Anteile, um die Komplexität zu reduzieren (Leonhard & Rihm, 2011, S. 244). Die vorliegende Arbeit setzte sich eine Öffnung dieser Einschränkung zum Ziel, um auch eine Bezugnahme der Reflektierenden auf ihre Gefühle Berücksichtigung finden zu

lassen. In die Kategorie „Gefühle“ wurden sowohl klassische Emotionen wie Freude, Trauer, Ärger, Angst oder Überraschung aufgenommen (Beispiel: „Diese Situation kam sehr überraschend, da ich alle technischen Geräte vor der Stunde noch geprüft habe.“ ANIJÜR03_3) als auch deren Abwesenheit, wenn sie direkt angesprochen wurde („Ich hatte keine Angst vor ihm und die Mitarbeit im Unterricht verlief sehr gut.“ ANDCLA19_2). In einem etwas weiter gefassten Sinn wurden hierunter auch Empfindungen, wie beispielsweise Verunsicherung, Irritation oder sich wohl bzw. unwohl fühlen, kodiert („Den für mich irritierenden Punkt stellte die Lautstärke bei der Gruppenarbeit dar.“ VERJOC08_3). In diese Kategorie wurden insgesamt 36 Textstellen (2.2 %) eingeordnet.

In Kapitel 2.1.2.8 wurde gezeigt, dass ein weiterer wesentlicher Anteil des Reflektierens darin besteht, Theorie und Praxis zu verknüpfen (z.B. Schön, 1983), sodass die Nennung und Einbindung (fach-)wissenschaftlicher Quellen, die in den Stufen nach Hatton und Smith keine Berücksichtigung finden, als eigene Kategorie erfasst wurde. In einem ersten Kodierdurchgang wurde unterschieden zwischen Bezügen mit eindeutiger Nennung der Forscher*innen bzw. der Theorie („Förderlich auf die Disziplin in der Klasse wirkt sich auch die Anwendung der Kouninschen Regeln des Classroom-Management aus [...].“ 1_BIRPET26) und Bezügen ohne eindeutige Nennung, aber mit erkennbarer Referenz zu fachwissenschaftlichen Theorien („Hierbei gilt grundsätzlich, dass es wirksamer ist, ein angemessenes Verhalten der Schüler anzuregen und zu verstärken, als ein negatives Verhalten zu sanktionieren.“ ULRAND22_1). Da allerdings nur vereinzelt explizite Bezüge mit Nennung von Namen und Theorien hergestellt wurden (insgesamt 7 Textstellen), wurden für die endgültige Kodierung beide Kategorien zusammengefasst und unter „Aussagen mit Theoriebezug“ erfasst. In diese Kategorie wurden 329 Textstellen (19.92 %) eingeordnet.

Diese beiden Kategorien wurden nach der Erfassung der Stufen von Hatton und Smith in einem zweiten Kodierdurchgang aus den vorliegenden Reflexionstexten herausgearbeitet (siehe Codebuch Anhang 3). Im Gegensatz zum Schema von Hatton und Smith, in das jeder einzelne Satz eingeordnet wurde, wurden diese weiteren Kategorien nicht für jeden Satz, sondern nur für auf die Kategorien zutreffende Sätze vergeben. Da bei diesen Aspekten keine Wertigkeit vorliegt, handelt es sich um nominale Kategorien.

Reliabilitätssicherung. In mehreren Iterationen mit verschiedenen geschulten Wissenschaftler*innen wurde der Kodierleitfaden geschärft und verfeinert. Die 92 Reflexionstexte wurden mithilfe des finalen Leitfadens vom ersten Rater anhand der beiden Kategoriensysteme kodiert. Für dieses Vorgehen ließ sich in Zusammenarbeit mit einem zweiten geschulten Rater, der 10 % der Reflexionstexte unabhängig kodierte, eine Übereinstimmung von Krippendorffs Alpha (ordinal) von $\alpha = .79$ für das Kategoriensystem nach Hatton und Smith (1995) und von $\alpha = .70$ (nominal) für die zusätzlich untersuchten Kategorien erzielen (Hayes & Krippendorff, 2007). Diese Übereinstimmungswerte können nach Cramer (2016) als berücksichtigungsfähig eingestuft werden, wofür ein Schwellenwert von $\alpha = .67$ vorgeschlagen wird (vgl. S. 115). Das Übereinstimmungsergebnis für die Stufen nach Hatton und Smith, die der Berechnung der Reflexionsmaße zugrunde liegen, verfehlen nur ganz knapp den Wert von $\alpha = .80$, ab dem das Ergebnis bereits als belastbar gelten kann.

4.4.1.2 Quantifizierung der Ergebnisse

Berechnung der Reflexionsmaße. Um die Ergebnisse der qualitativen Inhaltsanalyse rechnerisch erfassen und mit den Fragebogendaten in Beziehung setzen zu können, mussten diese quantifiziert werden. Für eine derartige Transformation von qualitativen in quantitative Daten stehen unterschiedliche Möglichkeiten zur Verfügung (vgl. Kuckartz, 2014, 130ff). Ziel der vorliegenden Studie war es, aus den Daten der qualitativen Textanalyse ein Maß der Reflexionsleistung für jeden Text zu ermitteln. Das wurde vereinzelt bereits in anderen Studien versucht (Kempin et al., 2019), eine anerkannte und bewährte Strategie steht dafür aber noch nicht zur Verfügung. Es wurden deshalb auf unterschiedliche Weisen zwei Reflexionsmaße ermittelt, mit denen im Vergleich weitergearbeitet wurde. Im Folgenden werden die beiden Reflexionsmaße beschrieben (Beispiel zur Veranschaulichung der Berechnung siehe Anhang 4).

Reflexionsmaß „Höchste erreichte Reflexionsstufe“. Als erstes Maß der Reflexionsleistung wurde die jeweils höchste erreichte Stufe nach dem Modell von Hatton und Smith (1995) im Dokument herangezogen.

Die Stufung, die bei den Reflexionen verwendet wurde, ähnelt anderen bekannten Textanalyseverfahren, wie beispielsweise der Stufung der moralischen Entwicklung (Kohlberg, 1981). Dort wird die vorliegende Stufe der moralischen Entwicklung im Text allerdings so ermittelt, dass nach der am häufigsten im Material vorkommenden Stufe gesucht wird. Dieses Verfahren wurde vereinzelt auch im Kontext von Reflexion eingesetzt (Hager, 2020). Die Anwendung dieses Verfahrens erwies sich bei den vorliegenden Reflexionstexten als nicht dienlich, da in überwiegendem Maße die unterste Stufe als häufigste ermittelt wurde (in 80 von 93 Texten, in den restlichen 13 Texten Stufe 1) und sich so keine Varianz erzeugen ließ. Bei Hager (2020) wurde zusätzlich zur am häufigsten vorkommenden Stufe auch noch die höchste erreichte Stufe im Text verwendet, so wie es auch in der vorliegenden Studie versucht wurde.

Reflexionsmaß „Durchschnittliche Reflexionsleistung“. Für das zweite Reflexionsmaß wurden die Kategorien nach Hatton und Smith mit aufsteigenden Punkten versehen (Punkte siehe Abb. 3). Da jeder Satz im Dokument einer Stufe zugewiesen wurde, konnte jedem Satz auch ein eindeutiger Wert zugeteilt werden. Diese einzelnen Wertungen wurden für jeden Reflexionstext addiert und zunächst ein Gesamtsummenscore gebildet. Da die Bildung einer Gesamtsumme Proband*innen bevorzugt, die eine große Menge an Text produziert haben, wurde ein durchschnittliches Reflexionsmaß berechnet, indem die ermittelte Summe der Reflexionspunkte durch die Satzzahl des Dokuments geteilt wurde. Jedoch unterschätzt dieses Maß diejenigen Reflexionen, in denen die erlebte Situation sehr ausführlich beschrieben wurde. Da Situationsbeschreibungen nach Hatton und Smith in die Kategorie des nicht-reflexiven Schreibens fallen und damit mit 0 Punkten bewertet wurden, sinkt der so berechnete Reflexionswert, je ausführlicher die Situation beschrieben wurde. Wie bereits erläutert, passt dieses Vorgehen nicht zur Theorie. Außerdem verlangte die Aufgabe 3 explizit die Beschreibung der erlebten Situation, sodass deren Ergebnisse dadurch eine systematische Abwertung erfahren hätten. Aufgrund dessen wurden Situationsbeschreibungen, wie bereits erläutert, separat erfasst und aus der Zählung genommen, um eine Verzerrung auszuschließen.

Da die beiden vorgestellten Maße mit gewissen Unschärfen behaftet sind, wurden beide weiterverfolgt und im Vergleich eingesetzt mit dem Ziel, aus den Ergebnissen eine Empfehlung für zukünftiges Vorgehen ableiten zu können und zu prüfen, ob

sich für die unterschiedlichen Berechnungen andere Ergebnisse herauskristallisieren.

Verteilung und Extremwerte der Daten aus der Inhaltsanalyse. Zusätzlich zur Berechnung der Reflexionsmaße wurden auch die Häufigkeiten des Vorkommens der inhaltsanalytischen Kategorien ermittelt und für statistische Berechnungen verwendet. Für die Ergebnisse beider Kodierschemata ergaben sich nicht-normalverteilte Daten mit zum Teil erheblichen Ausreißern. Die Ausreißer wurden nicht ausgeschlossen, da es sich bei diesen nicht um Fehler im Datensatz handelt. Es wurde deshalb bei der Auswertung immer dann auf nonparametrische Verfahren zurückgegriffen, wenn mit den Ergebnissen der Inhaltsanalyse gerechnet wurde, die nicht in die Reflexionsmaße überführt wurden. Darunter fallen beispielsweise die Ausprägungen der einzelnen Kategorien.

4.4.2 Auswertung des Kreativitätstests

In der vorliegenden Studie wurde der Guilford-Test eingesetzt, um drei Aspekte des divergenten Denkens zu ermitteln, nämlich die Fluidität, die Flexibilität und die Originalität der Studienteilnehmer*innen. Im Rahmen des Tests wurden im Fragebogen verschiedene alternative Einsatzmöglichkeiten eines Regenschirms erfragt. Die drei Aspekte divergenten Denkens mussten bei der Auswertung jeweils einzeln aus den Nennungen im Fragebogen berechnet werden. Wie diese Berechnung erfolgte, wird im Folgenden kurz dargestellt.

Das Maß der Fluidität entspricht der Gesamtzahl der Nennungen und zeigt, wie viele alternative Ideen insgesamt entwickelt werden konnten. Die Zahl der Nennungen reichte von minimal einer Nennung bis maximal 20. Der Mittelwert lag bei $M = 6.63$ ($SD = 4.27$).

Für das Maß der Flexibilität wurden inhaltliche Kategorien gebildet, die die einzelnen Nennungen zu Oberbegriffen zusammenfassen. Diese Kategorien oder Oberbegriffe waren beispielsweise „Schutzfunktion“, „Spiel und Sport“, „Transportfunktion“ oder „Werkzeug“. Insgesamt wurden elf Kategorien gebildet. Für die Berechnung des Flexibilitätsmaßes wurde ermittelt, auf wie viele unterschiedliche Kategorien die Versuchsteilnehmer*innen mit ihren Nennungen Bezug nahmen. Um

sicherzustellen, dass die Zuordnung der Nennungen zu den Kategorien reliabel erfolgt, wurden 10 % der Nennungen von einem zweiten Rater zugeordnet. Die Übereinstimmung erwies sich mit einem Krippendorffs $\alpha = .87$ als hoch.

Als Mittelwert der genutzten Kategorien ergab sich $M = 4.27$ ($SD = 1.83$). Minimal wurde sich aus einer Kategorie bedient, maximal aus zehn Kategorien.

Für das Maß der Originalität wurde zunächst ausgewertet, wie häufig jede einzelne Nennung im Gesamtdatensatz vorkam. „Als Sonnenschutz“ wurde beispielsweise von 14 % der Proband*innen genannt, „als Waffe“ von 2 % und „zum Nasebohren“ von 0.16 %. Den Prozenten wurden anschließend Punkte von 0 (für sehr häufige Nennung) bis 5 (sehr seltene Nennung) zugewiesen und mit diesen Punkten ein Gesamtscore der Originalität für jede Person gebildet. Aus diesem Gesamtscore wurde der Durchschnitt für jede Person berechnet, indem er durch die Anzahl der Nennungen geteilt wurde. So ergab sich für die Originalität ein Mittelwert von $M = 1.95$ ($SD = 0.90$). Der kleinste Wert war dabei 0, der größte Wert 3.86.

5 Ergebnisse

5.1 Reflexionsmaße

Textmenge. Es gab hinsichtlich der Satzzahl keinen signifikanten Unterschied zwischen den weiblichen Teilnehmerinnen ($M = 17.90$, $SD = 12.48$) und den männlichen Teilnehmern ($M = 18.55$, $SD = 13.01$), $t(89) = .23$, $p = .41$, $d = .05$.

Höchste erreichte Reflexionsstufe. Für das Reflexionsmaß „Höchste erreichte Reflexionsstufe“ ergab sich über alle Proband*innen hinweg ein Mittelwert von $M = 2.26$ ($SD = 1.14$, $Min = 0$, $Max = 5$). Es konnte die ganze verfügbare Skala von 0 („descriptive writing“: keine Reflexion) bis 5 („critical reflection“) in den Texten identifiziert werden. Dieses erste Reflexionsmaß war unabhängig vom Geschlecht, dem Fachsemester und der gewählten Praktikumsform (alle $ps \geq .065$). Es lag allerdings ein signifikanter Mittelwertunterschied zwischen den Hochschulstandorten vor, $t(67.39) = -2.64$, $p = .010$, $d = 0.58$, wobei der Wert der Passauer Studierenden ($M = 2.54$, $SD = 0.90$) höher lag, als der der Erlangen/Nürnberger ($M = 1.90$, $SD = 1.32$). Die Studierenden in Passau erreichten also im Durchschnitt höhere Reflexionsstufen als die Studierenden in Erlangen/Nürnberg. Außerdem lag eine kleine positive Korrelation dieses Reflexionsmaßes mit der Satzzahl in den Dokumenten vor ($r = .34$, $p = .001$). Je mehr Sätze in einem Dokument geschrieben wurden, desto höhere Reflexionsstufen wurden im Text erreicht und umgekehrt.

Tabelle 3 zeigt, wie viele Proband*innen insgesamt (in Prozent) welche jeweils höchste Stufe in ihren Reflexionstexten erreicht haben. Nur 2.2 %, das entspricht zwei Teilnehmer*innen, erreichten dabei die höchste Stufe des kritischen Reflektierens. Mit 3.3 % gab es allerdings auch nur sehr wenige, die auf der ganz unreflektierten Stufe eingeordnet wurden. Der Rest liegt im Mittelfeld mit der häufigsten Ausprägung auf Stufe 3. Fasst man die unterteilten Stufen wieder zusammen, so zeigt sich, dass insgesamt 47.3 % der Teilnehmer*innen der Stufe der beschreibenden Reflexion zuzuordnen sind (Stufen 1 und 2) und ebenfalls knapp die Hälfte, nämlich auch 47.3 %, der Stufe der dialogischen Reflexion (Stufen 3 und 4).

Tabelle 3. Vorkommen der jeweils höchsten im Text erreichten Reflexionsstufe in Prozent

Reflexionsstufe	Prozent
0 = Descriptive writing	3.3
1 = Descriptive reflection, one perspective	26.4
2 = Descriptive reflection, multiple perspectives	20.9
3 = Dialogic reflection, one perspective	38.5
4 = Dialogic reflection, multiple perspectives	8.8
5 = Critical reflection	2.2

Durchschnittliche Reflexionsleistung. Für das zweite Reflexionsmaß, die „Durchschnittliche Reflexionsleistung“, ergab sich ein Mittelwert von $M = 0.64$ ($SD = 0.43$, $Min = 0$, $Max = 2$). Der Wert zeigte sich unabhängig vom Fachsemester und es gab keine signifikanten Unterschiede zwischen den Geschlechtern, den Praktikumsformen oder den Hochschulstandorten (alle $ps \geq .50$).

Die beiden Maße korrelierten erwartungsgemäß hoch miteinander ($r = .60$, $p \leq .001$) aber in einem Grad, der darauf schließen lässt, dass hier trotzdem noch teilweise unterschiedliche Konstrukte gemessen werden.

5.2 Aufgabenbezogene Hypothesen

H 1.1: Je größer der über die Aufgabe induzierte persönliche Bezug zum Reflexionsproblem ist, desto höher ist die Reflexionsleistung.

Als erste Hypothese wurde formuliert, dass ein größerer persönlicher Bezug zur Reflexionsaufgabe eine höhere Reflexionsleistung zur Folge haben sollte. Da der persönliche Bezug von Aufgabe 1 bis Aufgabe 3 steigt, ist davon auszugehen, dass sich auch eine schrittweise Steigerung der Reflexionsleistung in Form der beiden Reflexionsmaße zeigt, wenn der vermutete Zusammenhang tatsächlich vorliegt.

Für das Maß „Höchste erreichte Reflexionsstufe“ stiegen die Mittelwerte erwartungsgemäß von Aufgabe 1 bis Aufgabe 3 an (Abb. 4).

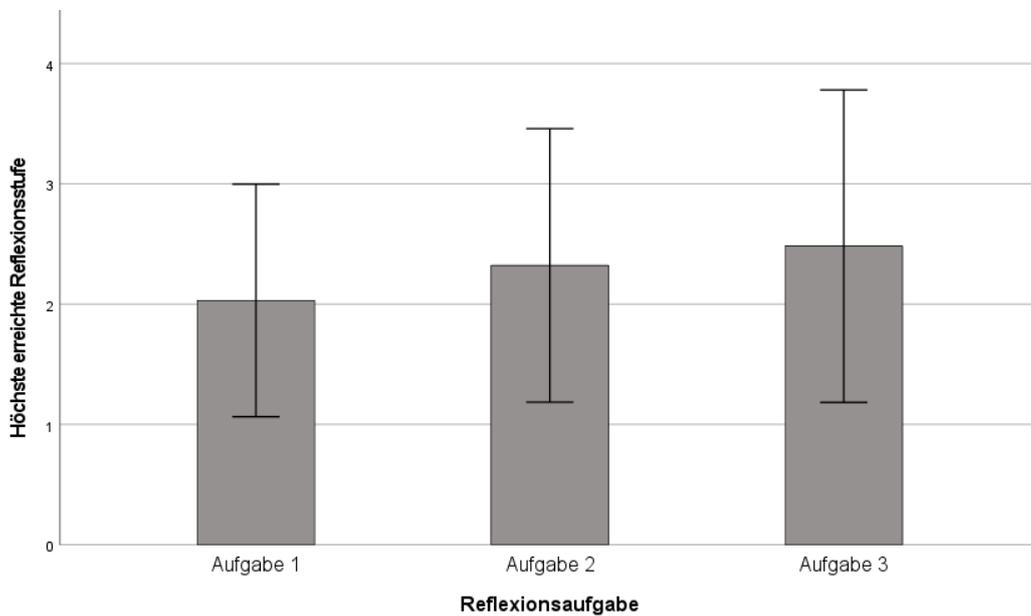


Abbildung 4. Höchste erreichte Reflexionsstufe in Abhängigkeit von der Reflexionsaufgabe. Die Fehlerbalken indizieren einfache Standardabweichungen.

Das gleiche Bild zeigte sich in der deskriptiven Analyse auch für das zweite Maß, die „Durchschnittliche Reflexionsleistung“ (Abb. 5).

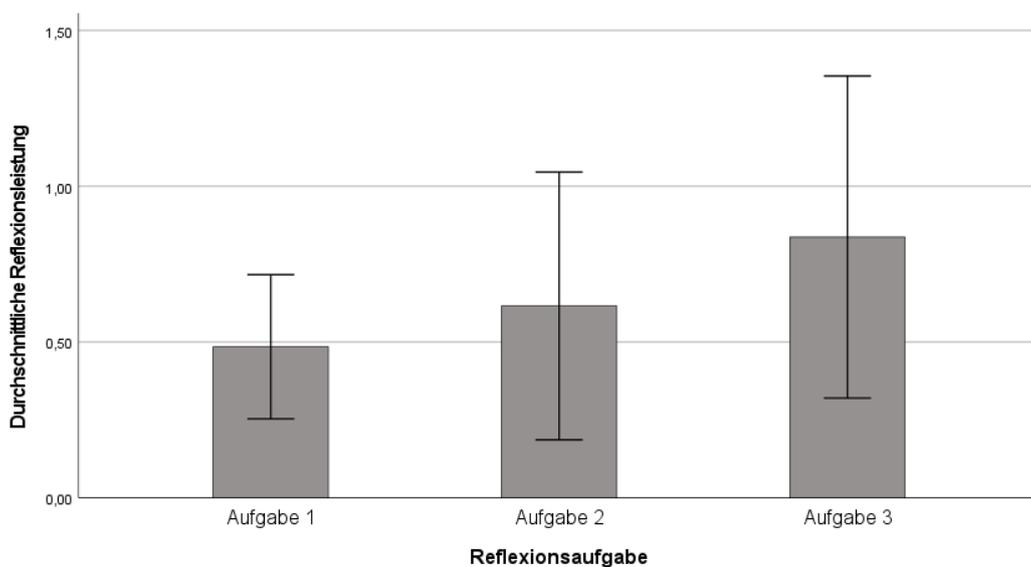


Abbildung 5. Durchschnittliche Reflexionsleistung in Abhängigkeit von der Reflexionsaufgabe. Die Fehlerbalken indizieren einfache Standardabweichungen.

Obwohl sich grafisch eine recht eindeutige Tendenz für beide Reflexionsmaße

zeigte, konnte diese inferenzstatistisch nur für das Maß „Durchschnittliche Reflexionsleistung“ nachgewiesen werden. Hier erwies sich der Unterschied als statistisch signifikant: Eine univariate Varianzanalyse ergab einen signifikanten Haupteffekt der Aufgabenstellung auf das Reflexionsmaß „Durchschnittliche Reflexionsleistung“, $F(2, 89) = 5.80, p = .004, \eta^2 = .12$.

Eine Post-hoc-Testung nach Bonferroni bestätigte den deskriptiven Befund der steigenden Mittelwertunterschiede zum Teil. Die durchschnittliche Reflexionsleistung lag bei Aufgabe 3 ($M = 0.84, SD = 0.53$) höher als bei Aufgabe 1 ($M = 0.49, SD = 0.23$) und der Unterschied war signifikant ($p = .003$). Zwischen den Aufgaben 1 und 2 ($M = 0.62, SD = 0.43; p = .61$) und den Aufgaben 2 und 3 ($p = .11$) zeigte sich zwar ebenfalls ein Anstieg, jedoch kein signifikanter Unterschied.

Beim Maß „Höchste erreichte Reflexionsstufe“ zeigten sich zwar ebenfalls hypotesenkonforme Mittelwertanstiege in den deskriptiven Daten (Aufgabe 1: $M = 2.03, SD = 0.97$, Aufgabe 2: $M = 2.32, SD = 1.14$, Aufgabe 3: $M = 2.48, SD = 1.30$). Diese Unterschiede erwiesen sich in einer univariaten Varianzanalyse aber als nicht signifikant, $F(2, 89) = 1.25, p = .29, \eta^2 = .03$.

Es ließ sich also nur für eines der beiden Maße, nämlich die „Durchschnittliche Reflexionsleistung“, zeigen, dass es einen signifikanten Unterschied zwischen den drei Aufgaben – und damit dem steigenden Grad der persönlichen Involviertheit – gab. Einschränkend ist zu sagen, dass dieser auch nicht zwischen allen drei Aufgaben, sondern nur zwischen den Extremausprägungen von Aufgabe 1 (kaum persönliche Relevanz) und Aufgabe 3 (hohe persönliche Relevanz) vorlag. Das Reflektieren zu einem eigenen Erlebnis erzeugte also wie erwartet eine signifikant höhere durchschnittliche Reflexionsleistung als das Reflektieren zu einem Problem einer unbekanntes Lehrkraft.

H 1.2 Je größer der über die Aufgabe induzierte persönliche Bezug zum Reflexionsproblem ist, desto mehr Motivation löst sie aus und desto höher wird ihr Nutzen eingeschätzt.

Zweitens wurde angenommen, dass der größere persönliche Bezug in den Reflexionsaufgaben sich auch auf die durch sie ausgelöste Motivation und den wahrge-

nommenen Nutzen der Aufgabe auswirkt. Auch hier wurde also vermutet, dass sowohl die aufgabenbezogene Motivation als auch der empfundene Nutzen der Aufgabe von Aufgabe 1 bis Aufgabe 3 steigen. Tatsächlich zeigte sich der Effekt wie vermutet in den deskriptiven Daten, wenngleich die Unterschiede klein ausfielen, wie Tabelle 4 zeigt.

Tabelle 4. *Aufgabenbezogene Mittelwerte und Standardabweichungen der Variablen Motivation und Nutzen*

	Aufgabe 1		Aufgabe 2		Aufgabe 3	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Motivation	3.65	0.68	3.81	0.64	3.83	0.50
Nutzen	3.94	0.67	4.13	0.59	4.28	0.47

Um die Hypothese inferenzstatistisch zu prüfen, wurde je eine univariate Varianzanalyse mit der unabhängigen Variable „Aufgabe“ und der abhängigen Variable „Motivation“ bzw. „Nutzen“ durchgeführt. Die drei Aufgaben unterschieden sich weder hinsichtlich der ausgelösten Motivation, $F(2, 87) = 0.80$, $p = .454$, $\eta^2 = .02$, noch hinsichtlich des wahrgenommenen Nutzens signifikant, $F(2, 87) = 2.40$, $p = .097$, $\eta^2 = .05$.

Der Grad des persönlichen Bezugs in der Aufgabenstellung, der sich von Aufgabe 1 bis Aufgabe 3 steigerte, hatte also in den vorliegenden Daten keinen nachweisbar signifikanten Effekt auf die Motivation der Teilnehmenden, diese zu bearbeiten und auch nicht auf den Nutzen, der der Bearbeitung der Aufgabe zugeschrieben wurde.

H 1.3 Je höher der durch die Aufgabe induzierte persönliche Bezug zum Reflexionsproblem ist, desto mehr emotionale und desto weniger kognitive Anteile finden sich in den Reflexionstexten.

Die dritte aufgabenbezogene Hypothese ging von der Annahme aus, dass der größere persönliche Bezug in den Aufgaben unterschiedliche Beantwortungstendenzen

zur Folge hat. Konkret wurde davon ausgegangen, dass ein höherer persönlicher Bezug dazu führt, dass sich in den Reflexionstexten mehr emotionale und weniger kognitive Anteile finden. Umgekehrt sollte ein geringerer Grad des persönlichen Bezugs zur Aufgabe zu einem eher kognitiven und weniger emotionalen Bearbeitungsmodus führen. Die emotionalen und kognitiven Anteile wurden inhaltsanalytisch aus den Texten herausgearbeitet und quantifiziert in die Analyse aufgenommen.

Für die kognitiven Anteile wurden die in den Reflexionstexten hergestellten Bezüge zur fachwissenschaftlichen Theorie als Indikator gewählt. Während die Zahl dieser Theoriebezüge bei Aufgabe 1 ($Mdn = 2.50$) und Aufgabe 2 ($Mdn = 4.00$) hoch sind, zeigt sich ein deutlicher Abfall derselben in Aufgabe 3 ($Mdn = 1.00$), wie sich in Abb. 6 deutlich zeigt.

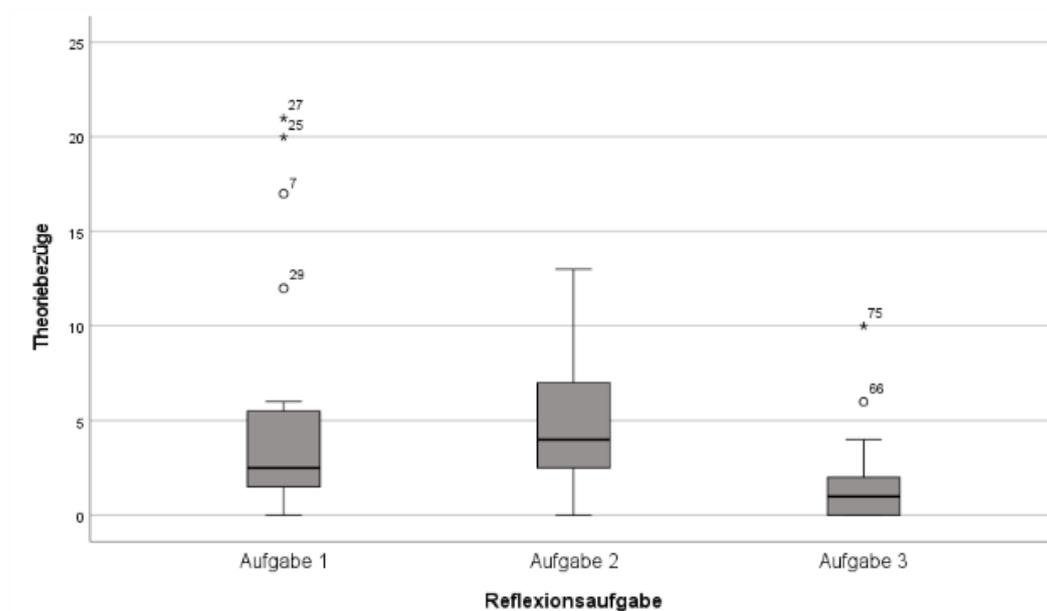


Abbildung 6. Boxplots der in den Reflexionstexten hergestellten Theoriebezüge in Abhängigkeit von der Reflexionsaufgabe.

Zur inferenzstatistischen Hypothesenprüfung wurde ein Kruskal-Wallis-Test durchgeführt, da die Ergebnisse der qualitativen Inhaltsanalyse für ein parametrisches Verfahren zu stark von der Normalverteilung abwichen und mit einigen Ausreißern versehen waren. Dieser zeigte, dass es einen signifikanten Unterschied zwischen den drei Aufgaben gab, $H(2) = 23.42, p \leq .001$.

Ein Dunn-Bonferroni-Test, der als Post-hoc-Test eingesetzt wurde, bestätigte den Befund der deskriptiven Daten, nämlich dass der Unterschied zwischen Aufgabe 3

und 1 ($z = 3.48$, $p = .001$, $r = .36$) und zwischen Aufgabe 3 und 2 ($z = 4.68$, $p \leq .001$, $r = .49$) signifikant war.

Bei der Reflexion eines eigenen Erlebnisses, wie sie in Aufgabe 3 gefordert wurde, konnten also in den Reflexionstexten signifikant weniger Bezüge zur fachwissenschaftlichen Theorie gefunden werden, als es bei den eher unpersönlicheren Aufgaben 1 und 2 der Fall war.

Neben den fachwissenschaftlichen Theoriebezügen wurden auch noch Bezüge zu Gefühlen der Proband*innen aus den Texten herausgearbeitet. Als Indikator für einen eher emotionalen Bearbeitungsmodus wurde die Menge der Erwähnungen von Gefühlen im Text gewertet. Wie Abbildung 7 zeigt, liegt auch die Zahl der Herstellung von Bezügen zu Gefühlen bei Aufgabe 3 ($Mdn = 1.00$) höher als bei den Aufgaben 1 ($Mdn = 0.00$) und 2 ($Mdn = 0.00$).

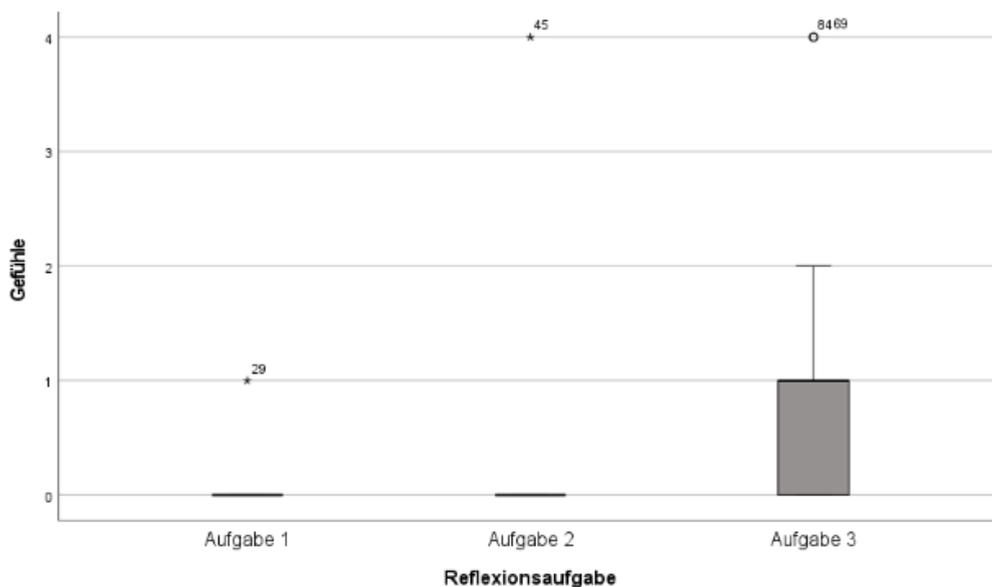


Abbildung 7. Boxplots der Bezüge zu Gefühlen in Abhängigkeit von der Reflexionsaufgabe.

Auch hier bestätigte sich der Unterschied in einem Kruskal-Wallis-Test als signifikant, $H(2) = 45.07$, $p \leq .001$ und der Dunn-Bonferroni-Test zeigte, wie bei den Theoriebezügen auch, den signifikanten Unterschied zwischen Aufgabe 1 und 3 ($z = 5.92$, $p \leq .001$, $r = .62$) und zwischen Aufgabe 3 und 2 ($z = -5.78$, $p \leq .001$, $r = .60$). In Aufgabe 3 wurden folglich häufiger die eigenen Gefühle einbezogen als in den Aufgaben 1 und 2.

Es zeigte sich also bei den Bezügen zu Gefühlen das genau gegenteilige Bild zur Herstellung von fachwissenschaftlichen Bezügen: Die persönlichere Aufgabe 3 regte signifikant mehr dazu an, sich mit eigenen Gefühlen auseinanderzusetzen, als die distanzierteren Aufgaben 1 und 2, die nichts mit eigenen Erfahrungen zu tun hatten.

Während das Reflektieren eines eigenen Erlebnisses also insgesamt mehr zu einem emotionalen Bearbeitungsmodus anregte, wurden gleichzeitig weniger fachwissenschaftliche Theorien eingebunden. Sich mit einer fiktiven oder fremden Situation auseinanderzusetzen, aktivierte hingegen eher eine kognitive, theoriebezogene Herangehensweise mit signifikant geringerem emotionalen Bezug.

5.3 Prädiktorbezogene Hypothesen

H 2.1 Motivation und wahrgenommener Nutzen der Aufgabe präzisieren die Reflexionsleistung.

In den aufgabenbezogenen Hypothesen wurde bereits untersucht, ob die Aufgaben sich hinsichtlich der ausgelösten Motivation und dem wahrgenommenen Nutzen unterscheiden. In einem weiteren Schritt wurde nun angenommen, dass diese beiden situationalen Faktoren auch die Reflexionsleistung präzisieren. Ein Effekt wurde dahingehend erwartet, dass ein höheres Messergebnis der Motivation bei der Aufgabenbearbeitung und ein als höher empfundener Nutzen der Aufgabe zu einem Ansteigen der beiden Reflexionsmaße führt.

Wie Tabelle 5 zu entnehmen ist, konnte eine signifikant positive Korrelation sowohl der aufgabenbezogenen Motivation als auch des eingeschätzten Nutzens der Aufgabe mit dem Reflexionsmaß „Höchste erreichte Reflexionsstufe“ festgestellt werden. Mit dem Reflexionsmaß „Durchschnittliche Reflexionsleistung“ korrelierte nur der wahrgenommene Nutzen signifikant, nicht aber die Motivation. Der Nutzen der Aufgabe und die Motivation korrelierten ebenfalls erwartungsgemäß hoch miteinander.

Tabelle 5. Mittelwerte, Standardabweichungen und Interkorrelationen der Prädiktoren für die Reflexionsmaße (N = 93)

Variable	<i>M</i>	<i>SD</i>	1	2
Höchste erreichte Reflexionsstufe	2.26	1.14	.286**	.307**
Durchschnittliche Reflexionsleistung	0.64	0.43	.145	.240*
Prädiktorvariablen				
1. Motivation	3.76	0.61	---	
2. Nutzen	4.13	0.59	.659**	---

Anmerkungen. * $p \leq .05$. ** $p \leq .01$.

Der Erklärwert der Variablen Motivation und empfundener Nutzen wurde für beide Reflexionsmaße robust geschätzt, da die Voraussetzung normalverteilter Residuen nicht erfüllt war. Für die abhängige Variable „Höchste erreichte Reflexionsstufe“ wurde eine schrittweise Regression gerechnet, bei der es zu einem Ausschluss des nicht signifikanten Prädiktors Motivation kam, sodass nur noch der Faktor Nutzen als Prädiktor im Modell blieb (siehe Tabelle 6). Je höher die Proband*innen also den Nutzen der Aufgabenbearbeitung für sich einschätzten, desto höhere Reflexionsstufen erreichten sie in den Texten. Die Motivation, die Aufgabe zu bearbeiten, hatte keinen Effekt auf die in den Texten erreichten Reflexionsstufen.

Tabelle 6. Regressionsanalyse für die Vorhersage des Maßes „Höchste erreichte Reflexionsstufe“

Prädiktor	<i>b</i> ^a	<i>SE b</i> ^a	β	<i>t</i>	<i>p</i>
Nutzen	0.583	0.796	0.307	3.04	.003
	[0.270, 0.946]				

Anmerkungen. $R^2 = .094$, $p = .003$, $N = 91$

^a Konfidenzintervalle und Standardfehler per Bootstrapping mit 10 000 Stichproben.

95 % Konfidenzintervalle.

Für die abhängige Variable „Durchschnittliche Reflexionsleistung“ erwies sich der Faktor Nutzen in einer Regressionsanalyse ebenfalls als signifikanter Prädiktor, wie der Tabelle 7 zu entnehmen ist. Auch hier ergab sich also analog zum ersten Reflexionsmaß: Je nützlicher die Bearbeitung der Aufgabe eingeschätzt wurde, desto höher fiel die durchschnittliche Reflexionsleistung aus. Auch hier gab es keinen signifikanten Einfluss der Motivation.

Tabelle 7. Regressionsanalyse für die Vorhersage des Maßes „Durchschnittliche Reflexionsleistung“

Prädiktor	b^a	SE b^a	β	t	p
Nutzen	0.173	0.065	0.240	2.33	.022
	[0.505, 0.309]				

Anmerkungen. $R^2 = .057$, $p = .022$, $N = 91$

^a Konfidenzintervalle und Standardfehler per Bootstrapping mit 10 000 Stichproben.

95 % Konfidenzintervalle.

Anzumerken ist an dieser Stelle noch, dass sich sowohl die aufgabenbezogene Motivation als auch der empfundene Nutzen der Aufgabe zwischen den Universitätsstandorten unterschieden. In Passau ($M = 4.24$, $SD = 0.52$) lag beispielsweise der empfundene Nutzen der Aufgabe signifikant höher als in Erlangen/Nürnberg ($M = 3.98$, $SD = 0.65$), $t(72.57) = -2.13$, $p = .036$, $d = 0.62$.

Außerdem ist noch festzuhalten, dass die aufgabenbezogene Motivation zwar keinen signifikanten Einfluss auf die beiden Reflexionsmaße hatte, sie aber 14 % der Varianz der Gesamtsatzzahl in den Dokumenten erklärte. Das heißt, dass bei höherer Motivation zwar längere Texte verfasst wurden, aber daraus keine höhere Reflexionsleistung folgte.

H 2.2: Die Persönlichkeitsmerkmale Ungewissheitstoleranz, Offenheit und Kreativität prädizieren die Reflexionsleistung.

Des Weiteren wurde angenommen, dass verschiedene Persönlichkeitsvoraussetzungen die Reflexionsleistung prädizieren. Überprüft wurde ein Einfluss der Faktoren „Ungewissheitstoleranz“, „Offenheit“ und „Kreativität“, letztere repräsentiert durch die drei Facetten „Fluidität“, „Flexibilität“ und „Originalität“. Es wurde erwartet, dass sich eine höhere Reflexionsleistung zeigt, je höher die Werte für diese spezifischen Persönlichkeitsvoraussetzungen ausgeprägt sind. Es lagen jedoch keine signifikanten Korrelationen mit den beiden Reflexionsmaßen vor, wie Tabelle 8 zu entnehmen ist.

Tabelle 8. Mittelwerte, Standardabweichungen und Korrelationen der Prädiktoren für die Reflexionsmaße (N = 93)

Variable	M	SD	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a
Durchschnittliche Reflexionsleistung	0.64	0.43	-.013 [-0.30, 0.20]	-.095 [-0.27, 0.11]	.157 [-0.52, 0.38]	.205 [-0.03, 0.37]	.170 [-0.05, 0.37]
Höchste erreichte Reflexionsstufe	2.26	1.14	.066 [-0.16, 0.29]	-.024 [-0.23, 0.20]	.197 [-0.02, 0.39]	.126 [-0.10, 0.32]	.135 [-0.11, 0.36]
Prädiktorvariablen							
1. Offenheit	3.66	0.68	---				
2. Ungewissheitstoleranz	3.41	0.69	.046	---			
3. Fluidität	6.63	4.27	.314**	.007	---		
4. Flexibilität	4.27	1.84	.280**	-.003	.811**	---	
5. Originalität	1.96	0.90	.368**	.055	.648**	.643**	---

Anmerkungen. ^a 95 % Konfidenzintervalle basierend auf 10 000 Bootstrapping-Stichproben.

* $p \leq .05$. ** $p \leq .01$.

Während die Persönlichkeitsmerkmale zum Teil untereinander Autokorrelationen aufwiesen (beispielweise Offenheit mit den drei Kreativitätsaspekten und letztere

auch untereinander), standen Offenheit, Ungewissheitstoleranz, Fluidität, Flexibilität und Originalität in keinem signifikanten Zusammenhang mit den Reflexionsmaßen „Höchste erreichte Reflexionsstufe“ (alle $ps \geq .081$) und „Durchschnittliche Reflexionsleistung“ (alle $ps \geq .071$).

H 2.3 Der empfundene Problemgehalt des Reflexionsthemas prädiziert die Reflexionsleistung.

H 2.3 vermutete eine höhere Reflexionsleistung, je problematischer die Proband*innen das Thema bzw. die Situation empfanden, die sie reflektieren sollten. Bei Aufgabe 1 und 2 war diese Situation in der Aufgabenstellung bereits vorgegeben, bei Aufgabe 3 sollte an ein eigenes Erlebnis gedacht werden. Deshalb könnte es bei Aufgabe 3 passiert sein, dass Proband*innen entweder kein problematisches Erlebnis erinnerten und ein unproblematisches schilderten, oder aber ein wesentlich problematischer empfundenes Ereignis erlebt haben, als das in Aufgabe 1 und 2 vorgegebene. Tatsächlich zeigte sich, dass die Varianz bei Aufgabe 3 am größten war, nur hier lag das volle Spektrum von Min = 1 bis Max = 5, also von *sehr wenig problematisch* bis *sehr problematisch* vor. Bei Aufgabe 1 (Min = 2, Max = 4) fiel sie am geringsten aus und die Tendenz zur Mitte war am größten. Aufgabe 2 (Min = 1, Max = 4) lag zwischen den beiden, wurde aber auch von niemandem als sehr problematisch empfunden. Es zeigt sich also in der deskriptiven Analyse, dass für das eigene Erlebnis, zu dem in Aufgabe 3 reflektiert wurde, von manchen sehr problematische Situationen ausgewählt wurden und manche schienen nur sehr wenig problematische Erlebnisse zu erinnern oder schildern zu wollen. Die vorgegebene und am wenigsten persönliche Situation, die in Aufgabe 1 geschildert wurde, wurde von den meisten Proband*innen als mittelmäßig problematisch eingeschätzt.

Dies spiegelt sich allerdings nicht in den Mittelwerten wider. Es zeigte sich, dass die drei Aufgaben hinsichtlich des eingeschätzten Problemgehalts nahezu identisch ausfielen. Die Mittelwerte bei Aufgabe 1 ($M = 3.10$, $SD = 0.71$), Aufgabe 2 ($M = 2.97$, $SD = 0.88$) und Aufgabe 3 ($M = 2.93$, $SD = 0.84$) unterschieden sich einer Varianzanalyse zufolge nicht signifikant voneinander, $F(2, 87) = 0.36$, $p = .70$, $\eta^2 = .008$.

Der empfundene Problemgehalt der Aufgabe korrelierte ferner mit keinem der beiden Reflexionsmaße signifikant (vgl. Tab. 9).

Tabelle 9. Mittelwert, Standardabweichung und Korrelationen des Prädiktors Problemgehalt der Aufgabe für die Reflexionsmaße ($N = 91$)

Variable	<i>M</i>	<i>SD</i>	Durchschnittliche Reflexionsleistung ^a	Höchste erreichte Reflexionsstufe ^a
Problemgehalt der Aufgabe	3.00	0.79	0.076 [-0.10, 0.24]	0.136 [-0.05, 3.16]

Anmerkungen. ^a 95 % Konfidenzintervalle basierend auf 10 000 Bootstrap-Stichproben.

* $p \leq .05$. ** $p \leq .01$

Wie problematisch die Situation eingeschätzt wurde, zu der reflektiert werden sollte, spielte also in den vorliegenden Daten keine Rolle für die resultierende Reflexionsleistung.

H 2.4 Die Erwartungen der Proband*innen an eine Reflexionsaufgabe prädictieren die in den Texten gefundene Menge an reflexionsspezifischen Elementen.

In der letzten Hypothese wurde angenommen, dass die Erwartungen der Proband*innen an eine Reflexionsaufgabe die Menge der tatsächlich in den Texten gefundenen reflexionsspezifischen Elemente prädictiert. Je wichtiger dabei ein bestimmter Aspekt einer Reflexionsaufgabe eingeschätzt wird, desto mehr Elemente dieses Aspekts werden im Reflexionstext erwartet. An einem konkreten Beispiel formuliert könnte das heißen: Je höher die Wichtigkeit des Einbezugs fachwissenschaftlicher Theorien bei der Bearbeitung einer Reflexionsaufgabe eingeschätzt wird, desto mehr Theoriebezüge finden sich tatsächlich bei der inhaltsanalytischen Auswertung der Texte.

Die folgenden Ausführungen sind dreigeteilt: In einem ersten Schritt wird betrachtet, welche Erwartungen an Reflexionsaufgaben die Studierenden generell aufweisen und ob diese sich in Abhängigkeit von den drei Reflexionsaufgaben unterscheiden.

den. In einem zweiten Schritt werden die Ergebnisse der inhaltsanalytischen Analyse anhand der Häufigkeiten der jeweiligen Kategorien dargestellt. Im letzten Schritt werden diese beiden Befunde, also die Erwartungen der Studierenden und die tatsächlich in den Texten gefundenen Elemente, zur Hypothesenprüfung zusammengeführt.

Betrachtet man also zunächst nur die Ergebnisse des Fragebogens, in dem erhoben wurde, welche Erwartungen die Proband*innen an Reflexionsaufgaben haben, so zeigte sich in den Daten für die meisten Items eine hohe Varianz (Abb. 8).

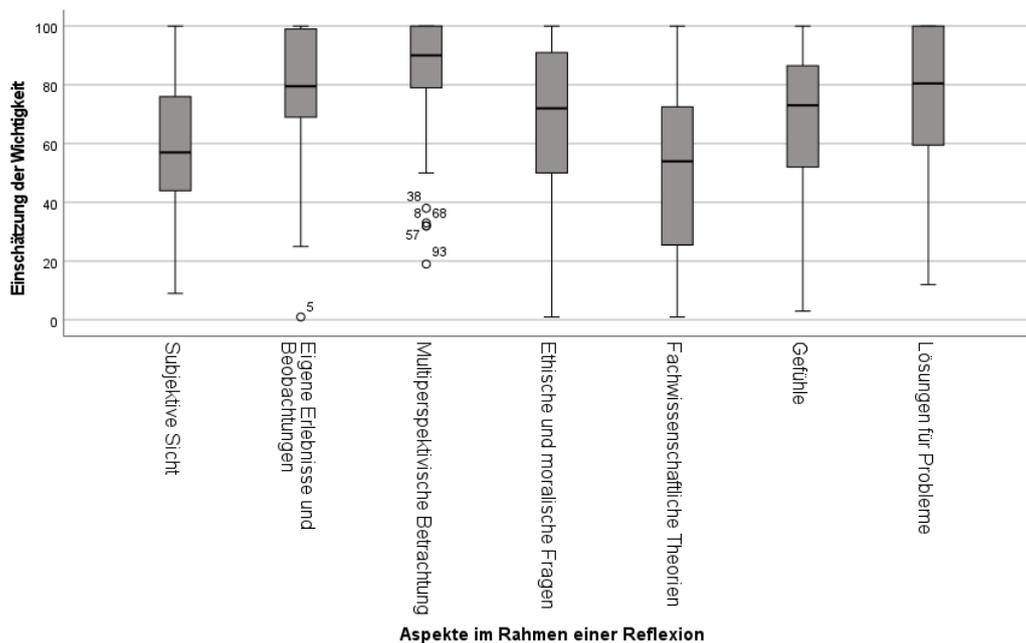


Abbildung 8. Einschätzung der Proband*innen zu Reflexionsaufgaben: Boxplots zu den „Wie wichtig ist... im Rahmen einer Reflexion“-Items.

Das heißt, die Einschätzungen der Wichtigkeit verschiedener Aspekte unterschieden sich sehr stark zwischen den einzelnen Personen. Bei den Items „Wie wichtig ist: Fachwissenschaftliche Theorien einzubeziehen“, „Wie wichtig ist: Sich auf eigene Erlebnisse und Beobachtungen beziehen“ und „Wie wichtig ist: Ethische und moralische Fragen zu erörtern“, zeigte sich beispielsweise sogar, dass die ganze Skala von 1 bis 100 genutzt wurde. Das gleiche Bild ergab sich zumindest annähernd bei der Wichtigkeit, Gefühle einzubeziehen (Min = 3, Max = 100) und die eigene subjektive Sicht auf einen Sachverhalt darzustellen (Min = 9, Max = 100).

Als für eine Reflexion am wichtigsten wurde insgesamt die Betrachtung eines Sachverhalts aus mehreren Perspektiven eingeschätzt ($M = 84.61$, $SD = 18.56$), gefolgt

vom Einbezug eigener Erlebnisse und Beobachtungen ($M = 78.97$, $SD = 19.89$), dem Finden von Lösungen für Probleme ($M = 75.06$, $SD = 24.90$), dem Nachdenken über ethische und moralische Fragen ($M = 68.89$, $SD = 24.72$) und der Auseinandersetzung mit Gefühlen ($M = 67.42$, $SD = 24.58$). Mit etwas größerem Abstand folgte noch die Darstellung der eigenen subjektiven Sicht auf einen Sachverhalt ($M = 57.82$, $SD = 23.58$) und als am wenigsten wichtig wurde der Einbezug fachwissenschaftlicher Theorien eingeschätzt ($M = 51.91$, $SD = 26.93$).

Um festzustellen, ob die drei Aufgaben einen Einfluss auf diese Erwartungen hatten, wurde mittels Varianzanalysen für Mittelwertunterschiede geprüft. Es gab keinen signifikanten Unterschied zwischen den drei Aufgaben für alle Einschätzungen (alle $ps \geq .10$) außer beim Item „Wie wichtig ist: Die eigene subjektive Sicht einzubeziehen“, $F(2, 85) = 6.01$, $p = .004$, $\eta^2 = .05$. Ein Post-hoc-Test nach Bonferroni ergab, dass Studierende, die Aufgabe 3 ($M = 69.24$, $SD = 3.76$) bearbeitet hatten, es signifikant wichtiger fanden, bei einer Reflexion ihre eigene subjektive Sicht der Dinge darzustellen, als diejenigen, die Aufgabe 1 ($M = 53.47$, $SD = 3.89$, $p = .025$) und diejenigen, die Aufgabe 2 ($M = 50.45$, $SD = 4.51$, $p = .005$) bearbeitet hatten, was sich durch die Formulierung der Aufgabenstellung und dem persönlichen Erlebnis als Reflexionsanlass erklären lässt.

Im Folgenden wird nun auf die Ergebnisse der durchgeführten Inhaltsanalyse für die entsprechenden Kategorien eingegangen. Ein Überblick findet sich in Abbildung 9.

Am häufigsten fanden sich Äußerungen der eigenen Meinung, die zum nicht-reflexiven Schreiben gehören ($M = 8.96$, $SD = 8.69$), gefolgt von „descriptive reflection“, der untersten Stufe des reflexiven Schreibens ($M = 5.15$, $SD = 4.46$), Theoriebezügen ($M = 3.53$, $SD = 4.15$) und Beschreibungen selbst erlebter Situationen ($M = 2.78$, $SD = 5.53$). Eher selten fanden sich abwägende, mehrperspektivische Reflexionen, bezeichnet als „dialogic reflection“ ($M = 0.74$, $SD = 0.99$) und Bezüge zu Gefühlen ($M = 0.39$, $SD = 0.86$). Sehr selten wurde kritisch reflektiert (Kategorie „critical reflection“, $M = 0.02$, $SD = 0.15$).

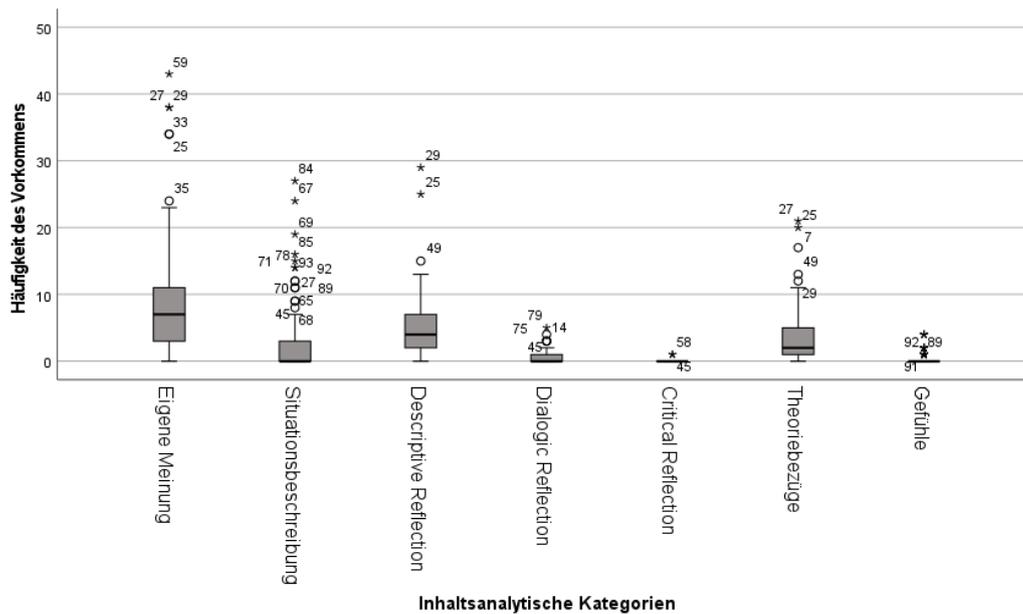


Abbildung 9. Boxplots zum Vorkommen der inhaltsanalytischen Kategorien (Teil 1 und 2) in den Texten.

Die Einschätzungen der Proband*innen zur Wichtigkeit bestimmter Aspekte in Reflexionen wurden nun in einem dritten Schritt mit dem tatsächlichen Vorkommen dieser Aspekte in den Reflexionstexten in Beziehung gesetzt. Aufgrund der Eigenheit dieser über die Inhaltsanalyse gewonnenen Daten, nämlich einer sehr geringen Varianz, einigen Ausreißern und dem Fehlen einer Normalverteilung, hätte bei der Berechnung einer Korrelation zwischen der Einschätzung der Wichtigkeit der einzelnen Faktoren und ihrem tatsächlichen Auftreten die Gefahr einer Verzerrung bestanden. Deshalb wurde ein anderes Verfahren eingesetzt, und zwar wurden die Fragebogendaten zu den Erwartungen an eine Reflexionsaufgabe jeweils mit einem Mediansplit in zwei Gruppen geteilt. So ergab sich jeweils eine Gruppe, die diesen Aspekt für weniger wichtig oder unwichtig hielt und eine Gruppe, die den Aspekt für wichtiger oder sehr wichtig hielt. Diese beiden Gruppen wurden nun hinsichtlich der Menge der tatsächlich im Text gefundenen reflexionspezifischen Elemente verglichen.

Die Einschätzung zur Frage „Wie wichtig ist: Fachwissenschaftliche Theorien in die Überlegungen miteinzubeziehen“ wurde verglichen mit der Menge an Bezügen zu fachwissenschaftlichen Theorien, die in den Texten gefunden wurden. Der t-Test für unabhängige Stichproben war in diesem Fall signifikant, $t(65.38) = 3.55$, $p = .001$, $d = 0.74$. Das heißt, die Gruppe, die den Einbezug fachwissenschaftlicher

Theorien weniger wichtig oder unwichtig fand ($M = 2.06$, $SD = 2.39$), bezog sich in den Texten auch signifikant weniger häufig auf fachwissenschaftliche Theorien, als das die Gruppe tat, die den Einbezug von Theorien wichtig oder sehr wichtig fand ($M = 5.04$, $SD = 4.98$).

Außerdem wurden gegenübergestellt: Die Einschätzung zur Frage „Wie wichtig ist: Sich auf eigene Erlebnisse und Beobachtungen zu beziehen?“ mit der Menge der Treffer in der inhaltsanalytischen Kategorie „Beschreibung erlebter Situationen“, $t(89) = 1.36$, $p = .18$, die Einschätzung zur Frage „Wie wichtig ist: Eine Situation aus verschiedenen Perspektiven zu beleuchten?“ mit der inhaltsanalytischen Kategorie „dialogic reflection“, $t(88) = 0.58$, $p = .57$, die Einschätzung zur Frage „Wie wichtig ist: Sich mit den eigenen Gefühlen auseinanderzusetzen?“ mit der inhaltsanalytischen Kategorie „Gefühle“, $t(89) = 0.58$, $p = .56$, und die Einschätzung zur Frage „Wie wichtig ist: Die eigene subjektive Sicht auf einen Sachverhalt darzustellen?“ mit der inhaltsanalytischen Kategorie „Eigene Meinung“, $t(89) = -0.60$, $p = .55$. Es zeigte sich also, dass für alle anderen Kategorien, abgesehen von dem Einbezug fachwissenschaftlicher Theorien, keine signifikanten Mittelwertunterschiede vorlagen. Für diese Kategorien, in denen mit Mediansplit keine Unterschiede gefunden wurden, wurden zur genaueren Kontrolle im Folgenden die beiden durch Mediansplit hergestellten Gruppen erneut unterteilt, sodass die Gruppe der untersten 25 % und die Gruppe der obersten 25 % verglichen werden konnten. Es wurden also noch einmal diejenigen, die einen Aspekt sehr unwichtig fanden, mit denjenigen verglichen, die den Aspekt sehr wichtig fanden, aber auch hier gab es keine signifikanten Unterschiede (alle $ps \geq .14$). Das Fragebogenitem „Wie wichtig ist: Über ethische und moralische Folgen und Fragen nachzudenken“ wäre der inhaltsanalytischen Kategorie „critical reflection“ zugeordnet gewesen, es wurde hier aber kein Mittelwertunterschied berechnet, weil in den Texten zu wenig kritische Textstellen für einen sinnvollen Vergleich gefunden wurden. Das Item „Wie wichtig ist: Lösungen für Probleme zu finden“ war keiner inhaltsanalytischen Kategorie zugeordnet.

Zusammenfassend kann man festhalten, dass die Erwartungen an Reflexionsaufgaben sich also nur im Fall des Einbezugs fachwissenschaftlicher Bezüge auch in den Texten widerspiegeln. Wenn diese als wichtiger eingeschätzt wurden, kamen sie

auch häufiger in den Texten vor und umgekehrt. Alle anderen inhaltsanalytisch ermittelten Textelemente waren in der Menge ihres Vorkommens unabhängig von den selbstberichteten Erwartungen an eine Reflexionsaufgabe.

6 Diskussion

6.1 Zusammenfassung und Einordnung der Ergebnisse

Die vorliegende Studie setzte es sich zum Ziel, ein bisher häufiger eingesetztes Verfahren zur Messung von Reflexionsfähigkeit zu prüfen: Die qualitative Inhaltsanalyse von Reflexionstexten mit dem Stufenschema von Hatton und Smith (1995). Dabei wurden zwei Teilfragestellungen beleuchtet, und zwar zum einen die Frage, inwiefern drei unterschiedliche Reflexionsanlässe Auswirkungen auf das Messergebnis zeigen und zum anderen, ob es andere situationale Voraussetzungen und Persönlichkeitsmerkmale gibt, die die Messergebnisse beeinflussen. Im Folgenden werden die Ergebnisse der Untersuchung zusammengefasst. Dabei wird zunächst auf die Ergebnisse der qualitativen Inhaltsanalyse eingegangen und in einem zweiten Schritt die Quantifizierung derselben beleuchtet. Anschließend werden die Ergebnisse zu den beiden Teilfragestellungen zusammengefasst und eingeordnet sowie die Forschungsfrage der Studie beantwortet.

Inhaltsanalyse mit dem Stufenschema von Hatton und Smith (1995). Zur Messung der Reflexionstiefe der Texte wurde das Stufenschema von Hatton und Smith (1995) eingesetzt. Es erhebt die Reflexionsqualität in vier Stufen, wobei die erste Stufe nicht-reflexives Schreiben erfasst. Auffällig ist, dass insgesamt zwei Drittel (66.10 %) des gesamten Textkorpus aus solchen Anteilen des nicht-reflektierenden Schreibens, also hier der Darstellung von eigener Meinung und Situationsbeschreibungen, besteht. Selbst wenn man die Situationsbeschreibungen aus dieser Stufe herausnimmt (vgl. Kapitel 4.4.1), bleibt noch die Hälfte (50.43 %) übrig, in der sich die Studierenden nur auf eigene Einschätzungen und Meinungen beziehen, ohne diese zu begründen oder den Gegenstand aus unterschiedlichen Perspektiven zu beleuchten. Auf die zweite Stufe von Hatton und Smith, also reflexives Schreiben mit Begründungen, entfiel noch ein knappes Drittel (29 %) der Textstellen. Auf die dritte Stufe, das dialogische, also mehrperspektivische Schreiben, entfielen nur

noch 4.18 % und auf die vierte Stufe, die kritische Reflexion, überhaupt nur zwei Textstellen (0.12 %). In Kapitel 2.2.2.1 wurden ausführlich die Befunde derjenigen Forscher*innen dargestellt, die ebenfalls das Schema von Hatton und Smith (1995) in ursprünglicher oder abgeänderter Form eingesetzt haben und verschiedene andere Untersuchungen zur Reflexionsqualität studentischer Texte vorgestellt. Die Ergebnisse der vorliegenden Studie bestätigen den bisherigen Forschungsstand, der fast durchgehend von Reflexionen auf niedrigen Stufen berichtet. Vor allem die fast vollständige Abwesenheit kritischer Reflexionen (El-Dib, 2007; Hall, 2018; Poom-Valickis & Mathews, 2013; Wyss, 2011) und der sehr geringe Anteil multiperspektivischer, erörternder Anteile (Gutzwiller-Helfenfinger et al., 2017; Penso et al., 2001; Poom-Valickis & Mathews, 2013) sind auffällig und wurden durch die vorliegenden Daten bestätigt. Es wurde bereits gezeigt, dass es nur sehr wenige Studien gab, die höhere Anteile kritischer Reflexionen verzeichneten (L. Andersen & Matkins, 2011; Behr & Welzel, 2009), allerdings nicht bei einer Analyse mit dem Schema von Hatton und Smith (1995). Ein Vergleich zwischen den Ergebnissen unterschiedlicher Messungen ist aus bereits erörterten Gründen schwierig. Ursachen für die unterschiedlichen Ergebnisse lassen sich kaum festmachen. Bei Fund et al. (2002) und Artmann et al. (2013) fanden sich ausschließlich oder überwiegend die beiden untersten Stufen in den Texten, wie es auch in den vorliegenden Reflexionstexten der Fall war. Dass diese Einschätzung nicht nur auf der Basis des Schemas von Hatton und Smith (1995) erfolgt, sondern sich auch mit anderen Stufenschemata im Rahmen qualitativer Untersuchungen ergab, wurde in Kapitel 2.2.2.1 ausführlich dargestellt. Das Ergebnis überrascht auch dahingehend wenig, als es sich bei den Studienteilnehmer*innen um Studierende zwischen dem dritten und fünften Semester handelt, die sich noch am Anfang ihrer Kompetenzentwicklung befinden. In Kapitel 2.2.2.2 wurde ein kurzer Abriss zur Entwicklung von Reflexionsfähigkeit gegeben, in dem gezeigt wurde, dass sich Reflexionsfähigkeit im Laufe der Professionalisierung steigern kann und gefördert werden kann (vgl. Kapitel 2.2.2.3). Hierzu haben sich möglicherweise im Studium der Teilnehmer*innen, die sich ja im ersten Praktikum befanden und ihr erstes Portfolio verfassten, bisher wenig gezielte Gelegenheiten ergeben. Ein Unterschied der Reflexionsergebnisse in Abhängigkeit vom Fachsemester zeigte sich in der vorliegenden Studie aber nicht. Daraus könnte geschlossen werden, dass sich die Reflexionsfähigkeit entweder nur in größeren Zeitabständen weiterentwickelt oder sich auch für

die weiter im Studium fortgeschrittenen Studierenden keine relevanten Lerngelegenheiten ergeben haben, was ein bedenkliches Ergebnis darstellen würde. Weitere mögliche Begründungen für das eher ernüchternde Ergebnis wurden in Kapitel 2.2.2.4 ausführlicher dargestellt: Ursächlich könnten auch Erwartungen der Studierenden sowie der generelle Rückgang der Bedeutung einer kritisch-moralischen Herangehensweise sein. Es könnte aber auch, zumindest zum Teil, an den Spezifika der Messmethode liegen, dass sich wiederholt die schlechten Befunde hinsichtlich der Reflexionsfähigkeit von Lehramtsstudierenden zeigen. Möglicherweise sind die Stufenmodelle noch nicht ausgereift bzw. differenziert genug oder erfassen nicht, was sie eigentlich zu erfassen behaupten. Eine weitere Erklärung liegt vielleicht in den Rahmenbedingungen der Messung, welche die bei den Studierenden eigentlich vorhandenen Fähigkeiten eventuell nicht vollständig abrufen.

In der vorliegenden Studie wurde außerdem versucht, aufgrund theoretischer Vorüberlegungen den Reflexionsbegriff komplexer zu fassen, als es bei Hatton und Smith (1995) geschah und zwei weitere Aspekte, Theoriebezüge und Bezüge zu Gefühlen, zusätzlich zu erheben. Auch andere, die die Stufen von Hatton und Smith eingesetzt haben, äußerten die Kritik, dass diese zu auflösungsschwach seien und wesentliche Aspekte des Reflektierens nicht erfassten, weshalb sie ebenfalls zusätzliche Aspekte wie die Reflexionsbreite erhoben oder die Originalstufen ausdifferenzierten (Artmann et al., 2013; Leonhard & Rihm, 2011; Lüsebrink & Grimminger, 2014). Die vorliegende Studie stimmt mit dieser Einschätzung überein. Während es bei anderen Studien aber häufig nicht gelungen ist, die zusätzlichen Untergliederungen reliabel zu kodieren, konnten mit der stark formal orientierten Kodierweise und der relativen Nähe zum Original akzeptable Reliabilitäten erreicht werden.

Reflexionsmaße. In der bisherigen Forschung wurden zwar bereits viele qualitative Inhaltsanalysen an Reflexionstexten vorgenommen, diese wurden aber selten auf individueller Ebene betrachtet oder in ein quantitatives Reflexionsmaß überführt (vgl. Kapitel 2.2.1.2). Die vorliegende Arbeit unternahm eine solche Quantifizierung der qualitativen Inhaltsanalyse anhand zweier verschiedener plausibler Mög-

lichkeiten der Berechnung. Diese zwei Varianten wurden beide im Vergleich eingesetzt: das Maß „Höchste erreichte Reflexionsstufe“ und das Maß „Durchschnittliche Reflexionsleistung“ (vgl. zur Berechnung der Maße Anhang 4 und Kapitel 4.4.1). Während das Maß „Höchste erreichte Reflexionsstufe“ eine breite Varianz aufwies und alle erreichbaren Stufen abdeckte, wiesen die Ergebnisse des Maßes „Durchschnittliche Reflexionsleistung“ einen sehr niedrigen Durchschnitt und eine geringe Standardabweichung auf. Das lässt sich zurückführen auf das bereits dargestellte Ergebnis der Mehrzahl von Textelementen auf den niedrigsten Stufen. Die „Durchschnittliche Reflexionsleistung“ nimmt demzufolge dieses wenig differenzierte Bild der Reflexionsqualität auf und ist nicht in der Lage, es auszugleichen, wie es mit dem Maß „Höchste erreichte Reflexionsstufe“ gelungen ist. Das Maß „Höchste erreichte Reflexionsstufe“ wiederum leidet unter dem Nachteil, dass es die Reflexionsqualität der Texte durch Zufallstreffer nach oben überschätzen kann (weitere Ausführungen zur Problematik der Quantifizierung, vgl. folgendes Kapitel 6.2). Während die beiden Maße einen recht hohen Überschneidungsbereich aufweisen, ergaben sich doch zum Teil unterschiedliche Ergebnisse abhängig vom gewählten Maß, sodass ein vergleichender Einsatz der beiden Maße als sinnvoll und gewinnbringend beurteilt werden kann.

Beide Maße sind durch die Art ihrer Berechnung nicht sensibel für die verfasste Textmenge, was ein wesentliches Qualitätskriterium darstellt. Das spielt eine Rolle, weil andere Studien (z.B. Lüsebrink & Grimminger, 2014) gezeigt haben, dass die Textmenge mit Faktoren wie beispielsweise dem Geschlecht zusammenhängen könnte und weibliche Teilnehmerinnen möglicherweise mehr Text produzieren. In den Daten der vorliegenden Studie zeigte sich zwar kein signifikanter Unterschied zwischen den Geschlechtern hinsichtlich der Textmenge, aber die Motivation der Teilnehmenden wirkte sich dahingehend aus, dass motiviertere Studierende längere Texte verfassten. Für ein quantifizierendes Vorgehen kann demzufolge festgehalten werden, dass es sinnvoll ist, für die produzierte Textmenge zu kontrollieren.

Einfluss der Aufgaben. Die vorliegende Arbeit setzte es sich als erste Teilfragestellung, den Einfluss des reflexionsauslösenden Anlasses zu untersuchen. Es wurde von der Vermutung ausgegangen, dass der in der Aufgabenstellung eingesetzte Reflexionsanlass, in diesem Fall graduiert von einem unpersönlichen Thema

bis zu einem persönlichen Thema, einen Einfluss auf die Reflexionsleistung in den Texten aufweist. In dieser Arbeit konnte gezeigt werden, dass die Aufgabenstellung tatsächlich einen signifikanten Effekt nicht nur auf die Reflexionsleistung, sondern vor allem auch auf die Art der Aufgabenbearbeitung hatte, also ob eine eher emotionale oder theoretisch-distanzierte Herangehensweise gewählt wurde.

Wie in Kapitel 2.2.3.1 gezeigt wurde, wurde in einigen Studien mit eigenen Erlebnissen als Reflexionsanlass gearbeitet, in anderen Studien wurden Erlebnisse und Situationen fremder Lehrkräfte gewählt. Letzteres steht aber eigentlich im Gegensatz zu den theoretischen Überlegungen vor allem von Dewey (1910; 1933) und Schön (1983), die postulierten, dass Reflexion nur von authentischen Problemen ausgelöst wird, die sich dem Individuum in der eigenen Erfahrungswelt stellen. Während einige Forscher*innen das Reflektieren eigener Erlebnisse als schwieriger einschätzten als das fremder Erfahrungen (Leijen et al., 2012; Poom-Valickis & Mathews, 2013; Risko, 2009) kamen andere zu dem Ergebnis, dass eigene Erlebnisse bessere Reflexionsergebnisse zur Folge hatten (L. Andersen & Matkins, 2011). Die Ergebnisse der vorliegenden Studie unterstützen die Annahme, dass persönliche Erlebnisse höhere Reflexionsleistungen zur Folge haben und dass das Reflektieren derselben den Studierenden nicht schwerer fällt als Reflexionsanlässe, zu denen sie mehr Distanz haben. Die Ergebnisse sind als deutlicher Hinweis darauf zu interpretieren, dass der Grad des persönlichen Bezugs zur Aufgabe das Reflexionsergebnis beeinflusst, und zwar im Sinne der theoretischen Grundlagenliteratur Deweys und Schöns. Es scheint allerdings auch nicht so zu sein, dass Situationen ohne persönlichen Bezug gar keine Reflexionen auslösen. Auch bei den unpersönlicheren Aufgaben 1 und 2 fanden sich ähnliche Reflexionsstufen und Elemente, wie bei der sehr persönlichen Reflexionsaufgabe 3. Das deutet darauf hin, dass es hier doch auch in den meistens Fällen zu einer „Übernahme und Bedeutungszuweisung durch das Subjekt“ (Häcker, 2017, S. 26) gekommen sein könnte (vgl. Kapitel 2.1.2.1), welche als Voraussetzung für eine Reflexion anlässlich nicht selbst erlebter Situationen erachtet wird.

Ferner gab es einige Hinweise aus der bisherigen Forschung (Göbel & Gösch, 2019; Kosinár, 2017; Seidel et al., 2011), dass eine höhere persönliche Involviertheit mit höheren Werten von aufgabenbezogener Motivation und einem höheren empfundenen Nutzen der Aufgabe einhergehen könnte. Während die deskriptiven Ergebnisse

der vorliegenden Studie diese Erwartungen abbildeten und die Werte für Motivation und Nutzen von Aufgabe 1 bis 3 anstiegen, erwiesen sich die Unterschiede letztlich als nicht signifikant. Trotzdem deuten die Daten darauf hin, dass der vermutete Zusammenhang bestehen könnte, auch wenn sie – möglicherweise aufgrund der Stichprobengröße – nicht eindeutig sind. Festzuhalten ist in jedem Fall, dass die mittleren Zustimmungswerte sowohl für aufgabenbezogene Motivation als auch den Nutzen der Aufgabe für alle drei Aufgaben über dem theoretischen Mittel lagen, was zeigt, dass die Aufgaben insgesamt gut gewählt waren und von den meisten Studierenden als sinnvoll und anregend beurteilt wurden.

In der vorangegangenen Forschung gab es Hinweise, dass sich die Bearbeitung der drei Aufgaben in qualitativer Weise unterscheiden könnte: Für einen steigenden Grad persönlicher Involviertheit wurde eine zunehmend emotional geprägte Bearbeitung erwartet, die sich in einer höheren Bezugnahme auf Gefühle in den Texten widerspiegelt (Poom-Valickis & Mathews, 2013). Gleichzeitig wurde die Abnahme einer theoretisch-distanzierten Haltung vermutet, die sich in der Menge fachwissenschaftlicher Bezüge ausdrückt und die bei der unpersönlichen Aufgabe 1 am höchsten erwartet wurde (Göbel & Gösch, 2019). Die vermuteten Zusammenhänge zeigten sich im Wesentlichen in den Daten: Bei der Bearbeitung der unpersönlichsten Aufgabe wurde sich deutlich häufiger auf fachwissenschaftliche Theorie bezogen als bei der persönlichsten Aufgabe, umgekehrt fanden sich bei der persönlichsten Aufgabe mehr Gefühlsäußerungen als in den unpersönlicheren. Dass die Aufgabenstellung beeinflusst, ob die Studierenden sich auf theoretisches Wissen beziehen, ist ein wichtiges Ergebnis hinsichtlich der großen Hoffnung, die in das Reflektieren gesetzt wird, wenn es darum geht, die Kluft zwischen Theorie und Praxis zu schließen (z.B. Herzog, 1995; Rodgers, 2002; Schön, 1983). Es wurde zwar gezeigt, dass in knapp 20 % der kodierten Einheiten Bezüge zur fachwissenschaftlichen Literatur hergestellt wurden. Es fehlt jedoch erstens ein Maßstab, um zu beurteilen, ob das viel oder wenig ist, und zweitens können die Qualität und der Erfolg der hergestellten Bezüge mit den verwendeten Messverfahren nicht eingeschätzt werden. Doch unabhängig davon deuten die Ergebnisse darauf hin, dass das Reflektieren zu den eigenen Erfahrungen eine Verknüpfung mit der fachwissenschaftlichen Theorie noch mehr zu erschweren scheint, als das ohnehin schon der Fall ist, wie

bereits herausgestellt wurde (Penso et al., 2001). Möglich wäre auch die Vermutung, dass sich eine eher emotionale und eine eher kognitive Herangehensweise gegenseitig ausschließen oder eine Verknüpfung sich zumindest schwierig gestaltet. Zwar zeigte sich hinsichtlich der Bezugnahme auf Gefühle, dass diese in Aufgabe 3 häufiger vorkamen, doch muss einschränkend darauf hingewiesen werden, dass nur 36 Textstellen, also 2.2 % des Textkorpus, Gefühle im Kontext des Reflektierens benannten, obwohl der Begriff ohnehin schon sehr weit gefasst wurde. Das ist insgesamt so wenig, dass kaum von einer sehr emotional behafteten Form des Schreibens gesprochen werden kann, auch nicht bei der Bearbeitung der persönlichsten Aufgabe. Interessant ist, dass sich auch hier der Unterschied wieder nur zwischen der ersten und dritten Aufgabe nachweisen ließ. Während die deskriptiven Daten einen hypothesenkonformen An- (emotionale Referenzen) bzw. Abstieg (fachwissenschaftliche Bezüge) zeigten, war dieser jeweils zur Aufgabe 2 nicht signifikant. Man kann also festhalten, dass im Rahmen der vorliegenden Daten im Grunde nur das Reflektieren einer wirklich erlebten Situation im Gegensatz zu einer fremden Situation einen Unterschied machte.

Einfluss weiterer Prädiktoren. Als zweite Forschungsfrage wurde untersucht, ob weitere Faktoren, die bisher nicht oder selten kontrolliert wurden, das Messergebnis beeinflussen. Einige vermutete Prädiktoren, einerseits situationale Voraussetzungen, andererseits Persönlichkeitsmerkmale, wurden im begleitenden Online-Fragebogen erhoben.

Untersucht wurden zunächst die situationalen Faktoren aufgabenbezogene Motivation und wahrgenommener Nutzen der Reflexionsaufgabe. Während sich Motivation nicht als Prädiktor für die Reflexionsleistung nachweisen ließ, zeigte sich der wahrgenommene Nutzen der Aufgabe durchaus als signifikanter Einflussfaktor und zwar jeweils hinsichtlich beider Reflexionsmaße. Wurden die Aufgaben also als nützlicher eingeschätzt, folgten daraus höhere Werte der ermittelten Reflexionsleistung in den Texten. Leonhard und Rihm (2011) beispielsweise äußerten die Vermutung, die unterschiedlichen Reflexionsergebnisse zu unterschiedlichen Messzeitpunkten, die sie beobachtet hatten, könnten sich durch die unterschiedliche Motivation der Teilnehmer*innen erklären lassen, die in ihrer Studie nicht kontrolliert worden war. Zumindest in den vorliegenden Daten und mit der Einschränkung,

dass nur die aufgabenbezogene Motivation erhoben wurde, zeigten sich keine Hinweise, die diese Vermutung stützen.

Ausgehend von den in Kapitel 2.1.2.6 dargestellten theoretischen Annahmen zu den Voraussetzungen von Reflexion wurde vermutet, dass höhere Ausprägungen der Persönlichkeitsmerkmale Offenheit, Ungewissheitstoleranz und Kreativität höhere Werte hinsichtlich der Reflexionsleistung nach sich ziehen. Diese drei Persönlichkeitsmerkmale hatten aber zumindest in den vorliegenden Daten keinen nachweisbaren Einfluss auf die Reflexionsleistung, und zwar beide Maße betreffend. Es ließ sich nicht einmal eine Korrelation zwischen den Persönlichkeitsmerkmalen und der Reflexionsleistung zeigen. Das lässt sich auf verschiedene Arten interpretieren: Bisher wurde empirisch noch nicht nachgewiesen, dass es sich bei diesen theoretisch postulierten Voraussetzungen von Reflexionsfähigkeit überhaupt tatsächlich um Prädiktoren handelt. Es könnte also sein, dass sie generell gar keinen empirisch nachweisbaren Zusammenhang zur Reflexionsleistung aufweisen. Es zeigt sich an dieser Stelle, dass beim Thema Reflexion die theoretische Basis und die empirische Forschung noch zu wenig zusammengeführt wurden. Eine weitere mögliche Erklärung liegt darin, dass zwar vielleicht ein Zusammenhang besteht, dass er sich aber aufgrund der kleinen Stichprobe oder der Art und Weise, wie die Reflexionsmaße gebildet wurden, nicht im Ergebnis niederschlägt. Das könnte an der Art der Ermittlung und Berechnung der Reflexionsmaße liegen oder an der Art und Weise, wie das Stufenschema von Hatton und Smith (1995) Reflexion erfasst. Die Arbeit ging von der Vermutung aus, dass die Validität der Messung durch den Einfluss der Persönlichkeitsmerkmale beeinträchtigt sein könnte. Dass kein Einfluss gezeigt werden konnte, kann aber auch nicht abschließend als Bestätigung dafür gewertet werden, dass die Messung valide ist, was sich aus den erörterten möglichen alternativen Erklärungen für das Ergebnis ableiten lässt.

Außerdem wurde angenommen, dass sich der empfundene Problemgehalt der Aufgabe positiv auf die Reflexionsleistung auswirkt. Es wurde also erwartet, dass die Reflexionsleistung zunimmt, wenn die Reflexionssituation der Aufgabe als problematischer eingeschätzt wird. Der Hintergrund für diese Annahme war, dass Reflexionen nach Dewey und Schön nur ausgelöst werden, wenn eine Person vor ein Problem gestellt und in ihrem bisherigen Handeln irritiert wird. Eine als unproblematisch oder wenig problematisch empfundene Aufgabe sollte demnach

keine oder oberflächliche Reflexionen zur Folge haben. Aber ähnlich wie bei den Persönlichkeitsmerkmalen hatte auch der empfundene Problemgehalt der Aufgabe in den vorliegenden Daten keinen signifikanten Einfluss auf die Reflexionsleistung. Auch hier ist unklar, ob es mit der vorliegenden Messung oder der Qualität der Daten nicht gelungen ist, den erwarteten Zusammenhang zu zeigen, oder ob die theoretischen Überlegungen empirisch betrachtet nicht haltbar sind. Möglicherweise wurde mit der Frage nach dem Problemgehalt der Aufgabe das, was von Dewey und Schön angenommen worden war, also eine Irritation, ein Stolperstein im Handlungsfluss, überhaupt nicht operationalisiert. Da an diesem Punkt Neuland betreten wurde, sollte sich zur Aufklärung weitere Forschung anschließen.

In Kapitel 2.1.1.2 wurde gezeigt, dass im Alltagsverständnis sehr Unterschiedliches unter dem Begriff Reflexion verstanden wird (Berndt & Häcker, 2017; Wyss & Mahler, 2021). Auch in den vorliegenden Daten zeigte sich, dass Studierende sehr unterschiedliche Vorstellungen davon haben, was bei einer Reflexionsaufgabe zu tun sei. Die sehr hohe Varianz der Einschätzungen von Aspekten, die bei einer Reflexionsaufgabe wichtig sind, reichten von *gar nicht wichtig* bis *sehr wichtig* und die großen Standardabweichungen zeigen, dass eine beachtliche Uneinigkeit darüber herrscht, was im Rahmen einer Reflexionsaufgabe verlangt wird und wie diese zu bearbeiten ist. Die Vermutung, dass sich die Einschätzung dessen, was Studierende glauben bei einer Reflexionsaufgabe tun zu müssen, auf das auswirkt, was sie tatsächlich schreiben, bestätigte sich allerdings nur teilweise. Je bedeutender Studierende es einschätzten, bei Reflexionsaufgaben Bezüge zu fachwissenschaftlichen Bezügen herzustellen, desto häufiger fanden sich entsprechende fachwissenschaftliche Bezüge in deren Reflexionstexten und umgekehrt. Der Zusammenhang zeigt sich allerdings nur bei diesem einen Aspekt. Dafür gibt es verschiedene mögliche Erklärungen. Erstens könnte es sein, dass die Studierenden etwas zwar als wichtig einschätzen, es aber in der konkreten Situation nicht umsetzen konnten oder aus anderen Gründen nicht umgesetzt haben. In diesem Fall wäre es wichtig, herauszufinden, woran es liegt, dass die Umsetzung nicht möglich war. Zweitens könnte das Ergebnis dafür sprechen, dass die Inhaltsanalyse mit dem Stufenschema von Hatton und Smith (1995) nicht valide ist in dem Sinne, dass sie nicht das erfasst, was die Studierenden glauben getan zu haben. Die Studierenden haben vielleicht ihrer Meinung nach unterschiedliche Perspektiven eingenommen, die aber durch

das Schema von Hatton und Smith oder die Kodiervorgaben der vorliegenden Studie nicht erfasst wurden. Es wäre aber auch möglich, dass die rationale Einschätzung dessen, was wichtig wäre, sich überhaupt nicht im konkreten Handeln niederschlägt.

Zur Kontrolle, ob die Aufgabenstellung „Wie könnten Sie dieses Problem lösen?“ von den Studierenden tatsächlich als Prompt für eine Reflexion wahrgenommen wurde, wurde die Frage mitaufgenommen, wie wichtig das Finden von Lösungen für Probleme im Kontext von Reflexionen eingeschätzt wird. Es zeigte sich eine sehr hohe Zustimmungsrage, sodass die Formulierung der Aufgabe insgesamt als erfolgreich beurteilt wird. Allerdings konnte sie auch nicht alle mitnehmen, was sich an einer breiten Streuung der Antworten zeigt, die auch bis in den Bereich „weniger wichtig“ reichen. An dieser Stelle zeigt sich in den Daten die Heterogenität der Reaktionen auf Reflexionsimpulse, die in Kapitel 2.2.2.3 beschrieben wurde (Hayden & Chiu, 2015; Leonhard & Rihm, 2011). Zur konkreten Formulierung der Aufgabenstellung gibt es zwar bereits Forschungsergebnisse (vgl. Kapitel 2.2.3.2), die sich aber mehr mit der Frage nach der Detailliertheit befassen und Fragen nach geeigneten Operatoren bisher weitgehend unbeantwortet lassen. Die Formulierung einer Problemlösung zeigt sich in Ansätzen als vielversprechend, hier ist aber noch systematische Forschung nötig. Zu bedenken ist auch wieder die bereits erwähnte Nähe zu anderen Konzepten, hier im Speziellen natürlich dem Problemlösen, welches ja selbst ein eigenes Forschungsfeld darstellt.

Zusammenfassende Beantwortung der Forschungsfrage. Innerhalb des Studiendesigns und der Forschungsfragen lieferte die qualitative Inhaltsanalyse mit dem Stufenschema von Hatton und Smith (1995) zufriedenstellende Ergebnisse, anhand derer die Hypothesen schlüssig beantwortet werden konnten. Es konnte jedoch auch gezeigt werden, dass die Ermittlung der Reflexionsleistung durch die Textanalyse noch nicht weit genug geht und dass es über die Erfassung von Stufen der Reflexionstiefe hinaus wichtig ist, sich mehr mit der inhaltlichen Analyse der Texte zu befassen, da Reflexionen aus vielen verschiedenen relevanten Aspekten bestehen (vgl. Kapitel 2.2.1.2). Dabei wird allerdings bezweifelt, ob es sinnvoll ist, eine Reflexionsbreite im Sinne einer weiteren Qualitätseinschätzung der Texte zu

erfassen. Eine zunächst eher explorative, beschreibende Erforschung von Textinhalten und thematischen Aspekten könnte jedoch durchaus noch einen wertvollen Erkenntnisgewinn liefern.

Ein wichtiges Ergebnis der vorliegenden Arbeit ist, dass die Beantwortung der Hypothesen zum Teil auch vom gewählten Reflexionsmaß abhängig war. Bisher gibt es kein einheitliches, empfohlenes oder bewährtes Verfahren, quantitative Reflexionsmaße aus den qualitativen Daten abzuleiten und zu berechnen. Deshalb werden aktuell noch sehr individuelle Ansätze verfolgt. Wenn die Art der Berechnung aber die Ergebnisse beeinflusst, schränkt das die Validität und Interpretierbarkeit der Untersuchungen ein und stellt damit einen wesentlichen und zu wenig beachteten Faktor der Messung von Reflexion dar.

Die erste Forschungsfrage, die sich mit dem Einfluss der Aufgabenstellung auseinandersetzte, kann zusammenfassend so beantwortet werden, dass die Aufgabenstellung das Ergebnis der Messung tatsächlich auf verschiedene Arten beeinflussen kann und sich die Hinweise einiger Wissenschaftler*innen auf diesen Zusammenhang mit kleinen Einschränkungen auch in den vorliegenden Daten bestätigten. Insgesamt wurde in der bisherigen Forschung auf diesen Aspekt häufig zu wenig Aufmerksamkeit gerichtet.

Die zweite Forschungsfrage untersuchte den Einfluss weiterer situationaler und persönlichkeitsbezogener Faktoren auf die Messung. Der vermutete Einfluss bestätigte sich in den vorliegenden Daten für die Einschätzung der Nützlichkeit der Aufgabe, nicht aber für die aufgabenbezogene Motivation oder die erhobenen Persönlichkeitsvoraussetzungen Offenheit, Ungewissheitstoleranz oder Kreativität. Auch die Frage, wie problematisch der Reflexionsanlass von den Studierenden eingeschätzt wurde oder was sie von einer Reflexionsaufgabe erwarten, stand, vom Einbezug fachwissenschaftlicher Theorie abgesehen, in keinem Zusammenhang mit der Reflexionsleistung.

Auch wenn natürlich nicht alle potentiellen Einflüsse untersucht werden konnten und sich letztlich auch nicht alle untersuchten Faktoren als Einflüsse auf die Reflexionsleistung nachweisen ließen, so zeigt die vorliegende Studie doch trotzdem deutlich, dass bei der Messung von Reflexion eine Erweiterung des Fokus auf Rahmenbedingungen der Messung zu richten ist. Es wurde außerdem evident, dass die

theoretischen Überlegungen zur Komplexität des Konstrukts sich noch zu wenig in der empirischen Forschung und den Messmethoden widerspiegeln.

Die Frage nach der Validität der Messung mit dem Schema von Hatton und Smith kann nicht abschließend beantwortet werden, sie wird von den Ergebnissen der Studie aber zumindest teilweise in Frage gestellt. Die Einschätzungen anderer Autor*innen, welche ebenfalls von Friktionen beim Einsatz der geprüften Messmethode berichten (Leonhard & Rihm, 2011), können bestätigt werden, sodass vom unkritischen Einsatz dieses und ähnlicher Messinstrumente zunächst abzuraten ist. Probleme und Grenzen des Vorgehens werden im nächsten Kapitel ausführlicher beleuchtet.

6.2 Grenzen

Sowohl aus dem Design und der Durchführung der Studie als auch aus den verwendeten Messinstrumenten und deren Auswertung ergeben sich bestimmte Grenzen, die die Belastbarkeit der Ergebnisse möglicherweise einschränken. Diese werden im Folgenden diskutiert.

Versuchsbedingungen und Methodik. Um die Stichprobe zu vergrößern, wurden zwei Erhebungen zu unterschiedlichen Zeitpunkten und an unterschiedlichen Universitäten durchgeführt. Daraus folgt auch, dass die Aufgaben in zwei verschiedene E-Portfolios eingebunden waren, die die Praktika an den Universitäten jeweils begleiteten. Die Erhebungszeitpunkte waren von äußeren Faktoren in den Praktika abhängig und fanden einmal am Semesterende des Sommersemesters, einmal in der Mitte des Wintersemesters statt. Dadurch entstanden unterschiedliche Erhebungsbedingungen, die eventuell Einfluss auf die Ergebnisse genommen haben könnten. Das Ergebnis, dass die Studierenden der Universität Erlangen/Nürnberg signifikant niedrigere Reflexionsstufen erreichten als die Passauer Studierenden, weniger motiviert waren und die Aufgaben als weniger nützlich einschätzten, könnte also auch auf den Zeitpunkt der Erhebung oder die jeweilige E-Portfolio-Umgebung zurückzuführen sein. Auch ist ein Versuchsleitereffekt an der Universität Passau nicht

ganz auszuschließen, da die Passauer Studierenden im Praktikum von der Autorin der Studie betreut wurden. Es wurde versucht, den Einfluss dadurch abzuschwächen, dass kein persönlicher Kontakt im Rahmen der Studie stattfand und sowohl in Passau als auch in Erlangen/Nürnberg alle Hinweise und Aufgaben schriftlich formuliert und in identischer Form eingesetzt wurden. Da aber bereits gezeigt wurde, dass das Vertrauen beim Verfassen von Reflexionen ein wichtiger Faktor ist (vgl. Kapitel 2.2.3.4), könnte es doch sein, dass das Wissen darum, wer die Reflexionen lesen wird, eine Rolle gespielt haben könnte.

Obwohl im Theorieteil darauf hingewiesen wurde, welchen Einfluss die Rahmenbedingungen auf das Reflektieren ausüben können, konnten im Rahmen der asynchronen und online durchgeführten Erhebung nicht alle kontrolliert bzw. berücksichtigt werden. In Kapitel 2.1.2.6 wurde beispielsweise gezeigt, dass für Reflexionen genügend Zeit vorhanden sein muss und kein Leistungsdruck bestehen sollte. Mayring (2015) betont, dass in der qualitativen Forschung die Entstehungssituation des Untersuchungsmaterials eine bedeutende Rolle spielt und beispielsweise kontrolliert werden sollte, in welcher emotionalen oder kognitiven Verfassung und vor welchem individuellen Hintergrund dieses verfasst wurde (vgl. S. 56). Ob die Aufgaben im individuellen Fall unter Zeitdruck oder in Ruhe bearbeitet wurden, ob es Störungen gab und andere Modalitäten der Entstehungssituation sind im Rahmen der online zur Verfügung gestellten Aufgaben und des Fragebogens nicht nachvollziehbar. Im Kapitel zu den Voraussetzungen wurden noch einige weitere genannt, die einen Einfluss auf das Reflexionsergebnis genommen haben könnten, beispielsweise Bildungshintergrund, Fachwissen oder Selbstwirksamkeitserwartungen, welche nicht alle kontrolliert werden konnten. Es ist also bei der Interpretation der Ergebnisse zu berücksichtigen, dass es weitere erklärende Variablen gegeben haben könnte.

In Kapitel 2.2.3.4 wurde betont, dass Freiwilligkeit ebenfalls eine wünschenswerte Rahmenbedingung darstellt, die zur Zuverlässigkeit der Ergebnisse beiträgt. In der vorliegenden Untersuchung konnte zwar die Teilnahme an der Studie verweigert werden, die Anfertigung der Reflexionsaufgaben war aber verpflichtend, da sie in das Portfoliokonzept eingearbeitet waren, welches für die erfolgreiche Ableistung des Praktikums vollständig vorgelegt werden musste. Dadurch konnte allerdings

wiederum eine möglicherweise auftretende Positivselektion bei freiwilliger Anfertigung der Aufgaben vermieden werden. Einige sehr kurze Texte, die teilweise nur aus wenigen Sätzen bestanden, deuten aber darauf hin, dass einzelne Studierende die Reflexionen vielleicht doch nur im „Modus des Abarbeitens“ (Berndt & Häcker, 2017, S. 248) erledigt haben. Dass das passiert sein könnte, ist aber auch für längere Texte prinzipiell nicht auszuschließen und kann nicht unbedingt nur an der Textlänge als Indikator festgemacht werden. Deshalb wurden sehr kurze Texte nicht aus der Auswertung ausgeschlossen. Zusätzlich war den Studierenden bekannt, dass ihre Reflexionstexte Teil einer Studie werden sollten, was die Aufrichtigkeit negativ im Sinne einer sozialen Erwünschtheit verändert haben könnte (Artmann et al., 2013). Es ist dem Studiendesign inhärent, dass die Reflexionen nicht natürlich entstanden, sondern in gewisser Weise forciert wurden. Dieser Faktor sollte zumindest teilweise mit der Frage nach dem empfundenen Problemgehalt der Aufgabe kontrolliert werden. Wäre kein Problemempfinden vorhanden gewesen, hätte dies unter anderem zeigen können, dass hier eine künstliche Reflexions-situation ohne echten Auslöser eingesetzt worden wäre. Die Auswertung des Items fiel aber uneindeutig und mit einer deutlichen Tendenz zur Mitte aus. Zum Problemgehalt der Aufgabe muss außerdem angemerkt werden, dass dieser eine sehr hohe Varianz bei Aufgabe 3, dem selbstgewählten Beispiel, aufwies. Das heißt, dass einige ihr selbstgewähltes Beispiel sehr problematisch fanden, andere allerdings ein als eigentlich unproblematisch empfundenen Beispiel gewählt hatten. Das deutet darauf hin, dass die Überlegungen, die angestellt wurden, um eine Vergleichbarkeit der individuellen Reflexionsanlässe herzustellen, nicht in jeder Hinsicht erfolgreich waren. Diejenigen, die ein eigenes Problem beschrieben haben, welches sie aber als eigentlich unproblematisch einstufen, konnten vielleicht kein problematisches Ereignis erinnern, wollten keines schildern oder haben im Praktikum kein Erlebnis als problematisch und damit reflexionswürdig wahrgenommen.

Eine mögliche Überlegung, diesen Schwierigkeiten zu begegnen, wäre der Ausschluss derjenigen, bei denen es aus den verschiedenen dargestellten Ursachen Hinweise darauf gibt, dass keine echte Reflexion ausgelöst werden konnte. Ein solches Vorgehen wurde für die vorliegende Studie nicht gewählt, denn es handelt sich nur um unbelegte Mutmaßungen und keine eindeutigen Belege dafür, dass in den geschilderten Situationen tatsächlich keine Reflexion ausgelöst wurde. Insgesamt

kann aber festgehalten werden, dass die Formulierung von Aufgaben gelungen ist, die von einem Großteil der Studierenden als ausreichend problematisch empfunden wurden, um weiterführende Überlegungen anzustellen.

Qualitative Inhaltsanalyse und Reflexionsstufen nach Hatton und Smith (1995). Wie in Kapitel 2.2.1.2 bereits kritisch angemerkt wurde, basieren die Zuweisungskriterien zu den Reflexionsstufen nach Hatton und Smith (1995) primär auf eher formalen bzw. sprachlichen Kriterien. Wird eine Aussage beispielsweise durch eine Begründung ergänzt, wird diese eine Stufe aufgewertet und als „descriptive reflection“ gewertet. Auf inhaltliche Qualität der Aussagen wird dabei keine Rücksicht genommen. Innerhalb der gleichen Kategorie findet sich bezüglich der inhaltlichen Qualität eine hohe Varianz unterschiedlicher Aussagen. Während beispielsweise eine sachlich oder fachlich falsche Begründung mit der Gewichtung 1 („descriptive reflection“) gewertet wird, bleibt ein treffender Theoriebezug ohne Begründung auf der untersten Ebene des nicht-reflexiven Schreibens und wird mit 0 Punkten gewertet. Zwar hat die Richtigkeit und fachliche Korrektheit nicht unbedingt etwas mit der Reflexionsqualität zu tun und wird in der Theorie auch nicht als relevanter Faktor erwähnt. Es ist aber auch schwer vorstellbar, dass fachliche falsche Annahmen und Aussagen im Prozess des Reflektierens zu einem erfolgreichen Reflexionsergebnis führen können. Ein wesentlicher Kritikpunkt des Schemas von Hatton und Smith ist demzufolge, dass es zu formalisiert ist und die Inhalte zu wenig berücksichtigt. In eine ähnliche Richtung geht die Kritik, dass in Reflexionen, so wie sie momentan durchgeführt werden, eigene Überzeugungen, Vorannahmen und Vorurteile vielleicht nur gerechtfertigt und verstärkt werden und die Reflexionen so dazu beitragen, dass problematische Einstellungen konserviert statt emanzipatorisch aufgebrochen werden (Cushion, 2018, S. 6; Fendler, 2003, 16f). Die vorliegende Studie kann nicht ausschließen, dass die Reflexionen hier so gewirkt haben könnten und dass auch als sehr hochwertig bewertete Reflexionen diesem Vorwurf ausgesetzt sein könnten, weil die inhaltliche Ebene nicht Teil der Analyse ist. Die Orientierung an der sprachlichen Form als Analysekriterium wurde trotz dieser inhaltlichen Bedenken verfolgt, weil nur so die Texte weitgehend reliabel und intersubjektiv nachvollziehbar kodiert werden konnten, was eine zentrale Voraussetzung für die Ableitung zuverlässiger quantitativer Reflexionsmaße aus

der qualitativen Textanalyse darstellt. Das Vorgehen einiger vorgestellter Studien, die Untersuchung der Reflexionstiefe um die Dimension der Reflexionsbreite zu erweitern (vgl. Kapitel 2.2.1.2), die sich einer eher inhaltlichen Auswertung widmen kann, scheint also begründet, ist aber in der vorliegenden Untersuchung bewusst unterblieben. Der Grund dafür war, dass bisher keine überzeugenden Kriterien vorgelegt wurden, anhand derer eine nachvollziehbare und theoretisch begründete Abstufung oder Wertung dieser Inhalte erfolgen könnte – ein Kritikpunkt, der bestehen bleibt.

Ferner war ja bereits nach den theoretischen Erwägungen deutlich, dass das Schema von Hatton und Smith (1995) einige für Reflexionen relevante Aspekte nicht erfasst (vgl. Kapitel 2.2.1.2). Deshalb wurde die Textanalyse noch um die Kategorien „fachwissenschaftliche Bezüge“ und „Bezüge zu Gefühlen“ ergänzt. Weitere möglicherweise relevante Aspekte, wie beispielsweise Bezüge zu subjektiven Theorien oder Entwicklungsmöglichkeiten (Hilzensauer, 2017) wurden aber nicht erfasst. Dass die zusätzliche Erhebung von Bezügen zu Gefühlen und Theorien einen wesentlichen Mehrwert bei der Analyse der Reflexionen brachte, konnte in dieser Arbeit gezeigt werden. Insofern liegt der Verdacht nahe, dass auch andere Aspekte, die vom Kategorienschema nicht erfasst wurden, eine Rolle spielen könnten und ein noch breiteres, differenzierteres Bild von den Reflexionen liefern könnten. Das Schema von Hatton und Smith (1995) erweist sich in der Analyse rückblickend als unterkomplex: Eine kategoriale Skalierung wie hier durchgeführt reicht wohl nicht aus, um ein komplexes Konstrukt wie Reflexion zu erfassen.

Die Arbeit ist insgesamt der gleichen Schwierigkeit ausgesetzt wie praktisch alle anderen Messungen von Reflexionstiefe, nämlich der fehlenden theoretischen Definition von Stufen der Reflexionsfähigkeit (Gutzwiller-Helfenfinger et al., 2017). Die herausgearbeiteten Stufen wurden zudem in der Literatur häufig als nicht trennscharf beschrieben: „Although we have separated these elements and stages [...] it is not possible to regard them as distinct and unrelated“ (Boud et al., 1985a, S. 27). Darüber hinaus wurden in Kapitel 2.2.1.4 bereits Forscher*innen vorgestellt, die generell Zweifel an der Hierarchisierbarkeit von Reflexionselementen geäußert haben (z.B. Sydnor et al., 2020) und die Sinnhaftigkeit von Stufenmodellen in Zweifel stellen (LaBoskey, 1993). Diese eher theoretische Frage entzieht sich allerdings

weitgehend einer empirischen Klärung und muss im Forschungsfeld noch eingehender diskutiert werden.

In der vorliegenden Studie wurde letztlich nur ein Reflektieren im Sinne eines ausschnittshaften, verschriftlichen Nachdenkens zu einer spezifischen Situation gemessen. Ob diese Reflexionen zu einer erfolgreichen Lösung des Problems in der Praxis geführt hätten, ob die Bereitschaft bei den Teilnehmer*innen vorhanden gewesen wäre, Reflexion auch ohne Anweisung einzusetzen, ob die Fähigkeiten, die hier ermittelt wurden, auch in anderen Kontexten volitional angewandt werden, bleibt offen. Insofern kann man auch nicht behaupten, man hätte mit einer solchen Messung eine *Reflexionskompetenz* erfasst, wie andere Studien zum Teil postulieren (vgl. Kapitel 2.1.1.2).

Quantifizierung und Reflexionsmaße. In der vorliegenden Studie wurde versucht, die analysierten Aspekte der Reflexionen in ein quantitatives Maß zu überführen, das über die individuelle Reflexionsleistung Auskunft geben kann. Ein Problem des quantifizierenden Vorgehens liegt darin, dass sich eine starke normative Wertung in der Gewichtung verbirgt. Dass im ursprünglichen Schema von Hatton und Smith (1995) Situationsbeschreibungen eine nicht theoriekonforme Abwertung erfahren hätten, wurde bereits gezeigt. Das wurde zwar in der vorliegenden Studie korrigiert, es bleiben aber noch andere Fragen offen: Ist eine kritische Reflexion (Gewichtung 5) beispielsweise so viel wert wie fünf Begründungen (Gewichtung 1)? Sollte ein Bezug zur fachwissenschaftlichen Theorie schwerer wiegen als eine präzise Situationsbeobachtung? Die Tätigkeiten auf den einzelnen Stufen sind im Grunde so unterschiedliche Aktivitäten, dass sie kaum wertend miteinander verglichen werden können. Die Reflexionsstufen sind außerdem naturgemäß nicht intervallskaliert und die Abstände zwischen den beschriebenen Stufen damit nicht gleich. Trotzdem werden sie hier künstlich so behandelt, was ein stark vereinfachendes Vorgehen darstellt. Damit die Quantifizierung gerechtfertigt wäre, müssten die Stufen eigentlich eine eindeutige Reihenfolge und eine klare Wertigkeit haben. Dass es daran Zweifel gibt, wurde auch schon im vorangegangenen Abschnitt deutlich gemacht.

Auch ist das Bilden eines Zahlenmaßes verallgemeinernd und reduziert die Facetten der Auswertung (vgl. Paseka et al., 2021, S. 157). Es gibt gute Argumente dafür, die Reflexionen in ihrer Komplexität und Breite zu betrachten und dadurch ein differenzierteres Profil der Reflektierenden zu erhalten, als es mit einer Zahl ausgedrückt werden kann. So bilanziert beispielsweise Gödeke in ihrer Studie:

Die Entscheidung, die fachspezifische Reflexionskompetenz der Studierenden nicht in Form eines Gesamturteils, sondern vielmehr mit Blick auf die einzelnen Teilbereiche beschreibbar zu machen, hat sich dabei als sinnvoll herausgestellt, da dadurch individuelle Reflexionsstärken und -schwächen in den Teilbereichen gezielt herausgearbeitet werden konnten. (Gödeke, 2020, S. 325)

Hier wird deutlich, dass durch das Bilden eines Gesamtscores eventuell relevante Teilinformationen nicht berücksichtigt werden oder sich im Reflexionsmaß nicht angemessen widerspiegeln, sodass die ermittelten Zahlenwerte den tatsächlichen Fähigkeiten letztlich vielleicht nicht angemessen entsprechen.

Eine Quantifizierung wurde in der vorliegenden Studie trotz dieser Bedenken versucht, weil sich nur durch diese die Chance ergibt, Forschungsfragen zu beantworten, die über eine rein qualitative Textanalyse hinausgehen, was auch gelungen ist. Inwiefern die Quantifizierung Gültigkeit beanspruchen kann, bleibt aber letztlich unklar, weil es an Möglichkeiten fehlt, das Gesamtmaß der Reflexionsleistung anhand anderer zuverlässiger Prüfgrößen zu validieren. Die Frage, ob Reflexion quantitativ erfassbar ist oder nicht, kann also nicht abschließend beantwortet werden. Sicher ist, dass die Quantifizierung mit einigen Problemen und offenen Fragen behaftet ist, zu denen die vorliegende Arbeit hinsichtlich verschiedener Aspekte einen wertvollen Diskussionsbeitrag leistet.

Validitätsprobleme der Validitätsprüfung. Die vorgelegte Arbeit versucht, die Validität eines Reflexionsmessverfahrens zu prüfen. Zur Verfügung stehen ihr dafür aber keine erprobten und abgesicherten Methoden oder Messverfahren, anhand derer eine Validitätsprüfung zuverlässig durchzuführen gewesen wäre. Die gemachten Annahmen basieren primär auf theoretischen Überlegungen, sind aber selbst auch fehleranfällig, weil sie empirisch noch gar nicht bestätigt wurden. So wurde, wie bereits erläutert wurde, empirisch zuvor beispielsweise noch nie geklärt,

ob eine Korrelation von Reflexionsleistung und den hier untersuchten Persönlichkeitsmerkmalen überhaupt vorliegt. Das Fehlen oder Vorliegen einer solchen Korrelation ist also kein sicherer Marker einer mangelnden Validität der Messmethode. Dass gesichertes Grundlagenwissen über Reflexion in so vielen Aspekten fehlt (vgl. Kapitel 2.1.3), erschwert die Prüfung der Validität des inhaltsanalytischen Verfahrens, was Krippendorff (1980) als „Trilemma“ bezeichnet:

Wenn der Inhaltsanalytiker kein direktes Wissen über seinen Gegenstand besitzt, dann kann er tatsächlich nichts über die Validität seiner Ergebnisse aussagen. Wenn er einiges Wissen über den Kontext des Materials besitzt und er dies zur Entwicklung seiner analytischen Konstrukte benutzt, dann ist dieses Wissen nicht länger unabhängig von seiner Untersuchung und kann nicht zur Validierung der Ergebnisse benutzt werden. Und wenn er es fertigbringt, das Wissen über den Untersuchungsgegenstand unabhängig von seiner Untersuchung zu bewahren, dann ist die Anstrengung, diese durch das Material zu erschließen, eigentlich überflüssig und liefert bestenfalls einen Fall zur Generalisierung des Untersuchungsverfahrens. (Krippendorff, 1980, S. 165)

Besonders bei der zweiten Forschungsfrage, die sich primär auf theoretische Annahmen stützte, wird das Problem evident. Es gibt eine Kluft zwischen normativ-geisteswissenschaftlicher und empirischer Herangehensweise an das Thema Reflexion, die es erst noch zu überwinden gilt, damit das Wissen aus beiden Bereichen konsolidiert werden kann. Einige Schritte zu einer Zusammenführung wurden in der vorliegenden Studie bereits unternommen.

Aus diesen Erwägungen leiten sich Folgerungen für die zukünftige Forschung ab, die im nächsten Kapitel dargestellt werden.

6.3 Ausblick

6.3.1 Folgerungen für zukünftige Forschung

Führt man die Argumente aus den vorherigen Kapiteln weiter, ergeben sich aus ihnen verschiedene Schlussfolgerungen hinsichtlich der Frage, wie zukünftige Forschung an die vorliegende Studie anknüpfen kann und welche Probleme und Fragestellungen es im Forschungsfeld zu lösen gibt.

Zukünftige Messungen mit der Inhaltsanalyse von Reflexionstexten. Wie bereits dargestellt können ähnliche Messungen, wie sie hier untersucht wurden, nicht uneingeschränkt empfohlen werden, wenn es darum geht, Aussagen über ein Mehr oder Weniger an Reflexion in Texten zu treffen oder den Erfolg von Interventionen zu beurteilen. Das Schema von Hatton und Smith (1995) und andere Messverfahren müssen zunächst noch systematischer erforscht werden.

Eine Möglichkeit, diese Messmethoden noch weiter zu überprüfen, als es in dieser Arbeit geschehen ist, ist der Einsatz weiterer Validierungsoptionen. Es wurde zwar bereits darauf hingewiesen, dass keine Messmethode bisher als abgesichert und zuverlässig gelten kann, sodass solche Untersuchungen zunächst unter der gleichen Unsicherheit der Ergebnisse leiden werden, wie die vorliegende. Es wäre aber trotzdem empfehlenswert, verschiedene Verfahren zur Messung von Reflexion im Vergleich einzusetzen, also zum Beispiel Fragebögen mit Stufenschemata zur Textanalyse zu vergleichen (vgl. Korthagen & Wubbels, 2000). Denn über eine solche Triangulation und eine genaue Betrachtung, wo es Übereinstimmungen und wo Abweichungen gibt, könnte man sich an den Kern von Reflexionen herantasten und feststellen, welche Aspekte sich eher quantitativ oder eher qualitativ erfassen lassen.

Eine vielleicht noch gewinnbringendere, bisher kaum genutzte Chance zur Validierung könnte auch darin bestehen, nachträglich Interviews mit den Studienteilnehmer*innen zu führen und diese selbst zu Wort kommen zu lassen (vgl. Häcker, 2017, S. 31, Kempin et al. 2019, S. 441). Auf diese Weise könnte Verschiedenes erfragt werden: Einerseits könnten die Reflektierenden berichten, ob sie selbst die Reflexion als erfolgreich und gewinnbringend erlebt haben und ob sich für sie ein

Erkenntniszuwachs eingestellt hat. Es könnte sein, dass der bisherige Ansatz, die Reflexionstiefe oder -breite einer Reflexion zu ermitteln, eigentlich in keinem Verhältnis dazu steht, ob die Reflexion tatsächlich klärend, hilfreich oder – im konstruktivistischen Sinne – viabel war. Da Reflektieren aber kein Selbstzweck, sondern ein Mittel zum Zweck ist, könnte diese Dimension ein solches Ziel darstellen, an dem sich ihr Erfolg messen lassen muss. Es lässt sich bisher nicht zuverlässig beurteilen, ob in Reflexionen bisher nur Bekanntes und Gewohntes wiederholt wird oder ob tatsächlich neue, handlungsleitende Impulse erarbeitet werden: “When reflection is understood as a turning back upon the self, the danger is that it will reveal no more than what is already known” (Cushion, 2018, S. 6). Wir können nicht eindeutig erkennen, ob die Reflexion neue Einsichten zutage gefördert oder neue Beziehungen hergestellt hat, die als hilfreich empfunden werden. Die Studienteilnehmer*innen könnten aber möglicherweise darüber Auskunft geben. Andererseits könnte man aber auch in solchen nachträglichen Interviews beispielsweise direkt Feedback zu den ermittelten Reflexionsstufen geben und im Gespräch klären, ob die Teilnehmer*innen die Einschätzung nachvollziehbar finden, oder ob sie sich selbst anders einschätzen würden.

Ein Forschungsdesiderat, das sich unmittelbar aus den in der vorliegenden Studie gewonnen Erkenntnissen ergibt, ist die weitergehende Untersuchung von Rahmenbedingungen des Reflektierens. Das heißt einerseits, dass die hier untersuchten Faktoren noch genauer betrachtet werden müssen. Es gab teilweise uneindeutige Befunde, beispielsweise die Rolle der Motivation, aber auch die weiteren Voraussetzungen betreffend. Die Reflexionsanlässe wurden nur hinsichtlich des persönlichen Bezugs variiert und die konkrete Aufgabenformulierung wurde gar nicht verändert, sodass hier noch weitere Forschungsfragen denkbar wären. Zu deren Beantwortung wären Within-Subjects-Designs mit Messwiederholungen interessant, um noch besser feststellen zu können, welchen Einfluss unterschiedliche Reflexionsanlässe oder Aufgabenformulierungen auf das individuelle Reflexionsergebnis nehmen oder ob dieses doch eher konstant bleibt. Von Interesse wären auch noch darüber hinausgehende Fragen: Wann und unter welchen Umständen wird Reflexion abgewehrt und haben das Problemempfinden und der persönliche Bezug darauf einen Einfluss? Welche Rolle spielt die Freiwilligkeit? Welche weiteren Faktoren

könnten relevante Einflussgrößen auf die Reflexionsleistung darstellen? Der Messzeitpunkt wurde in der vorliegenden Studie zwar nicht systematisch untersucht, es gab aber Hinweise darauf, dass dieser einen Einfluss nehmen könnte und vermutlich mit der zur Verfügung stehenden Zeit und Muße bzw. Prüfungsdruck in Zusammenhang steht.

Natürlich wäre es wünschenswert, dass solche Fragen anhand größerer Stichproben noch einmal beleuchtet werden. Wenn das aber unter Einsatz der qualitativen Inhaltsanalyse geschehen soll, besteht weiterhin das Problem, dass die Textanalysen so aufwändig sind, dass sich kaum größere Fallzahlen erreichen lassen. Hier müssten Wege gefunden werden, größere Stichproben zuverlässig auszuwerten, ggf. mit den in Kapitel 2.2.1.3 bereits vorgestellten Verfahren der automatisierten Textauswertung (z.B. Ullmann, 2019). Auch ist weiterhin fraglich, inwiefern es sinnvoll ist, die so gewonnenen Erkenntnisse im Sinne der Ermittlung eines individuellen Reflexionsmaßes zu quantifizieren und wie ein solches am zuverlässigsten zu ermitteln wäre.

Ein Ansatzpunkt, um Studien vergleichbar zu machen, wäre die Entwicklung kompletter Instrumente, die die Reflexionsanlässe, Aufgaben, Auswertung und ggf. Quantifizierung normieren. Zwei solche Instrumente wurden in Kapitel 2.2.1.2. vorgestellt und auch gezeigt, dass diese im aktuellen Entwicklungsstand noch nicht unverändert eingesetzt werden können (Leonhard & Rihm, 2011; Weinberger & Seyfried, 2009). Grenzen solcher Instrumente könnten darin bestehen, dass sie nicht für die Anwendung im internationalen Bereich geeignet und nicht in den sehr unterschiedlichen Forschungskontexten anwendbar sind, die angepasste Aufgaben oder Reflexionsanlässe erfordern. Trotzdem zeigen die Ergebnisse der vorliegenden Studie den Einfluss der Aufgabenstellung so deutlich, dass eine vergleichbare Messung von Reflexion nur über solche Instrumente möglich erscheint, die möglichst viele Aspekte der Messung normieren.

Es wurde auch angesprochen, dass bisher kaum Längsschnittuntersuchungen durchgeführt wurden. Um die Reliabilität der Messungen noch besser einschätzen zu können, wären wiederholte Untersuchungen nach einem längeren Zeitraum interessant. Die Frage, inwiefern Reflexionsfähigkeit veränderbar ist und in welchen Anteilen sie doch über einen längeren Zeitraum konstant bleibt, ist bisher noch ungeklärt.

Zusammenfassend kann man sagen, dass es für zuverlässige Messungen von Reflexion eine unabdingbare Voraussetzung ist, zunächst mehr Grundlagenforschung durchzuführen. Erst wenn geklärt ist, was das Reflektieren eigentlich für ein Vorgang ist, welche „kognitive[n], motivationale[n], ethische[n], volitionale[n] und soziale[n] Komponenten“ (Lang-von Wins, 2007, S. 763) in dieser Handlung stecken, aber auch, welche metakognitiven Fertigkeiten, welche Überzeugungen und Werte, welche Wissensanteile (Baumert & Kunter, 2013) und Persönlichkeitseigenschaften (Schaper & Hochholdinger, 2006) eine Rolle im Kontext von Reflexion spielen, könnte man den Vorgang tatsächlich auch über eine Messung erfassen. Damit wird deutlich: Erst wenn die Theorie der Reflexion geklärt ist, kann eine Kompetenz modelliert und operationalisiert werden.

Entwicklung alternativer Messverfahren. Die Messung von Reflexion mit dem hier vorgestellten Verfahren ist, wie gezeigt wurde, in mehrfacher Hinsicht kritisch zu sehen. Deshalb sollten auch andere Messverfahren beforscht und entwickelt werden. Sie könnten bestehende Verfahren erweitern oder sich zur Alternative für diese entwickeln.

Ein möglicherweise erfolgversprechender Ansatz könnte darin bestehen, Reflexionskompetenz als tatsächliche Handlungskompetenz zu erfassen. Damit ist weniger die Handlung des Reflektierens an sich gemeint, sondern die Erfassung der letzten Phase der Reflexion: Die Umsetzung der Überlegungen ins Handeln. In Kapitel 2.1.2.4 zum Ablauf einer Reflexion wurde ausführlich erläutert, dass Reflexion nicht beim bloßen Nachdenken endet, sondern – zumindest in den theoretischen Modellen – untrennbar mit einer praktischen Umsetzung verbunden ist. Auffällig ist aber, dass bisher fast nie die Handlungsebene in die Messungen einbezogen wurde, während sich der Erfolg einer Reflexion doch eigentlich in einer Handlung oder Handlungsänderung, also den Outcomes, widerspiegeln müsste. Gerade in dieser Phase könnte sich zeigen, ob die angestellten Überlegungen erstens überhaupt handlungswirksam werden und zweitens zu einem erfolgreicherem Handeln führen. Das Verfassen einer schriftlichen Reflexion zeigt im Grunde nicht den Erfolg derselben, weil diese theoretischen Ausführungen vielleicht nie Anwendung im tatsächlichen Leben und Handeln finden und wirkungslos bleiben

(Cornford, 2002). Über die Erhebung behavioraler Daten, beispielsweise in Beobachtungsstudien, könnten diese Ergebnisse der Reflexion sichtbar gemacht werden. Aber auch dieses Vorgehen beinhaltet Schwierigkeiten, für die Lösungen gefunden werden müssten: Für beobachtete Verhaltensänderungen könnte es auch andere Gründe als die Reflexion geben; auch eine gelungene Reflexion muss nicht unbedingt zu einer Verhaltensänderung führen und eine mögliche Verhaltensänderung könnte natürlich auch außerhalb des Beobachtungszeitraums auftreten. Zuletzt bestünde noch das Problem der Wertung: Ist eine Reflexion nur dann gelungen, wenn sie zu einer erfolgreichen Handlung führt und anhand welcher Kriterien könnte der Erfolg beurteilt werden? Auch dieser Ansatz kann also nur in Verbindung mit einer Befragung der Reflektierenden selbst erfolgen, die darüber Auskunft geben könnte, ob das Reflexionsergebnis als erfolgreich empfunden wurde und ob verändertes Handeln eine Folge des Reflektierens darstellt.

Vielfalt statt Vereinheitlichung. Es wurde gezeigt, dass verschiedene Messmethoden schon seit Jahrzehnten erarbeitet werden, ohne dass dabei ein roter Faden erkennbar wäre. Genauso lange wird auch schon über deren Probleme berichtet und darüber, dass keine einheitliche Definition und keine übereinstimmende Operationalisierung des Begriffs Reflexion gefunden werden kann. Immer noch herrscht „viel Ungewissheit und wenig Gewissheit“ (Paseka et al., 2021, S. 143). Es wäre denkbar, dass Reflexion vielleicht tatsächlich kein objektiv erfassbares theoretisches Konstrukt darstellt, das man entschlüsseln und letztlich als Kompetenz modellieren kann. Es wäre möglich, dass Reflexion ein genuin normatives und höchst bedeutungsoffenes Konstrukt ist, das tatsächlich nur mit individueller Schwerpunktsetzung erschlossen werden kann und dass es sich ferner als Phänomen einer vergleichbaren empirischen Messung entzieht. Davon ausgehend wäre das Ziel möglicherweise keine Vereinheitlichung der Forschung und der Messmethoden, sondern man könnte es beim Stand der Dinge belassen, nämlich, dass die Forschenden Reflexion so definieren und operationalisieren, wie es ihren Forschungsinteressen und ihrem Kontext entspricht. Die Folge daraus müsste dann allerdings sein, dass nicht mehr der Anspruch erhoben wird, allgemeingültige Aussagen aus den Studien ableiten zu können – möglich wären folglich nur noch evaluierende Ansätze auf der Ebene von Einzelprojekten. Aussagen über die

Reflexionsfähigkeit von Studierenden im Allgemeinen, deren Entwicklung oder die Wirksamkeit von Methoden (vgl. Kapitel 2.2.2) würden sich folglich verbieten, da deutlich gemacht wurde, dass die Studienergebnisse in dieser Form in keiner Weise vergleichbar oder generalisierbar sind.

Bevor man sich allerdings auf diese Option zurückziehen sollte, stehen noch viele Möglichkeiten offen, neue Messmethoden zu entwickeln, die bestehenden Messmethoden von Reflexion zu prüfen und sie mit dem Ziel weiterzuentwickeln, zu Ergebnissen zu gelangen, die im Forschungsfeld überzeugen können. Einige solcher Möglichkeiten wurden in diesem Kapitel vorgeschlagen.

6.3.2 Folgerungen für die Lehrkräftebildung

Abschließend werden aus den Ergebnissen der vorliegenden Studie Folgerungen für die Lehrkräftebildung abgeleitet.

Zu Beginn der Arbeit wurde gezeigt, dass der Reflexionsbegriff nicht nur in der Wissenschaft, sondern auch im Alltagsverständnis sehr uneinheitlich interpretiert wird. Letzteres bestätigte sich deutlich in der vorliegenden Studie: Studierende verstehen den Begriff Reflexion sehr verschieden und mit individuell gesetzten Schwerpunkten. An eine Reflexionsaufgabe haben sie folglich sehr unterschiedliche Erwartungen. Es ist also zunächst wichtig, dass Lehrerbildner*innen den Reflexionsbegriff für sich selbst klären, festlegen, was sie darunter verstehen und sich bewusst machen, welche Erwartungen sie mit ihm verbinden. Diese Erwartungen müssen den Studierenden transparent kommuniziert werden. Denn diese wünschen sich mehr Orientierung und klarere Anweisungen beim offenen Begriff Reflexion: „Someone should have told us what reflection is at the beginning of the course. Everyone wants us to reflect and no one explains exactly what reflection should be“ (ein*e Studienteilnehmer*in in: L. Andersen & Matkins, 2011, S. 35). Dabei könnte die „Perspektivendifferenz“ (Paseka et al., 2021, S. 158) im Verständnis des Begriffs Reflexion gewinnbringend zum Ausgangspunkt einer Diskussion über Reflexion gemacht und Leitlinien und Ziele derselben auch gemeinsam erarbeitet werden. Trotz des Appells, expliziter zu thematisieren, was bei Reflexionen erwartet wird und wie vorzugehen ist, sollte dies nicht in einem Zuviel an Formalismus und

technischer Vermittlung enden, damit Motivation und individueller Zugang nicht abhandenkommen. Die Folge könnte sonst der Verlust einer frischen und unvoreingenommenen Herangehensweise sein, „since reflection is reduced to a set of techniques, instead of being a higher-order cognitive/affective/socially conscious activity“ (Akbari, 2007, S. 201).

Ein solcher Reduktionismus könnte auch dazu führen, dass das kritisch-emanzipatorische Potential von Reflexion nicht genug zum Tragen kommt, welches ohnehin bereits gering ausgeprägt ist, wie die vorliegende Studie zeigt. Das fast gänzliche Fehlen kritischer Reflexion und moralisch-gesellschaftlicher Bezüge geht auch mit dem Stand der Forschung insgesamt konform (vgl. Kapitel 2.2.2.1). Falls dies die Ursache darin hat, dass den Studierenden solche eher abstrakteren, aus einer gewissen Distanz stattfindenden Denkbewegungen noch schwer fallen, kann eine Antwort darauf eigentlich nur darin bestehen, möglichst viele Gelegenheiten zu schaffen, sich gemeinsam mit ethischen, politischen, moralischen oder sozial relevanten Inhalten und Konsequenzen von Schule, Unterricht und Erziehung auseinanderzusetzen. Seminare oder Prüfungsformate sollten mehr zur eigenen Meinungsbildung und zur kontroversen, ergebnisoffenen Diskussion anregen. Es wurde gezeigt, dass die kritische, emanzipative Dimension des Reflexionsbegriffs gegenwärtig an Aufmerksamkeit verliert und sich auch wenig in den vor allem quantitativen Messmethoden abbildet – eine Entwicklung, der hinsichtlich einer sich schnell wandelnden Welt und der Bedeutung von Verantwortung und Mitgestaltung in dieser entgegengewirkt werden sollte. Das könnte sich im Rahmen eines didaktischen Wandels weg von einer dozierendenzentrierten Didaktik hin zu einer Aufwertung diskursiver Anteile und kooperativem Austausch vollziehen. Reflexion könnte so nicht nur als individuelle Aufgabe, sondern als kooperative, soziale Tätigkeit verstanden werden, die vom Diskurs und dem Austausch unterschiedlicher Perspektiven lebt (vgl. Kapitel 2.1.2.5). Da die Ergebnisse dieser Arbeit zum studentischen Reflexionsstand außerdem ergeben haben, dass die Reflexionsprodukte wenig Multiperspektivität und alternative Ideen enthalten, könnte mit solchen vom Austausch lebenden Ansätzen eine offener geistige Haltung geprägt werden. Die Offenheit könnte auch mit einer zukunftsorientierteren und kreativeren Herangehensweise an Reflexion verknüpft werden, als es bisher der Fall ist, denn aktuell finden sich diese Aspekte in Reflexionen kaum wieder. Das könnte sich beispielsweise über den Einsatz kreativer und freier Aufgabenstellungen realisieren lassen,

die auch schöpferisch-assoziative Gestaltungsformen wie Zeichnung, Fotografie, Mediengestaltung oder die Arbeit mit Metaphern zulassen (vgl. Korthagen, 1993b). Es wäre wünschenswert, dass sich eine gewisse Offenheit auch in den Prüfungsformaten abbildet.

Im Zuge dieser Öffnung sollte das Reflektieren also nicht auf einer rein intellektuell-theoretischen Ebene bleiben, sondern dazu angeregt werden, sich mit eigenen Gefühlen, Werten und Haltungen auseinanderzusetzen, um auch diesen Dimensionen von Reflexion gerecht zu werden, die immer noch eher ein Schattendasein führen. Die vorliegende Arbeit zeigt, dass es bei eigenen Erlebnissen zumindest ansatzweise gelingt, sich auf diese zu beziehen und dass solche Bezüge praktisch gänzlich verschwinden, sobald es um eine distanziertere, eher theoretische Auseinandersetzung mit Fallbeispielen geht. Dazu anzuregen, eigene Gefühle und individuelle Haltungen in Reflexionen einzubeziehen, könnte also vor allem bei eher distanziert-theoretischen Aufgaben gewinnbringend sein. In Kapitel 2.1.2.7 wurde bereits gezeigt, dass gewinnbringende Reflexionen eigentlich untrennbar und auf vielen Ebenen mit Emotionen verknüpft sind.

Gleichzeitig sollten aber auch Lerngelegenheiten für diejenigen geschaffen werden, die sich eher auf eigene Gefühle, Erlebnisse und Eindrücke und weniger auf wissenschaftliches Wissen beziehen, damit diese die Fähigkeit erwerben, auch die fachwissenschaftliche Theorie in ihre Überlegungen miteinzubeziehen. Da in der vorliegenden Untersuchung von den Studierenden vor allem der Punkt „Fachwissenschaftliche Theorien einbeziehen“ als am wenigsten wichtig im Rahmen einer Reflexion erachtet wurde und sich auch gezeigt hat, dass sich in den Texten eher wenig auf Theorien rückbezogen wurde, wäre es von Bedeutung, den Nutzen theoretischen Wissens beim Reflektieren herauszustellen und zu verdeutlichen. Da vor allem die Verknüpfung eigener Erlebnisse und persönlicher Erfahrungen mit theoretischen Inhalten schwerfiel, könnte diese beispielsweise in Praktikumsbegleitveranstaltungen oder -portfolios stärker eingeübt und angeleitet werden. Auch hier geht es aber wieder nicht primär darum, Reflexion nur instrumentell einzusetzen, um die erziehungswissenschaftliche Theorie effizienter zu vermitteln, wie beispielsweise Valli (1992) kritisiert. Wenn die Studierenden nicht die Erfahrung ma-

chen, dass sie durch den Einbezug wissenschaftlicher Theorie tatsächlich profitieren und durch diesen das Reflektieren über Praxiserfahrungen bereichert wird, wird sich an der relativen Geringschätzung der Theorie kaum etwas ändern.

Wie sich in Kapitel 2.2.2.3 zur Förderung von Reflexionsfähigkeit gezeigt hat, waren praktisch alle vorgestellten methodischen Möglichkeiten, Reflexionsfähigkeit zu fördern, bei einigen Studierenden erfolgreich, bei anderen nicht. Diese Heterogenität deutete sich in Ansätzen auch für die vorliegende Studie an, wenngleich sie kein zentrales Forschungsinteresse darstellte. Daraus kann abgeleitet werden, dass der Schlüssel zum Erfolg nicht in einer einzigen Methode, sondern einer abwechslungsreichen Kombination liegt, die für die Einzelnen vielfältige individuelle Anknüpfungspunkte bietet. Wichtig scheint außerdem eine möglichst individuelle und längerfristige Begleitung durch Mentor*innen (Leonhard & Bolle, 2015), da Professionalisierungsprozesse nicht unbedingt von selbst ablaufen (vgl. Kapitel 2.2.2.2).

Die vorliegende Studie deutet auf einen Zusammenhang des eingeschätzten Nutzens der Aufgabe mit der Reflexionsleistung hin. Dies gilt es beim Stellen von Reflexionsaufgaben zu beachten. Das könnte beispielsweise durch bedeutungsinduzierende Prompts geschehen, die ins Bewusstsein rufen, was durch Reflexion zu gewinnen ist. Wie bereits ausführlich dargestellt wurde, zeigten die Daten, die in dieser Studie erhoben wurden, keinen signifikanten Unterschied zwischen den drei Aufgaben hinsichtlich der ausgelösten Motivation, obwohl die deskriptiven Daten in Übereinstimmung mit anderen Studien (Göbel & Gösch, 2019; Kosinár, 2017; Seidel et al., 2011) darauf hindeuten, dass persönlichere Reflexionssituationen evtl. eine höhere Motivation auslösen könnten. Insofern wäre es ratsam, eigene Erfahrungen oder Praxiserlebnisse der Studierenden in die Lehre zu integrieren und aufzugreifen.

Im Kapitel 2.2.3 wurde herausgestellt, dass die Rahmenbedingungen des Reflektierens einen wesentlichen Einfluss auf die Qualität der Reflexionen darstellen, was sich teilweise auch in dieser Studie zeigen ließ. Bedingungen wie genügend Zeit, eine vertrauensvolle Atmosphäre, die Wahl der Reflexionsanlässe und die Formulierung der Aufgabenstellung könnten diesbezüglich eine Rolle spielen. Insofern sollten Lehrerbilder*innen mit diesem differenzierten Wissen an die Formulierung von Reflexionsaufgaben und die Herstellung von Lernumgebungen gehen und sich im Klaren sein, dass das Reflexionsergebnis der Studierenden zumindest teilweise

auch von den Vorgaben und Lernumgebungen der Dozierenden abhängt. Dazu gehört beispielsweise auch, Reflexionen möglichst nicht als Prüfungsformate einzusetzen und sie von Leistungsdruck, Zeitdruck und Beurteilung freizuhalten. Eine vertrauensvolle und fehlertolerante Begleitung und Beratung könnte die Voraussetzung für aufrichtige Reflexionen schaffen. Wenn das nicht gewährleistet werden kann, muss in Betracht gezogen werden, dass Reflexionsaufgaben nur „abgearbeitet“ werden und nicht nur unaufrichtig sind, sondern dann auch ihr Potential verfehlen (Artmann et al., 2013; Berndt & Häcker, 2017; Boud, 2010; Rosenberger, 2017). Wünschenswert wäre die Schaffung institutioneller Freiräume und strukturell verankerter Möglichkeiten zur Reflexion – nicht nur in der Lehrkräftebildung, sondern auch im Lehrberuf.

7 Literaturverzeichnis

- Adorno, T. W. (2017). *Erziehung zur Mündigkeit: Vorträge und Gespräche mit Hellmut Becker 1959 - 1969* (26. Aufl.). Suhrkamp Taschenbuch: Bd. 11. Suhrkamp.
- Aeppli, J. & Lötscher, H. (2017). Charakterisierung der Reflexionskategorien zum Rahmenmodell für Reflexion EDAMA. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 159–175). Julius Klinkhardt.
- Afshar, H. S. & Farahani, M. (2015). Reflective thinking and reflective teaching among Iranian EFL teachers: Do gender and teaching experience make a difference? *Procedia - Social and Behavioral Sciences*, 192, 615–620. <https://doi.org/10.1016/j.sbspro.2015.06.107>
- Akbari, R. (2007). Reflections on reflection: A critical appraisal of reflective practices in L2 teacher education. *System*, 35(2), 192–207. <https://doi.org/10.1016/j.system.2006.12.008>
- Akbari, R., Behzadpoor, F. & Dadvand, B. (2010). Development of English language teaching reflection inventory. *System*, 38(2), 211–227. <https://doi.org/10.1016/j.system.2010.03.003>
- Allas, R., Leijen, Ä. & Toom, A. (2017). Supporting the construction of teacher's practical knowledge through different interactive formats of oral reflection and written reflection. *Scandinavian Journal of Educational Research*, 61(5), 600–615. <https://doi.org/10.1080/00313831.2016.1172504>
- Allen, K. L., Brodeur, K., Israelson, M. H., Martin-Kerr, K.-G., Ortmann, L. & Peterson, D. S. (2018). Developing reflective practice in teacher candidates through program coherence. *Teaching & Learning Inquiry*, 6(2), 81–96.

- Alwin, D. (2010). How good is survey measurement? Assessing the reliability and validity of survey measures. In P. V. Marsden & J. D. Wright (Hrsg.), *Handbook of survey research* (S. 405–434). Emerald.
- Andersen, L. & Matkins, J. J. (2011). Web 2.0 tools and the reflections of preservice secondary science teachers. *Journal of Digital Learning in Teacher Education*, 28(1), 27–38.
- Andersen, N. B., O'Neill, L., Gromsen, L. K., Hvidberg, L. & Morcke, A. M. (2014). A validation study of the psychometric properties of the Groningen Reflection Ability Scale. *BMC Medical Education*.
- Armutcu, N. & Yaman, S. (2010). ELT pre-service teachers' teacher reflection through practicum. *Procedia - Social and Behavioral Sciences*, 3, 28–35. <https://doi.org/10.1016/j.sbspro.2010.07.009>
- Arrastia-Chisholm, M. C., Torres, K. M. & Tackett, S. (2017). Using Reflection to Increase Self-Regulation Among Pre-Service Teachers. In Z. Djoub (Hrsg.), *Advances in Higher Education and Professional Development. Fostering reflective teaching practice in pre-service education* (S. 148–165). IGI Global. <https://doi.org/10.4018/978-1-5225-2963-7.ch008>
- Artmann, M. (2019). "Es ist mir wichtig, dass die Studierenden sehen, dass Reflexion ohne Theorie ja gar nicht funktioniert." Epistemologische Zugänge von Hochschullehrenden zum Theorie-Praxis-Problem in der Lehrer*innenbildung. *Frontiers in human neuroscience*, 20(3). <https://doi.org/10.17169/fqs-20.3.3129>
- Artmann, M., Herzmann, P., Hoffmann, M. & Proske, M. (2013). Wissen über Unterricht - Zur Reflexionskompetenz von Studierenden in der ersten Phase der Lehrerbildung. In A. Gehrman, B. Kranz, S. Pelzmann & A. Reinartz (Hrsg.), *Formation und Transformation der Lehrerbildung: Entwicklungstrends und Forschungsbefunde* (S. 134–150). Klinkhardt.
- Atkins, S. & Murphy, K. (1993). Reflection: A review of the literature. *Journal of Advanced Nursing*, 18(8), 1188–1192. <https://doi.org/10.1046/j.1365-2648.1993.18081188.x>

- Aufschnaiter, C. von, Fraij, A. & Kost, D. (2019). *Reflexion und Reflexivität in der Lehrerbildung*. Justus-Liebig-Universität Gießen.
- Auge, T. (2019). *Reflexivität von Studierenden als Teil der Qualitätsoffensive PSI: Eine qualitative Analyse reflexiver Inhalte in Praxisphasen der universitären Lehramtsausbildung am Beispiel des WAT-Praxissemesters der Universität Potsdam. Lern dich glücklich - Arbeiten zum prozessorientierten Lehren und Lernen: Bd. 2*. Universitätsverlag Potsdam. <https://doi.org/10.25932/publishup-43091>
- Aukes, L. C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R. P. & Slaets, J. P. J. (2007). The development of a scale to measure personal reflection in medical practice and education. *Medical teacher*, 29(2-3), 177–182. <https://doi.org/10.1080/01421590701299272>
- Ayan, D. & Seferoğlu, G. (2011). Using electronic portfolios to promote reflective thinking in language teacher education. *Educational Studies*, 37(5), 513–521. <https://doi.org/10.1080/03055698.2010.539782>
- Bain, J. D., Ballantyne, R., Packer, J. & Mills, C. (1999). Using journal writing to enhance student teachers' reflectivity during field experience placements. *Teachers and Teaching*, 5(1), 51–73. <https://doi.org/10.1080/1354060990050104>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037//0033-295X.84.2.191>
- Baumert, J. & Kunter, M. (2013). Professionelle Kompetenz von Lehrkräften. *Stichwort: Zeitschrift für Erziehungswissenschaft*, 277–337.
- Beauchamp, C. (2006). *Understanding reflection in teaching: A framework for analyzing the literature*. Dissertation.
- Beauchamp, C. (2015). Reflection in teacher education: Issues emerging from a review of current literature. *Reflective Practice*, 16(1), 123–141. <https://doi.org/10.1080/14623943.2014.982525>
- Behr, F. & Welzel, M. (2009). Entwicklung der Reflexionskompetenz als Metakompetenz von Erzieherinnen: Videoanalyse eines

- Fortbildungsprogramms. In D. Höttecke (Hrsg.), *Chemie- und Physikdidaktik für die Lehramtsausbildung* (S. 119-121). LIT-Verl.
- Benko, S. L., Guise, M., Earl, C. E. & Gill, W. (2016). More than social media: Using twitter with preservice teachers as a means of reflection and engagement in communities of practice. *Contemporary Issues in Technology and Teacher Education (CITE Journal)*, 16(1), 1–21.
- Bergman, D. (2015). Comparing the effects of classroom audio-recording and video-recording on preservice teachers' reflection of practice. *Teacher Educator*, 50(2), 127–144.
- Berndt, C. & Häcker, T. (2017). Der Reflexion auf der Spur. Über den Versuch, Reflexionen von Lehramtsstudierenden zum Forschungsgegenstand zu machen. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 240–253). Julius Klinkhardt.
- Birmingham, C. (2004). Phronesis. *Journal of Teacher Education*, 55(4), 313–324. <https://doi.org/10.1177/0022487104266725>
- Blomberg, G., Sherin, M. G., Renkl, A., Glogger, I. & Seidel, T. (2014). Understanding video as a tool for teacher education: Investigating instructional strategies to promote reflection. *Instructional Science*, 42(3), 443–463. <https://doi.org/10.1007/s11251-013-9281-6>
- Boer, H. de & Reh, S. (2012). *Beobachtung in der Schule - Beobachten lernen. Lehrbuch*. Springer VS. <https://doi.org/10.1007/978-3-531-18938-3>
- Bogner, K. & Landrock, U. (2015). *Antworttendenzen in standardisierten Umfragen*. https://doi.org/10.15465/gesis-sg_016
- Bohdick, C. (2016). Denken und schreiben lernen. Konzeption und Evaluation eines Workshops für Lehramtsstudierende. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 34(1), 65–77. <https://doi.org/10.25656/01:13920>
- Bonanati, S., Westphal, P. & Wiethoff, C. (2020). Theoriebasierte Fallreflexion (TFR) im Praxissemester: Didaktische Umsetzung und Evaluation. *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption*,

- Gestaltung und Diskussion*, 3(1), 461–479. <https://doi.org/10.4119/HLZ-3142>
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler; mit 87 Tabellen* (4. Aufl.). Springer-Lehrbuch Bachelor, Master. Springer-Medizin-Verl.
- Boud, D. (2010). Relocating reflection in the context of practice. In H. Bradbury, N. Frost, S. Kilminster & M. Zukas (Hrsg.), *Beyond reflective practice: New approaches to professional lifelong learning* (S. 25–36). Routledge.
- Boud, D., Keogh, R. & Walker, D. (1985a). Promoting reflection in learning: A model. In D. Boud (Hrsg.), *Reflection: Turning experience into learning* (S. 18–40). Kogan Page.
- Boud, D., Keogh, R. & Walker, D. (1985b). What is reflection in learning? In D. Boud (Hrsg.), *Reflection: Turning experience into learning* (S. 7–17). Kogan Page.
- Boud, D., Keogh, R. & Walker, D. (2013). *Reflection: Turning experience into learning*. Routledge.
- Braun, J. A. & Crumpler, T. P. (2004). The social memoir: An analysis of developing reflective ability in a pre-service methods course. *Teaching and Teacher Education*, 20(1), 59–75.
<https://doi.org/10.1016/j.tate.2003.09.006>
- Brockbank, A. & McGill, I. (2000). *Facilitating reflective learning in higher education* (Reprinted.). Society for Research into Higher Education & Open University Press.
- Brosziewski, A. (2013). Die Zweiseitigkeit der Reflexion: Portfolios als Selbstreflexion der Lehrerinnen- und Lehrerbildung - Eine empirische Untersuchung zur Portfolioarbeit an drei Standorten. In B. Koch-Priewe, T. Leonhard, A. Pineker & C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung: Konzepte und empirische Befunde* (S. 212–225). Klinkhardt.
- Carrington, S. & Selva, G. (2010). Critical social theory and transformative learning: Evidence in pre-service teachers' service-learning reflection logs.

- Higher Education Research & Development*, 29(1), 45–57.
<https://doi.org/10.1080/07294360903421384>
- Carter Andrews, D. J., Brown, T., Castillo, B. M., Jackson, D. & Vellanki, V. (2019). Beyond damage-centered teacher education: Humanizing pedagogy for teacher educators and preservice teachers. *Teachers College Record*, 21(6), 1–28.
- Cavilla, D. (2017). The effects of student reflection on academic performance and motivation. *SAGE Open*, 7(3), 1-13.
<https://doi.org/10.1177/2158244017733790>
- Charness, G., Gneezy, U. & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1–8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Cheng, G. & Chau, J. (2013). An approach to identify levels of reflection using latent semantic analysis. In *2013 International Conference on IT Convergence and Security (ICITCS 2013): Macao, 16 - 18 December 2013* (S. 1–3). IEEE. <https://doi.org/10.1109/ICITCS.2013.6717800>
- Cherrington, S. & Loveridge, J. (2014). Using video to promote early childhood teachers' thinking and reflection. *Teaching and Teacher Education*, 41, 42–51. <https://doi.org/10.1016/j.tate.2014.03.004>
- Chittum, J. R. (2018). The theory-to-practice ePortfolio: An assignment to facilitate motivation and higher order thinking. *International Journal of ePortfolio*, 8(1), 27–42.
- Choy, S. C. & Oo, P. S. (2012). Reflective thinking and teaching practices: A precursor for incorporating critical thinking into the classroom? *International Journal of Instruction*, 5(1), 167–182.
- Christianson, S.-A. (2014). *The handbook of emotion and memory*. Psychology Press. <https://doi.org/10.4324/9781315807454>
- Chye, S., Zhou, M., Koh, C. & Liu, W. C. (2021). Levels of reflection in student teacher digital portfolios: A matter of sociocultural context? *Reflective Practice*, 22(5), 577–599. <https://doi.org/10.1080/14623943.2021.1937976>

- Clarà, M. (2015). What is reflection? Looking for clarity in an ambiguous notion. *Journal of Teacher Education*, 66(3), 261–271.
<https://doi.org/10.1177/0022487114552028>
- Collin, S. & Karsenti, T. (2011). The collective dimension of reflective practice: The how and why. *Reflective Practice*, 12(4), 569–581.
<https://doi.org/10.1080/14623943.2011.590346>
- Collin, S., Karsenti, T. & Komis, V. (2013). Reflective practice in initial teacher training: Critiques and perspectives. *Reflective Practice*, 14(1), 104–117.
- Contreras, J. A., Edwards-Maddox, S., Hall, A. & Lee, M. A. (2020). Effects of reflective practice on baccalaureate nursing students' stress, anxiety and competency: An integrative review. *Worldviews on Evidence-Based Nursing*, 17(3), 239–245. <https://doi.org/10.1111/wvn.12438>
- Cornford, I. R. (2002). Reflective teaching: Empirical research findings and some implications for teacher education. *Journal of Vocational Education & Training*, 54(2), 219–236. <https://doi.org/10.1080/13636820200200196>
- Cramer, C. (2016). *Forschung zum Lehrerinnen- und Lehrerberuf: Systematisierung und disziplinäre Verortung eines weiten Forschungsfeldes*. Verlag Julius Klinkhardt.
- Cushion, C. J. (2018). Reflection and reflective practice discourses in coaching: A critical analysis. *Sport, Education and Society*, 23(1), 82–94.
<https://doi.org/10.1080/13573322.2016.1142961>
- Czerwionka, T., Knutzen, S. & Bieler, D. (2010). Mit ePortfolios selbstgesteuert lernen. Ein Ansatz zur Reflexionsförderung im Rahmen eines hochschulweiten ePortfoliosystems. *MedienPädagogik*(18), 1–21.
<http://dx.doi.org/10.21240/mpaed/18/2010.03.26.X>
- Dalbert, C. (1999). *Die Ungewißheitstoleranzskala: Skaleneigenschaften und Validierungsbefunde*. *Hallesche Berichte zur pädagogischen Psychologie: Bd. 1*. Universitäts- und Landesbibliothek Sachsen-Anhalt; Martin-Luther-Univ. Halle-Wittenberg.
- Damasio, A. R. (2015). *Descartes' Irrtum: Fühlen, Denken und das menschliche Gehirn* (8. Aufl.). *List-Taschenbuch: Bd. 60443*. List.

- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences, 12*(1), 91–142.
- Davis, E. A. (2006). Characterizing productive reflection among preservice elementary teachers: Seeing what matters. *Teaching and Teacher Education, 22*(3), 281–301. <https://doi.org/10.1016/j.tate.2005.11.005>
- Deci, E. L., Eghrari, H., Patrick, B. C. & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62*(1), 119–142. <https://doi.org/10.1111/j.1467-6494.1994.tb00797.x>
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior. Perspectives in social psychology*. Plenum Press.
- Decker, A.-T., Kunter, M. & Voss, T. (2015). The relationship between quality of discourse during teacher induction classes and beginning teachers' beliefs. *European Journal of Psychology of Education, 30*(1), 41–61. <https://doi.org/10.1007/s10212-014-0227-4>
- Deing, P. (2019). Selbstreguliertes Lernen. Theoretische Grundlagen und Förderempfehlungen. In S. Rietmann & P. Deing (Hrsg.), *Psychologie der Selbststeuerung* (S. 319–345). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-24211-4_14
- Dewey, J. (1910). *How we think*. Heath.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D. C. Heath and Company.
- Dick, A. (1999). *Vom Ausbildungs- zum Reflexionswissen in der LehrerInnenbildung*. Pädagogisches Institut der Universität.
- Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press.
- Dudenredaktion. (o.J.). "Reflexion" auf Duden online, <https://www.duden.de/node/119540/revision/462249>.

- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Edwards, S. (2017). Reflecting differently. New dimensions: Reflection-before-action and reflection-beyond-action. *International Practice Development Journal*, 7(1), 1–14. <https://doi.org/10.19043/ipdj.71.002>
- El-Dib, M. A. B. (2007). Levels of reflection in action research. An overview and an assessment tool. *Teaching and Teacher Education*, 23(1), 24–35. <https://doi.org/10.1016/j.tate.2006.04.002>
- Epler, C. M., Drape, T. A., Broyles, T. W. & Rudd, R. D. (2013). The influence of collaborative reflection and think-aloud protocols on pre-service teachers' reflection: A mixed methods approach. *Journal of Agricultural Education*, 54(1), 47–59.
- Etscheidt, S., Curran, C. M. & Sawyer, C. M. (2012). Promoting reflection in teacher preparation programs. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 35(1), 7–26. <https://doi.org/10.1177/0888406411420887>
- Faddar, J., Vanhoof, J. & Maeyer, S. de (2017). School self-evaluation instruments and cognitive validity. Do items capture what they intend to? *School Effectiveness and School Improvement*, 28(4), 608–628. <https://doi.org/10.1080/09243453.2017.1360363>
- Farahian, M. & Parhamnia, F. (2021). Knowledge sharing through WhatsApp: Does it promote EFL teachers' reflective practice? *Journal of Applied Research in Higher Education*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/JARHE-12-2020-0456>
- Felten, R. von. (2005). *Lernen im reflexiven Praktikum: Eine vergleichende Untersuchung*. Zugl.: Bern, Univ., Diss., 2004. *Internationale Hochschulschriften: Bd. 441*. Waxmann.
- Fendler, L. (2003). Teacher reflection in a hall of mirrors: Historical influences and political reverberations. *Educational Researcher*, 32(3), 16–25. <https://doi.org/10.3102/0013189X032003016>

- Fenigstein, A., Scheier, M. F. & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 43(4), 522–527. <https://doi.org/10.1037/h0076760>
- Ferguson, H. (2018). How social workers reflect in action and when and why they don't: The possibilities and limits to reflective practice in social work. *Social Work Education*, 37(4), 415–427. <https://doi.org/10.1080/02615479.2017.1413083>
- Filipp, S.-H. & Freudenberg, E. (1989). *Der Fragebogen zur Erfassung dispositionaler Selbstaufmerksamkeit: (SAM-Fragebogen)*. Verlag für Psychologie Dr. C. J. Hogrefe.
- Fook, J., White, S. & Gardner, F. (2008). Critical reflection: A review of contemporary literature and understandings. In S. White (Hrsg.), *Critical reflection in health and social care* (S. 1–18). Open University Press.
- Forster, E. (2014). Reflexivität. In C. Wulf & J. Zirfas (Hrsg.), *Handbuch. Handbuch pädagogische Anthropologie* (S. 589–597). Springer VS. https://doi.org/10.1007/978-3-531-18970-3_54
- Foucault, M. (2016). *Überwachen und Strafen: Die Geburt des Gefängnisses* (16. Aufl.). *Suhrkamp-Taschenbuch: Bd. 2271*. Suhrkamp.
- Fox, R. K., Dodman, S. & Holincheck, N. (2019). Moving beyond reflection in a hall of mirrors: Developing critical reflective capacity in teachers and teacher educators. *Reflective Practice*, 20(3), 367–382. <https://doi.org/10.1080/14623943.2019.1617124>
- Fraij, A. (2018). *Skalendokumentation der Gießener Offensive Lehrerbildung zur Reflexionsbereitschaft*. Giessener Elektronische Bibliothek.
- Frankfurt, H. G. (2016). *Sich selbst ernst nehmen*. Suhrkamp Verlag.
- Freisler-Mühlemann, D., Winkler, A., Schafer, Y. & Böhlen, L. (2021). Die Bedeutung des Mentorats für die Reflexivität von Lehrpersonen am Anfang ihrer beruflichen Laufbahn. Vorab-Onlinepublikation. <https://doi.org/10.25656/01:22110>
- Fuertes-Camacho, M. T., Dulsat-Ortiz, C. & Álvarez-Cánovas, I. (2021). Reflective practice in times of covid-19: A tool to improve education for

- sustainable development in pre-service teacher training. *Sustainability*, 13(11), 6261. <https://doi.org/10.3390/su13116261>
- Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, 6(2), 207–226.
- Fund, Z., Court, D. & Kramarski, B. (2002). Construction and application of an evaluative tool to assess reflection in teacher-training courses. *Assessment & Evaluation in Higher Education*, 27(6), 485–499.
- Gagné, R. M. (1968). Learning hierarchies. In M. D. Merrill (Hrsg.), *Instructional design: Readings* (S. 118–131). Prentice-Hall.
- Garza, R. & Smith, S. F. (2015). Pre-service teachers' blog reflections: Illuminating their growth and development. *Cogent Education*, 2(1).
- Gehrmann, A. (2007). Kompetenzentwicklung im Lehramtsstudium. Eine Untersuchung an der Universität Rostock. In M. Lüders (Hrsg.), *Forschung zur Lehrerbildung* (S. 85–102). Waxmann Verlag.
- George, J. M. & Dane, E. (2016). Affect, emotion, and decision making. *Organizational Behavior and Human Decision Processes*, 136, 47–55. <https://doi.org/10.1016/j.obhdp.2016.06.004>
- Ghaye, T. (2007). Is reflective practice ethical? (The case of the reflective portfolio). *Reflective Practice*, 8(2), 151–162. <https://doi.org/10.1080/14623940701288859>
- Göbel, K. & Gösch, A. (2019). Die Nutzung kollegialer Reflexion von Unterrichtsvideos im Praxissemester. In M. Degeling, N. Franken, S. Freund, S. Greiten, D. Neuhaus & J. Schellenbach-Zel (Hrsg.), *Herausforderung Kohärenz: Praxisphasen in der universitären Lehrerbildung: Bildungswissenschaftliche und fachdidaktische Perspektiven* (S. 277–288). Klinkhardt, Julius.
- Göbel, K., Wyss, C., Neuber, K. & Raaflaub, M. (2021). Student feedback as a source for reflection in practical phases of teacher education. In W. Rollett, H. Bijlsma & S. Röhl (Hrsg.), *Student Feedback on Teaching in Schools* (S. 173–189). Springer International Publishing. https://doi.org/10.1007/978-3-030-75150-0_11

- Gödecke, G. (2020). *Gestaltung eines e-Portfolios in der Fremdsprachenlehrkräfteausbildung zur Förderung fachspezifischer Reflexionskompetenz: Eine empirische Studie*. Dissertation. *Studien zur Fremdsprachendidaktik und Spracherwerbsforschung: Bd. 18*. WVT Wissenschaftlicher Verlag Trier.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde (Hrsg.), *Personality psychology in Europe, volume 7: Selected papers from the Eighth European Conference on Personality held in Ghent, Belgium, July 1996* (S. 7–28). Tilburg Univ. Press.
- Gorski, P. C. & Dalton, K. (2020). Striving for critical reflection in multicultural and social justice teacher education: Introducing a typology of reflection approaches. *Journal of Teacher Education*, 71(3), 357–368.
<https://doi.org/10.1177/0022487119883545>
- Grant, A. M., Franklin, J. & Langford, P. (2002). The Self-Reflection and Insight Scale: A new measure of private self-consciousness. *Social Behavior and Personality*, 30(8), 821–836. <https://doi.org/10.2224/sbp.2002.30.8.821>
- Griffin, M. (2003). Using critical incidents to promote and assess reflective thinking in preservice teachers. *Reflective Practice*, 4(2), 207–220.
<https://doi.org/10.1080/14623940308274>
- Griffiths, M. & Tann, S. (1992). Using reflective practice to link personal and public theories. *Journal of Education for Teaching*, 18(1), 69–84.
<https://doi.org/10.1080/0260747920180107>
- Gröning, K. (2011). *Pädagogische Beratung: Konzepte und Positionen* (2., aktualisierte und überarbeitete Auflage). VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92892-0>
- Gruber, H. (2021). Reflexion. Der Königsweg zur Expertise-Entwicklung. *journal für lehrerInnenbildung jlb 01-2021 Mythos Reflexion*, 108–117.
<https://doi.org/10.35468/jlb-01-2021-10>
- Guardiera, P., Podlich, C. & Reimer, A. (2018). Zum Aufbau einer reflexiv-forschenden Grundhaltung mittels Fallarbeit - ausgewählt Ergebnisse eines

- Lehrkonzepts im Rahmen des Praxissemesters. *Zeitschrift für Studium und Lehre in der Sportwissenschaft*, 1(1), 21–26.
<https://doi.org/10.25847/ZSLS.2018.002>
- Guilford, J. P. (1967). *The nature of human intelligence. McGraw-Hill series in psychology*. McGraw-Hill.
- Gutzwiller-Helfenfinger, Aeppli, J. & Lötscher, H. (2017). Lehramtsstudierende reflektieren eine Praxiserfahrung im Bereich "Beurteilen & Fördern". Qualität der Erfahrungsdarstellung und Multiperspektivität. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 133-146). Julius Klinkhardt.
- Haan, G. de (2010). The development of ESD-related competencies in supportive institutional frameworks. *Internationale Zeitschrift für Erziehungswissenschaft*, 56(2/3), 315–328.
<https://doi.org/10.1007/s11159-010-9157-9>
- Häcker, T. (2017). Grundlagen und Implikationen der Forderung nach Förderung von Reflexivität in der Lehrerinnen- und Lehrerbildung. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 21–45). Julius Klinkhardt.
- Hager, A. (2020). *Potenziale ausgewählter Reflektiermethoden für die gezielte Förderung von Reflektierfähigkeiten im Rahmen eines Blogprojekts innerhalb einer Bildung für nachhaltige Entwicklung* [Dissertation]. Universität Bayreuth. https://doi.org/10.15495/EPUB_UBT_00004687
- Hahn, A. (1987). Identität und Selbstthematisierung. In V. Kapp (Hrsg.), *Selbstthematisierung und Selbstzeugnis: Bekenntnis und Geständnis* (S. 9–24). Suhrkamp.
- Hall, L. A. (2018). Using blogs to support reflection in teacher education. *Literacy Research and Instruction*, 57(1), 26–43.
<https://doi.org/10.1080/19388071.2017.1367055>

- Harrington, R. & Loffredo, D. A. (2010). Insight, rumination, and self-reflection as predictors of well-being. *The Journal of Psychology*, 145(1), 39–57. <https://doi.org/10.1080/00223980.2010.528072>
- Harry, A. G. & Salvador, K. (2021). Fostering critical reflection and productive dialogue in music teacher education. *Journal of Music Teacher Education*. Vorab-Onlinepublikation. <https://doi.org/10.1177/10570837211034136>
- Hartig, J., Jude, N. & Rauch, W. (2003). Entwicklung und Erprobung eines deutschen Big-Five-Fragebogens auf Basis des International Personality Item Pools (IPIP40). *Institut für Psychologie der Johann Wolfgang Goethe-Universität*(1).
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik: Mit 18 Tabellen* (S. 126–143). Springer.
- Hartung-Beck, V. & Schlag, S. (2020). Lerntagebücher als Reflexionsinstrument im Praxissemester. *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 3(2), 75–90. <https://doi.org/10.4119/HLZ-2492>
- Hatton, N. & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49. [https://doi.org/10.1016/0742-051X\(94\)00012-U](https://doi.org/10.1016/0742-051X(94)00012-U)
- Hauser, B. (2021). Können sie es nachher besser? Oder können sie nur besser darüber reden? *journal für lehrerInnenbildung jlb 01-2021 Mythos Reflexion*, 26–34. <https://doi.org/10.35468/jlb-01-2021-02>
- Hayden, H. E. & Chiu, M. M. (2015). Reflective teaching via a problem exploration--teaching adaptations--resolution cycle: A mixed methods study of preservice teachers' reflective notes. *Journal of Mixed Methods Research*, 9(2), 133–153. <https://doi.org/10.1177/1558689813509027>
- Hayes, A. F. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>

- Hays, R. & Gay, S. (2011). Reflection or 'pre-reflection': What are we actually measuring in reflective practice? *Medical education*, 45(2), 116–118.
<https://doi.org/10.1111/j.1365-2923.2010.03813.x>
- Helsper, W. (2002). Lehrerprofessionalität als antinomische Handlungsstruktur. In M. Kraul, W. Marotzki & C. Schweppe (Hrsg.), *Biographie und Profession* (S. 64–102). Klinkhardt.
- Herbart, J. F. (1997). *Grundlagen der Pädagogik: Bd. 1,1. Systematische Pädagogik* (D. Benner, Hg.). Dt. Studien-Verl.
- Hernandez, D. (2020). *Examining the effects of reflection on professional identity development in community college preservice teachers* [Dissertation]. City University of New York, CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/4013/
- Herzmann, P., Artmann, M. & Wichelmann, E. (2017). Theoriegeleitete Reflexionen videographierten Unterrichts. Eine Typologie studentischer Perspektiven auf universitäre Theorie-Praxis-Bezüge. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 176–189). Julius Klinkhardt.
- Herzmann, P. & Proske, M. (2014). Unterrichtsvideographien als Medium der Beobachtung und Reflexion von Unterricht im Lehramtsstudium. Ein Forschungsbericht. *Journal für LehrerInnenbildung*, 14(1), 33–38.
- Herzog, W. (1995). Reflexive Praktika in der Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 13(3), 235–273.
- Herzog, W. (2006). *Zeitgemäße Erziehung: Die Konstruktion pädagogischer Wirklichkeit* (Studienausg.). Velbrück Wiss.
- Hilzensauer, W. (2017). *Wie kommt die Reflexion in den Lehrberuf? Ein Lernangebot zur Förderung der Reflexionskompetenz bei Lehramtsstudierenden*. Waxmann.
- Hinzke, J.-H. (2020). Praktikumsberichte als Reflexionsanlässe? Reflexion zwischen Professionalisierungshoffnung und Praxisbewältigung in der ersten Phase der Lehrer_innenbildung. *Herausforderung*

- Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 3(2), 91–107. <https://doi.org/10.4119/HLZ-2505>
- Hosenfeld, A. (2010). *Führt Unterrichtsrückmeldung zu Unterrichtsentwicklung? Die Wirkung von videographischer und schriftlicher Rückmeldung bei Lehrkräften der vierten Jahrgangsstufe. Empirische Erziehungswissenschaft: Bd. 22*. Waxmann.
- Husu, J., Toom, A. & Patrikainen, S. (2008). Guided reflection as a means to demonstrate and develop student teachers' reflective competencies. *Reflective Practice*, 9(1), 37–51. <https://doi.org/10.1080/14623940701816642>
- Idel, T.-S. & Schütz, A. (2017). Praktiken der Reflexion in der Lehrerinnen- und Lehrerbildung. Praxistheoretische Überlegungen zur Routinisierung professioneller Reflexivität am Beispiel studentischer Portfolioarbeit. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 201–213). Julius Klinkhardt.
- Jain, A. K. (2013). Die Dialektik des Automatismus - Deflexion oder das Andere der Reflexion. In H. Bublitz, I. Kaldrack, T. Röhle & M. Zeman (Hrsg.), *Schriftenreihe des Graduiertenkollegs "Automatismen". Automatismen - Selbst-Technologien* (S. 181–192). Wilhelm Fink.
- Jay, J. K. & Johnson, K. L. (2002). Capturing complexity: A typology of reflective practice for teacher education. *Teaching and Teacher Education*(18).
- Jenert, T. (2008). Ganzheitliche Reflexion auf dem Weg zu Selbstorganisiertem Lernen. *Bildungsforschung*, 5(2). <https://doi.org/10.25656/01:4596>
- Jenson, J. D. (2011). Promoting self-regulation and critical reflection through writing students' use of electronic portfolio. *International Journal of ePortfolio*, 1(1), 49–60. <https://files.eric.ed.gov/fulltext/EJ1107586.pdf>

- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E. & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain*, *125*(8), 1808–1814. <https://doi.org/10.1093/brain/awf181>
- Junkins, M. B. & Haeffel, G. J. (2017). Rumination: Reflection can amplify the depressogenic effects of brooding. *International Journal of Cognitive Therapy*, *10*(1), 34–46. https://doi.org/10.1521/ijct_2016_09_19
- Kaasila, R. & Lauriala, A. (2012). How do pre-service teachers' reflective processes differ in relation to different contexts? *European Journal of Teacher Education*, *35*(1), 77–89. <https://doi.org/10.1080/02619768.2011.633992>
- Kalk, K., Luik, P., Taimalu, M. & Täht, K. (2014). Validity and reliability of two instruments to measure reflection: A confirmatory study. *Trames. Journal of the Humanities and Social Sciences*, *18*(2), 121. <https://doi.org/10.3176/tr.2014.2.02>
- Kasztner, A. (2009). *Fördert Reflexion die Anwendbarkeit des Gelernten? Eine empirische Untersuchung eines Ausbildungsganges*. Books on Demand.
- Kauffman, D. F., Ge, X., Xie, K. & Chen, C.-H. (2008). Prompting in web-based environments: Supporting self-monitoring and problem solving skills in College students. *Journal of Educational Computing Research*, *38*(2), 115–137. <https://doi.org/10.2190/EC.38.2.a>
- Kayapinar, U. & Erkus, A. (2009). Measuring teacher reflection: Development of TRS. *Egitim Arastirmalari-Eurasian Journal of Educational Research*, *37*, 144–158.
- Keller-Schneider, M. (2021). Multiperspektivische Reflexion als dynamisierender Prozess zur Förderung von Professionalität. Vorab-Onlinepublikation. <https://doi.org/10.25656/01:22107>
- Kember, D., Leung, D. Y. P., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F. K., Wong, M. & Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education*, *25*(4), 381–395. <https://doi.org/10.1080/713611442>

- Kember, D., McKay, J., Sinclair, K. & Wong, F. K. Y. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & Evaluation in Higher Education*, 33(4), 369–379. <https://doi.org/10.1080/02602930701293355>
- Kempin, M., Kulgemeyer, C. & Schecker, H. (2017). *Reflexionsperformanz: Lehramtsstudierende reflektieren Physikunterricht*.
- Kempin, M., Kulgemeyer, C. & Schecker, H. (2019). Wirkung von Professionswissen und Praxisphasen auf die Reflexionsfähigkeit von Physiklehramtsstudierenden. In S. Habig (Vorsitz), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen*. Symposium im Rahmen der Tagung von Gesellschaft für Didaktik der Chemie und Physik, Wien.
- Killeavy, M. & Moloney, A. (2010). Reflection in a social space: Can blogging support reflective practice for beginning teachers? *Teaching and Teacher Education*, 26(4), 1070–1076. <https://doi.org/10.1016/j.tate.2009.11.002>
- Killion, J. P. & Todnem, G. R. (1991). A process for personal theory building. *Educational Leadership*, 48(6), 14–16.
- King, S. (2008). Inspiring critical reflection in preservice teachers. *Physical Educator*, 65(1), 21–29.
- Knipfer, S., Scholl, A. & Kump, B. (2014). Lernen durch Reflexion: Lösungsansätze für die Zukunft. *Wirtschaftspsychologie aktuell. Zeitschrift für Personal und Management*(3), 19–23.
- Knowles, J. G. (1993). Life-history accounts as mirrors: A practical avenue for the conceptualization of reflection in teacher education. In J. Calderhead & P. Gates (Hrsg.), *Conceptualizing reflection in teacher development* (S. 70–92). Falmer Press.
- Kohlberg, L. (1981). *The meaning and measurement of moral development. Heinz Werner Lectures: Bd. 13*.
- Kolb, D. A. (1984). *Experience as the source of learning and development*. Prentice Hall.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A. & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and

- interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, 38, 76–88.
<https://doi.org/10.1016/j.tate.2013.11.004>
- Körkkö, M., Kyrö-Ämmälä, O. & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education*, 55, 198–206. <https://doi.org/10.1016/j.tate.2016.01.014>
- Korthagen, F. (1985). Reflective teaching and preservice teacher education in the Netherlands. *Journal of Teacher Education*, 9(3), 317–326.
- Korthagen, F. (1993a). Measuring the reflective attitude of prospective mathematics teachers in the Netherlands. *European Journal of Teacher Education*, 16(3), 225–236. <https://doi.org/10.1080/0261976930160305>
- Korthagen, F. (1993b). Two modes of reflection. *Teaching and Teacher Education*, 9(3), 317–326. [https://doi.org/10.1016/0742-051X\(93\)90046-J](https://doi.org/10.1016/0742-051X(93)90046-J)
- Korthagen, F. (1999). Linking reflection and technical competence: The logbook as an instrument in teacher education. *European Journal of Teacher Education*, 22(2-3), 191–207. <https://doi.org/10.1080/0261976899020191>
- Korthagen, F. (2001). A reflection on reflection. In F. Korthagen (Hrsg.), *Linking practice and theory: The pedagogy of realistic teacher education*. Routledge.
- Korthagen, F., Kessels, J., Koster, B., Lagerwerf, B. & Wubbels, T. (2001). *Linking practice and theory: The pedagogy of realistic teacher education*. Routledge. <https://doi.org/10.4324/9781410600523>
- Korthagen, F. & Wubbels, T. (2000). Are reflective teachers better teachers? In Willems, Gerard M., J. H. Stakenborg & W. Veugelers (Hrsg.), *Trends in Dutch Teacher Education* (S. 131–142). Garant.
- Korthagen, F. & Wubbels, T. (2001). Evaluative research on the realistic approach and on the promotion of reflection. In F. Korthagen (Hrsg.), *Linking practice and theory: The pedagogy of realistic teacher education* (S. 88–107). Routledge.
- Kosinár, J. (2017). Reflexion eigener Handlungspraxis im Prozess des Erfahrungslernens. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.),

- Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 147–158). Julius Klinkhardt.
- Kowalczyk-Walędziak, M., Korzeniecka-Bondar, A., Danilewicz, W. & Lauwers, G. (2019). *Rethinking teacher education for the 21st century*. Verlag Barbara Budrich. <https://doi.org/10.3224/84742241>
- Krippendorff, K. (1980). *Content Analysis. An introduction to its methodology*. Sage.
- Kuckartz, U. (2014). *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Lehrbuch*. Springer VS.
- Kufner, S. (2012). *Diagnose und Prognose von Handlungskompetenz im Bereich adaptiven Lehrens bei Studierenden-eine Videostudie*. Dissertation. Universität Passau.
- Kultusministerkonferenz. (2004, i.d.F. von 2019). *Standards für die Lehrerbildung: Bildungswissenschaften*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf
- LaBoskey, V. K. (1993). A conceptual framework for reflection in preservice teacher education. In J. Calderhead & P. Gates (Hrsg.), *Conceptualizing reflection in teacher development* (S. 23–38). Falmer Press.
- Lambert, M. D., Sorensen, T. J. & Elliott, K. M. (2014). A comparison and analysis of preservice teachers' oral and written reflections. *Journal of Agricultural Education*, 55(4), 85–99. <https://doi.org/10.5032/jae.2014.04085>
- Lambrecht, J. & Bosse, S. (2020). Lässt sich die Reflexionsfähigkeit von angehenden Lehrkräften verändern? *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 3(2), 137–150. <https://doi.org/10.4119/HLZ-2497>
- Landes, M. & Steiner, E. (2014). *Psychologische Auswirkungen von Change Prozessen: Widerstände, Emotionen, Veränderungsbereitschaft und Implikationen für Führungskräfte. Essentials*. Springer VS.

- Lang-von Wins, T. (2007). Kompetenzhaltigkeit von Methoden moderner psychologischer Diagnostik-, Personalauswahl- und Arbeitsanalyseverfahren sowie aktueller Management-Diagnostik-Ansätze. In J. Erpenbeck & L. von Rosenstiel (Hrsg.), *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2. Aufl., S. 758–789). Schäffer-Poeschel Verlag.
- Larrivee, B. (2000). Transforming teaching practice: Becoming the critically reflective teacher. *Reflective Practice*, 1(3), 293–307.
<https://doi.org/10.1080/713693162>
- Larrivee, B. (2008). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice*, 9(3), 341–360.
<https://doi.org/10.1080/14623940802207451>
- Lee, H.-J. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching and Teacher Education*, 21(6), 699–715.
<https://doi.org/10.1016/j.tate.2005.05.007>
- Leijen, Ä., Valtna, K., Leijen, D. A. & Pedaste, M. (2012). How to determine the quality of students' reflections? *Studies in Higher Education*, 37(2), 203–217. <https://doi.org/10.1080/03075079.2010.504814>
- Leng, E. Y., Ali, Wan Zah bte Wan, Baki, R. & Mahmud, R. (2010). Stability of the Intrinsic Motivation Inventory (IMI) for the use of Malaysian form one students in ICT literacy class. *Eurasia Journal of Mathematics, Science & Technology Education*, 6(3), 215–226.
- Leonhard, T. (2013). Portfolioarbeit zwischen Reflexion und Leistungsbewertung: Empirische Befunde zur Messbarkeit von Reflexionskompetenz. In B. Koch-Priewe, T. Leonhard, A. Pineker & C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung: Konzepte und empirische Befunde* (S. 180–192). Klinkhardt.
- Leonhard, T. (2020). Reflexion in zwei Praxen. *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 3(2), 14–28. <https://doi.org/10.4119/HLZ-2482>

- Leonhard, T. & Abels, S. (2017). Der "reflective practitioner". Leitfigur oder Kategorienfehler einer reflexiven Lehrerinnen- und Lehrerbildung? In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 46–55). Julius Klinkhardt.
- Leonhard, T. & Bolle, R. (2015). Wenn Studierende reflektieren sollen. Portfolioarbeit und der Anspruch von Reflexion. In R. Bolle (Hrsg.), *Schulpraktische Studien: Bd. 10. Schulpraktische Studien 2015 zwischen Standards, Alltag und Zukunftsvisionen* (S. 31–46). Leipziger Univ.-Verl.
- Leonhard, T., Nagel, N., Rihm, Thomas, Strittmatter-Haubold, Veronika & Wengert-Richter, P. (2010). Zur Entwicklung von Reflexionskompetenz bei Lehramtsstudierenden. In A. Gehrman, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 111–124). Verlag Julius Klinkhardt.
- Leonhard, T. & Rihm, T. (2011). Erhöhung der Reflexionskompetenz durch Begleitveranstaltungen zum Schulpraktikum? Konzeption und Ergebnisse eines Pilotprojekts mit Lehramtsstudierenden. *Lehrerbildung auf dem Prüfstand*, 4(2), 240–270.
- Lethbridge, K., Andrusyszyn, M.-A., Iwasiw, C., Laschinger, H. K. & Fernando, R. (2013). Assessing the psychometric properties of Kember and Leung's reflection questionnaire. *Assessment & Evaluation in Higher Education*, 38(3), 303–325.
<https://doi.org/10.1080/02602938.2011.630977>
- Lew, M. D. N. & Schmidt, H. G. (2011). Self-reflection and academic performance: Is there a relationship? *Advances in health sciences education: theory and practice*, 16(4), 529–545.
<https://doi.org/10.1007/s10459-011-9298-z>
- Li, A. & Peng, T. (2018). Observing "myself" in the video: Fostering reflective practice in oral presentation training. *Advances in Language and Literary Studies*, 9(3), 138–144. <https://doi.org/10.7575/aiac.all.v.9n.3p.138>

- Linninger, C. (2016). *Reflexion bei angehenden Lehrkräften: Bedeutung und Förderung im Professionalisierungsprozess: Dissertation.*
- Liston, D. P. & Zeichner, K. M. (1987). Reflective teacher education and moral deliberation. *Journal of Teacher Education*, 38(6), 2–8.
<https://doi.org/10.1177/002248718703800601>
- Liston, D. P. & Zeichner, K. M. (1990). Reflective teaching and action research in preservice teacher education. *Journal of Education for Teaching*, 16(3), 235–254. <https://doi.org/10.1080/0260747900160304>
- Liu, K. & Ball, A. F. (2019). Critical reflection and generativity: Toward a framework of transformative teacher education for diverse learners. *Review of Research in Education*, 43(1), 68–105.
<https://doi.org/10.3102/0091732X18822806>
- Liu, P. (2022). Reflection in alignment to professional standards: What did the student teachers highlight? *Journal of Global Education and Research*, 6(1), 67–81. <https://doi.org/10.5038/2577-509X.6.1.1111>
- Lohse-Bossenz, H., Schönknecht, L. & Brandter, M. (2019). Entwicklung und Validierung eines Fragebogens zur Erfassung Reflexionsbezogener Selbstwirksamkeit von Lehrkräften im Vorbereitungsdienst. *Empirische Pädagogik*, 33(2), 164–179.
- Lüsebrink, I. & Grimminger, E. (2014). Fallorientierte Lehrer/innenausbildung evaluieren - Überlegungen zur Modellierung von unterrichtsbezogener Reflexionskompetenz. In I. Pieper, P. Frei, K. Hauenschild & B. Schmidt-Thieme (Hrsg.), *Was der Fall ist* (S. 201–211). Springer Fachmedien Wiesbaden.
- Luttenberg, J. & Bergen, T. (2008). Teacher reflection: The development of a typology. *Teachers and Teaching*, 14(5-6), 543–566.
<https://doi.org/10.1080/13540600802583713>
- Mann, K., Gordon, J. & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in health sciences education: theory and practice*, 14(4), 595–621.
<https://doi.org/10.1007/s10459-007-9090-2>

- Martin, M. (2005). Reflection in teacher education: How can it be supported? *Educational Action Research*, 13(4), 525–542.
<https://doi.org/10.1080/09650790500200303>
- Matsuo, M. (2020). Reflection on success and failures: Scale development and validation. *Journal of Workplace Learning*, 32(8), 615–626.
<https://doi.org/10.1108/JWL-07-2020-0117>
- Mauri, T., Clarà, M., Colomina, R. & Onrubia, J. (2017). Patterns of interaction in the processes of joint reflection by student teachers. *Journal of Education for Teaching: International Research and Pedagogy*, 43(4), 427–443.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Beltz Pädagogik. Beltz.
- McAuley, E., Duncan, T. & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1), 48–58. <https://doi.org/10.1080/02701367.1989.10607413>
- McKay, L. & Barton, G. (2018). Exploring how arts-based reflection can support teachers' resilience and well-being. *Teaching and Teacher Education*, 75, 356–365. <https://doi.org/10.1016/j.tate.2018.07.012>
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232. <https://doi.org/10.1037/h0048850>
- Meyer-Siever, K. & Levin, A. (2018). Entwicklung der Reflexionskompetenz im Rahmen eines fächerübergreifenden e-Portfolios. *Resonanz. Magazin für Lehre und Studium an der Universität Bremen. Sonderausgabe*, 24–31.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.
- Michalsky, T. & Kramarski, B. (2015). Prompting reflections for integrating self-regulation into teacher technology education. *Teachers College Record*(117).
- Moghaddam, R. G., Davoudi, M., Adel, S. M. R. & Amirian, S. M. R. (2020). Reflective teaching through journal writing: A study on EFL teachers' reflection-for-action, reflection-in-action, and reflection-on-action. *English*

- Teaching & Learning*, 44(3), 277–296. <https://doi.org/10.1007/s42321-019-00041-2>
- Mönig, S. (2013). *Wissenschaftliche Reflexionsfähigkeit von Lehrenden und Lehramtsstudierenden - theoretische Kompetenzmodellierung und empirische Überprüfung*. Dissertation. Technische Universität Dortmund.
- Monteiro, V., Mata, L. & Peixoto, F. (2015). Intrinsic Motivation Inventory: Psychometric properties in the context of first language and mathematics learning. *Psicologia: Reflexão e Crítica*, 28(3), 434–443. <https://doi.org/10.1590/1678-7153.201528302>
- Moore-Russo, D. A. & Wilsey, J. N. (2014). Delving into the meaning of productive reflection: A study of future teachers' reflections on representations of teaching. *Teaching and Teacher Education*, 37, 76–90. <https://doi.org/10.1016/j.tate.2013.10.002>
- Moran, J. M. (2016). Cognitive neuroscience of self-reflection. In J. R. Absher & J. Cloutier (Hrsg.), *Neuroimaging Personality, Social Cognition, and Character* (S. 205–219). Elsevier Science. <https://doi.org/10.1016/B978-0-12-800935-2.00010-5>
- Mühlhausen, U. (Hrsg.). (2006). *Unterrichten lernen mit Gespür: Szenarien für eine multimedial gestützte Analyse und Reflexion von Unterricht; mit DVD-Beilage mit multimedialen Unterrichtsdokumenten* (2., überarb. Aufl.). Schneider Verl. Hohengehren.
- Mulryan-Kyne, C. (2020). Supporting reflection and reflective practice in an initial teacher education programme: An exploratory study. *European Journal of Teacher Education*, 1–18. <https://doi.org/10.1080/02619768.2020.1793946>
- Münste-Goussar, S. (2017). Künste der Existenz oder Reflexion. Zweierlei Modi der Beziehung zu sich. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 74–88). Julius Klinkhardt.

- Navin, L. (2018). Does seeing matter? Exploring pre-service teachers' use of self-video as a tool for self-reflection in the study of their own practice. *ProQuest LLC* (Dissertation).
- Neuber, K. & Göbel, K. (2016). *Schülerrückmeldungen zum Unterricht und Unterrichtsreflexion: Dokumentation der entwickelten Erhebungsinstrumente im Projekt "Schülerrückmeldungen zum Unterricht und ihre Beitrag zur Unterrichtsreflexion im Praxissemester (ScRiPS)". Erste Skalenanalysen*. DuEPublico. Universität Duisburg-Essen.
- Neuweg, G. H. (2017). Herrlich unreflektiert. Warum Könner weniger denken, als man denkt. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 89–101). Julius Klinkhardt.
- Niggli, A. (2004). Welche Komponenten reflexiver beruflicher Entwicklung interessieren angehende Lehrerinnen und Lehrer? - Faktorenstruktur eines Fragebogens und erste empirische Ergebnisse. *Schweizerische Zeitschrift für Bildungswissenschaften*, 26(2), 343–364.
- Noffke, S. E. & Brennan, M. (1988). *The dimensions of reflection: A conceptual and contextual analysis*. New Orleans. Paper presented at the Annual Meeting of the American Educational Research Association.
- Paradis, E. & Sutkin, G. (2017). Beyond a good story: From Hawthorne Effect to reactivity in health professions education research. *Medical education*, 51(1), 31–39. <https://doi.org/10.1111/medu.13122>
- Paseka, A., Hinzke, J.-H., Feld, I. & Kuckuck, K. (2021). Zur Entwicklung studentischer Reflexionskompetenz im Rahmen Forschenden Lernens: Viel Ungewissheit und wenig Gewissheit. In D. Kemethofer, J. Reitinger & K. Soukup-Altrichter (Hrsg.), *Beiträge zur Bildungsforschung: Band 7. Vermessen? Zum Verhältnis von Bildungsforschung, Bildungspolitik und Bildungspraxis* (S. 143–160). Waxmann.
- Patry, J.-L. (2014). Theoretische Grundlagen des Theorie-Praxis-Problems in der Lehrerbildung. In K.-H. Arnold (Hrsg.), *Schulpraktika in der*

- Lehrerbildung: Theoretische Grundlagen, Konzeptionen, Prozesse und Effekte* (S. 29–42). Waxmann.
- Penso, S., Shoham, E. & Shiloah, N. (2001). First steps in novice teachers' reflective activity. *Teacher Development*, 5(3), 323–338.
<https://doi.org/10.1080/13664530100200147>
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Klett.
- Pineker, A. & Störtländer, C. (2013). Gestaltung von praktikumsbezogenen Reflexionsanlässen im Rahmen des "Bielefelder Portfolio Praxisstudien" - zwei hochschuldidaktische Varianten und ihre Evaluation. In B. Koch-Priewe, T. Leonhard, A. Pineker & C. Störtländer (Hrsg.), *Portfolio in der LehrerInnenbildung: Konzepte und empirische Befunde* (S. 193–211). Klinkhardt.
- Pongratz, L. A., Wimmer, M., Nieke, W. & Masschelein, J. (2004). *Nach Foucault: Diskurs- und machtanalytische Perspektiven der Pädagogik. Schriftenreihe der Kommission Bildungs- und Erziehungsphilosophie der DGfE*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-322-80625-3>
- Poom-Valickis, K. & Mathews, S. (2013). Reflecting others and own practice: An analysis of novice teachers' reflection skills. *Reflective Practice*, 14(3), 420–434. <https://doi.org/10.1080/14623943.2013.767237>
- Pörksen, U. (2011). *Plastikwörter: Die Sprache einer internationalen Diktatur* (7. Aufl.). Klett-Cotta.
- Priddis, L. & Rogers, S. L. (2018). Development of the reflective practice questionnaire: Preliminary findings. *Reflective Practice*, 19(1), 89–104.
- Prieto, L. P., Magnuson, P., Dillenbourg, P. & Saar, M. (2020). Reflection for action: Designing tools to support teacher reflection on everyday evidence. *Technology, Pedagogy and Education*, 29(3), 279–295.
<https://doi.org/10.1080/1475939X.2020.1762721>
- Procee, H. (2006). Reflection in education: A Kantian epistemology. *Educational Theory*, 56(3), 237–253. <https://doi.org/10.1111/j.1741-5446.2006.00225.x>

- Quinton, S. & Smallbone, T. (2010). Feeding forward: Using feedback to promote student reflection and learning – a teaching model. *Innovations in Education and Teaching International*, 47(1), 125–135.
<https://doi.org/10.1080/14703290903525911>
- Raaflaub, M., Wyss, C. & Hüsler, N. (2019). Kollegiale Unterrichtsreflexion im Lehramtsstudium. *journal für lehrerInnenbildung jlb 3-2019 Peers*. Vorab-Onlinepublikation. https://doi.org/10.35468/jlb-03-2019_04
- Rauin, U. & Meier, U. (2007). Subjektive Einschätzung des Kompetenzerwerbs in der Lehramtsausbildung. In M. Lüders (Hrsg.), *Forschung zur Lehrerbildung* (S. 103–132). Waxmann Verlag.
- Richieri, C. (2017). The role of reflection in L2 teacher education. Student teacher's lesson plans as tools for promoting professional development. In C. Battisti, P. Vettorel & C. Richieri (Hrsg.), *Teacher education - English Language, Literatur and Culture - Good Practices* (S. 319–349). Pensa MultiMedia.
- Rihm, T. (2017). Verständigungsarbeit in der Erwachsenenbildung. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 228–239). Julius Klinkhardt.
- Risko, V. (2009). Detailing reflection instruction: The efficacy of a guided instructional procedure on prospective teachers' pedagogical reasoning. *Action in Teacher Education*, 31(2), 47–60.
- Roberts, C. & Stark, P. (2008). Readiness for self-directed change in professional behaviours: Factorial validation of the Self-Reflection and Insight Scale. *Medical education*, 42(11), 1054–1063. <https://doi.org/10.1111/j.1365-2923.2008.03156.x>
- Roberts, P. (2016). Reflection: A renewed and practical focus for an existing problem in teacher education. *Australian Journal of Teacher Education*, 41(7), 19–35.
- Roberts, P. (2018). Developing reflection through an ePortfolio-based learning environment: Design principles for further implementation. *Technology*,

- Pedagogy and Education*, 27(3), 313–326.
<https://doi.org/10.1080/1475939X.2018.1447989>
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842–866.
<https://doi.org/10.1111/1467-9620.00181>
- Roessger, K. M. (2014). The effect of reflective activities on instrumental learning in adult work-related education: A critical review of the empirical research. *Educational Research Review*, 13, 17–34.
<https://doi.org/10.1016/j.edurev.2014.06.002>
- Rogers, S. L., Priddis, L. E., Michels, N., Tieman, M. & van Winkle, L. J. (2019). Applications of the reflective practice questionnaire in medical education. *BMC Medical Education*, 19(1), 47. <https://doi.org/10.1186/s12909-019-1481-6>
- Rominger, C., Reitinger, J., Seyfried, C., Schneckenleitner, E. & Fink, A. (2017). The reflecting brain: Reflection-competence in an educational setting is associated with increased EEG activity in the alpha band: THE REFLECTING BRAIN. *Mind, Brain, and Education*, 11(2), 54–63.
- Rosenberger, K. (2017). Schreibend reflektieren. Zur Praxis institutionalisierter Reflexionstexte in der Lehramtsausbildung. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen - Zugänge - Perspektiven* (S. 190–200). Julius Klinkhardt.
- Rosseel, J. & Anseel, F. (2021). When reflection hinders creative problem-solving: A test of alternative reflection strategies. *Journal of Business and Psychology*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s10869-021-09741-8>
- Roters, B. (2011). *Professionalisierung durch Reflexion in der Lehrerbildung: Eine empirische Studie an einer deutschen und einer US-amerikanischen Universität*. Dissertation. *Studien zur international und interkulturell vergleichenden Erziehungswissenschaft: Bd. 12*. Waxmann.

- Runhaar, P., Sanders, K. & Yang, H. (2010). Stimulating teachers' reflection and feedback asking: An interplay of self-efficacy, learning goal orientation, and transformational leadership. *Teaching and Teacher Education*, 26(5), 1154–1161. <https://doi.org/10.1016/j.tate.2010.02.011>
- Russell, T. (2013). Has reflective practice done more harm than good in teacher education? *Phronesis*, 2(1), 80–88. <https://doi.org/10.7202/1015641ar>
- Russell, T. (2018). Paradigmatic changes in teacher education: The perils, pitfalls, and unrealized promise of the reflective practitioner. In J. S. Johnston & R. Bruno-Jofre (Hrsg.), *Teacher Education in a Transnational World* (S. 158–176). University of Toronto Press. <https://doi.org/10.3138/9781442619999-011>
- Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education*, 16(1), 99–111. <https://doi.org/10.1080/13562517.2010.507311>
- Santagata, R. & Angelici, G. (2010). Studying the impact of the lesson analysis framework on preservice teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education*, 61(4), 339–349. <https://doi.org/10.1177/0022487110369555>
- Schaal, S. (2006). *Fachintegratives Lernen mit digitalen Medien: Die theoriegeleitete Entwicklung und Evaluation einer hypermedialen Lernumgebung für den naturwissenschaftlichen Unterricht an der Realschule*. Kovac.
- Schaarschmidt, U. & Kieschke, U. (2013). Beanspruchungsmuster im Lehrerberuf. Ergebnisse und Schlussfolgerungen der Potsdamer Lehrerstudie. In M. Rothland (Hrsg.), *Belastung und Beanspruchung im Lehrerberuf: Modelle, Befunde, Interventionen* (2. Aufl., S. 81–98). Springer VS.
- Schäfer, S., Blomberg, G. A. & Seidel, T. (2012). Der Einsatz von Lerntagebüchern in der universitären Lehrerbildung - Welchen Effekt haben strukturierende Leitfragen auf die Reflexionsfähigkeit von Lehramtsstudierenden? *Empirische Pädagogik*, 26(2), 271–291.

- Schaper, N. (2014). Validitätsaspekte von Kompetenzmodellen und -tests für hochschulische Kompetenzdomänen. In G. Spöttl & F. Musekamp (Hrsg.), *Berufliche Bildung in Forschung, Schule und Arbeitswelt / Vocational Education and Training: Bd. 12. Kompetenz im Studium und in der Arbeitswelt: Nationale und internationale Ansätze zur Erfassung von Ingenieurkompetenzen* (1. Aufl., S. 21–48). Peter Lang GmbH Internationaler Verlag der Wissenschaften.
<https://doi.org/10.25656/01:12883>
- Schaper, N. & Hochholdinger, S. (2006). (Arbeits-)Psychologische Konzepte zur Modellierung und Messung von Kompetenzen in der Lehrerbildung. In A. H. Hilligus & H.-D. Rinkens (Hrsg.), *Standards und Kompetenzen - neue Qualität in der Lehrerbildung? Neue Ansätze und Erfahrungen in nationaler und internationaler Perspektive* (S. 215–222). LIT Verlag Münster.
- Schmelzing, S., Wüsten, S., Sandmann, A. & Neuhaus, B. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions* (1. Aufl.). *The Jossey-Bass higher education series*. Jossey-Bass.
- Schön, D. A. (Hrsg.). (1991). *The Reflective turn: Case studies in and on educational practice*. Teachers College Press.
- Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M. & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, 27(2), 259–267.
<https://doi.org/10.1016/j.tate.2010.08.009>

- Seyfried, C., Weinberger, A. & Reitinger, J. (2009). DINE (Dispositions Inventory for Education): Entwicklung eines Inventars zur Entscheidungsfindung bei Aufnahmeverfahren von Studienbewerber/-innen an pädagogischen Hochschulen. In A. Weinberger (Hrsg.), *Reflexion im pädagogischen Kontext* (S. 113–134). LIT Verlag.
- Seyfried, C., Weinberger, A. & Reitinger, J. (2013). "DINE (Dispositions Inventory for Education): Entwicklung eines Inventars zur Entscheidungsfindung bei Aufnahmeverfahren von Studienbewerber/-innen an pädagogischen Hochschulen. In A. Weinberger (Hrsg.), *Austria Erziehungswissenschaft: Band 19. Reflexion im pädagogischen Kontext* (S. 113–134). LIT.
- Shoffner, M. (2011). Considering the first year: Reflection as a means to address beginning teachers' concerns. *Teachers and Teaching*, 17(4), 417–433. <https://doi.org/10.1080/13540602.2011.580518>
- Silvia, P. J. & Phillips, A. G. (2011). Evaluating self-reflection and insight as self-conscious traits. *Personality and Individual Differences*, 50(2), 234–237.
- Slepcevic-Zach, P. & Stock, M. (2018). ePortfolio as a tool for reflection and self-reflection. *Reflective Practice*, 19(3), 291–307. <https://doi.org/10.1080/14623943.2018.1437399>
- Snyder, D. W. (2011). Preparing for teaching through reflection. *Music Educators Journal*, 97(3), 56–60.
- Sobral, D. T. (2000). An appraisal of medical students' reflection-in-learning. *Medical education*, 34, 182–187.
- Sobral, D. T. (2005). Medical students' mindset for reflective learning: A revalidation study of the reflection-in-learning scale. *Advances in health sciences education: theory and practice*, 10(4), 303–314. <https://doi.org/10.1007/s10459-005-8239-0>
- Spalding, E. & Wilson, A. (2002). Demystifying reflection: A study of pedagogical strategies that encourage reflective journal writing. *Teachers College Record*, 104(7), 1393–1421. <https://doi.org/10.1111/1467-9620.00208>

- Sparks-Langer, G. M. & Colton, A. (1991). Synthesis of research on teacher's reflective thinking. *Educational Leadership*, 48(6), 37–44.
- Sparks-Langer, G. M., Simmons, J. M., Pasch, M., Colton, A. & Starko, A. (1990). Reflective pedagogical thinking: How can we promote it and measure it? *Journal of Teacher Education*, 41(5), 23–32.
- Stenberg, K., Rajala, A. & Hilppo, J. (2016). Fostering theory–practice reflection in teaching practicums. *Asia-Pacific Journal of Teacher Education*, 44(5), 470–485. <https://doi.org/10.1080/1359866X.2015.1136406>
- Stender, J., Watson, C., Vogelsang, C. & Schaper, N. (2021). Wie hängen bildungswissenschaftliches Professionswissen, Einstellungen zu Reflexion und die Reflexionsperformanz angehender Lehrpersonen zusammen? *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 4(1), 229–248. <https://doi.org/10.11576/HLZ-4057>
- Subramaniam, K. (2013). Examining the content of preservice teachers' reflections of early field experiences. *Research in Science Education*, 43(5), 1851–1872. <https://doi.org/10.1007/s11165-012-9337-7>
- Sumsion, J. & Fleet, A. (1996). Reflection: can we assess it? Should we assess it? *Assessment & Evaluation in Higher Education*, 21(2), 121–130. <https://doi.org/10.1080/0260293960210202>
- Sung, Y.-T., Chang, K.-E., Yu, W.-C. & Chang, T.-H. (2009). Supporting teachers' reflection and learning through structured digital teaching portfolios. *Journal of Computer Assisted Learning*, 25(4), 375–385. <https://doi.org/10.1111/j.1365-2729.2009.00312.x>
- Sydnor, J., Daley, S. & Davis, T. (2020). Video reflection cycles: Providing the tools needed to support teacher candidates toward understanding, appreciating, and enacting critical reflection. *Contemporary Issues in Technology and Teacher Education*, 20(2), 369–391.
- Taylor, I. & Getzels, J. W. (1975). *Perspectives in creativity*. Taylor and Francis.

- Teasdale, J. D. & Green, H. A. (2004). Ruminative self-focus and autobiographical memory. *Personality and Individual Differences*, 36(8), 1933–1943. <https://doi.org/10.1016/j.paid.2003.08.022>
- Teddlie, C. & Tashakkori, A. (2010). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences* [Nachdr.]. SAGE Publ.
- Thompson, N. & Pascal, J. (2012). Developing critically reflective practice. *Reflective Practice*, 13(2), 311–325. <https://doi.org/10.1080/14623943.2012.657795>
- Tiainen, O., Korkeamäki, R.-L. & Dreher, M. J. (2018). Becoming reflective practitioners: A case study of three beginning pre-service teachers. *Scandinavian Journal of Educational Research*, 62(4), 586–600. <https://doi.org/10.1080/00313831.2016.1258673>
- Torrance, E. P. (1966). *Torrance tests of creative thinking: Norms-technical manual*. Personnel press.
- Totter, A., Egger, N., Schmitz, F., Hermann, T. & Petko, D. (2016). Online-Reflexion mittels Weblogs: Ein Mittel zur Auseinandersetzung mit belastenden Situationen in der berufspraktischen Ausbildung von Lehrpersonen? *Beiträge zur Lehrerinnen- und Lehrerbildung*, 34(2), 203–211. <https://doi.org/10.25656/01:13943>
- Trapnell, P. D. & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, 76(2), 284–304.
- Tremmel, R. (1993). Zen and the art of reflective practice in teacher education. *Harvard Educational Review*, 63(4), 434–459. <https://doi.org/10.17763/haer.63.4.m42704n778561176>
- Tripp, T. & Rich, P. (2012). Using video to analyze one's own teaching. *British Journal of Educational Technology*, 43(4), 678–704. <https://doi.org/10.1111/j.1467-8535.2011.01234.x>

- Turhan, B. & Kirkgoz, Y. (2018). Towards becoming critical reflection writers: A case of English language teacher candidates. *Reflective Practice, 19*(6), 749–762. <https://doi.org/10.1080/14623943.2018.1539651>
- Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education, 29*(2), 217–257. <https://doi.org/10.1007/s40593-019-00174-2>
- Valli, L. (Hrsg.). (1992). *SUNY series, Teacher preparation and development. Reflective teacher education: Cases and critiques*. State University of New York Press.
- van Beveren, L., Roets, G., Buysse, A. & Rutten, K. (2018). We all reflect, but why? A systematic review of the purposes of reflection in higher education in social and behavioral sciences. *Educational Research Review, 24*, 1–9. <https://doi.org/10.1016/j.edurev.2018.01.002>
- van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry, 6*(3), 205–228.
- Volmer, M., Pawelzik, J., Todorova, M. & Windt, A. (2021). Reflexionsfähigkeit im Praxissemester fördern. Ein Konzept universitärer Begleitung. Vorab-Onlinepublikation. <https://doi.org/10.25656/01:22104>
- Volpert, W. (2003). *Wie wir handeln - was wir können: Ein Disput als Einführung in die Handlungspsychologie* (3. Aufl.). *Buchreihe Positionen: Bd. 1*. Artefact-Verl. Weber.
- Walz, M. & Roth, J. (2017). Professionelle Kompetenzen angehender Lehrkräfte erfassen - Zusammenhänge zwischen Diagnose-, Handlungs- und Reflexionskompetenz. In U. Kortenkamp & Kuzle A. (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 1367–1370). WTM-Verlag.
- Ward, J. R. & McCotter, S. S. (2004). Reflection as a visible outcome for preservice teachers. *Teaching and Teacher Education, 20*(3), 243–257. <https://doi.org/10.1016/j.tate.2004.02.004>

- Wegner, C., Remmert, K. & Ohlberger, S. (2017). Evaluation of a wiki for lesson reflection in the project "Kolumbus-Kids" at Bielefeld University. *Electronic Journal of Science Education*, 21(4), 1–40.
- Weichold, M. (2014). *Zwischen Reflex und Reflexion*. Dissertation. *Ideen & Argumente* [XII, 347 Seiten].
- Weinberger, A. & Seyfried, C. (2009). RIFE: Reflection Instrument for Education. *Salzburger Beiträge zur Erziehungswissenschaft*, 1(2), 83–94.
- Weinert, F. E. (2014). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3. Aufl., S. 17–32). Beltz.
- Wigfield & Eccles (2000). Expectancy-value theory of achievement motivation. *Contemporary educational psychology*, 25(1), 68–81.
<https://doi.org/10.1006/ceps.1999.1015>
- Williams, R. & Grudnoff, L. (2011). Making sense of reflection: A comparison of beginning and experienced teachers' perceptions of reflection for practice. *Reflective Practice*, 12(3), 281–291.
<https://doi.org/10.1080/14623943.2011.571861>
- Wyss, C. (2011). *Unterricht und Reflexion* [Dissertation]. GBV Gemeinsamer Bibliotheksverbund.
- Wyss, C. & Mahler, S. (2021). Mythos Reflexion. Theoretische und praxisbezogene Erkenntnisse in der Lehrer*innenbildung. *journal für lehrerInnenbildung jlb 01-2021 Mythos Reflexion*, 16–25.
<https://doi.org/10.35468/jlb-01-2021-01>
- Zahn, L. (2007). Reflexion. In J. Ritter, K. Gründer & R. Eisler (Hrsg.), *Historisches Wörterbuch der Philosophie, Band 8* (S. 396–405). Schwabe.
- Zeichner, K. M. & Liston, D. P. (2014). *Reflective teaching: An introduction. Reflective teaching and the social conditions of schooling*. Routledge.

Anhang

- 1 Online-Fragebogen (durchgeführt auf SoSci-Survey)
- 2 Codebuch – Teil 1
- 3 Codebuch – Teil 2
- 4 Beispiel zur Berechnung der Reflexionsmaße

A-1 Online-Fragebogen (durchgeführt auf SoSci-Survey)

Seite 01

Soziodemographische Fragen

1. Geben Sie zuerst bitte noch einmal Ihren individuellen Code an, damit der Fragebogen Ihrem Reflexionstext zugeordnet werden kann.

Die ersten drei Buchstaben des Vornamens der Mutter (z.B. Helga: HEL),
die ersten drei Buchstaben des Vornamens des Vaters (z.B. Thomas: THO) und
der Tag Ihres Geburtsdatums (z.B. 26.03.1998: 26)

(Beispielcode würde lauten: HELTHO26)

2. Wir bitten Sie zunächst um einige Angaben zu Ihrer Person.

Geschlecht?

- m
 w
 d

3. Welches Praktikum absolvieren Sie?

- Pädagogisch-didaktisches Praktikum
 Exerctium Paedagogicum

4. In welchem Fachsemester studieren Sie?

Seite 02
MK

5. Für wie wichtig halten Sie die folgenden Aspekte bei der Bearbeitung einer Reflexionsaufgabe?

Verschieben Sie den Regler jeweils an die für Sie passende Stelle.

ganz unwichtig

sehr wichtig

Sich auf eigene Erlebnisse und Beobachtungen zu beziehen.

Lösungen für Probleme zu finden.

Die eigene subjektive Sicht auf einen Sachverhalt darzustellen.

Fachwissenschaftliche Theorien in die Überlegungen miteinzubeziehen.

Sich sachlich und objektiv mit einer Sache auseinanderzusetzen.

Über ethische und moralische Folgen und Fragen nachzudenken.

Eine Situation aus verschiedenen Perspektiven zu beleuchten.

Sich mit eigenen Gefühlen und Werten auseinanderzusetzen.

6. Wie schwerwiegend erschien Ihnen das Problem, zu dem Sie reflektiert haben?

- überhaupt nicht problematisch
- ein wenig problematisch
- teilweise problematisch
- ziemlich problematisch
- sehr problematisch

Seite 03
IMI

7. Die folgenden Aussagen beziehen sich auf die Reflexionsaufgabe im Portfolio.
Bitte entscheiden Sie, in welchem Maß die folgenden Aussagen jeweils auf Sie zutreffen!

	trifft überhaupt nicht zu	trifft nicht zu	teils, teils	trifft zu	trifft völlig zu
Mir gefiel das Bearbeiten der Reflexionsaufgabe sehr gut.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich glaube, auf diese Weise über Probleme nachzudenken, könnte von einigem Wert für mich sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Das Reflektieren hat Spaß gemacht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, dass solche Reflexionen nützlich sind für meine eigene Weiterentwicklung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfand das Nachdenken über die Situation als langweilig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, es ist wichtig zu reflektieren, denn es kann einem helfen, Lösungen für Probleme zu finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Diese Reflexionsaufgabe konnte meine Aufmerksamkeit überhaupt nicht binden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wäre bereit, noch einmal so eine Reflexionsaufgabe zu bearbeiten, weil das einigen Nutzen für mich hat.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich würde solche Reflexionen als sehr interessant bezeichnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, das Reflektieren kann mir dabei helfen, eine professionelle Lehrkraft zu werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfand das Nachdenken über das Problem als recht angenehm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich glaube, mir Zeit für Reflexion zu nehmen, könnte nützlich für mich sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Seite 05

9. Bitte entscheiden Sie, in welchem Maß die folgenden Aussagen jeweils auf Sie zutreffen.

	sehr unzutreffend	eher unzutreffend	weder noch	eher zutreffend	sehr zutreffend
Ich mag Kunst nicht.	<input type="radio"/>				
Ich vermeide philosophische Diskussionen.	<input type="radio"/>				
Ich denke, dass Kunst wichtig ist.	<input type="radio"/>				
Ich bringe Gespräche auf ein höheres Niveau.	<input type="radio"/>				
Ich gehe nicht gerne zu Kunstausstellungen.	<input type="radio"/>				
Ich tendiere zu einer konservativen politischen Einstellung.	<input type="radio"/>				
Abstrakte Ideen interessieren mich nicht.	<input type="radio"/>				
Ich tendiere zu einer alternativen politischen Einstellung.	<input type="radio"/>				

Seite 06

AlternativeUses

10. Letzte Aufgabe zum Abschluss des Fragebogens:

Wozu könnte man einen Regenschirm (außer zum Schutz vor Regen) noch benutzen?

Zählen Sie so viele alternative Nutzungsmöglichkeiten auf, wie Sie können.

Nutzen Sie bitte für jede Idee ein eigenes Eingabefeld!

Letzte Seite

Vielen Dank für Ihre Teilnahme! Falls Sie Fragen zur Studie, Ihren Daten oder der Auswertung haben, melden Sie sich gerne:
martina.riel@uni-passau.de

A-2 Codebuch – Teil 1

Code	Name der Kategorie	Beschreibung	Anwendung	Beispiel
	1. Descriptive writing	Keine Reflexion. Bloß beschreibende Textanteile.	Wird nicht vergeben, sondern nur die Unterkategorien.	
0	Persönliche Meinung/ fiktive Beschreibung	- Meinung zu einer Situation ohne Begründung oder Rechtfertigung. - Wiederholung der Aufgabenstellung - Beschreibung einer fiktiven Situation: Was könnte die Lehrkraft/der Praktikant/die Praktikantin tun?		„Unterschiedliche Aufgaben sind gut geeignet, Störungen zu vermeiden.“ „Ich finde, man kann viel mit einem Gespräch erreichen.“ → <i>persönliche Meinung</i> Ich überlege also, wie ich mit Paul umgehen würde. → <i>Wiederholung der Aufgabe</i> „Ich würde unterschiedliche Aufgaben mit verschiedenen Schwierigkeitsstufen einsetzen.“ → <i>fiktive Beschreibung</i>
1	Situationsbeschreibung (erlebt)	Bloße Schilderung einer erlebten Situation ohne Nennung von Gründen, Interpretationen oder Folgerungen: Situationen aus dem Praktikum oder Beobachtetes aus dem Praktikum oder Erinnerungen an eigene Schulzeit.		"Im Praktikum ging es mir so, dass ein Schüler massiv gestört hat."

	2. Descriptive reflection	Nicht nur Beschreibung, sondern auch Begründung oder Rechtfertigung für Ereignisse oder Handlungen.	Wird nicht vergeben, sondern nur die Unterkategorien.	
2	One perspective/ factor	Nennung von nur einer Begründung/einer Rechtfertigung.	Typische Formulierungen: denn, weil, da, somit Wird aber auch kodiert, wenn <u>kein</u> sprachlicher Begründungszusammenhang (denn/weil) erkennbar ist, sondern nur ein inhaltlicher!	"Ich habe mich für dieses Vorgehen entschieden, weil ich wollte, dass die Kinder aktiv werden." "Es sollten Verhaltensregeln aufgestellt und unbedingt auch eingefordert werden. (Code: Persönliche Meinung) Nur so nehmen die Lernenden die Lehrkraft auch ernst. (Code: Descriptive reflection)"
3	Multiple perspectives/ factors	Nennung weiterer Begründungen/Rechtfertigungen.	Mehrere Begründungen können in einem Satz stecken. Oder in getrennten Sätzen. Die erste genannte Begründung wird mit "one perspective" codiert. Ab der zweiten Begründung mit "multiple perspectives".	„Ich habe es noch einmal erklärt, denn ich hatte die Sorge, dass es noch nicht alle verstanden haben und außerdem hatte ich ohnehin noch zehn Minuten in der Stunde übrig.“ “Ich habe mich für dieses Vorgehen entschieden, weil ich wollte, dass die Kinder aktiv werden. (Code: One perspective) Zudem dachte ich, dass es damit gelingen könnte, auch die Außenseiter in die Gruppe zu integrieren. (Code: Multiple perspectives).“

	<p>3. Dialogic reflection</p>	<p>Nicht nur Aneinanderreihung von Ideen, sondern eine intensivere Auseinandersetzung mit einem Gedanken, eine mehrperspektivische Betrachtung. Aus einer gewissen Distanz wird eine Art Diskussion mit einem selbst geführt, ein Abwägen verschiedener Optionen findet statt. Es werden Inkonsistenzen und Schwierigkeiten, ein Für und Wider, einerseits und andererseits erkannt und thematisiert.</p>	<p>Wird nicht vergeben, sondern nur die Unterkategorien.</p>	
4	<p>One perspective/factor</p>	<p>Nennung von nur einem Für und Wider, einem Gegenargument, einer Schwierigkeit.</p>	<p>Typische Formulierungen: Einerseits – andererseits, eigentlich – aber, dafür spricht – dagegen spricht, obwohl, allerdings, jedoch, dennoch, dagegen</p>	<p>„Nach den Klassenregeln hätte der Schüler zwar bestraft werden müssen, aber andererseits wollte ich kein Drama aus der Situation machen und habe ihn zunächst ignoriert.“</p> <p>„Es braucht eine klare Linie und Regeln, um von den Schülern akzeptiert zu werden. (Code: Descriptive reflection) Jedoch muss man in Einzelfällen diese auch anpassen und darf nicht nur stur auf ihnen beharren. (Code: Dialogic reflection)</p>

5	Multiple perspectives/factors	Nennung weiterer Für-und-Wider-Aspekte, Gegenargumente, Schwierigkeiten.	<p>Mehrere Aspekte können in einem Satz stecken.</p> <p>Oder weitere Aspekte werden erst in folgenden Sätzen genannt. Die erste genannte Begründung wird dann mit "one perspective" codiert. Ab der zweiten Begründung mit "multiple perspectives".</p>	<p>„Man kann Schüler auch ohne Meldung aufrufen, dagegen spricht jedoch, dass sie das blamieren könnte, obwohl man es auch so sehen kann, dass manche die unangenehme Situation auch auf ihr Fehlverhalten hinweisen könnte.“</p> <p>„Ich hatte eigentlich geplant, verschiedene Texte lesen zu lassen, um zu differenzieren, aber habe mich dann doch aus Zeitgründen dagegen entschieden. (Code: One perspective). Dadurch habe ich es zwar geschafft, pünktlich fertig zu werden, allerdings waren einige Schüler dadurch überfordert. (Code: Multiple perspectives)“</p>
6	4. Critical reflection	Ausführungen zeigen, dass ein Bewusstsein darüber besteht, dass Handlungen und Ereignisse Teil eines größeren Ganzen sind: historische, politische, systemische Zusammenhänge, Fragestellungen und Überlegungen. Auch moralische/ethische Erwägungen , z.B. zu Gerechtigkeit.	Da es hier keine Untergliederung in a) und b) gibt, wird die Oberkategorie vergeben.	„Es ist die Frage, ob Bestrafungen tatsächlich einen Platz in Klassenführungsstrategien haben sollten, denn diese Form der Machtausübung hat zwar eine lange historische Tradition, aber ist nicht mehr zeitgemäß.“

A-3 Codebuch – Teil 2

Name der Kategorie	Beschreibung	Anwendung	Beispiel
Weitere Aspekte	Sowohl in der Theorie als auch in den Texten selbst finden sich weitere Aspekte, die für eine Reflexion relevant erscheinen, aber bei Hatton & Smith keine explizite Berücksichtigung finden.	Es wird nicht jeder einzelne Satz mit einer der beiden Unterkategorien kodiert, sondern nur, wenn sie zutrifft.	
1. Aussagen mit Theoriebezug	Herstellen eines Bezugs zur Fachliteratur oder theoretischen Hintergründen für die Entscheidungen oder Einschätzungen.	<ul style="list-style-type: none"> - Wird verwendet, wenn der Name der Theorie oder des/der Wissenschaftler/in genannt wird. - Wird verwendet, wenn Fachbegriffe aus einschlägigen Theorien verwendet werden (z.B. Strukturierung, Gruppenfokus, Reibungslosigkeit). 	<p>"Störungen können durch flüssige Übergänge und eine gute Strukturierung behoben werden."</p> <p>"Die Lehrkraft kann Störungen vorbeugen, indem sie eine emotionale Objektivität behält."</p> <p>"Nach Canter könnte man hier auch versuchen, ihn für angemessenes Verhalten zu loben."</p> <p>"Um Paul zu aktivieren, könnte man auch nach dem Rechenschaftsprinzip vorgehen (Kounin)."</p>

2. Situationsbezo- gene Gefühle	Eigene Gefühle und Empfindungen bezüglich des Falls, der Erlebnisse oder Gedanken werden geäußert.	Sowohl Gefühle, die während der Reflexion auftreten, als auch Gefühle, die in vergangenen Situationen aufgetreten sind und geschildert werden.	„Heute ärgert es mich, dass ich die Situation damals nicht besser gelöst habe.“ „Ich war schockiert, wie laut die Klasse wurde, obwohl die Lehrerin anwesend war.“ „Im empfand das als sehr unangemessen.“
--	--	--	--

A-4 Beispiel zur Berechnung der Reflexionsmaße

Kodier- einheiten	Studentischer Reflexionstext	Code bzw. Bepunktung
1	Ich würde vermutlich zunächst ein Gespräch mit Paul suchen.	0
2	Unter vier Augen könnte man die Ursache klären, weil er sich vielleicht vor der Gruppe keine Blöße geben will.	1
3	Einerseits glaube ich also, dass man das Problem nicht vor der Gruppe klären sollte, andererseits kann man problematisches Verhalten auch nicht einfach stehen lassen.	3*
4	Man sollte ihn auch im Klassenzimmer irgendwie zurechtweisen.	0
5**	Schreien oder Schimpfen sind aber nicht das beste Mittel, denn das kann die ganze Klasse negativ beeinflussen und vielleicht noch weitere Störungen erzeugen, die den Unterricht verzögern.	2
	Summe	6**

*Reflexionsmaß „**Höchste im Text erreichte Stufe**“: 3

Reflexionsmaß „Durchschnittliche Reflexionsleistung**“:

$$\frac{\text{Summe Bepunktung}}{\text{Summe der Kodiereinheiten}} = \frac{6}{5} = 1.2$$