

**Erstellung von problemhaltigen Prüfungsaufgaben –
eine Herausforderung für angehende Lehrkräfte in der kaufmännisch-
beruflichen Bildung?**

*Inaugural-Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Wirtschaftswissenschaften
der Johann Wolfgang Goethe-Universität
Frankfurt am Main*

vorgelegt von Lütfiye Turhan
aus Erlenbach am Main

Frankfurt am Main, 2023



Publiziert unter der Creative Commons-Lizenz Namensnennung (CC BY) 4.0 International. Published under a Creative Commons Attribution (CC BY) 4.0 International License.
<https://creativecommons.org/licenses/by/4.0/>

Danksagung

Von einer bayerischen Hauptschule startend war es mein Ziel, einen akademischen Grad zu erlangen. Dieses Ziel hinderte mich daran, vorherige Abschlüsse als „Abschluss“ zu betrachten, da sie für mich Meilensteine waren. **Elhamdulillah – Gott sei Dank**, fühle ich mich heute am Ziel angekommen, dem das Lebenslange Lernen folgen soll.

Der Weg zum Ziel wurde begleitend von wertvollen Personen. Beginnend von Lehrkräften an der Hauptschule, die mich motivierten, bis hin zu Universitätsprofessorinnen und -professoren, die zur Freude an der Wissenschaft beitrugen. Zu diesen gehört Herr Prof. Dr. Karl Wilbers (FAU Erlangen-Nürnberg), der geduldig meine Fragen beantwortete und hilfreiche Tipps zum weiteren Verlauf nach Ende des Studiums gab. Besonders danken möchte ich meiner Doktormutter Frau Prof. Dr. Eveline Wuttke (Goethe-Universität Frankfurt) und meinem Zweitbetreuer Herrn Prof. Dr. Helmut Niegemann (Goethe-Universität Frankfurt). Auch danke ich Frau Prof. Dr. Susan Seeber (Georg-August-Universität Göttingen) und meiner ehemaligen Kollegin Frau Dr. Carolin Geiser (ehem. Georg-August-Universität Göttingen) für die wertvolle Unterstützung bei der Datenerhebung. Vielen Dank an die Masterstudierenden der Wirtschaftspädagogik der Universitäten Frankfurt und Göttingen, den Lehrkräften im Vorbereitungsdienst, (vertretenden) Studiengangsleitungen sowie Fach- und Seminarleitungen, die an meinen Studien teilnahmen und/oder unterstützten. Danke auch an die Freunde und Förderer der Goethe-Universität, durch deren finanzielle Förderung es mir möglich war, einigen Probandinnen und Probanden für ihre Teilnahme an Studien ein kleines Dankeschön zu geben. Mein Dank geht auch an Leonie Kapitza, Philip Padberg und Ilias Hauas, die mich beim Transkribieren der Materialien und/oder durch ihre Zweitkodierung unterstützen und zur Qualität der Ergebnisse beitrugen. Danken möchte ich auch meinen ehemaligen Frankfurter Kolleginnen und Kollegen Katharina Bushyla, Jule Hangen, Lukas Meier und Dr. Christin Siegfried, die während der Promotionszeit hilfreiche Tipps, motivierende Worte und einen kritischen Blick auf die Arbeit schenkten! Unter meinen ehemaligen Frankfurter Kolleginnen und Kollegen möchte ich Kremena Eneva (ehem. Goethe-Universität Frankfurt) einen ganz besonderen Dank für ihre Unterstützung, den tollen Austausch und für die Konstruktivität aussprechen. Ein besonderer Dank gilt auch meiner ehemaligen Kollegin Lisa Gregorius (ehem. Universität Duisburg-Essen) für den fachlichen Austausch und die freundschaftlichen, aber dennoch kritischen Rückmeldungen.

Mein großer Dank gilt meiner Familie, allen voran meinen Eltern Elmas und Ahmet. Sie hörten mir meinen Plänen, Ideen, Sorgen und Herausforderungen geduldig zu und halfen mir, wo immer es ihnen möglich war. Ohne ihre Hilfe, Geduld und Unterstützung wäre es mir kaum möglich, mein hoch gesetztes Ziel zu erreichen. Ebenso danke ich mich meinen Geschwistern Ali Riza, Sümeyye und Muhammed Enes – danke für euren geduldigen Umgang auch dann, wenn stressbedingt schwer mit mir wurde. Es macht mich besonders glücklich, den Erfolg mit ihnen zusammen feiern zu dürfen.

Die im Text häufig genutzten Wörter *Geduld* und *Dank* kann ich durch kaum andere ersetzen und würde diese noch häufiger nutzen, wenn ich die Liste der *geduldigen* Menschen, die einen *Dank* verdienen, fortsetzen würde. Ich bitte alle Personen, die in diesem Text ihren Namen suchen, sich angesprochen zu fühlen: Vielen herzlichen Dank!

Frankfurt am Main, November 2023

Lütfiye Turhan

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
Abkürzungsverzeichnis.....	VIII
1 Problemstellung	1
2 Kompetenzanforderung an (angehende) Lehrkräfte zur Erstellung problemhaltiger Prüfungsaufgaben.....	7
2.1 Modellierung der Kompetenzen (angehender) Lehrkräfte	7
2.2 Curricular begründete Anforderung an (angehende) Lehrkräfte.....	13
2.3 Praktisch begründete Anforderung an (angehende) Lehrkräfte	19
2.4 Pädagogische Diagnostik als Teil allgemein-pädagogischer und fachdidaktischer Kompetenz.....	20
2.4.1 Begriffsverständnis und Funktion	20
2.4.2 Bewerten und Beurteilen	22
2.4.3 Formen des Assessments.....	24
2.4.4 Bezeichnungen und Einsatz der Instrumente und Verfahren	33
2.4.5 Gütekriterien im summativen Assessment.....	35
2.5 Problemorientierung als didaktischer Zugang und Problemlösekompetenz	37
2.5.1 Kompetenz und Kompetenzorientierung	37
2.5.2 Problemdefinition und Problemlösen.....	38
2.5.3 Merkmale problemhaltiger Aufgaben	41
2.5.4 Einflussfaktoren auf das Problemlösen	43
2.5.5 Problem- vs. Handlungsorientierung.....	46
2.6 Erstellung problemhaltiger Prüfungsaufgaben	49
3 Stand der Forschung zu betrachteten Kompetenzen (angehender) Lehrkräfte	54
3.1 Diagnostische und fachdidaktische Kompetenz (angehender) Lehrkräfte	54
3.2 Hinweise aus der schriftlich-kaufmännischen Prüfungspraxis	56

4	Studien zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte.....	61
4.1	Überblick über die Fragestellungen und das Design der Studien	61
4.2	Studien zur Angebotsperspektive.....	63
4.2.1	Methodisches Vorgehen in den Dokumentenanalysen zur ersten und zweiten Phase der Lehrer*innenbildung	63
4.2.2	Auswertungsverfahren in den Dokumentenanalysen.....	76
4.2.3	Methodisches Vorgehen in der Fragebogenstudie zur ersten Phase der Lehrer*innenbildung und Auswertung.....	91
4.2.4	Beschreibung der Stichprobe der Teilnehmenden an der Fragebogenstudie und deren Studienangebot	92
4.2.5	Methodisches Vorgehen bei der Interviewstudie zur zweiten Phase der Lehrer*innenbildung und Auswertung.....	94
4.2.6	Beschreibung der Stichprobe der Teilnehmenden der Interviewstudie zur zweiten Phase der Lehrer*innenbildung.....	96
4.3	Studie zur Nachfrageperspektive.....	97
4.3.1	Methodisches Vorgehen in der Interviewstudie mit angehenden Lehrkräften	97
4.3.2	Auswertung der Interviewstudie mit angehenden Lehrkräften	99
4.3.3	Beschreibung der Stichprobe	102
5	Ergebnisse, Limitationen und Implikationen der Studien zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte.....	103
5.1	Ergebnisse der Studien zur Angebotsperspektive	103
5.1.1	Vorwort zur Ergebnisdarstellung der Studien zur Angebotsperspektive	103
5.1.2	Ergebnisse zum wirtschaftspädagogischen Studium.....	103
5.1.3	Ergebnisse zum Referendariat für das Lehramt an kaufmännisch-beruflichen Schulen.....	114
5.1.4	Limitationen der Studien zur Angebotsperspektive	126
5.2	Ergebnisse der Studie zur Nachfrageperspektive	131
5.2.1	Ergebnisse der Interviewstudie	131
5.2.2	Limitationen der Interviewstudie mit angehenden Lehrkräften.....	138

5.3	Zusammenfassende Implikationen aus den Studien zu den Ausbildungsinhalten	142
6	Stand der Trainingsforschung.....	145
6.1	Hinführung zu den Einflussfaktoren	145
6.2	Teilnehmendenbedingte Faktoren	145
6.3	Trainingsbedingte Faktoren.....	147
6.4	Kontextbezogene Faktoren.....	150
6.5	Trainingsevaluation	152
7	Trainingsstudie mit Studierenden der Wirtschaftspädagogik und LiV an kaufmännisch-berufsbildenden Schulen	155
7.1	Relevanz des Trainingsinhalts und Vorstellung des Trainingskonzepts	155
7.2	Ziel und Beschreibung des methodischen Vorgehens der Trainingsevaluation..	160
7.3	Methodisches Vorgehen in der quasi-experimentellen Studie	163
7.3.1	Eingesetzte Instrumente	163
7.3.2	Durchführung und Sampling	165
7.3.3	Beschreibung der Stichprobe der Interventions- und Kontrollgruppe	170
7.3.4	Auswertung der quantitativen Daten zur Trainingsevaluation.....	171
7.4	Methodisches Vorgehen in der Interviewstudie zur Trainingsevaluation.....	178
7.4.1	Durchführung und Sampling	178
7.4.2	Auswertung der qualitativen Daten zur Trainingsevaluation.....	179
8	Ergebnisse, Limitationen und Implikationen der Trainingsstudie mit Studierenden der Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen	183
8.1	Ergebnisse der quasi-experimentellen Studie und der Interviewstudie.....	183
8.1.1	Vorwort zur Ergebnisdarstellung	183
8.1.2	Ergebnisse der Trainingsevaluation auf Ebene 1	183
8.1.3	Ergebnisse der Trainingsevaluation auf Ebene 2	191
8.1.4	Limitationen der Trainingsevaluation	204
8.2	Implikationen aus der Trainingsstudie, insb. -evaluation.....	210

9 Diskussion und Forschungsdesiderat	214
Literaturverzeichnis.....	219
Rechtsquellenverzeichnis.....	266
Anhang A: Kodierleitfaden zu den Ausbildungsinhalten aus der Angebotsperspektive	268
Anhang B: Leitfaden für das Interview mit Fachleiter*innen an Studienseminaren	281
Anhang C: Zuordnung der Fragen des teilstandardisierten Interviewleitfadens mit angehenden Lehrkräften (Ausbildungsinhalte aus der Nachfrageperspektive) zu den Forschungsfragen und Kategorien des Kodierleitfadens	284
Anhang D: Kodierleitfaden zur Interviewstudie mit angehenden Lehrkräften zur Ausbildung aus der Nachfrageperspektive.....	286
Anhang E: Raster zu den Merkmalen problemhaltiger Aufgaben	294
Anhang F: Zuordnung der Forschungsfragen zu den Fragen des teilstandardisierten Interviewleitfadens zur Trainingsevaluation und den Kategorien des Kodierleitfadens...	295
Anhang G: Kodierleitfaden zur Interviewstudie mit angehenden Lehrkräften zur Trainingsevaluation	297

Abbildungsverzeichnis

Abbildung 1:	<i>Das Kompetenzmodell von COACTIV mit Spezifikationen für das Professionswissen</i>	9
Abbildung 2:	<i>Facetten generischen pädagogischen Wissens und Könnens</i>	12
Abbildung 3:	<i>Qualitätskontrolle und -sicherung in Schule und Unterricht.....</i>	27
Abbildung 4:	<i>Gegenüberstellung der formativen und summativen Assessments</i>	29
Abbildung 5:	<i>Der IDEAL-Ansatz</i>	39
Abbildung 6:	<i>Kompetenzmodell</i>	45
Abbildung 7:	<i>Schritte der vollständigen Handlung.....</i>	47
Abbildung 8:	<i>Logic-Assessment-Modell.....</i>	51
Abbildung 9:	<i>Durchgeführte Studien zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte</i>	62
Abbildung 10:	<i>Ablauf der Dokumentenanalyse</i>	64
Abbildung 11:	<i>Grafische Darstellung der Kategorien zu den Ausbildungsinhalten</i>	90
Abbildung 12:	<i>Hauptkategorien nach Forschungsfragen</i>	100
Abbildung 13:	<i>Code-Relations-Modell (MAXQDA) zum bisherigen Vorgehen beim Klausurerstellen.....</i>	133
Abbildung 14:	<i>Trainingsstudie des Forschungsprojektes „TeKoP“</i>	157
Abbildung 15:	<i>Hauptkategorien zur Trainingsevaluation nach Forschungsfragen.....</i>	179
Abbildung 16:	<i>Code-Relations-Modell (MAXQDA) zum eingeschätzten Lernzuwachs (L) und zur Anwendung (A)</i>	198

Tabellenverzeichnis

Tabelle 1:	<i>Standards für die theoretischen und praktischen Ausbildungsabschnitte für die Ausbildung von Lehrer*innen</i>	17
Tabelle 2:	<i>Diagnostikarten nach Gegenstand, Methoden, Ziel(e) und Assessment</i>	31
Tabelle 3:	<i>Universitäten in der BRD, an denen Wirtschaftspädagog*innen ausgebildet werden, inkl. Markierung der Standorte</i>	65
Tabelle 4:	<i>Anzahl der Module zum wirtschaftspädagogischen Studium</i>	68
Tabelle 5:	<i>Anzahl der analysierten Module zum wirtschaftspädagogischen Studium</i>	70
Tabelle 6:	<i>Beschreibung des analysierten Materials zum Bachelor- und Masterstudium Wirtschaftspädagogik</i>	72
Tabelle 7:	<i>Beschreibung des analysierten Materials zum Referendariat</i>	76
Tabelle 8:	<i>Reliabilitäten der Kategorien in der Dokumentenanalyse</i>	88
Tabelle 9:	<i>Beschreibung der Stichprobe und allgemeine Angaben zu den Standorten</i>	93
Tabelle 10:	<i>Verteilung der Module auf den Pflicht- und Wahlbereich des Bachelor- und Masterstudiums</i>	94
Tabelle 11:	<i>Kappa-Werte (κ) zu den Kategorien des Interviews zu den Ausbildungsinhalten aus Nachfrageperspektive</i>	101
Tabelle 12:	<i>Übersicht der deskriptiven Werte der Kodierungen in der Dokumentenanalyse zum Studium</i>	105
Tabelle 13:	<i>Anzahl der in der Dokumentenanalyse kodierten Module im Studium zu den systematisch angelegten Kategorien der Eingangs-, Prozess- und Outputdiagnostik</i>	107
Tabelle 14:	<i>Inhalte zur Prüfungsaufgabenerstellung und -bewertung in den Bachelor- und Mastermodulen</i>	109
Tabelle 15:	<i>Übersicht der deskriptiven Werte der Kodierungen in der Dokumentenanalyse zum Referendariat</i>	115
Tabelle 16:	<i>Anzahl der kodierten Module in der Dokumentenanalyse zum Referendariat zu den systematisch angelegten Kategorien der Eingangs-, Prozess- und Outputdiagnostik</i>	118
Tabelle 17:	<i>Skalen und Beispielitems zur Evaluation auf der ersten Ebene</i>	164
Tabelle 18:	<i>Skala und Beispielitems zur Evaluation auf der zweiten Ebene</i>	165
Tabelle 19:	<i>Durchführungs- und Erhebungszeiträume der Trainings, inkl. Pre- und Posttestbefragung (Interventionsgruppe)</i>	168
Tabelle 20:	<i>Zeiträume der Pre- und Posttestbefragung (Kontrollgruppe)</i>	168

Tabelle 21:	<i>Beschreibung der Stichprobe der Masterstudierenden</i>	170
Tabelle 22:	<i>Beschreibung der Stichprobe der LiV</i>	171
Tabelle 23:	<i>Reliabilitätswerte der eingesetzten Skalen</i>	172
Tabelle 24:	<i>Voraussetzungsprüfung für ANCOVA</i>	173
Tabelle 25:	<i>Ergebnisse zur Multikollinearitätsprüfung</i>	176
Tabelle 26:	<i>Kappa-Werte (κ) zu den Kategorien des Interviews zur Trainingsevaluation</i>	182
Tabelle 27:	<i>Deskriptive Ergebnisdarstellung zu Ebene 1</i>	184
Tabelle 28:	<i>Punkte im Pre- und Posttest zum Wissen der Studierenden</i>	192
Tabelle 29:	<i>Ergebnis der Regressionsanalyse zur Erklärung der erreichten Punkte im Wissenstest zu den Merkmalen problemhaltiger Aufgaben</i>	194
Tabelle 30:	<i>Ergebnis der Regressionsanalyse zur Erklärung der erreichten Punkte im Wissenstest zu den diagnostischen Gütekriterien</i>	195
Tabelle 31:	<i>Punkte im Pre- und Posttest zum Wissen der LiV</i>	196
Tabelle 32:	<i>Code-Relation-Browser (MAXQDA) zum Lernzuwachs (L) durch und zur Reaktion (R) auf einzelne Bausteine</i>	204
Tabelle 33:	<i>Angenommene und abgelehnte Hypothesen</i>	210

Abkürzungsverzeichnis

α	Cronbachs Alpha
ANCOVA	Analysis of Covariance
AkA	Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen
B. A.	Bachelor of Arts
B. Sc.	Bachelor of Science,
BWL	Betriebswirtschaftslehre
BBiG	Berufsbildungsgesetz
B. Sc.	Bachelor of Science
COACTIV	Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die Entwicklung mathematischer Kompetenz
DBWA	Didaktik der beruflichen Aus- und Weiterbildung
F _[Nummer]	Forschungsfrage [Nummer]
Herv. i. Original	Hervorhebung im Original
IG	Interventionsgruppe
Herv. d. Verf.	Hervorhebung durch Verfasserin
Verf.	Verfasserin
IHK	Industrie- und Handelskammer
DQR	Deutscher Qualifikationsrahmen
EQR	Europäischer Qualifikationsrahmen
KG	Kontrollgruppe
k. K.	keine Kodierung
KMK	Kultusministerkonferenz
KW	Kalenderwoche

IX

LiV	Lehrkräfte im Vorbereitungsdienst (gemeint sind damit auch Referendar*innen)
M	Mittelwert
M. A.	Master of Arts
M. Ed.	Master of Education
M. Sc.	Master of Science
Max	Maximum
Min	Minimum
PE	Personalentwicklung
SD	Standardabweichung
TeKoP	Technologiebasiertes kompetenzorientiertes Prüfen
VFD	Vertiefende Fachdidaktik und Unterrichtsforschung Wirtschaftswissenschaften
VIF	Variance Inflation Factor
WI	Wirtschaftsinformatik
WiPäd	Wirtschaftspädagogik
WiWi	Wirtschaftswissenschaften
ZPA	Zentralstelle für Prüfungsaufgaben

1 Problemstellung

Kompetenzorientierung ist ein Begriff, der von allgemeinbildenden Institutionen bis hin zu Hochschulen in allen Bildungseinrichtungen thematisiert wird (Bethscheider, Höhns & Münchhausen, 2011; Klein, 2016; Reusser, 2014, S. 325-326; Schaper, 2012; Weyer, Wachendorf & Mört, 2017, S. 6). Gestärkt wurde die Kompetenzorientierung durch die Formulierung europäischer und nationaler Qualifikationsrahmen (Achtenhagen & Baethge, 2007, S. 52; Bethscheider et al., 2011, S. 11; Schaper, 2012, S. 24), die sich nicht nach Länge, Dauer und Methode einer Bildung richten – Voraussetzung für eine bestimmte Qualifikation sind die erworbenen Kompetenzen¹ (Fritz & Lauermann, 2019, S. 63).

Die Berufsausbildung hat das Ziel, angehende Fachkräfte auf die Anforderungen eines dynamischen Arbeitsmarktes vorzubereiten und die berufliche Handlungsfähigkeit bzw. *-kompetenz*² dieser zu fördern (§ 1 Abs. 3 BBiG, vgl. auch Bundesinstitut für Berufsbildung, 2015, S. 7). Die *umfassende Handlungskompetenz* wird im Deutschen Qualifikationsrahmen für lebenslanges Lernen (2011) definiert als „die Fähigkeit und Bereitschaft des Einzelnen, Kenntnisse und Fertigkeiten sowie persönliche, soziale und methodische Fähigkeiten zu nutzen und sich durchdacht sowie individuell und sozial verantwortlich zu verhalten“ (S. 4). Gerade in Zeiten des technologischen Wandels und einem Substituierbarkeitspotenzial beruflicher Tätigkeiten wird zudem die Relevanz beruflicher Problemlösekompetenz deutlich (Abele et al., 2021, S. 13; Bonekamp & Sure, 2015, S. 34; Dengler & Matthes, 2015, S. 12; Frey & Osborne, 2013, 2017; Funke, Fischer & Holt, 2018, S. 41; Goos, 2013; Hämäläinen, Cincinnato, Malin & Wever, 2014, S. 59; Seeber et al., 2019; Vuori, Helander & Okkonen, 2019, S. 238; Wuttke et al., 2015, S. 190). Am Beispiel des Ausbildungsberufs des*r Industriekaufmanns*frau wird deutlich, dass in dem Beruf weniger routinierte Aufgaben wahrgenommen werden müssen und der Trend u. a. in Richtung Problemlösung geht (Bundesinstitut für Berufsbildung, 2019, S. 442-443). So ermöglicht die berufliche Problemlösekompetenz – als Teil beruflicher Handlungskompetenz (z. B. Badura, 2015, S. 61; Ressel & Kiesecker, 2019, S. 22, vgl. Kapitel 2.5.5) mit dem Fokus auf das Problemlösen – (angehenden) Fachkräften die kompetente Lösung von Problemen, die nicht automatisiert bearbeitet werden können. Die Problemlösekompetenz wird

¹ Im Rahmen der Arbeit wird auf das Kompetenzverständnis von Weinert zurückgegriffen. Nach Weinert (2014) ist Kompetenz zu verstehen als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (S. 27-28).

² Der im Berufsbildungsgesetz genutzte Begriff der *beruflichen Handlungsfähigkeit* ist, nach dem Bundesinstitut für Berufsbildung (2015, S. 23) als adäquat mit dem Begriff der *Handlungskompetenz* des DQR zu verstehen.

im Rahmen der vorliegenden Arbeit verstanden als das „Konglomerat aus Fachwissen, Fertigkeiten und Fähigkeiten sowie Interessen und Bereitschaften, [zur] ... Bewältigung komplexer beruflicher Anforderungen“ (Siegfried et al., 2019, S. 990).

Problemlösekompetenz ist curricular verankert (§ 1, § 38 BBiG; am Beispiel des Berufs Industriekaufmann*frau: vgl. Bundesinstitut für Berufsbildung, 2002; Kultusministerkonferenz [KMK], 2002). Die curricularen Ansprüche richten sich sowohl an die Gestaltung von Lehr-Lern-Situationen als auch an die Prüfungspraxis. Dies ist im Sinne der Curriculum-Instruktion³-Assessment-Triade (Pellegrino, 2010, S. 4), wonach die drei Elemente aufeinander abgestimmt bzw. im Einklang sein sollen.⁴ Im Unterricht an beruflichen Schulen finden sich Ansätze, die die Problemlösekompetenz fördern⁵ (vgl. Anforderung aus: Sekretariat der KMK, 2007, S. 10; Methodisch über z. B. Lernfabriken, vgl. z. B. Fasshauer, Wilbers & Windelband, 2021; sowie als Planspiele, vgl. z. B. Beiträge in Blötz, 2008). Am Lernort Betrieb werden Auszubildende mit beruflichen Problemen aus der Praxis konfrontiert⁶, weshalb davon auszugehen ist, dass damit ihre berufliche Problemlösekompetenz gefördert wird. Durch die Abschlussprüfung als summatives Assessment (Vogel, 2014, S. 16) soll schließlich festgehalten werden, inwiefern die Auszubildenden das Ziel der beruflichen Bildung erreicht haben (§ 38 BBiG; vgl. auch Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen [AkA], 2023a, 2023b; ZPA Nord-West, 2021). Zur validen Messung der Problemlösekompetenz bedarf es demnach Prüfungsaufgaben, die berufliche Probleme (authentisch⁷) abbilden. Zur Beurteilung der Problemhaltigkeit von Aufgaben können Merkmale problemhaltiger⁸ Aufgaben herangezogen werden (z. B. Betsch, Funke & Plessner, 2011; Dörner, 1987; Jonassen, 2000; Jonassen & Hung, 2012; Sembill, 1992, vgl. Kapitel 2.5.3). Ergebnisse der Analyse von 1.454 schriftlichen Prüfungsaufgaben für angehende Industriekaufleute und Kaufleute für Büromanagement zeigen, dass bestehende Prüfungsaufgaben noch nicht hinreichend berufliche Probleme abbilden (Wuttke, Seeber, Geiser & Turhan, 2022). Somit ist die Validität der Prüfungsaufgaben gefährdet und

³ Die Instruktion steht für die Aktivitäten der Lehrenden und Lernenden (Pellegrino, 2010, S. 5) und werden hier als Lehr-Lern-Situationen verstanden.

⁴ In der vorliegenden Arbeit werden nicht alle drei Elemente der Triade bearbeitet, sondern der Fokus auf das Assessment, hier die Prüfung, und darunter die schriftliche Prüfung, gelegt. Zu begründen ist dies insbesondere mit der Kritik an bestehenden schriftlichen Prüfungsaufgaben für kaufmännische Berufe, auf die in Kapitel 3.2 eingegangen wird.

⁵ In Kapitel 2.5.5 wird auf Parallelen und die Beziehung von Handlungs- und Problemorientierung eingegangen.

⁶ Sicherlich ist davon auszugehen, dass Problemsituationen im beruflichen Alltag eines Auszubildenden je nach Unternehmen, Branche etc. unterschiedlich sind und unterschiedlich häufig vorkommen.

⁷ Authentizität wird als wichtig bei der Konzeption von Problemsituationen betrachtet, wenngleich davon auszugehen ist, dass es nicht Problemhaltigkeit bedingt (Wuttke et al., 2022, S. 32).

⁸ Unter *problemhaltig* bzw. dem Begriff der *Problemhaltigkeit* wird die Ausprägung der Merkmale problemhaltiger Aufgaben in einer (Prüfungs-)Aufgabe verstanden. Die im Rahmen der Arbeit berücksichtigten Merkmale werden in Kapitel 2.5.3 vorgestellt.

die von Pellegrino (2010) vorgestellten Ansprüche der Curriculum-Instruktion-Assessment-Triade sind nicht erfüllt. Die Kritik an der unzureichenden Passung von Curriculum und Assessment ist nicht neu (vgl. Klotz & Winther, 2016), auch wenn der Fokus dabei nicht immer auf dem Prüfen der Problemlösekompetenz lag. Dieses widersprüchliche Vorgehen führt zu einem Spagat: Während auf der einen Seite eine praxisnahe, handlungs- und problemorientierte Methode gefordert wird, besteht auf der anderen Seite die Notwendigkeit, Lernende auf Prüfungssituationen und -aufgaben vorzubereiten, die in der Regel wenig problemhaltig sind (Wuttke et al. 2022, S. 45; Wuttke & Wolf, 2007, S. 101). Das nicht-komplementäre Vorgehen in der Prüfungspraxis und den Anforderungen der Curricula kann dazu führen, dass Lehrkräfte ausschließlich Prüfungsrelevantes unterrichten bzw. auf die Prüfung vorbereiten, womit das Phänomen des *teaching to the test* angesprochen wird (Oehmann & Blumschein, 2019, S. 44-45; vgl. auch Wuttke et al., 2022, S. 45; Wuttke & Wolf, 2007, S. 101).

Kaufmännisch-schriftliche Prüfungsaufgaben sind bundesweit einheitlich (Ausnahme: Baden-Württemberg) und werden über die AkA bzw. ZPA-Nord-West entwickelt. Am Entwicklungsprozess sind drittparitätisch Arbeitgeber- und Arbeitnehmervertreter*innen beteiligt sowie auch Lehrkräfte beruflicher Schulen (AkA, 2022; ZPA Nord-West, 2021) bzw. Lehrkräfte und Prüfende sowie Landesfachausschüsse (Land Baden-Württemberg, vertreten durch das Institut für Bildungsanalysen Baden-Württemberg [IBBW], 2022, S. 3). Die Gruppe der Lehrkräfte kaufmännisch-beruflicher Schulen wird im Rahmen der vorliegenden Dissertation fokussiert.

Die Erstellung von schriftlichen problemhaltigen Prüfungsaufgaben setzt insbesondere pädagogisch-diagnostische und fachdidaktische Kompetenzen voraus. Diese findet sich in der Modellierung zu den Kompetenzen von Lehrkräften wieder (vgl. Baumert & Kunter, 2006, S. 485; 2011, S. 32; Shulman, 1986, S. 9-10; vgl. Kapitel 2.1). Ebenso begründet sich die notwendige Kompetenz der Lehrkräfte curricular (z. B. Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2003, 2014; Terhart, 2002; vgl. Kapitel 2.2) und praktisch (Klausurerstellung: z. B. Besser, Rau & Traum, 2022, S. 57; Dubs, 2008, S. 17; Schwaighofer, Heene & Bühner, 2019, S. 472; Prüfungserstellung: z. B. AkA, 2022; ZPA Nord-West, 2021, vgl. Kapitel 2.3). In Kapitel 2.4 werden Komponenten und wesentliche Aspekte pädagogischer Diagnostik beleuchtet, bevor auf die Problemorientierung eingegangen wird (Kapitel 2.5). In Letzterem werden neben der Problemdefinition auch das Problemlösen und Merkmale problemhaltiger Aufgaben thematisiert sowie Problem- und Handlungsorientierung gegenübergestellt. Das Kapitel schließt mit der Erläuterung zum Erstellen problemhaltiger Prüfungsaufgaben ab (Kapitel 2.6).

Der Stand der Forschung zu den relevanten Kompetenzen (angehender) Lehrkräfte, die notwendig sind, um entsprechende Aufgaben zu erstellen, ist rar (vgl. Kapitel 3). Zu den diagnostischen Kompetenzen von Lehrkräften besteht Forschungspotential (vgl. z. B. Baumert & Kunter, 2006, S. 489 zur Entwicklung und Stabilität diagnostischer Kompetenz). Bisherige Forschungsarbeiten beschränken sich zumeist auf die Diagnosegenauigkeit (Schrader, 2013, S. 162). Zur Erstellung von problemhaltigen Aufgaben zu summativen Assessmentzwecken, insbesondere schriftlichen Prüfungsaufgaben, ist der Forschungsstand ausbaufähig.

Um systematisch der Frage nachzugehen, wie angehende Lehrkräfte kaufmännisch-beruflicher Schulen auf das Erstellen problemhaltiger (Prüfungs-)Aufgaben vorbereitet werden, wurden im ersten Schritt der Forschungsarbeit *Ausbildungsinhalte* der ersten⁹ und zweiten^{10,11} Phase der Lehrer*innenbildung (vgl. Sekretariat der KMK, 2019, S. 4; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004, S. 4) kaufmännisch-beruflicher Schulen untersucht. Dabei wurde der Fokus auf die hierfür notwendige Kompetenz zur pädagogischen Diagnostik und Fachdidaktik gelegt. Die Untersuchung erfolgte zum einen über die Beleuchtung der Angebotsperspektive (vgl. Kapitel 4.2) und zum anderen der Nachfrageperspektive (vgl. Kapitel 4.3). Auf Seiten der Angebotsperspektive wurden Dokumente der ausbildenden Institute, die die Ausbildungsinhalte der jeweiligen Ausbildungsphasen abbilden, strukturiert-inhaltsanalytisch (Mayring, 2015, S. 98; 2016, S. 120) analysiert. Diese setzten sich aus Dokumenten von 22 Standorten zum Bachelor- und 21 Standorten zum Masterstudium zusammen. Zum Referendariat wurden Dokumente von sechs Bundesländern betrachtet. Zur Validierung der Ergebnisse der Dokumentenanalyse wurde eine bundesweite Fragebogenstudie mit (vertretenden) Studiengangsleiter*innen der Wirtschaftspädagogik (erste Phase der Lehrer*innenbildung) sowie eine Interviewstudie mit Fachleiter*innen der kaufmännisch-beruflichen Studienseminare (zweite Phase der Lehrer*innenbildung) durchgeführt. Um auch die Wahrnehmung angehender Lehrkräfte in der ersten und zweiten Phase der Lehrer*innenbildung zu erfahren (Nachfrageperspektive), wurden 13 Studierende des Masterstudiengangs Wirtschaftspädagogik und drei Lehrkräfte im Vorbereitungsdienst/Referendar*innen

⁹ In Deutschland ist es an 25 staatlichen Universitäten möglich, Wirtschaftspädagogik im Bachelor und/oder im Master zu studieren. Diese werden in Kapitel 4.2.1 vorgestellt.

¹⁰ Die zweite Phase der Lehrer*innenbildung wird unterschiedlich bezeichnet, wobei die Bezeichnung *Referendariat* geläufiger scheint. Daher wird im Rahmen der Arbeit von *Referendariat* gesprochen, wobei hiermit auch der Vorbereitungsdienst und insgesamt die zweite Phase der Lehrer*innenbildung gemeint ist.

¹¹ Zur zweiten Phase der Lehrer*innenbildung wurde der Fokus auf folgende Bundesländer gelegt: Baden-Württemberg, Bayern, Hessen, Niedersachsen, Nordrhein-Westfalen und Rheinland-Pfalz. Die Auswahl der Bundesländer wird in Kapitel 4.2.1 begründet.

(LiV) an kaufmännisch-beruflichen Schulen interviewt. Die Auswertung der Interviews erfolgte nach der zusammenfassend-qualitativen Inhaltsanalyse nach Mayring (2016, S. 96). In Kapitel 4 wird das methodische Vorgehen zu den Studien vorgestellt. In Kapitel 5 werden die Ergebnisse der Studien dargestellt sowie Limitationen aufgezeigt und zusammenfassende Implikationen abgeleitet.

Aufbauend auf der begründeten Notwendigkeit, problemhaltige Prüfungsaufgaben zu erstellen (Kapitel 2.1 bis 2.3) und dem ermittelten Optimierungsbedarf zu den Ausbildungsinhalten (Kapitel 5.1 bis 5.3), wurde ein Trainingskonzept mit der Zielgruppe der angehenden Lehrkräfte begründet ausgewählt, durchgeführt und evaluiert. Als Trainingskonzept wurde auf ein bestehendes und zum Einsatzzeitpunkt – (erstmalig) evaluiertes Konzept des Forschungsprojekts *TeKoP*¹² mit guten bis sehr guten Ergebnissen in Bezug auf beispielsweise die Trainings- und Instruktionsszufriedenheit, den Trainingsnutzen und den eingeschätzten Lernzuwachs (z. B. Geiser, Turhan, Wuttke, Seeber & Meiners, 2021, Folie 11-17; Wuttke, Seeber, Meiners & Turhan, 2023) zurückgegriffen. Das Konzept ist modular aufgebaut und setzt sich aus den Bausteinen *Erstellen problemhaltiger Aufgaben*, *Mediendidaktik*, *Einführung in die Lehr-Lernplattform*, *Feedback und Reflexion*, *Transfer* und *Diagnostik* zusammen. Beim Einsatz des Konzepts wurden die teilnehmendenbedingten (z. B. Lipowsky & Rzejak, 2021; Richter, Engelbert, Weirich & Pant, 2013, S. 195-196; Vigerske, 2016, vgl. Kapitel 6.2) sowie trainingsbedingten (z. B. Krille, 2019; Lipowsky, 2004, Lipowsky & Rzejak, 2021, vgl. Kapitel 6.3) Faktoren berücksichtigt. Die *Trainingsevaluation* erfolgte auf den ersten beiden Ebenen (Reaktion und Lernen) nach Kirkpatrick und Kirkpatrick (2006), vgl. Kapitel 6.5. Hierfür wurde ein quasi-experimentelles Design genutzt (vgl. Döring & Bortz, 2016, S. 199; McMillan & Schumacher, 2010, S. 22; Kapitel 7.3), indem, neben biografischen Daten, das Wissen der Interventions- und Kontrollgruppe mittels eines Pre- und Posttests abgefragt wurde (Studierende: N_{IG} = 58; N_{KG} = 30; LiV: N_{IG} = 5; N_{KG} = 5). Die Interventionsgruppe wurde im Posttest zusätzlich zum Training (Instruktionsqualität, Zufriedenheit, Nutzen, eingeschätzter Lernzuwachs) befragt. Die Daten wurden deskriptiv ausgewertet, mit studentischen Daten konnte zudem eine Kovarianzanalyse berechnet und eine Regressionsanalyse durchgeführt werden. Um tiefere Einblicke zu gewinnen, wurden zudem Interviews mit 13 Studierenden und drei LiV, als

¹² BMBF gefördertes Forschungsprojekt „Technologiebasiertes kompetenzorientiertes Prüfen“, Förderkennzeichen: 21AP001A, 21AP001B

Teilnehmenden der Interventionsgruppe, geführt (vgl. Kapitel 7.4). Diese wurden zusammenfassend-inhaltsanalytisch nach Mayring (2016, S. 96) ausgewertet. Die Ergebnisse und Limitationen der Studien sowie die daraus abgeleiteten Implikationen finden sich in Kapitel 8.

Die mittels qualitativer und quantitativer Daten erzielten Ergebnisse der Trainingsevaluation zeigen eine überwiegende Passung des Konzepts zu den Bedarfen der Zielgruppe. Dies spricht für eine Implementierung des Konzepts in die Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte. Die mittels der Interviewstudie mit Teilnehmenden der Interventionsgruppe gewonnenen Ergebnisse zeigen Verbesserungspotentiale des Konzepts, die bei der Implementierung berücksichtigt werden sollten. Dies betrifft insbesondere den Baustein zur Diagnostik. Der modulare Aufbau erlaubt die Anpassung bzw. Verbesserung einzelner Bausteine. Zudem ist es dadurch möglich, entsprechend den Bedarfen und organisatorischen Rahmenbedingungen des ausbildenden Standorts, einzelne Bausteine zu nutzen und sie ggf. zu adaptieren.

Die Ergebnisse der Vorstudien sowie der Trainingsevaluation leisten einen Beitrag dazu, die Ausbildung angehender Lehrkräfte zu optimieren. Zudem soll langfristig zur Verbesserung der Prüfungspraxis beigetragen werden¹³, was wiederum eine Verbesserung der beruflichen Ausbildung insgesamt nach sich ziehen soll. Die vorliegende Arbeit schließt mit einer Diskussion und der Darstellung des weiteren Forschungsbedarfs ab (Kapitel 9).

¹³ Dieses Ziel überschneidet sich mit dem des TeKoP-Projekts (vgl. Goethe-Universität Frankfurt & Georg-August-Universität Göttingen, 2023, S. 21).

2 Kompetenzanforderung an (angehende) Lehrkräfte zur Erstellung problemhaltiger Prüfungsaufgaben

2.1 Modellierung der Kompetenzen (angehender) Lehrkräfte

In der Literatur zu den notwendigen Kompetenzen von (angehenden) Lehrkräften findet sich keine einheitliche Definition (z. B. Dubs, 2008, S. 17; Weinert, Schrader & Helmke, 1990; vgl. hierzu auch die Beschreibung des DFG-Schwerpunktprogramms in Klieme & Leutner, 2006; auch Ansätze spezifisch für bestimmte Fächer/Domänen, fachübergreifend und/oder für bestimmte Schulstufen, vgl. z. B. Beiträge in Koch-Priewe, Köker, Seifried & Wuttke, 2015). Dies ist insofern nachvollziehbar, als dass die Handlungsaufgabe einer Lehrkraft, verbunden mit den hierfür notwendigen Kompetenzen, nicht klar festgelegt ist (Tenorth, 2006, S. 586). Baumert und Kunter (2006) ergänzen, dass die Aufgabe der Lehrkraft „unter Berücksichtigung der jeweiligen Umstände des Einzelfalls und der jeweiligen Situation ausdeutungs- und konkretisierungsbedürftig“ (S. 476) sei. Auch Weinert et al. (1990, S. 164-165) halten fest, dass etwa das Fach sowie die Situation der Lernenden zu Unterschieden führen. Zwar gibt es Modellierungen, die auf normativen Vorgaben und empirischen Forschungsarbeiten aufbauen (wie etwa Reynolds, 1995). Dennoch sind diese nicht als statisch zu betrachten. Insofern werden Ansätze zur Modellierung als Hilfestellungen verstanden, die beispielsweise für die Erstellung von Ausbildungsprogrammen nützlich sind. In der Vielfalt der (Ansätze zur) Modellierung scheint das Modell von Baumert und Kunter (2011, S. 32) das in den letzten Jahren am häufigsten genutzte bzw. zitierte zu sein (vgl. z. B. Bürger et al., 2021, S. 1089; Frey, 2014, S. 716-717; Greiwe, 2020; Holtsch, 2011; Niermann, 2017; Wuttke & Seifried, 2017).

Das *Modell professioneller Handlungskompetenz* unterteilt die Kompetenzaspekte in *Überzeugungen/Werthaltungen/Ziele*, *Motivationale Orientierungen*, *Selbstregulation* und *Professionswissen* (z. B. Baumert & Kunter, 2011, S. 32). Im Rahmen dieser Arbeit liegt der Fokus auf dem *Professionswissen*. In Kompetenzbereichen werden Kompetenzfacetten unterschieden, „die durch konkrete Indikatoren operationalisiert werden“ (Baumert & Kunter, 2011, S. 33). Diese wiederum umfassen deklaratives, prozedurales bzw. konzeptuelles Wissen (Krauss et al., 2004, S. 35). Im Kompetenzmodell werden die Kompetenzfacetten konkretisiert, insbesondere für das Fach Mathematik (vgl. z. B. Baumert & Kunter, 2011, S. 32; Krauss et al., 2004, S. 34-35; Kunter, Klusmann & Baumert, 2009, S. 154-156). Da der Schwerpunkt der vorliegenden Arbeit nicht auf dem Fach Mathematik liegt, wird auf eine solche Spezifikation verzichtet. Gleichwohl wird zu Zwecken der Vollständigkeit das Modell in seiner ursprünglichen Version in Abbildung 1 aufgenommen.

Baumert und Kunter (2006) fassen in ihrer Arbeit zusammen, dass ein breiter Konsens¹⁴ darüber besteht, „dass *Wissen und Können* [Herv. i. Original] – also deklaratives, prozedurales und strategisches Wissen – zentrale Komponenten der professionellen Handlungskompetenz von Lehrkräften darstellen“ (S. 481). Hingegen gibt es „in Bezug auf die Struktur und Topologie professionellen Wissens und Könnens, die unterschiedlichen Wissenstypen und ihren epistemischen Status, die mentale Repräsentation dieser Wissenstypen und die Genese professionellen Wissens und Könnens“ (Baumert & Kunter, 2006, S. 481) keinen derartigen Konsens. Die Definitionen nach Shulman (1986)¹⁵ haben sich jedoch durchgesetzt, wodurch in der Literatur Wissensdomänen von Lehrkräften meist in *allgemein pädagogisches Wissen*, *Fachwissen* sowie *fachdidaktisches Wissen* unterteilt werden¹⁶ (Baumert & Kunter, 2006, S. 482). Weiterhin bestehen *Organisationswissen* und *Beratungswissen* als Kompetenzbereiche des Professionswissens (Baumert & Kunter, 2006, S. 482). Von den Beschreibungen ausgehend, gehen die im Modell genannten Aspekte über das reine Wissen und Können hinaus und zielen auf Kompetenzen (z. B. Baumert & Kunter, 2011, S. 31).

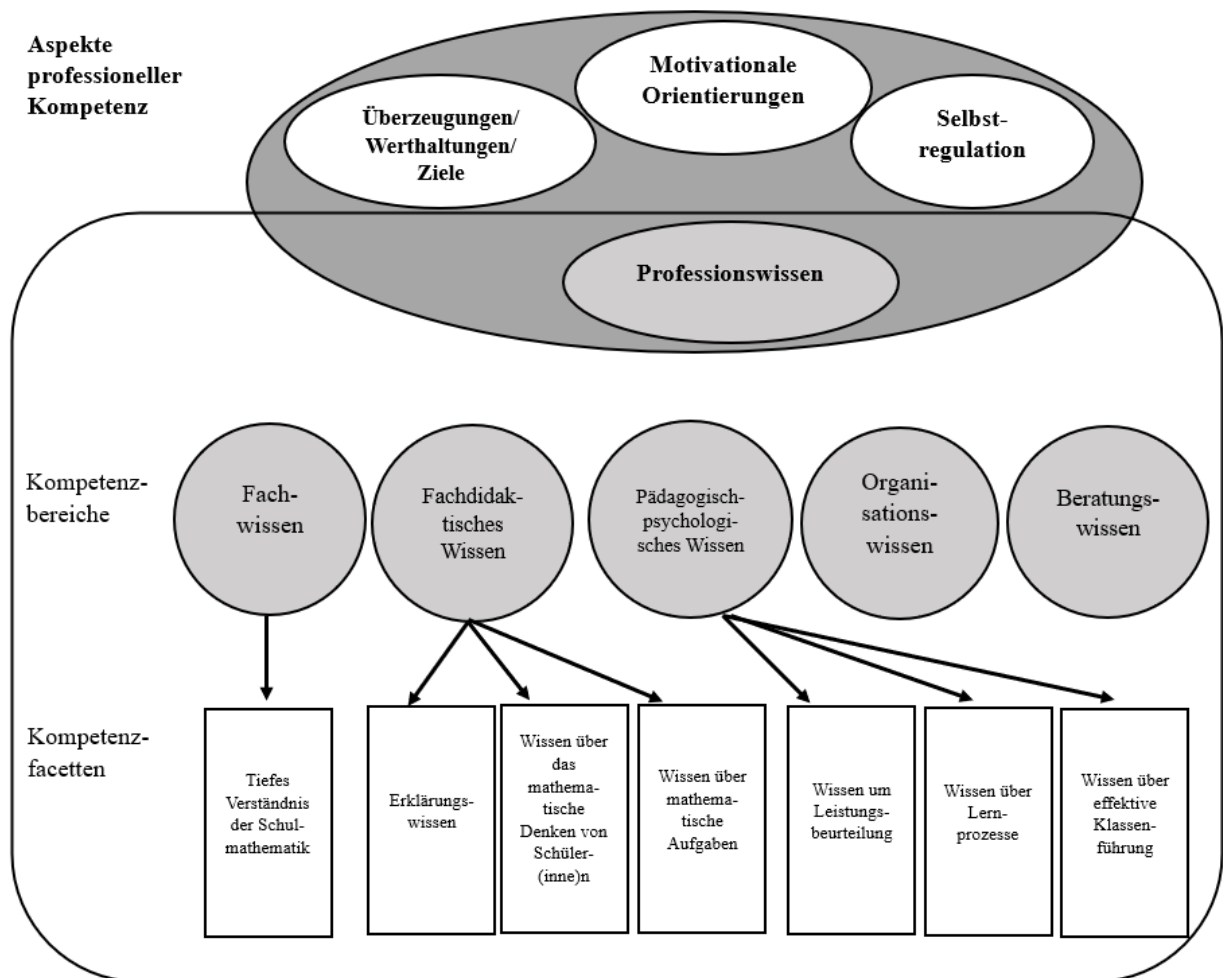
¹⁴ Mit dem Ziel der Formulierung von Standards zu dem, was (angehende) Lehrkräfte wissen und können sollten, stellte beispielsweise die National Conference of State Legislatures (2011, S. 4) fünf zentrale Anforderungen bzw. Aussagen auf: „Teachers are committed to student learning, Teachers know the subjects they teach and how to teach those subjects to students, Teachers are responsible for managing and monitoring student learning, Teachers think systematically about their practice and learn from experience, Teachers are members of learning communities [im Original stichpunktartig aufgezählt. Hier wurden die Punkte zur besseren Lesbarkeit mit einem Komma getrennt]“. Diese Standards sollen gleichzeitig valide gemessen werden [vgl. hierzu Zitat von Carnegie Forum on Education (1986) in National Conference of State Legislatures (2011, S. 4)].

¹⁵ Für eine genauere Angabe der Quellen: Shulman (1986, S. 9-10, 1987, S. 8).

¹⁶ Zu den genannten drei Domänen ergänzte Shulman (1986, S. 9-10, 1987, S. 8) weitere. Zu diesen zählen das *curriculare Wissen*, das *Wissen über die Lernenden und deren Charakteristiken*, das *Wissen über die Bildungskontexte* und das *Wissen über die Ziele, Zwecke und Werte der Bildung und ihrer philosophischen sowie historischen Grundlagen* (Shulman, 1987, S. 8; die kursiven Bezeichnungen wurden wörtlich aus dem Text übersetzt). Interessant scheint zudem, dass Shulman (1987, S. 8) diese als *Minimum* (wörtliche Übersetzung aus dem Original) bezeichnet. Auf eine Ausführung bzw. Erläuterung der genannten Domänen wird aufgrund des Schwerpunktes der Arbeit verzichtet und auf die Primärquellen Shulman (1986, 1987) verwiesen.

Abbildung 1:

Das Kompetenzmodell von COACTIV mit Spezifikationen für das Professionswissen



(Nachzeichnung nach Baumert & Kunter, 2011, S. 32)

Shulman (1986) beschreibt das Fachwissen (*Subject matter Content knowledge*) wie folgt:

Teachers must not only be capable of defining for students the accepted truths in a domain. They must also be able to explain why a particular proposition is deemed warranted, why it is worth knowing, and how it relates to other propositions, both within the discipline and without, both in theory and in practice. (S. 9)

Daraus ist zu schließen, dass das Fachwissen mehr als die Reproduktion von Wissen ist. Es erfordert vielmehr die Kompetenz, dieses Wissen zu verorten, abzugrenzen und die Relevanz dessen aufzuzeigen und zu begründen.

Fachdidaktischen Wissen¹⁷ (*pedagogical content knowledge*) beschreibt Shulman (1986) als „subject matter knowledge for teaching“ (S. 9) und führt etwa die Erklärungen, Beispiele und nützliche Darstellungen von Ideen auf. Damit ist die Art und Weise zu verstehen, wie ein Thema für andere Personen verständlich formuliert und dargestellt wird. Gleichzeitig soll es das Wissen bzw. Verständnis beinhalten, was das Erlernen bestimmter Themen vereinfacht oder erschwert sowie Strategien zur Unterstützung der Lernenden (Shulman, 1986, S. 9-10). Daher scheint es nachvollziehbar, dass Fachdidaktik als „die unabdingbare Voraussetzung für die inhaltliche Bewältigung des Unterrichts“ (Dubs, 2008, S. 26) definiert wird. Die Fachdidaktik umfasst zudem das Wissen in Bezug auf Aufgaben (Artelt & Kunter, 2019, S. 400). Auch wenn zum fachdidaktischen Wissen die Vermittlung von Inhalten – also den Lehr-Lern-Prozess – gehört (Artelt & Kunter, 2019, S. 400-401), ist im Sinne der Curriculum-Instruktion-Assessment-Triade (Pellegrino, 2010) davon auszugehen, dass fachdidaktische Aspekte auch bei der Gestaltung von Aufgaben zu (summativen) Assessmentzwecken berücksichtigt werden sollten. Ein tieferes Verständnis zum Wissenstypen und -niveau in Bezug auf das fachdidaktische Wissen scheint es bislang nicht zu geben (vgl. hierzu auch Baumert & Kunter, 2006, S. 492 nach Bromme, 1995).

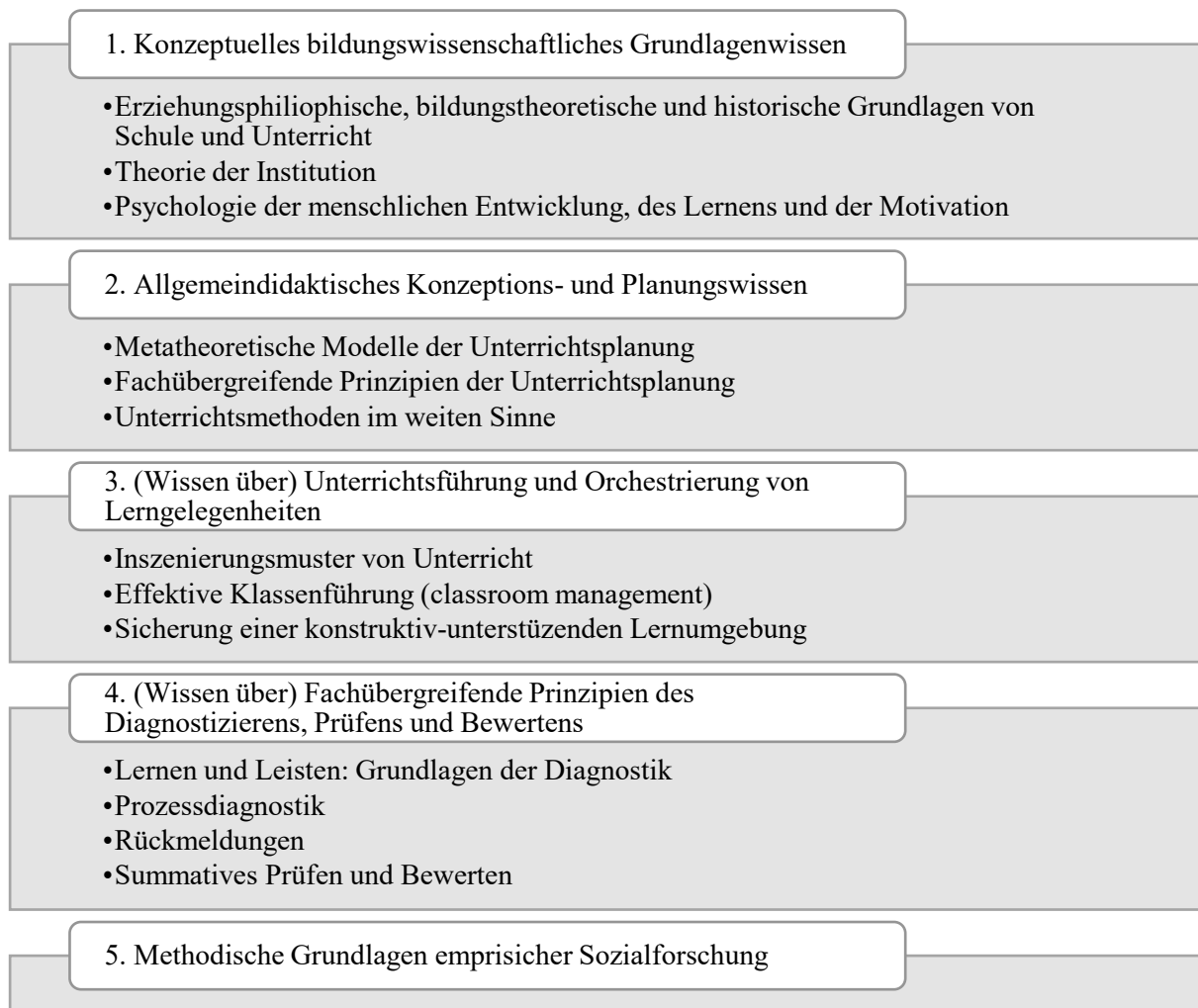
Im hier vorgestellten Modell ist die *diagnostische Kompetenz* als Facette fachdidaktischen Wissens enthalten (vgl. Krauss et al., 2004, 35). Krauss et al. (2004, S. 43) merken an, dass es sich dabei um die Einschätzung der Schüler*innenfähigkeiten in der betrachteten Domäne (Mathematik) geht. Neben *diagnostischer Kompetenz* werden hierzu auch das „Wissen über kognitive Anforderungen und Potenziale von Mathematikaufgaben“, die „Flexible Nutzung von vernetztem Wissen“ sowie der „Umgang mit Fehlern“ aufgeführt (Krauss et al., 2004, S. 42-43). Interessant scheint in dieser Aufführung insbesondere, dass „Diagnostische Kompetenz“ und das „Wissen über kognitive Anforderungen und Potenziale“ der Aufgaben separat aufgeführt werden. Dies führt zur Annahme, dass diagnostische Kompetenz im Kompetenzbereich des fachdidaktischen Wissens des hier betrachteten Modells liegt, anders als von anderen Autor*innen erläutert (vgl. z. B. Helmke, 2021, S. 132)¹⁸, und Aufgabenmerkmale nicht inbegriffen sind.

¹⁷ Genauso wie die Übersetzung der anderen englischen Begriffe zu den Wissensdomänen nach Shulman (1986, 1987), wurde die Übersetzung als fachdidaktisches Wissen auch aus Baumert und Kunter (2006, S. 490, 2011, S. 33-34) übernommen. Diese Übersetzung(en) findet sich in anderen Quellen wieder (z. B. Niemann, 2017, S. 9). Bromme (1995, S. 105) nutzt eine ähnliche Übersetzung als „didaktisches Fachwissen“. *Pedagogical content knowledge* wird in der Literatur aber auch als „pädagogisches Inhaltswissen“ (z. B. Greiwe, 2020, S. 33) übersetzt. Die Entscheidung für die Übersetzung liegt in der Erläuterung des Begriffs von Shulman (1986, S. 9) selbst, die in dem Absatz wiedergegeben wird und als didaktisches Wissen aufgefasst wird.

¹⁸ Auf das Begriffsverständnis und die Funktion pädagogischer Diagnostik wird im Kapitel 2.3.1 genauer eingegangen.

Dies wird auch durch die Erläuterung der Studie zur Erhebung diagnostischer Kompetenz bzw. Leistungsbeurteilung deutlich (vgl. Krauss et al., 2004, S. 43).

Allgemeines pädagogisches Wissen und Können werden als „zweifellos zum Kern der professionellen Kompetenz von Lehrkräften“ (Baumert und Kunter, 2006, S. 485) gehörend beschrieben. Unter allgemein pädagogischem Wissen (*General pedagogical knowledge*) wird das *Wissen zum Lehren* verstanden (Shulman, 1986, S. 9). Dieses umfasst die Art und Weise der Erklärung(en) bzw. Darstellung(en) eines Themas, um es für Lernende verständlich darzustellen. Auch das Wissen darüber, was ein Thema für Lernende erschwert oder vereinfacht, ist hierunter subsummiert. Zu Letzterem zählen auch die Bedingungen der Lernenden. Ebenso sind dazu die Erkenntnisse aus der Lehr-Lernforschung zum *pädagogischen Wissen* zu zählen (Shulman, 1986, S. 9-10). Aufbauend auf u. a. den Erkenntnissen von Shulman (1987) sowie z. B. Reynolds, Tannenbaum und Rosenfeld (1992), Rosenfeld und Tannenbaum (1991), Darling-Hammond und Bransford (2005) und Terhart (2002), haben Baumert und Kunter (2006) „die weitgehend konsensfähigen“ (S. 484) Punkte strukturiert. Nachfolgende Abbildung stellt die *Facetten generischen pädagogischen Wissens und Könnens* nach Baumert und Kunter (2006, S. 485, 2011, S. 39) dar.

Abbildung 2:*Facetten generischen pädagogischen Wissens und Könnens*

(Eigene Darstellung nach Baumert & Kunter, 2006, S. 485 sowie Baumert & Kunter, 2011, S. 39)

Das *allgemein-pädagogische Wissen* haben Baumert und Kunter (2011, S. 34) in ergänzter Form in das Kompetenzmodell aufgenommen, es findet sich unter der Bezeichnung *Pädagogisch-psychologisches Wissen* wieder. In Kapitel 2.4 werden die Komponenten zum „(Wissen über) Fachübergreifende Prinzipien des Diagnostizierens, Prüfens und Bewertens“ (Baumert & Kunter, 2006, S. 485, 2011, S. 39) betrachtet. Der Punkt *Rückmeldungen* wird dabei bewusst vernachlässigt, da der Fokus der vorliegenden Arbeit auf das Erstellen problemhaltiger Prüfungsaufgaben liegt. Die anderen Punkte werden zum Teil zusammengefasst und/oder erweitert dargestellt, um die Verständlichkeit und Möglichkeit zur Abgrenzung zu verbessern. Zudem werden im genannten Kapitel Gütekriterien als eigenständiges Unterkapitel aufgenommen.

2.2 Curricular begründete Anforderung an (angehende) Lehrkräfte

In diesem Kapitel sollen ausgewählte Dokumente zur Begründung der hier betrachteten Anforderungen aufgeführt werden. Zu den untersuchten Dokumenten gehört das Dokument *Standards für die Lehrerbildung* von Terhart (2002), welches sich auf die erste und zweite Phase der Lehrer*innenbildung bezieht (Terhart, 2002, S. 5). Hierin werden die Anforderungen an die Absolvent*innen der jeweiligen Phase als Standards vorgestellt (Terhart, 2002, S. 30). Daneben werden die Dokumente der KMK zu den *Standards für die Lehrer*innenbildung* in Bezug auf die *Bildungswissenschaften* (2004, 2019) sowie die *Basiscurricula* der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (2003, 2014) einbezogen. Die sich aus den Dokumenten ergebenden Anforderungen werden unterteilt in die erste und zweite Phase der Lehrer*innenbildung. Dabei werden ausschließlich Anforderungen dargestellt, die Kompetenzen bzw. Kompetenzbereiche betreffen, die dem Fokus der Arbeit entsprechen.

*Zur ersten Phase der Lehrer*innenbildung*

Die hier fokussierten pädagogisch-diagnostischen und fachdidaktischen Kompetenzen sind in der Aufstellung von Terhart (2002, S. 34-35) zur ersten Phase der Lehrer*innenbildung insbesondere den Bereichen der *Erziehungswissenschaft* und *Fachdidaktik* zuzuordnen. Anforderungen bezüglich der pädagogischen Diagnostik finden sich hierbei primär¹⁹ in den Punkten *Lern-diagnostik und Lernförderung* sowie *Lernschwierigkeiten, Heterogenität, Leistungsbeurteilung* (Terhart, 2002, S. 34-35). Jedoch kann nicht davon ausgegangen werden, „dass das erziehungswissenschaftliche Studium innerhalb der universitären Lehrerbildung (1. Phase) den Absolventen gewissermaßen berufsfertig entlässt“ (Terhart, 2002, S. 32). Bezüglich der betrachteten fachdidaktischen Anforderungen, insbesondere in Bezug auf die Gestaltung problemhaltiger Aufgaben, scheinen besonders die Punkte zum *Aufbau und zur Inhaltlichkeit des fachspezifischen Lehrplans*, zur *Fachdidaktischen Konzeptionen und Lehr-Lern-Forschung*, zu *Schulbücher/Unterrichtsmaterialien/Informationstechnologien im Fach* sowie zum fachspezifischen *zum Lernen und Lernschwierigkeiten von Schülern*, zur *Leistungsbeurteilung und Lernförderung* sowie zu *Methodischen Formen/Lehr-Lern-Formen* (Terhart, 2002, S. 34) relevant. Zur Definition der Fachdidaktik nutzt Terhart (2002) in der Taxonomiestufe höher einzuordnenden Operatoren, in dem er die Anforderung an Absolvent*innen stellt, „*fachdidaktisch analysieren*

¹⁹ Es wird nicht ausgeschlossen, dass auch in anderen Punkten relevante Inhalte vorkommen. Dies gilt auch in Bezug auf die Punkte zur Fachdidaktik. Terhart spricht (2002) „die Problematik der Leistungsbeurteilung“ ebenso unter dem „Bereich des *erziehungswissenschaftlichen und schulpädagogisch-didaktischen Wissens* [Herv. i. Original]“ (S. 32) an.

und argumentieren [zu] können [Herv. i. Original]“ (S. 32). Auch in Bezug auf die Fachdidaktik schreibt der Autor, dass „[d]ies ... am Ende der ersten Phase noch nicht vollständig unterrichtspraktisch durchdekliniert sein [kann]“ (Terhart, 2002, S. 32). In der ersten Phase der Lehrer*innenbildung werden demnach die Grundlagen in den verschiedenen Bereichen gelegt.

Für einen speziellen Blick auf die Ausbildung von (Berufs- und) Wirtschaftspädagog*innen²⁰ in der ersten Phase der Lehrer*innenbildung wurden die Basiscurricula der Jahre 2003 und 2014²¹ herangezogen. Die entwickelten Basiscurricula haben zwar keinen rechtsverbindlichen Charakter, bieten jedoch etwa den Vorteil, Studiengänge danach auszurichten und ein vergleichbares Niveau zu sichern (vgl. hierzu auch Dubs, 2008, S. 15-16)²². In den Basiscurricula (2003, 2014) finden sich Ansätze dazu, wie Wirtschaftspädagog*innen zum Erstellen und Bewerten von Assessments²³ vorbereitet werden bzw. Grundlagen hierzu erwerben sollen. Laut der Kommission der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (2014) gehören dazu folgende Punkte:

- „Strukturen, Institutionen, Rahmenbedingungen der beruflichen Bildung und deren Paradigmen, Theorien und Modellierungen sowie auf berufliche Schul- und Organisationsentwicklung, Lernortkooperation, berufliche Ordnungsmittel, Zertifizierungs- und Prüfungswesen – auch im internationalen Vergleich (*Strukturen/Bildungspolitik/Bildungsmanagement im beruflichen Bereich*) [Herv. i. Original]“ (S. 7)
- „individuelle Lernergebnisse (Diagnose) ... (*Assessment im beruflichen Bereich*) [Herv. i. Original]“ (S. 7)
- „Assessment (inkl. formatives und summatives Assessment, Coaching, Kursevaluationen in der beruflichen Bildung)“ (S. 11)

Ähnliche Kriterien finden sich im Basiscurriculum aus dem Jahr 2003 (Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2003). Die „Entwicklung der Rahmenbedingungen für die berufliche Bildung (Schul- und Organisationsentwicklung, Lernortkooperation, Ordnungsmittel, Zertifikations- und Prüfungswesen)“ (S. 6) ist eine der von Wirtschaftspädagog*innen zu erwerbenden Kompetenzen. Zu den Lehrinhalten

²⁰ Die genannten Basiscurricula beziehen sich auf die Ausbildung von Berufs- und Wirtschaftspädagog*innen. In der Arbeit werden Wirtschaftspädagog*innen in den Blick genommen, Berufspädagog*innen nicht. Deshalb werden Berufspädagog*innen ausgeklammert und im Folgenden nicht aufgeführt.

²¹ In diesem wurde u. a. der Beschluss der KMK berücksichtigt Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (2014, S. 2).

²² Um den Fokus der Arbeit nicht zu verlieren, wird an dieser Stelle nicht weiter auf die Ideen und Argumente für das Basis- bzw. Kerncurriculum eingegangen. Hierzu sei beispielsweise auf den Beitrag von Beck (2002) verwiesen.

²³ Auf die Unterscheidung von Assessments wird in Kapitel 2.4.3 eingegangen.

der Ausbildung zählen zudem die Themenfelder „Zensuren und Zeugnisse, Beurteilungen: schriftliche, mündliche, handlungsorientierte Prüfungen“ (S. 17). Auch gehören „Prüfungen, Zertifizierung und Durchlässigkeit im (beruflichen) Bildungswesen; Muster beruflicher Bildungskarrieren“ (S. 19) dazu.

Auf Grundlage der Curricula und den genannten Standards der KMK ist davon auszugehen, dass (angehende) Lehrkräfte an kaufmännisch-beruflichen Schulen, speziell Wirtschaftspädagog*innen, in ihrer Ausbildung auf die relevanten Anforderungen vorbereitet und in ihren Kompetenzen gefördert werden. Im Speziellen kann davon ausgegangen werden, dass dies auch in Bezug auf die Erstellung von kompetenzorientierten Aufgaben zu Assessmentzwecken gilt und Absolvent*innen „kriteriengeleitet[e]“ und „adressatengerecht[e]“ (Sekretariat der KMK, 2019, S. 12) Aufgaben erstellen können.

Zum Studium wird betont, dass Studierende „unter Rückgriff auf Theorien und empirische Befunde Probleme bearbeiten und erste theoriegeleitete Praxiserfahrungen sammeln [sowie] ... ein Kompetenzprofil [entwickeln], das es ihnen ermöglicht, wechselnde Situationen der vielfältigen komplexen Tätigkeitsfelder kompetent und kritisch wahrzunehmen, selbstbestimmt und professionell zu handeln, Verantwortung für eigenes Handeln und für die Gemeinschaft zu übernehmen sowie unter einer ethisch legitimierten und selbstreflexiv kontrollierten Perspektive Urteile zu bilden und entsprechende individuelle sowie gemeinschaftliche Entscheidungen zu treffen“ (Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2014, S. 5). Die Entwicklung wird allerdings durch das Studium nicht abgeschlossen, vielmehr werden hierin Grundlagen gesetzt (Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2014, S. 7-8; vgl. hierzu auch Artelt & Kunter, 2019, S. 407, 409; Kunter, Kleickmann, Klusmann & Richter, 2011, S. 60-61). Anlehnend an die Beschreibung von Siegfried (2017, S. 26) kann die erste Phase der Lehrer*innenbildung als die, wenn auch nicht ausschließliche, Phase verstanden werden, in der formelle Lerngelegenheiten zum Wissenserwerb bzw. -aufbau bestehen. Zwar bezieht sich die Autorin dabei auf fachliches und fachdidaktisches Wissen, doch kann dieses Verständnis auch auf andere Wissens- bzw. Kompetenzfacetten erweitert werden, wie etwa das hier betrachtete pädagogisch-psychologische bzw. allgemein-pädagogische Wissen (Baumert & Kunter, 2006, S. 485, 2011, S. 32; Shulman, 1986, S. 9-10).

*Zur zweiten Phase der Lehrer*innenbildung*

Die hier betrachteten Kompetenzen finden sich in den Standards für die zweite Phase der Lehrer*innenbildung im Punkt *Beurteilung, Diagnose und Förderung* (Terhart, 2002, S. 36). Fachdidaktische Anforderungen, die für das problemhaltige Gestalten einer Aufgabe relevant sind, sind den Punkten *Einsatz eines breiten Methodenrepertoire* und *Einsatz neuer Informationstechnologien* (Terhart, 2002, S. 36) zuzuordnen. Da allerdings die einzelnen Punkte nicht im Detail beschrieben werden, kann hierzu nicht abschließend beantwortet werden, was genau unter den formulierten Punkten zu verstehen ist. Zudem kann nicht ausgeschlossen werden, dass die hier betrachteten Anforderungen sich nicht auch in anderen Punkten wiederfinden²⁴. Neben formellen Lerngelegenheiten bietet die zweite Phase auch *informelle Lerngelegenheiten*, so Siegfried (2017, S. 26). Dabei wird das *unterrichtspraktische Wissen* fokussiert. Vorrangiges Ziel dieser Phase ist eine „Handlungssicherheit bei den angehenden Lehrkräften zu erzeugen“, indem sie im Prozess „professionelle Begleitung“ (Artelt & Kunter, 2019, S. 407) erfahren.

*Zu beiden Phasen der Lehrer*innenbildung*

Das Sekretariat der KMK (2019, S. 2) setzt Standards²⁵ fest, die in allen drei Phasen der Lehrer*innenbildung, also auch Studium und Referendariat, berücksichtigt werden sollen. In diesen werden Kompetenzen adressiert, die verstanden werden als „Fähigkeiten, Fertigkeiten und Einstellungen, über die eine Lehrkraft zur Bewältigung der beruflichen Anforderungen verfügen soll“ (Sekretariat der KMK, 2019, S. 4). In den Standards wird die Relevanz „pädagogisch-psychologische[r] und diagnostische[r] Kompetenzen von Lehrkräften“ (Sekretariat der KMK, 2019, S. 3; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004, S. 3) verdeutlicht (vgl. hierzu auch Sekretariat der KMK, 2019, S. 5). Im Kompetenzbereich *Beurteilen* wird auf die Kompetenz der Lehrkräfte zum Diagnostizieren von Lernvoraussetzungen und -prozessen sowie das Erfassen von Leistungsentwicklungen der Schüler*innen eingegangen. Auch das Beurteilen von Lernprozessen sowie von Leistungen zählen dazu. Als weitere Bestandteile der Kompetenz wird zudem die Förderung und Beratung, sowohl von Schüler*innen als auch ihrer Eltern, genannt (Sekretariat der KMK, 2019, S. 11-12; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004, S. 11). Unterteilt in Ausbildungsabschnitten, werden die Standards zusammengefasst in Tabelle 1 dargestellt:

²⁴ Zwar erläutert Terhart (2002, S. 31-33) die Bereiche, durch die eine Abgrenzung möglich ist, jedoch werden die einzelnen aufgeführten Punkte zu den Bereichen kaum erläutert.

²⁵ In den hier zitierten Werken der KMK (2004, 2019) wird auf die Standards in den Bildungswissenschaften eingegangen.

Tabelle 1:

*Standards für die theoretischen und praktischen Ausbildungsabschnitte für die Ausbildung von Lehrer*innen*

Standards für die...	
...theoretischen Ausbildungsabschnitte	...praktischen Ausbildungsabschnitte
Die Absolvent*innen	
<ul style="list-style-type: none"> • kennen Begriff und Merkmale von Heterogenität bzw. Diversität. • wissen um die Vielfalt von Einflussfaktoren auf Lernprozesse und deren Auswirkungen auf Lernleistungen. • wissen, wie unterschiedliche Lernvoraussetzungen Lehren und Lernen beeinflussen und wie diese im Unterricht in heterogenen Lerngruppen positiv nutzbar gemacht werden können. • kennen Formen von Hoch- und Sonderbegabung. • kennen die Grundlagen und Formen der Lernprozessdiagnostik einschließlich technischer Realisierungen, deren Möglichkeiten und Grenzen sowie datenschutzrechtlicher Bestimmungen. • kennen Prinzipien und Ansätze der Beratung von Schüler*innen sowie Eltern. • kennen die für Schule relevanten unterschiedlichen Kooperationspartner und wissen um die unterschiedlichen Perspektiven bei der Kooperation mit anderen Professionen und Einrichtungen. • kennen unterschiedliche Formen und Wirkungen der Leistungsbeurteilung und -rückmeldung, ihre Funktionen und ihre Vor- und Nachteile. • kennen verschiedene Bezugssysteme der Leistungsbeurteilung und wägen sie gegeneinander ab. • kennen das Spannungsverhältnis von lernförderlicher Rückmeldung und gesellschaftlicher Funktionen von Leistungsbeurteilungen. • kennen Potenziale und Grenzen technologiebasierter Aufgaben- und Prüfungsformate. 	<ul style="list-style-type: none"> • erkennen Entwicklungsstände, Lernpotenziale, Lernhindernisse und Lernfortschritte. • erkennen individuelle Lernausgangslagen und setzen spezielle Fördermöglichkeiten ein. • erkennen Begabungen und kennen Möglichkeiten der Begabungsförderung. • stimmen Lernmöglichkeiten und Lernanforderungen aufeinander ab. • setzen unterschiedliche Beratungsformen situationsgerecht ein und unterscheiden Beratungsfunktion und Beurteilungsfunktion. • kooperieren bei der Diagnostik, Förderung und Beratung inner- und außerschulisch mit Kolleg*innen sowie mit anderen Professionen und Einrichtungen. • nutzen digitale Lernprozessdiagnostik im Wissen um ihre Möglichkeiten und Grenzen, auch zur Reflexion über die eigene Unterrichtstätigkeit. • konzipieren Aufgabenstellungen kriteriengeleitet und formulieren sie adressatengerecht. • wenden Bewertungsmodelle und Bewertungsmaßstäbe fach- und situationsgerecht an. • verständigen sich auf Beurteilungsgrundsätze mit Kolleg*innen. • begründen Bewertungen und Beurteilungen adressatengerecht und zeigen Perspektiven für das weitere Lernen auf. • nutzen reflektiert digitale Möglichkeiten zur Unterstützung der Leistungserfassung, -feststellung und -bewertung. • nutzen Leistungsüberprüfungen als konstruktive Rückmeldung über die eigene Unterrichtstätigkeit.

(Eigene Darstellung²⁶ aus Kompetenz 7 und 8 in Anlehnung an Sekretariat der KMK, 2019, S. 11-12)

²⁶ Bei der Zusammenstellung wurde nur das Dokument der KMK (2019) genutzt, da diese die Neufassung des Dokuments der KMK (2004) ist. Demzufolge finden sich in dem Dokument gleiche, ähnliche und/oder erweiterte Punkte, die zu Redundanzen führen würden. In der Tabelle wurde mit einem „*“ gegendert, statt, wie im ursprünglichen Dokument, die weibliche und männliche Form auszuschreiben.

Bei der Betrachtung der Anforderungen der Standards fällt auf, dass die im Bereich des Beurteilens genannten Kompetenzen diverse diagnostische Tätigkeiten zu unterschiedlichen Zwecken umfassen. Hierzu gehört auch das Diagnostizieren, was im Speziellen auch das Erstellen von Aufgaben beinhaltet und worauf im Rahmen dieser Arbeit der Fokus gelegt wird. Auch gehört das Diagnostizieren von Lernvoraussetzungen und -prozessen sowie von Lernergebnissen/Leistungen dazu. Der Darstellung des KMK (2019, S. 11-12) zufolge, ist das Entwickeln von Aufgaben Bestandteil des Kompetenzbereichs des Beurteilens und somit auch Teil diagnostischer Kompetenz²⁷.

Die Ansätze der Fachdidaktik finden sich in den in Tabelle 1 gezeigten Bereichen wieder. An anderen Stellen finden sich ebenfalls Aspekte hierzu, auch wenn sie sich auf die Unterrichtsgestaltung bzw. die Gestaltung des Lehr-Lern-Prozesses beziehen (vgl. hierzu auch Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004, S. 8). Da Absolvent*innen Aufgaben und (weitere) Materialien konzipieren können sollen (Sekretariat der KMK, 2019, S. 8, 12), können Synergieeffekte aus den Kompetenzen angenommen werden. Zu begründen ist dies damit, dass die notwendigen Kompetenzen hierfür ebenso wichtig als auch übertragbar auf die Erstellung von Aufgaben zu Zwecken des summarischen Assessments sind.

Das Zusammenspiel der ersten und zweiten Phase definiert die KMK (2019) wie folgt: „Ausgehend von dem Schwerpunkt Theorie erschließt die erste Phase die pädagogische Praxis, während in der zweiten Phase diese Praxis und deren theoriegeleitete Reflexion im Zentrum stehen“ (S. 4). Dies bestärkt die Erkenntnisse aus vorherigen Dokumenten, dass die erste Phase der Lehrer*innenbildung den Fokus auf die Theorie legt und somit elementare Grundlagen vermittelt, während insbesondere in der zweiten Phase die Praxis erfolgt. So kann angenommen werden, dass die Verknüpfung von Theorie und Praxis es den angehenden Lehrkräften ermöglicht, das erworbene Wissen in ihren Handlungen umzusetzen (vgl. Dubs, 2008, S. 13-14).

²⁷ Dies ist damit zu begründen, dass in den Standards des Sekretariats der KMK (2019) zur „*Beurteilungs- und Beratungsaufgabe* [Herv. i. Original]“ der Lehrer*innen „hohe pädagogisch-psychologische und diagnostische Kompetenzen“ (S. 3) als notwendig erachten. Die zusammengeführten Kompetenzen 7 und 8 werden zum Kompetenzbereich des *Beurteilens* gefasst (Sekretariats der KMK, 2019, S. 11-12).

2.3 Praktisch begründete Anforderung an (angehende) Lehrkräfte

Als eine der Aufgaben von Lehrkräften wird die Aufgabe zum *Beurteilen und Beraten* definiert (Sekretariat der KMK, 2019, S. 3). Hierzu gehört die faire und verantwortungsvolle Handlungskompetenz zu Selektionszwecken²⁸. Damit ist die Aufgabe der Lehrkräfte gemeint, Klausuren bzw. Klassenarbeiten²⁹ zu erstellen, Leistungen bzw. Lernergebnisse festzustellen sowie Schüler*innenleistungen zu bewerten und zu beurteilen (vgl. z. B. Besser, Rau & Traum, 2022, S. 57; Dubs, 2008, S. 17; Lankes, 2008, S. 41; Schwaighofer, Heene & Bühner, 2019, S. 472). Aufgaben zu Zwecken des (summativen) Assessments zu erstellen, geht über die eigene Schüler*innenschaft der Klasse bzw. Klausuren hinaus (Besser et al., 2022, S. 57). So sind Lehrkräfte berufsbildender Schulen auch an der Erstellung von kaufmännischen Prüfungsaufgaben beteiligt (Aka, 2022; ZPA Nord-West, 2021; vgl. Kapitel 3.2). Diese Prüfungen sollen bestimmten Güte- und Qualitätskriterien entsprechen (Objektivität, Validität, Reliabilität, Ökonomie, Transparenz, Justiziabilität, Effizienz; vgl. Aka, 2023b; ZPA Nord-West, 2021). Der Anspruch an Güte- und Qualitätskriterien ist vor dem Hintergrund, dass Noten bzw. Zensuren nicht unbedeutend für den weiteren Werdegang und gar die Stellung in der Gesellschaft sind (Sauer & Gamsjäger, 2010, S. 651; Tent & Birkel, 2010, S. 950), nachvollziehbar.

Das Handeln einer Lehrkraft ist nicht willkürlich, sondern wird durch Inputvariable, z. B. Lehrpläne, normativ, sozial, organisatorisch und rechtlich reguliert (Baumert & Kunter, 2006, S. 472; Halbheer & Reusser, 2008, S. 255). Dies gilt auch für Prüfungen, die zum Überprüfen beruflicher Handlungskompetenz Einsatz finden (Aka, 2023a; ZPA Nord-West, 2021; § 38 BBiG). Die Problemlösekompetenz ist Bestandteil der Handlungskompetenz³⁰ und die Prüfung dieser soll in der Ausbildung erfolgen³¹. Schließlich werden im Rahmen der Ausbildung „für die Ausübung einer qualifizierten beruflichen Tätigkeit in einer sich wandelnden Ar-

²⁸ Konkret wird hierbei von davon gesprochen, dass Lehrkräfte „ihre Beurteilungs- und Beratungsaufgabe im Unterricht und bei der Vergabe von Berechtigungen für Ausbildungs- und Berufswege kompetent, gerecht und verantwortungsbewusst aus[üben]“ (Sekretariat der KMK, 2019, S. 3).

²⁹ Zur besseren Lesbarkeit wird im Folgenden von *Klausuren* gesprochen. Da allerdings unterschiedliche Bezeichnungen genutzt werden, ist nicht auszuschließen, dass stellenweise auch von Klassenarbeiten gesprochen wird, insbesondere von Interviewteilnehmer*innen (vgl. Kapitel 5.1.3, 5.2.1 und 8.1).

³⁰ Auf die Einordnung der Problemlösekompetenz, auch im Zusammenhang mit der Handlungskompetenz, wird in Kapitel 2.5.5 eingegangen.

³¹ Im Eingangskapitel wurde bereits auf die Relevanz der Problemlösekompetenz im beruflichen Alltag eingegangen. Auf eine Wiederholung wird verzichtet.

beitswelt notwendigen beruflichen Fertigkeiten, Kenntnisse und Fähigkeiten (berufliche Handlungsfähigkeit)“ (§ 1 Abs. 3 BBiG) vermittelt. Entsprechend müssen Aufgaben³² (fachdidaktisch) so gestaltet werden, dass sie Probleme abbilden und somit die Kompetenz valide messen.

2.4 Pädagogische Diagnostik als Teil allgemein-pädagogischer und fachdidaktischer Kompetenz

2.4.1 Begriffsverständnis und Funktion

Aktuell ist es kaum möglich, von einem bestimmten Konstrukt diagnostischer Kompetenz zu sprechen (Schrader, 2013, S. 159; Spinath, 2005, S. 93). Vielmehr sind es Definitionen, die den Begriff *diagnostische Kompetenz* beschreiben (vgl. hierzu z. B. Spinath, 2005, S. 88). Nach dem Kompetenzmodell von Baumert und Kunter (2011, S. 32) bzw. Krauss et al. (2004, S. 35, 43) ist diagnostische Kompetenz bzw. das Wissen und Können zur pädagogischen Diagnostik Bestandteil des fachdidaktischen Wissens und der Allgemeinpädagogik. In diesem und den nachfolgenden Kapiteln wird auf die Definition und die für diese Arbeit ausgewählten Bestandteile eingegangen^{33,34}, wobei der Fokus auf den *Grundlagen der Diagnostik* (vgl. Abbildung 2) liegt. Diese werden genutzt, um kritisch auf die Definition pädagogischer Diagnostik und die dazugehörigen Tätigkeiten einzugehen.

Laut Ingenkamp und Lissmann (2008) umfasst pädagogische Diagnostik

... alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren. Zur Pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen sowie die mehr gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen zum Ziel haben. (S. 13)

³² Im Sinne der Curriculum-Instruktion-Assessment-Triade von Pellegrino (2010), auf die eingangs eingegangen wurde, sind, entsprechend den curricularen Anforderungen neben dem Assessment auch die Instruktion so zu gestalten, dass diese stimmig sind. In dem hier betrachteten konkreten Fall bedeutet das, dass die curricular geforderte Problemlösekompetenz in der Instruktion gefördert und im Assessment geprüft/diagnostiziert werden.

³³ Im Rahmen der Arbeit wird nicht auf die Diskussion zur pädagogischen versus psychologischen Diagnostik eingegangen, um den Schwerpunkt beizubehalten. Hierzu kann beispielsweise in Leutner (2010, S. 624-625) nachgelesen werden.

³⁴ Im Unterschied zur pädagogischen Diagnostik geht es in der pädagogischen Forschung darum, „allgemeinere Erkenntnisse z. B. über die Angemessenheit bestimmter didaktischer Vorgehensweisen, Medien usw. für Schüler mit bestimmten Merkmalen zu gewinnen“ (Ingenkamp & Lissmann, 2008, S. 14). Auf die Vertiefung der Unterscheidung wird verzichtet, um den Fokus der Arbeit zu wahren.

Dieses Verständnis von diagnostischen Tätigkeiten der Lehrkräfte überschneidet sich mit der Zielformulierung von Schrader (2013, S. 154). Er beschreibt diese als das Informationsgewinnung über Lernergebnisse, -voraussetzungen und -vorgängen der Schüler*innen, welche für diverse pädagogische Entscheidungen relevant sind. Zu den pädagogischen Entscheidungen zählen beispielsweise Noten sowie das Planen von Unterricht (vgl. hierzu auch van Ophuysen & Behrmann, 2015, S. 86-87). Diagnostische Kompetenz definiert Schrader (2010, S. 102) hingegen als Fähigkeit zur zutreffenden Beurteilung einer anderen Person. Wie dies geschieht bzw. wie das Instrument zum Beurteilen erstellt bzw. genutzt werden soll, scheint kaum Beachtung zu finden. Nach Hascher (2008, S. 74) wird in der pädagogischen Diagnostik auf das Lernen sowie das Lernergebnis fokussiert. Diagnostische Tätigkeiten sind nach Ingenkamp und Lissmann (2008, S. 13) Beobachtungen und Befragungen, deren Ergebnisse interpretiert und anschließend berichtet werden. Ziel sei eine Verhaltensbeschreibung, -vorhersagung sowie Erläuterung der Verhaltensgründe (Ingenkamp & Lissmann, 2008, S. 13). Welche Instrumente bei der Beobachtung und Befragung eingesetzt werden, ist nicht festgelegt. Vielmehr gehen die Autoren darauf ein, dass die wissenschaftlichen Gütekriterien dabei beachtet werden sollen. Lissmann (2010b, S. 737) schreibt, dass Lehrkräfte die Instrumente nutzen können sollen, wobei unklar ist, ob bei der Nutzung auch die Erstellung der Aufgaben zu zählen ist. Insgesamt findet die Instrumentenentwicklung, um die gesetzten Ziele pädagogischer Diagnostik zu erreichen, in Definitionen sowie in Beschreibungen diagnostischer Tätigkeiten kaum Beachtung. Diese sind jedoch als Informationslieferanten für die pädagogische Diagnostik (Lissmann, 2010b, S. 737) von hoher Relevanz. Von Lehrkräften wird gefordert, dass sie „unterschiedliche Verfahren zur Diagnostik von Schülerkompetenz begründet auswählen, an den ‚richtigen‘ Stellen einsetzen sowie selbst in Anlehnung an dokumentierte Instrumente Aufgaben zur Diagnostik entwickeln und deren Güte hinterfragen können [sollten]“ (Aufschnaiter et al., 2015, S. 739). Auch Schlömerkemper (2014, S. 320) führt u. a. die *Entwicklung von Instrumenten*, unter denen hier auch das Entwickeln von Aufgaben zu Zwecken des (summativen) Assessments³⁵ verstanden wird, als Voraussetzung zur Leistungsmessung auf. Im Unterschied dazu ordnen Baumert und Kunter (2006, S. 489; in Anlehnung an Nagasaki & Becker, 1993; White & Gunstone, 1992; Wilcox & Lanier, 2000) diese Tätigkeiten weniger den diagnostischen Tätigkeiten, sondern vielmehr die Auswahl und Formulierung von Aufgaben bzw. Arbeitsaufträgen der fachdidaktischen Kompetenz/dem fachdidaktischen Können zu. In der Literatur wird von der *aufgabenbezogenen Diagnostik* gesprochen. Dies beschreibt Karst (2012, S. 88) damit,

³⁵ Dass dies auch für formative Assessments wichtig ist, um entsprechend den Voraussetzungen der Lernenden u. a. Aufgaben für der Lehr-Lern-Prozess vorzubereiten, zeigen z. B. Ohle, McElvany, Horz und Ullrich (2015, S. 15-16).

dass entsprechend den Bedingungen der Schüler*innen, im Sinne der Vorwissensanknüpfung, u. a. passende Aufgaben ausgewählt werden. Dabei bezieht sie sich auf den Unterricht. Dieses Verständnis kann auf Aufgaben zum (formativen³⁶ und summativen) Assessment³⁷, zumindest für durch Kammern organisierte Prüfungen, übertragen werden. Es ist jedoch zu beachten, dass sich Lern- und Prüfungsaufgaben unterscheiden, worauf in den folgenden Kapiteln eingegangen wird – insbesondere mit Blick auf die Bezugsnorm³⁸. Auch Helmke (2021, S. 132) unterscheidet bei den diagnostischen Leistungen von Lehrkräften zwischen Personen- und Aufgabenmerkmalen. Zu Letzteren führt er als Beispiel den Schwierigkeitsgrad einer Aufgabe auf. Im Gegensatz zu Karst (2012, S. 88) beschränkt Helmke (2021, S. 132) die Diagnostik von u. a. Aufgabenmerkmalen nicht auf den Unterricht, wodurch die vorher aufgeführte Annahme zur Übertragung des Verständnisses für Aufgaben zu Assessmentzwecken bestärkt wird. Ein umfassenderes Verständnis pädagogischer Diagnostik liefert Weinert (2000, S. 16) und beschreibt diese als „ein Bündel von Fähigkeiten, um den Kenntnisstand, die Lernfortschritte und die Leistungsprobleme der einzelnen Schüler sowie die Schwierigkeiten verschiedener Lernaufgaben im Unterricht fortlaufend beurteilen zu können, sodass das didaktische Handeln auf diagnostischen Einsichten aufgebaut werden kann“. Er führt ein Beispiel auf, das sich auf eine von der Lehrkraft erstellte Prüfungsaufgabe bezieht (Weinert, 2000, S. 16) und zeigt damit, dass sich die Kompetenz nicht nur auf *Lernaufgaben* beschränkt, sondern auch Aufgaben zum summativen Assessment einschließt – entgegen der Definition.

2.4.2 Bewerten und Beurteilen

Leistungsbeurteilung wird beschrieben als ein nachgelagerter Vorgang zur Leistungsmessung³⁹ bzw. -feststellung, in dem die Interpretation des Ergebnisses der Leistungsmessung bzw. -feststellung erfolgt (Ministerium für Kultus, Jugend und Sport Baden-Württemberg & Landesinstitut für Schulentwicklung, 2016, S. 5). Maier (2015, S. 16) zufolge umfasst Leistungsbeurteilung bzw. -bewertung oft auch die Leistungsmessung, wodurch die gewonnenen Werte bzw. Daten nach Referenzmaßstäben eingeordnet, interpretiert und bewertet werden. Zur

³⁶ Es ist gewollt, hier von beiden Assessments, dem formativen und summativen Assessment, zu sprechen, im Bewusstsein dessen, dass diese auch für den Unterricht relevant werden können und somit nicht gänzlich im Widerspruch zur Unterrichtsbezug von Karst (2012, S. 88) stehen. Auf die genaue Differenzierung der Assessmentformen wird in Kapitel 2.4.3 eingegangen.

³⁷ Es gibt Hinweise darauf, dass *pädagogische Diagnostik* als Synonym für *Assessment* verstanden wird (vgl. z. B. Wilbers, 2019b, S. 251).

³⁸ Mit der Orientierung am Vorwissen der Lernenden (Karst, 2012, S. 88) wird die individuelle bzw. soziale Bezugsnorm angesprochen. Im Unterschied hierzu folgen Prüfungsaufgaben der kriterialen Bezugsnorm (vgl. hierzu Kapitel 2.4.3).

³⁹ Auf die Begrifflichkeit wird in Kapitel 2.4.4 eingegangen.

Beurteilung können Hilfen in Form von beispielsweise Checklisten und Kriterienkatalogen genutzt werden (Wilbers, 2019b, S. 267).

Die Grundlage diagnostischer Urteile im Unterricht stellen Interpretationen von Beobachtungen schriftlicher bzw. mündlicher Leistungen dar (Schrader & Helmke, 2014, S. 47-49). Lissmann (2010a, S. 6) beschreibt Leistungsbeurteilung als das deskriptive Vorgehen und Bewertung der Leistung. Ergänzend beinhaltet die Beurteilung von Leistungen mehrere Bewertungen (Jürgens & Lissmann, 2015, S. 69). Somit ist zunächst unklar, wo die Unterschiede zwischen Bewertung und Beurteilung liegen. Klein (2009, S. 19) hingegen definiert *Beurteilung* als „eine abwägende, komplexe Stellungnahme, bei der verschiedene Sichtweisen und Bezüge zu einer Meinungsbildung führen, ohne dabei zwangsläufig Wertmaßstäbe und Rangskalen aufzustellen“. *Bewertung* wird als „eine Beurteilung, die eine eindeutig feststehende Bezugsnorm braucht und einen Werte-Maßstab erfordert, der Zuordnungsmöglichkeiten bieten, wie z. B. Zensurenkala bei der Leistungsbewertung“ (Klein, 2009, S. 19) beschrieben. Dabei wird davon ausgegangen, dass die Beurteilung von Leistungen vor der Bewertung dieser stattfindet. Bei der Bewertung geht es darum, das Ergebnis (einer Leistung) zu beziffern, etwa in Form einer Note (Klein, 2009, S. 19-21). Dieses Verständnis wird auch von Jürgens und Lissmann (2015, S. 69) gestützt. Wird die Definition der Autoren verglichen, fällt auf, dass Lissmann (2010a, S. 6) zur Definition der *Beurteilung* das Wort *Bewertung* nutzt, Klein (2009, S. 19) das Gegenteil. Daraus kann geschlossen werden, dass die Begriffe *Beurteilung* und *Bewertung* nicht grundsätzlich als unterschiedlich verstanden werden und sich nicht gegenseitig ausschließen; sie können sogar einander bedingen.

Folglich werden die Begriffe *Bewertung* und *Beurteilung* nicht immer eindeutig genutzt. Trotz der Unterschiede in den Definitionen der Begrifflichkeiten lassen sich Autor*innen finden, die darauf verweisen, die Begriffe synonym zu verwenden (z. B. Jürgens & Lissmann, 2015, S. 69; Lissmann, 2010a, S. 7). Dies kann mit der Übereinstimmung der Aufgaben in großen Teilen erklärt werden (Jürgens & Lissmann, 2015, S. 69). In der vorliegenden Arbeit wird dem gefolgt und auf eine strikte Unterscheidung verzichtet. Begründet wird es damit, dass die nicht immer einheitliche (im Sinne der Definition) Nutzung der Begriffe bei der Literaturrecherche einen permanenten Abgleich erfordern würde, um Verzerrungen zu vermeiden. Dies würde jedoch voraussetzen, dass jede*r Autor*in zunächst eine Definition vornimmt. Der durchgeführte Abgleich würde zu schwer verständlichen Texten führen und den Lesefluss behindern. Ein solches Vorgehen scheint zudem nicht zielführend, da sich die Begriffe im Wesentlichen kaum vonei-

inander unterscheiden und sich sogar in den Definitionen überschneiden. *Bewertung* und *Beurteilung* stehen nicht im Fokus der vorliegenden Arbeit, sondern das Erstellen von Aufgaben – auch wenn dies nicht unabhängig voneinander ist.

2.4.3 Formen des Assessments

Assessment wird unterteilt in *formatives* und *summatives Assessment* (z. B. Aufschnaiter et al., 2015, S. 741-742; Jürgens & Lissmann, 2015, S. 52-53; Wilbers, 2018, S. 241, 2019b, S. 252). Durch das *formative Assessment* ist es möglich, den Stand im (Lehr-)Lernprozess zu erfassen und das Lernangebot pädagogisch und didaktisch anzupassen (Aufschnaiter et al., 2015, S. 741; Jürgens & Lissmann, 2015, S. 50-53; Maier, 2010, S. 294; Wilbers, 2018, S. 241). Es wird auch als *Assessment for learning* (Maier, 2012, S. 72; Wilbers, 2018, S. 241, 2019b, S. 252) und *Prozessdiagnostik* (Jürgens & Lissmann, 2015, S. 53) bezeichnet. *Prozessdiagnostik* wird von Wild und Krapp (2006, S. 532) als *handlungsbegleitender Diagnostik* vorgestellt, in der z. B. Informationen zu Veränderungen gewonnen werden. Dieses Verständnis überschneidet sich mit der *unterrichtsbezogenen Diagnoseleistung*, wonach die *Adaptivität* im Vordergrund steht, durch welche der Unterricht an die Lernvoraussetzungen der Lernenden angepasst wird (Schradler, 2013, S. 155). Die Diagnostik geschieht also im Lehr-Lern-Prozess. Zum Prozess zählen dabei (die Nutzung von) Lerngelegenheiten, Schulkultur sowie Lehr- und Lernpraktiken (Halbheer und Reusser, 2008, S. 255). Die Evaluation auf Prozessebene geschieht demnach in Bezug auf die Qualität von Schule und Unterricht (in Anlehnung an Ditton, 2000, S. 79, vgl. Abbildung 3), wobei die Angebotsperspektive um die der Nachfrage erweitert wird (in Anlehnung an Halbheer & Reusser, 2008, S. 254).

Summatives Assessment findet im Gegensatz zum *formativen Assessment* am Ende bzw. zum Abschluss eines Lehr-Lern-Prozesses statt (Jürgens & Lissmann, 2015, S. 53; Knight & Yorke, 2003, S. 16; Maier, 2010, S. 299; Winther, 2006, S. 93) und wird als *Assessment of learning*⁴⁰ (Wilbers, 2018, S. 241) sowie *Ergebnis- oder Produktdiagnostik* (Jürgens & Lissmann, 2015, S. 53) bezeichnet. Es setzt den Fokus auf den Lernoutput, wodurch das Kompetenzniveau nach Bildungsabschluss erfasst werden soll (Winther, 2006, S. 93). *Summatives Assessment* wird weiter unterteilt in interne und externe Prüfungen. Zu einer internen Prüfung zählen z. B. Klausuren in der Schule. Abschlussprüfungen der Kammern hingegen werden zu externen Prüfungen gezählt (Wilbers, 2018, S. 241, 2019b, S. 251). Typischerweise mündet das Ergebnis des

⁴⁰ Nach Winter (2018, S. 15) kann Assessment of Learning auch im Lernprozess geschehen, nicht nur nach Abschluss einer Lerneinheit. Zur Wahrung der Trennschärfe zwischen *Assessment of Learning* und *Assessemnt for Learning* wird allerdings das engere Verständnis genutzt.

summative Assessments in einem Zeugnis oder Zertifikat (Knight & Yorke, 2003, S. 16). Dieses kann zu Selektionszwecken genutzt werden, etwa zur Bedingungs- oder Personenauswahl (Leutner, 2010, S. 625)⁴¹. Damit können – im Sinne der Abschlussprüfung zum Ende der Berufsausbildung – Personen, die über kein Zertifikat verfügen, von gewissen Positionen oder Bildungsgängen ausgeschlossen werden (Hollmann, Kiedrowski, Lorig & Schürger, 2021, S. 9). Daneben haben Abschlussprüfungen eine Allokationsfunktion (Hollmann et al., 2021, S. 9); sie ermöglichen einen Einstieg in den Beruf und die damit verbundenen beruflichen Möglichkeiten (vgl. auch Klotz, 2015, S. 44).

Nach dem Modell von Ditton (2000, S. 79) sind u. a. Leistungen dem Output zuzuordnen und zählen zum sogenannten *erreichten Curriculum* (vgl. Abbildung 4). „Die erzielten Wirkungen sind ... auf die Intentionen, die verfolgt werden sollen (Faktor 2), zu beziehen“ (Ditton, 2000, S. 80). Hieraus kann geschlossen werden, dass die Zielgröße der Bildung durch das *intendierte Curriculum* bestimmt wird. Auch Schrader (2013, S. 155) schreibt, dass der Soll-Zustand durch Standards zur Bildung, Lehrziele und Curricula gesetzt wird. Als Inputvariablen sind diese sowohl für den schulischen (Halbheer & Reusser, 2008, S. 254-255) als auch für den betrieblichen Teil der Ausbildung bindend. So sind in der beruflichen Bildung nicht nur Vorgaben zum schulischen Teil der Ausbildung zu finden (Rahmenlehrplan), sondern auch für den betrieblichen Teil (Ausbildungsordnung). Weiter wird das Lehrpersonal als *personale Ressource*⁴² zum Input gezählt⁴³ (Halbheer & Reusser, 2008, S. 255; Wilbers, 2019b, S. 165). Diese Ressource führt Ditton (2000, S. 79) als eine Bedingung unter den Voraussetzungen auf (vgl. Abbildung 3).

Der in der Literatur zu findende Begriff *Outcome* ist nicht gleichzusetzen mit *Output*. Während *Output* als das direkte Resultat eines Bildungsprozesses bezeichnet wird, ist unter *Outcome* das langfristige Resultat zu verstehen (Ditton, 2000, S. 79; Ebbinghaus, 2006, S. 33; Klieme & Tippelt, 2008b, S. 8). Frommberger (2013, S. 7) erklärt *Output* als das Resultat des Lehr-Lern-Prozesses und *Outcome* als die private und berufliche Nutzung der Lernergebnisse. Beide Begriffe sind jedoch nicht unabhängig voneinander; sie werden beide als Teil des Ergebnisses eines Bildungsprozesses gesehen (Ditton, 2000, S. 76, 79; Klieme & Tippelt, 2008a, S. 8). Ferner wird von *Outcome* gesprochen, wenn der *Output* reproduziert werden kann (Frommberger, Held, Milolaza, Reinisch & Steib, 2013, S. 133). In der vorliegenden Arbeit liegt der Fokus auf *Output*.

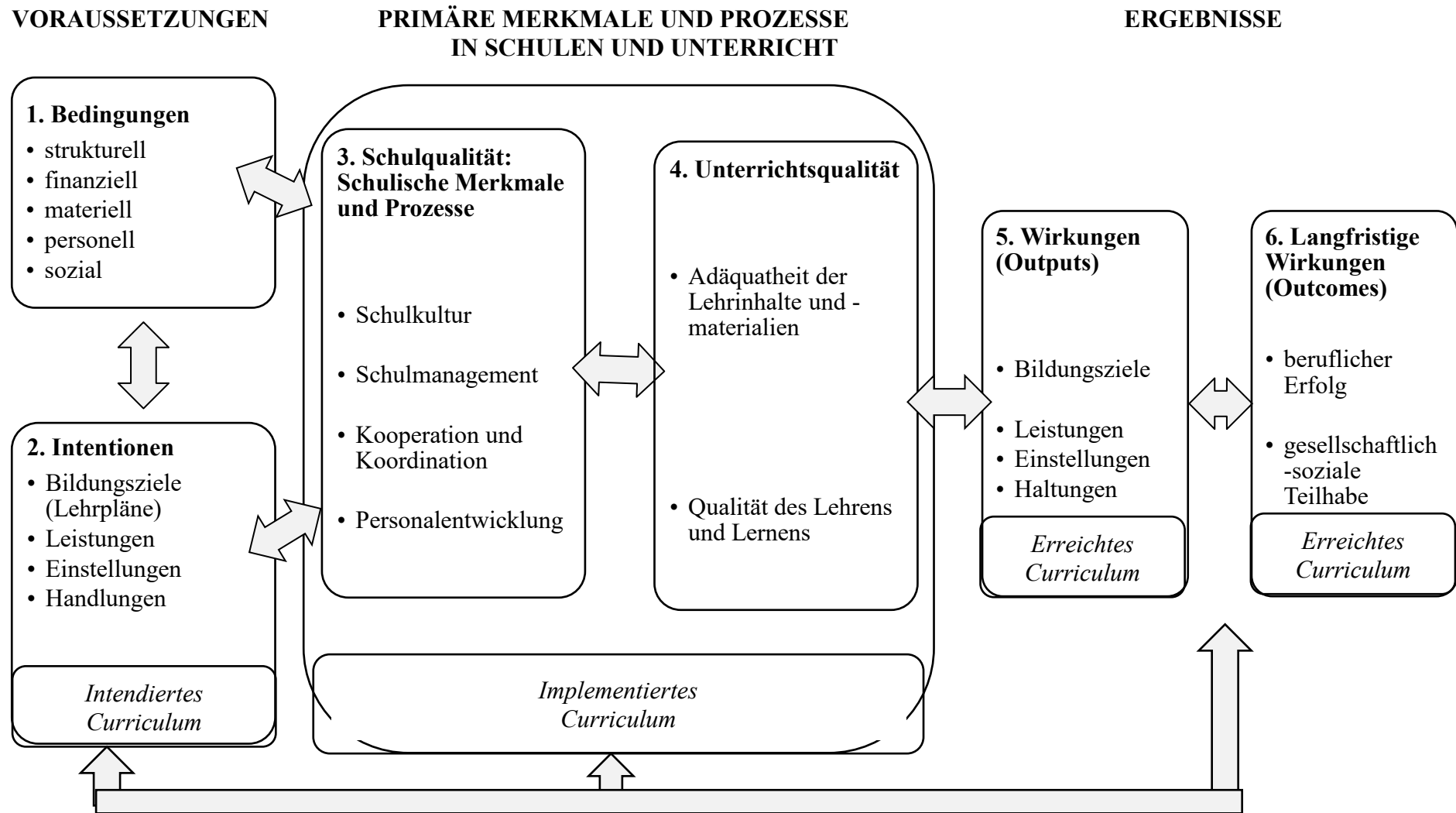
⁴¹ Leutner (2010, S. 625) schreibt hier von einer *Selektions- bzw. Auslesediagnostik*.

⁴² Bezeichnung aus Halbheer und Reusser (2008, S. 255).

⁴³ Dieses Verständnis wird von Kunter (2010) kritisiert, da Lehrkräfte über Input-Variablen hinaus „im Bildungswesen aktiv handelnde Personen“ (S. 307) sind.

Die Qualität des *Outputs* wird über die Leistung bzw. das Resultat des Lehr-Lern-Prozesses festgelegt, etwa in Form einer Zensur bei einer Abschlussprüfung (Büchter, Hahn, Jastrzebski, Kräenbring & Wölk, 2011, o. A.; Ebbinghaus, 2006, S. 33; Windelband, Spöttl & Becker, 2014, S. 301). (Abschluss-)Prüfungen gelten demnach als ein wichtiger Indikator, der Aufschluss über die Outputqualität liefert. Nach Weiß (2011, S. 38) erfolgt dadurch nicht nur die Leistungsmessung von Auszubildenden. Vielmehr lässt das Ergebnis auch Rückschlüsse auf die Leistungsqualität der an der beruflichen Bildung Beteiligten (Institutionen) zu (Weiß, 2011, S. 38).

Abbildung 3:
Qualitätskontrolle und -sicherung in Schule und Unterricht



(Nachzeichnung nach Ditton, 2000, S. 79)

Die Diagnostik des Outputs (*Outputdiagnostik*) kann als eine besondere Form der *Statusdiagnostik* bezeichnet werden. In der *Statusdiagnostik* wird „die Ausprägung der interessierenden Eigenschaft zum gegebenen Zeitpunkt“ (Wilhelm & Kunina-Habenicht, 2021, S. 313; vgl. auch Aufschnaiter et al., 2015, S. 744; Leutner, 2010, S. 626) erhoben. Im berufsschulischen Kontext, insbesondere in der dualen Ausbildung, werden die Lernenden durch ihre Bewerbungen bei Ausbildungsbetrieben selektiert. Da die Ansprüche der Betriebe unterschiedlich sein können, ist von einer heterogenen Gruppe von Lernenden in Berufsschulklassen auszugehen. Die Statusdiagnostik kann eingesetzt werden, um etwa zu Beginn eines neuen Schuljahres zur Bedingungsanalyse das Vorwissen und die Voraussetzungen der Lernenden zu erfassen. Diese erfolgt aber auch dann, wenn eine bestimmte Maßnahme abgeschlossen wird und der Stand zum Ende dieser erhoben wird, um etwa die Teilnehmenden einer weiteren Maßnahme zuzuordnen (vgl. Aufschnaiter et al., 2015, S. 744) oder ein Zertifikat auszustellen, wie es in der Abschlussprüfung der Fall ist. Insofern wird die Statusdiagnostik auch zur Diagnostik des Outputs eingesetzt. Da allerdings Outputdiagnostik – im Gegensatz zur Statusdiagnostik – zu einem bestimmten (vorgegebenen) Zeitpunkt eingesetzt wird, werden die Begriffe nicht synonym verwendet. Dies gilt auch für die Diagnostik zur Bestimmung der Eingangsvoraussetzungen, z. B. Lernvoraussetzungen. Hierfür wird der Begriff der *Eingangsdagnostik* verwendet und stellt ebenfalls eine besondere Form der Statusdiagnostik dar. Während Outputdiagnostik dem summativen Assessment zuzuordnen ist, wird Eingangsdagnostik dem formativen Assessment zugeordnet (vgl. Tabelle 2).

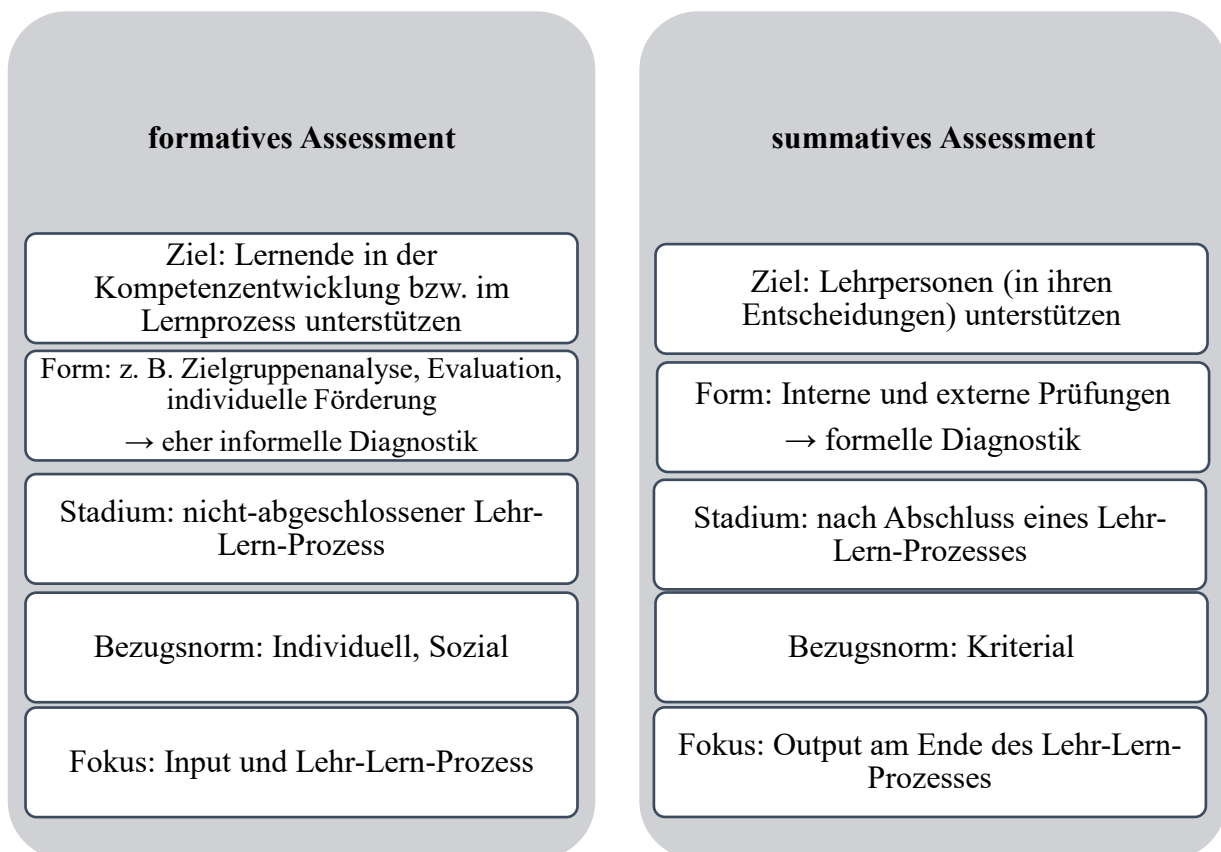
Zur Leistungsbeurteilung werden Bezugsnormen genutzt (Klein, 2009, S. 19; Schrader & Helmke, 2014, S. 50-51). Dem formativen Assessment werden *individuelle Bezugsnormen* zugrunde gelegt (Baumert, 2001, S. 19; Schrader, 2013, S. 156; Winther, 2006, S. 92). Durch die Formulierung individueller Ziele findet eine Differenzierung statt, wobei die Ziele, Standards bzw. Anforderungen den personenbezogenen Gegebenheiten angepasst werden (Jürgens & Lissmann, 2015, S. 55). Des Weiteren ist es möglich, *soziale Bezugsnormen* im formativen Assessment, etwa zu Zwecken der Förderdiagnostik, zu nutzen (Jungjohann & Gebhardt, 2022, S. 27). Dabei erfolgt die Bewertung durch den Vergleich innerhalb einer Gruppe (Baumert, 2001, S. 19). Hingegen sind summative Assessments standardisiert, indem Kriterien als Bezugsnorm genutzt werden (Baumert, 2001, S. 19-21; Schmidt, 2010, S. 20). Das heißt, die in den Curricula formulierten Ziele werden als Kriterien definiert, anhand derer überprüft wird, ob das Ziel nach Ende eines Lehr-Lern-Prozesses erreicht wurde (Jungjohann & Gebhardt, 2022, S. 27-28). Dies wird als *kriteriumsorientierte Diagnostik* bezeichnet (Schwaighofer et al., 2019, S. 475).

Weiterhin wird zwischen *formeller* und *informeller Diagnostik* unterschieden. Formelle Diagnostik wird empfohlen, wenn die anhand des Ergebnisses zu treffende Entscheidung langfristige bzw. erhebliche Folgen nach sich zieht. Hingegen wird für die Lehr-Lern-Prozess beeinflussenden Entscheidungen die informelle Diagnostik empfohlen (Schrader, 2013, S. 161). Insofern ist die Outputdiagnostik der formellen und Prozess- und Eingangsdiagnostik der informellen Diagnostik zuzuordnen.

In Abbildung 4 sind die wesentlichen Merkmale summativer und formativer Assessments zusammengefasst.

Abbildung 4:

Gegenüberstellung der formativen und summativen Assessments



(Eigene Darstellung nach Aufschnaiter et al., 2015, S. 741-744; Baumert, 2001, S. 19-21; Büchter et al., 2011, o. A.; Ditton, 2000, S. 76, 79; Ebbinghaus, 2006, S. 33; Frommberger, 2013, S. 7; Frommberger et al., 2013, S. 133; Halbheer & Reusser, 2008, S. 254-255; Hollmann et al., S. 9; Jungjohann & Gebhardt, 2022, S. 27-28; Jürgens & Lissmann, 2015, S. 50-55; Klein, 2009, S. 19; Klieme & Tippelt, 2008a, S. 8, 2008b, S. 8; Klotz, 2015, S. 44; Knight & Yorke, 2003, S. 16; Leutner, 2010, S. 625-626; Maier, 2010, S. 72, 294, 299; Schmidt, 2010, S. 20; Schrader, 2013, S. 155-156, 161; Schrader & Helmke, 2014, S. 50-51; Schwaighofer et

al., 2019, S. 475; Weiß, 2011, S. 38; Wilbers, 2018, S. 241, 2019b, S. 165, 251-252; Wild & Krapp, 2006, S. 532; Wilhelm & Kunina-Habenicht, 2021, S. 313; Windelband, Spöttl & Becker, 2014, S. 301; Winther, 2006, S. 92-93)

Nach Winther (2006, S. 94) werden die Ergebnisse summatives Assessment als Anlass genommen, um formative Assessments zu gestalten – und vice versa. Dieser Ansatz kann im Unterricht Anwendung finden, indem beispielsweise zum Üben mit Altprüfungen, Fehler besprochen und als Anlass zum weiteren Lernen genommen werden, etwa zur Prüfungsvorbereitung. Abschlussprüfungen folgen diesem Ansatz nicht, mit der Ausnahme der Fälle, in denen Auszubildende die Prüfung nicht bestehen. Dann kann das summative Assessment, das nicht bestanden wurde, Anlass zum formativen Assessment werden. Auch ist denkbar, dass Lehrpersonen nach Abgabe des summativen Assessments, also hier der Abschlussprüfung, die Lösungen mit den Lernenden besprechen. Es wird dabei zwar auch eine Lerngelegenheit angeboten, allerdings wird es, sofern die Lernenden nicht durchgefallen sind, keine zweite Prüfung in diesem Ausbildungsabschnitt als summatives Assessment geben. Am summativen Assessment wird kritisiert, dass dessen Ergebnis in der Regel nicht als Feedback genutzt wird. Ihre Aussage, dass Feedback dann am nützlichsten ist, wenn Lernende Schwächen zeigen dürfen (Knight & Yorke, 2003, S. 17), ist kritisch zu betrachten. Da das Ergebnis eines summativen Assessments u. a. zu Selektionszwecken genutzt wird, ist dieses nicht zwingend mit einer Feedbackfunktion verbunden (vgl. Knight & Yorke, 2003, S. 16).

Bei der Beschreibung von Assessments finden sich Diagnostikarten. Auf diese und weitere, noch nicht genannten Diagnostikarten, soll im Folgenden eingegangen und der Versuch einer kurzen, aber differenzierten Darstellung vorgenommen werden. In der Literatur findet sich primär eine Unterscheidung zwischen *Status-*, *Prozess-*, *Veränderungs-*, *Verlaufs-* und *Outputdiagnostik* (z. B. Aufschnaiter et al., 2015, S. 744-747). Dabei ist nicht ausgeschlossen, dass auch weitere Begriffe bestehen.⁴⁴ Zur Gegenüberstellung der hier beschriebenen Diagnostikarten wird in Tabelle 2 der Vorschlag von Aufschnaiter et al. (2015, S. 742) aufgegriffen und nach *Gegenstand*, *Methoden* und *Zielen* differenziert. Dabei werden die Ziele lediglich beispielhaft aufgeführt, da davon ausgegangen wird, dass die Verfahren mit spezifischeren und/oder an die Situation angepassten Zielsetzungen durchgeführt werden können. In Tabelle 2 wird zusätzlich eine Spalte zur Zuordnung zu formativem bzw. summativem Assessment aufgenommen.

⁴⁴ So finden sich neben den in Tabelle 2 beschriebenen Arten der Diagnostik in der Literatur weitere Bezeichnungen, wie etwa die *Modifikationsdiagnostik* nach Schwaighofer, Heene und Bühner (2019, S. 474). Die Darstellung hat keinen Anspruch auf Vollständigkeit.

Tabelle 2:*Diagnostikarten nach Gegenstand, Methoden, Ziel(e) und Assessment*

Art	Gegenstand	Methoden	Ziel(e), z. B.	Assessment
Statusdiagnostik	„Erfassung (der Ausprägung) zu einem bestimmten Zeitpunkt vorliegende Kompetenz (bzw. eines vorliegenden Merkmals)“ (Aufschnaiter et al., 2015, S. 744)	i.d.R. schriftliche Erfassung, kann aber auch mündlich sein (Aufschnaiter et al., 2015, S. 744)	„Erfassung der Lernausgangslagen“ → „Fördermaßnahmen oder auch binnendifferenzierende Maßnahmen“ „am Ende von bestimmten Bildungsabschnitten“ → „Maßnahmen der Zuweisung zu weiteren Bildungsabschnitten“ (Aufschnaiter et al., 2015, S. 744)	Formativ & Summativ
Eingangsdagnostik	Fokus auf (die Ausprägung) einer Kompetenz zu Beginn einer Maßnahme bzw. einer Einheit (in Anlehnung an Aufschnaiter et al., 2015, S. 744; Wilhelm & Kunina-Habenicht, 2021, S. 313), verstanden als eine besondere Form der Statusdiagnostik	Mündlich und schriftlich, z. B. in Form einer Analyse der Voraussetzungen der Lernenden	„Erfassung der Lernausgangslagen“ → „Fördermaßnahmen oder auch binnendifferenzierende Maßnahmen“ (Aufschnaiter et al., 2015, S. 744) ⁴⁵	Formativ
Prozessdiagnostik	Erfassung „aktuell vorliegender Kompetenz, jedoch liegt der Fokus nicht auf dem Produkt, der Aufgabenlösung, sondern vielmehr auf dem Lösungsprozess“ (Aufschnaiter et al., 2015, S. 745), auch verstanden als Bezeichnungen „handlungsbegleitende Diagnostik“ (Wild, K.-P. & Krapp, 2006, S. 532)	z. B. „Videoaufzeichnungen oder Verfahren des lauten Denkens, evtl. auch Verfahren der Verschriftlichung von Denkprozessen“ (Aufschnaiter et al., 2015, S. 745)	„Selektion“; „Entwicklung von spezifisch abgestimmten Fördermaßnahmen“ (Aufschnaiter et al., 2015, S. 745)	Formativ (Jürgens & Lissmann, 2015, S. 53)

Veränderungs- diagnostik	„wie sich eine Kompetenz von einem Messzeitpunkt zum nächsten verändert hat“ (Aufschnaiter et al., 2015, S. 745)	Vergleich von „(mindestens) zwei Diagnosen“, Methodik „entsprechen dabei denen der Status- bzw. der Prozessdiagnostik“	„Aussage über Zuwachs, Stagnation oder Abnahme eine Kompetenz treffen“ (Aufschnaiter et al., 2015, S. 745) →„Einschätzung der Wirkung einer Fördermaßnahme“ (Aufschnaiter et al., 2015, S. 746)	Formativ
Verlaufdiag- nostik	„alle [Herv. i. Original] Handlungs- und Denkprozesse, die auf den Aufbau bzw. die (Weiter-)Entwicklung einer bestimmten Kompetenz gerichtet sind“ (Aufschnaiter et al., 2015, S. 746)	Beobachtungen, insbesondere „Video- und Audioaufzeichnungen“ (Aufschnaiter et al., 2015, S. 746)	„Ableitung von Aussagen über die Dynamik (fachspezifischen) Lernens“ →„Ableitung von Fördermaßnahmen“, „Differenzierung“ (Aufschnaiter et al., 2015, S. 746)	Formativ
Output-/Ergebnis-/ Produkt diagnostik ⁴⁶	Fokus auf (die Ausprägung) einer Kompetenz zum Abschluss einer Maßnahme bzw. einer Einheit (in Anlehnung an Winther, 2006, S. 93), verstanden als eine besondere Form der Statusdiagnostik	Mündlich und schriftlich, z. B. in Form von (Abschluss-)Prüfungen (z. B. Büchter et al., 2011; Wilbers, 2018, S. 241; Windelband et al., 2014, S. 301)	Überprüfung der Zielerreichung (Ditton, 2000, S. 79; Wild, K.-P. & Krapp, 2006, S. 532) ⁴⁷ mittels Zensuren bzw. Noten (z. B. Knight & Yorke, 2003, S. 16; Wild & Krapp, 2006, S. 532) →Selektion (z. B. Leutner, 2010, S. 625)	Summativ (Jürgens & Lissmann, 2015, S. 53)

(Eigene Darstellung nach Aufschnaiter et al., 2015, S. 744-746; Ditton, 2000, S. 79; Jürgens & Lissmann, 2015, S. 53; Knight & Yorke, 2003, S. 16; Leutner, 2010, S. 625; Wild & Krapp, 2006, S. 532; Wilhelm & Kunina-Habenicht, 2021, S. 313 Winther, 2006, S. 93)

⁴⁵ Aufschnaiter et al. (2015, S. 744) nutzen die Beschreibung des Ziels zur *Statusdiagnostik* (siehe auch vorherige Zeile). Da die Eingangsdiagnostik im Rahmen dieser Arbeit als eine spezielle Form der Statusdiagnostik verstanden wird, wird dieses Ziel hierauf übertragen.

⁴⁶ Die in dieser Zeile genutzten Erläuterungen stammen meist aus Beschreibungen zum summativen Assessment, wie etwa Knight und Yorke (2003, S. 16); Winther (2006, S. 93). Da allerdings diese Beschreibung auch passend zur Outputdiagnostik scheint, wird diese hier übernommen.

⁴⁷ Wild und Krapp (2006, S. 532) nutzen diese Beschreibungen im Zusammenhang mit der *summativen Evaluation*. Dennoch ist die Beschreibung auch für Outputdiagnostik zutreffend, weshalb diese verwendet wird. Auf die Unterscheidung *Evaluation* und *Diagnostik* wird im Nachfolgenden eingegangen.

In Werken zur pädagogisch-psychologischen Diagnostik wird ferner eine Unterscheidung zwischen formativer und summativer *Evaluation* vorgenommen. Werden die Definitionen von summativen bzw. formativen Assessments und summativer bzw. formativer Evaluation verglichen, zeigt sich, dass diese teilweise synonym verwendet werden können. Nach der Definition und Zielsetzung der *Evaluation* stehen die Verbesserung und Optimierung einer Intervention im Vordergrund (Wottawa, 2006, S. 662). So wird die summative Evaluation als eine Diagnostik zum bzw. nach Ende einer Intervention verstanden, in der die erzielten Ergebnisse überprüft werden (Wild & Krapp, 2006, S. 532; Wottawa, 2006, S. 662). Zum Vergleich steht beim summativen Assessment vielmehr der Lernende im Fokus, und das Assessment wird etwa zu Selektionszwecken genutzt. Aus diesem Grund werden die Begriffe Evaluation und Assessment nicht als Synonyme betrachtet. Um den Fokus der Arbeit zu wahren, wird im Folgenden zur Erstellung von Prüfungsaufgaben lediglich auf das summative Assessment Bezug genommen.

2.4.4 *Bezeichnungen und Einsatz der Instrumente und Verfahren*

In Dokumenten und andere Informationsquellen zu Instrumenten bzw. Verfahren zu diagnostischen Zwecken sind unterschiedliche Bezeichnungen zu finden (vgl. Kapitel 4.2.2). Zwecks eines Überblicks und des Verständnisses werden ausgewählte Begriffe aufgegriffen und erläutert. Die Auswahl richtet sich nach den Instrumenten und Verfahren, nach denen aufgrund ihres Vorkommens in den Teilstudien kodiert wurde (vgl. Kapitel 4).

Kompetenzmessung bzw. -diagnostik

Kompetenzen, als Ziele von Bildungsprozessen, sind nicht direkt beobachtbar. Vielmehr wird die Performanz als Indiz für Kompetenzen gesehen (Bayer, 2015, S. 3; Helmke, 2021, S. 120; Hollmann et al., 2021, S. 9; Wilbers, 2019b, S. 249-250, vgl. Kapitel 2.5.1). Die Performanz gilt als Indiz dafür, dass Kompetenzen vorliegen, jedoch unter der Annahme, dass die Performanz nicht zufällig – etwa durch Raten bei einer Multiple-Choice-Aufgabe – entstanden ist (Wilbers, 2014a, S. 63-64).

In der pädagogischen Diagnostik wird die Möglichkeit bzw. die Notwendigkeit des Lernens als eine Eigenschaft der Kompetenzen verstanden (Hartig & Klieme, 2006a, S. 130). Laut Seeber und Nikolaus (2010, S. 11) wird in der beruflichen Bildung eine Kompetenzmessung zu Beginn eines Ausbildungsabschnitts (*Berufseignungs- und Berufseingangsdiagnostik*) und am Ende der Ausbildungszeit bzw. des Ausbildungsabschnitts eingesetzt. Anhand von Skalen kann das (Kompetenz-)Niveau bzw. Stufe des Geprüften bestimmt werden, welche wiederum aus

Kompetenzanforderungen abgeleitet werden (vgl. Wilbers, 2019b, S. 75-76; Jürgens & Lissmann, 2015, S. 18).

Leistungsmessung bzw. -diagnostik

Held (2014) beschreibt Leistung als das Ergebnis einer Verhaltensbeurteilung. Leistung selbst sei „kein beobachtbares Verhalten, keine aus diesem Verhalten ableitbare Größe und auch keine Eigenschaft eines Verhaltens“ (S. 142). Demnach ist eine beobachtbare Performanz nicht mit Leistung gleichzusetzen. Die Bewertung einer Leistung setzt eine Norm voraus (Held, 2014, S. 142; Klein, 2009, S. 9-10). Sie ist also kriterienorientiert, wobei die Kriterien sowohl deskriptiv als auch wertend sind (Held, 2014, S. 143). Die Einordnung der Leistung ergibt sich aus dem Grad der Annäherung einer Person an das Handlungsziel. Performanz stellt im weitesten Sinne eine Leistung dar, welche sich aus dem vorgestellten Verständnis der Kompetenzentwicklung und Lernen ergibt. Performanzen von Prüflingen werden als Leistungen betrachtet, die kriterienorientiert bewertet werden. Somit ergeben sich Parallelen zur oben aufgeführten Kompetenzmessung und -diagnostik.

Jürgens und Lissmann (2015, S. 69) zufolge ist unter Leistungsfeststellung eine Beobachtung oder Abfrage zu verstehen, die es dem Beobachter ermöglicht, deskriptiv vorzugehen. Unter der *Leistungsmessung* wird ein operationalisiertes Verfahren verstanden, welches ebenfalls zum Abschluss einer Lerneinheit eingesetzt wird. Hier werden testtheoretische Gütekriterien genutzt, um die Leistung der Lernenden zu erfassen bzw. zu kontrollieren (Jürgens & Lissmann, 2015, S. 69). Durch eine Leistungsdiagnostik ist es nach Maier (2015, S. 40-41) möglich, über den Lernerfolg eines*r Schüler*in zu sprechen, was zum nächsten Begriff führt.

Lernerfolg und Lernergebnis

Erfolg und *Leistung* sind nach Held (2014, S. 142) nicht gleichzusetzen. Als Beispiel führt der Autor einen Versicherungsbetrüger auf, dessen Leistung erfolgreich ist, wenn er den Betrug erfolgreich abschließen konnte. Dieser Erfolg müsste entsprechend als Leistung anerkannt werden, wenn Erfolg und Leistung gleichzusetzen wären. Mit dem Beispiel wird jedoch die Bedeutung der Norm-/Kriterienorientierung deutlich. Held (2014, S. 142) versteht Erfolg als ein subjektiv gesetztes Ziel. Im Kontext der Pädagogik ist dies allerdings nicht immer zutreffend. Aus den Erklärungen des Begriffs *Lernerfolg* in der Literatur wird deutlich, dass Lernen dann als erfolgreich angenommen wird, wenn die gesetzten Lernziele erreicht werden (Schlicht, 2012, S. 100-101). Der Erfolg bzw. das (Lern-/Kompetenz-)Ziel wird demnach nicht unbedingt

ausschließlich subjektiv gesetzt, sondern auch objektiv – etwa basierend auf Ausbildungsordnungen und Rahmenlehrplänen (vgl. Schlicht, 2012, S. 97-98). Unter der Annahme, dass Kompetenzen erlernt werden können und Performanz als Indiz für Kompetenz betrachtet wird (vgl. Erläuterung zu *Kompetenzmessung bzw. -diagnostik*), ist die in Assessments ermittelte Performanz auch als Indiz des Lernerfolgs zu betrachten. Lernerfolgskontrollen können zu summativen und formativen Zwecken Einsatz finden (Tribelhorn, 2014, S. 207-208; vgl. hierzu die Nutzung des Begriffs in z. B. Euler & Pätzold, 2004, S. 4; Hanft & Müskens, o. A.; Severing, 2011; Winter, 2018, S. 15).

In den Erläuterungen der vorgestellten Begriffe zeigen sich viele Überschneidungen. Insbesondere haben sie gemeinsam, zu diagnostischen Zwecken verwendet zu werden. Im Rahmen der Literaturrecherche fällt auf, dass die Begrifflichkeiten häufig ohne eine jeweilige Definition und zusätzlichen Erläuterungen genutzt werden. Zudem ist es nicht immer möglich, Verfahren explizit einer bestimmten Assessmentform (summativ oder formativ) zuzuweisen. Vielmehr findet sich der Hinweis, dass für das formative und summative Assessment dieselben Verfahren Anwendung finden können (Aufschnaiter et al., 2015, S. 741-742). So wird beispielsweise über die Messung nicht nur im summativen Assessment gesprochen, sondern auch im formativen (z. B. Förster, Kuhn & Souvignier, 2017; Schütze, Souvignier & Hasselhorn, 2008). Daraus ist zu schließen, dass „nicht das Instrument, mit dem ein Assessment durchgeführt wird, formativ oder summativ [ist], sondern die Nutzung [dessens]“ (Schütze et al., 2008, S. 700).

2.4.5 Gütekriterien im summativen Assessment

In der Literatur werden als Hauptgütekriterien *Objektivität*, *Reliabilität* und *Validität* aufgeführt (z. B. Grabowski, Castello & Brodersen, 2022, S. 247; Ingenkamp & Lissmann, 2008, S. 51-60; Jürgens & Lissmann, 2015, S. 90-91; Schwaighofer et al., 2019, S. 476-480; Wilhelm & Kunina-Habenicht, 2021, S. 319-321). *Objektivität* wird unterteilt in *Durchführungs-*, *Auswertungs-* und *Interpretationsobjektivität* (z. B. Ingenkamp & Lissmann, 2008, S. 51-54; Schwaighofer et al., 2019, S. 476-477). Knight und Yorke (2003, S. 17-18) beschreiben verschiedene Rahmenbedingungen, etwa die Bearbeitungsdauer (Durchführungsobjektivität) und die Auswertung des Materials anhand von festgelegten Kriterien (Auswertungsobjektivität), die bei der Objektivität von Assessments bedeutend sind. Die Interpretation ist dann als objektiv zu bezeichnen, wenn alle Interpretierenden aus dem Ergebnis dieselbe Konsequenz ziehen bzw. zur selben Interpretation gelangen (Schwaighofer et al., 2019, S. 477).

Mit *Reliabilität* werden die Genauigkeit, Stabilität und Konsistenz der Testergebnisse bzw. der Messung betrachtet (z. B. Jürgens & Lissmann, 2015, S. 90; Wilhelm & Kunina-Habenicht,

2021, S. 319-320). Knight und Yorke (2003, S. 18) zufolge ist die Reliabilität bei einfachen Prüfungen am stärksten. Reliabilitäten können etwa durch die Testwiederholung und -halbierung festgestellt werden (Klein, 2009, S. 26; Schwaighofer et al., 2019, S. 477-478).

Die *Validität* von Aufgaben ist ein Indikator dafür, dass mit den Aufgaben das gemessen wird, was beabsichtigt wurde zu messen. Validität wird unterteilt in *Inhalts-, Konstrukt- und Kriteriumsvalidität* (Schwaighofer et al., 2019, S. 478-480). Ein Instrument bzw. ein Verfahren wird dann als inhaltlich valide bezeichnet, wenn dadurch das zu messende Merkmal gemessen werden kann; d. h. es findet ein Abgleich der Items⁴⁸ mit den curricularen Forderungen statt (Klein, 2009, S. 26; Schwaighofer et al., 2019, S. 478-479). Bei der Konstruktvalidität wird das Gemessene mit theoretischen Annahmen bzw. Modellierungen verglichen (Klein, 2009, S. 26; Moosbrugger & Kelava, 2008, S. 16). Ein Testinstrument ist dann als konstruktvalide zu bezeichnen, „wenn die postulierten, prozessualen und inhaltlichen Strukturen sinnvoll operationalisiert wurden und sich zudem empirisch in den Testergebnissen widerspiegeln“ (Klotz & Winther, 2012, S. 4). Kriteriumsvalide ist ein Test dann, „wenn vom Verhalten der Testperson innerhalb der Testsituation erfolgreich auf ein »Kriterium«, nämlich auf ein Verhalten außerhalb der Testsituation, geschlossen werden kann“ (Moosbrugger & Kelava, 2008, S. 18). Am Beispiel eines Mathematiktests erläutert bedeutet dies folgendes: Gibt es im Aufgabentext ein sprachliches Problem – etwa ein Wort, das Geprüfte nicht verstehen – so kann dies das Ergebnis verfälschen und der Test verfehlt das eigentliche Ziel, mathematische Kompetenzen zu überprüfen. Der Test ist somit nicht valide (Grabowski et al., 2022, S. 247-248; La Marca, Redfield & Winter, 2000, S. 6).

Die Hauptgütekriterien sind nicht unabhängig voneinander, denn Reliabilität und Validität setzen Objektivität voraus. Zur Validität muss neben der Objektivität die Reliabilität gegeben sein, auch wenn diese keine hinreichende Bedingung ist (Schwaighofer et al., 2019, S. 480). Neben den Hauptgütekriterien werden weitere Gütekriterien, beispielsweise *Normierung, Ökonomie* und *Nützlichkeit*, formuliert (z. B. Schwaighofer et al., 2019, S. 480-482; Wilhelm & Kunina-Habenicht, 2021, S. 321-322), welche in der vorliegenden Arbeit jedoch aus Gründen der Übersichtlichkeit nicht weiter erläutert werden.

Von Assessments wird weiter gefordert, dass sie Aufgaben enthalten, durch welche Prüflinge das eigene Wissen und die eigenen Fähigkeiten bzw. Kompetenzen auf jeglichen Stufen zeigen

⁴⁸ In diesem Sinne sind auch Aufgaben einer Prüfung als Items zu verstehen.

können (La Marca et al., 2000, S. 26-27; vgl. hierzu auch ZPA Nord-West, 2021). Dies erfordert eine Variation der Schwierigkeiten von Aufgaben. Gleichzeitig wird von Prüfungen auch gefordert, dass sie justizabel sind (AkA, 2023b; Badura, 2015, S. 18; Weiß, 2014, S. 3).

„Je relevanter eine Entscheidung für den Entscheidungsempfangenden ist, umso höher sind auch die Anforderungen an die Gütekriterien der erhobenen Information“ (Van Ophuysen & Behrmann, 2015, S. 87). Wie in Kapitel 2.4.3 dargestellt, sind schriftliche kaufmännische Prüfungsaufgaben ein Instrument zur Entscheidungsfindung und können weitreichende Konsequenzen für Prüflinge haben. Aus diesem Grund müssen die Aufgaben die Gütekriterien berücksichtigend gestaltet werden (vgl. AkA, 2023b; Hollmann et al., 2021, S. 10-11; ZPA Nord-West, 2021). Die geforderten Gütekriterien an Prüfungsaufgaben werden in Kapitel 3.2 aufgegriffen und in Bezug auf bestehende schriftlich-kaufmännische Prüfungsaufgaben diskutiert.

2.5 Problemorientierung als didaktischer Zugang und Problemlösekompetenz

2.5.1 Kompetenz und Kompetenzorientierung

In der Literatur finden sich diverse Ansätze zur Modellierung von Kompetenzen⁴⁹. Es gibt Kompetenzniveau- und Kompetenzstrukturmodelle. Erstere bilden Kompetenzniveaustufen mittels einer vertikalen Darstellung ab (Wilbers, 2014b, S. 68). Sie finden Anwendung in bestimmten Domänen bzw. Situationen und dienen der Messung einer erbrachten Leistung. Kompetenzniveaustufen legen Leistungsbereiche fest, die kriteriumsorientiert interpretiert bzw. beschrieben werden (z. B. Hartig & Klieme, 2006b, S. 133-136; Winther, 2010, 37, 40; vgl. hierzu auch Sälzer, Reiss, Schiepe-Tiska, Prenzel & Heinze, 2013, S. 61 zur Niveaumodellierung in der Domäne Mathematik in PISA). Kompetenzstrukturmodelle zeigen in einer horizontalen Darstellung verschiedene Kompetenzfacetten (Wilbers, 2014a, S. 68). Es wird an dieser Stelle beispielhaft auf die KMK (2007, S. 10-11) verwiesen, die als Bestandteile der Handlungskompetenz die Fach-, Human-, Sozial-, Methoden-, Lern- und kommunikative Kompetenz aufführt. Ein weiteres Beispiel aus dem kaufmännischen Bereich ist das *Kompetenzstrukturmodell für die kaufmännische Bildung* von Winther (2010, S. 54). Auf eine ausführliche Darstellung der diversen Modellierung und Ausführung der Entwicklung dieser wird verzichtet, da diese im Detail für die vorliegende Arbeit nicht von Relevanz sind. Auf die Modellierung der hier betrachteten Problemlösekompetenz wird jedoch in Kapitel 2.5.2 näher eingegangen.

⁴⁹ In der vorliegenden Arbeit wird das eingangs vorgestellte Kompetenzverständnis nach Weinert (2014, S. 27-28) genutzt (vgl. auch Kapitel 1). Ebenso sei auf Kapitel 2.4.4 verwiesen, in der auf die Kompetenzmessung bzw. -diagnostik eingegangen wurde.

Mit dem Begriff der *Kompetenzorientierung* wird der Wandel von der Input- zur Outputorientierung bzw. der Orientierung am Outcome⁵⁰ deutlich (Wilbers, 2019b, S. 149). Kompetenzorientierung verfolgt das Ziel, dass „die Lernenden in die Lage versetzt werden, reale Probleme in ihrer Lebenswelt angemessen und erfolgreich zu bewältigen“ (Frank & Iller, 2013, 34). Dabei spielen (weiterhin) das Fachwissen (Reusser, 2014, S. 326) sowie die Übertragbarkeit der Kompetenz (Körber, 2010, S. 10) eine bedeutende Rolle. Der Fokus der vorliegenden Arbeit richtet sich explizit auf die berufliche Problemlösekompetenz.

2.5.2 Problemdefinition und Problemlösen

In dieser Arbeit werden Probleme bzw. das Problemlösen⁵¹ nicht im klinisch-psychologischen Sinne verstanden, sondern im kognitiven und analytischen. Dörner (1987) zufolge bestehen Probleme aus einem „Unerwünschten Anfangszustand s_α “, einem „Erwünschten Endzustand s_ω “ und einer „Barriere, die die Transformation von s_α in s_ω im Moment verhindert“ (S. 10). Auch Bransford und Stein (1993, S. 7) beschreiben ein Problem als eine Situation, in der eine Diskrepanz zwischen dem aktuellen Zustand und dem erwünschten besteht. Routine ist als Gegenpol von Problemen zu verstehen (vgl. Betsch, Funke & Plessner, 2011, S. 138-139; Funke et al., 2018, S. 45-46). Im Sinne des hier betrachteten Kontexts, werden *Routinen* auf *Aufgaben* erweitert. Demnach unterscheidet sich eine problemhaltige Aufgabe dahingehend von einer nicht-problemhaltigen Aufgabe, dass zwischen dem Ist- und dem Soll-Zustand eine Barriere besteht. Weiter werden Probleme damit beschrieben, dass in diesen etwas Unbekanntes liegt. Das Finden oder das Lösen des Unbekannten hat einen sozialen, kulturellen oder intellektuellen Wert (Jonassen, 2000, S. 65, 2011, S. 1; Jonassen & Hung, 2012, S. 2680).

In der Literatur finden sich Ansätze, den Problemlöseprozess zu beschreiben. Nachfolgend werden diese exemplarisch vorgestellt. Dabei wird kein Anspruch auf Vollständigkeit erhoben. Vielmehr geht es darum, Gemeinsamkeiten der Ansätze herauszuarbeiten, um ein Grundverständnis für den Ablauf des Problemlöseprozesses aufzubauen.

Sembill (1992, S. 109) beschreibt den Problemlöseprozess in einer Problem- bzw. Handlungssituation als eine Bewertung des Ist- und Ziel- bzw. Soll-Zustands. Anschließend wird geprüft, ob zwischen den Zuständen eine Diskrepanz liegt. Wird diese Frage verneint, wird die Situation verlassen. Wird die Frage bejaht, werden diverse Schritte zur Behebung der Diskrepanz und

⁵⁰ Zur Differenzierung des Outputs und Outcomes sei auf Kapitel 2.4.3 verwiesen.

⁵¹ Zum Problemlösen bedarf es Problemlösekompetenz. Die hierfür genutzte Definition nach Siegfried et al. (2019, S. 990). wurde bereits in Kapitel 1 vorgestellt. Auf die Facetten der Problemlösekompetenz wird in Kapitel 2.5.4 eingegangen.

Lösung des Problems durchlaufen. Diese beinhalten die „Planung von Handlungsentwürfen; Auswahl von geeigneten Handlungsteilnehmern (Personen, Informationen, Gegenständen) und Zuordnung zu einem Handlungsschema; Aktionsprogramm; Problemlöseprogramm“ (Bezeichnungen wörtlich aus Abbildung 28 in Sembill, 1992, S. 109 entnommen). Die Wirkungen der Schritte werden mit Blick auf ihre erwarteten (Neben-)Effekten eingeschätzt. Anschließend folgt das Handeln und es stellen sich die Effekte ein. Zuletzt werden wieder die Zustände verglichen. Ist die Diskrepanz weiterhin vorhanden, müssen die Schritte erneut durchlaufen werden; ansonsten wird die Situation verlassen. Das Problem gilt sodann als gelöst.

Weiter ist in der Literatur der *IDEAL*-Ansatz zur Problemlösung von Bransford und Stein (1993, S. 19–20, 38) zu finden. Die Problemlösung findet in fünf Schritten statt, wobei die Reihenfolge der Problemlöseschritte nicht immer vorgeschrieben ist. Jedoch kommen alle Komponenten in Problemlöseprozessen vor und sind somit relevant. In Abbildung 5 werden die zugrundeliegenden Schritte des Modells dargestellt.

Abbildung 5:

Der IDEAL-Ansatz

<p>I = Identify problems and opportunities D = Define goals E = Explore possible strategies A = Anticipate outcomes and Act L = Look back and Learn</p>

(Eigene Darstellung nach Bransford & Stein, 1993, S. 20)

Unter *Identifikation der Probleme und Möglichkeiten* (I) ist zu verstehen, dass neben der Identifikation des Problems auch gleichzeitig eine Chance besteht, Neues zu erkennen, welches es ohne das Problem nicht geben würde. Dies bietet die Möglichkeit der Kreativität. Das Problem wird vom Problemlösenden erkannt und verortet sowie Potenziale dessen entdeckt. Warum das Erkennen von Problemen nicht immer gelingt, erklären die Autoren damit, dass unangenehme Situationen von Menschen als gegeben akzeptiert und somit Verbesserungspotenziale und Kreativität nicht gesehen bzw. genutzt werden (Bransford & Stein, 1993, S. 20-21). Bei der *Zieldefinition* (D) besteht die Annahme, dass sich die Ziele verschiedener Personen zum selben Problem unterscheiden können, wobei auch mehrere Ziele definiert werden können. Dabei ist zu beobachten, dass alternative Ziele von vielen Problemlösenden nicht beachtet werden

(Bransford & Stein, 1993, S. 24-26). Bei der *Suche nach möglichen Strategien* (E) geht es darum, alternative Ansätze der Problemlösung zu suchen. So werden die Ziele erneut betrachtet, die mit Hilfe einer Strategie erreicht werden sollen. Dieser Schritt gelingt nicht immer, etwa dann, wenn Menschen aus der Situation raus wollen und die Konfrontation mit dem Problem vermeiden. Auch scheitert die Problemlösung, wenn zwar allgemeine Problemlösestrategien bekannt sind, die spezifische Strategie allerdings noch nicht gelernt wurde. Im Gegensatz zu sich wiederholenden Problemen, dessen Lösung bekannt ist (wie z. B. 2×2), werden in komplexeren Aufgaben bzw. Problemen diese in Teilstücke zerlegt und nacheinander bearbeitet. Zudem gibt es die Möglichkeit, die Informationen zu einem Problem zu externalisieren und sich einen Überblick zu verschaffen, indem diese beispielsweise grafisch auf Papier gebracht werden oder ein Computer genutzt wird. In solchen Fällen kann das Ziel des Problemlösenden sein, eine Strategie zu finden, um die Informationen im Blick zu behalten. Welche Repräsentationsform die effektivere Strategie ist, hängt dabei von den Zielen und der Problemart ab. Neben den vorgestellten Strategien existieren noch weitere, die auch als allgemeine Problemlösestrategien gelten, wie etwa das Rückwärtsbearbeiten eines Problems und die Vereinfachung bzw. Spezialisierung des Problems (z. B. Modellbauten). Des Weiteren gibt es spezifische Problemlösestrategien, die Wissen in einem spezifischen Bereich für die Lösung des Problems voraussetzen, wie etwa Probleme in der Physik, Mathematik und Biologie. Das Spezialwissen, das zum Lösen des Problems nötig ist, kann meist nicht durch anderes Wissen ersetzt werden (Bransford & Stein, 1993, S. 27-33).

Nach der Strategiewahl folgt die *Vorwegnahme der Ergebnisse und das Handeln* (A). Die Problemlösenden sollen durch die Vorwegnahme der Ergebnisse davor bewahrt sein, Handlungen zu tätigen, die sie im Nachhinein bereuen. Beispiele hierfür sind die Einführung neuer Sicherheitssysteme in Unternehmen oder die Einführung neuer Produkte. Um die Konsequenz der Handlung, also das Ergebnis, vorherzusehen bzw. zu erraten, sollten vorab Befragungen, Berechnung oder Ähnliches getätigt werden. So ist es möglich, schon im Voraus oder aber während der Handlung Ideen für alternative Strategien zu generieren und Fehler rechtzeitig zu erkennen (Bransford & Stein, 1993, S. 33-36). *Sehen und Lernen* (L) bedeutet, dass die tatsächliche Auswirkung der Strategie gesehen und aus der Erfahrung gelernt wird. Die Autoren führen als Beispiel Studierende auf, die lediglich Interesse an der Note einer Prüfung haben und nicht die korrigierte Prüfung ansehen, um ihre Fehler zu verstehen und daraus zu lernen (Bransford & Stein, 1993, S. 36-37).

2.5.3 Merkmale problemhaltiger Aufgaben

Neben der Unterscheidung nach Problemtypen (z. B. Dörner, 1987, S. 11-13; Jonassen, 2000, 2012), wird in der einschlägigen Literatur zudem nach Problemmerkmalen unterschieden (vgl. Dörner, 1987, S. 15-25; Funke, 2003, S. 126-135, 2006, S. 379-380; Jonassen, 2000; Jonassen & Hung, 2012; Wuttke et al., 2022). Das Auftreten von bestimmten Problemtypen ist situationsbedingt und/oder typisch für bestimmte Domänen bzw. Disziplinen (Dörner, 1987, S. 11-13; Jonassen & Hung, 2008, S. 7). Die Unterscheidung von Problemtypen ist jedoch nicht trivial und erschwert die Anwendbarkeit bei der Beurteilung von Aufgaben. Daher werden im Folgenden statt Problemtypen, Merkmale von Problemsituationen bzw. problemhaltigen Aufgaben thematisiert. Jonassen und Hung (2012, S. 2681) unterscheiden Problemtypen nach Charakteristika, die u. a. für die Arbeit als Merkmale genutzt werden: *Strukturierung (well-/ill-structured)*, *Komplexität (simple/complex)* und *Dynamik (static/dynamic)*. Dadurch unterscheiden sich auch die Schwierigkeit und die Ressourcen zum Lösen des Problems (Jonassen & Hung, 2012, S. 2681).

Wohl- und *schlecht-strukturierte Probleme* sind nach Dörner (1987, S. 13) auf McCarthy (1956) zurückzuführen, der *wohldefinierte Probleme* wie folgt beschreibt: „We call a problem well-defined if there is a test which can be applied to a proposed solution. In case the proposed solution is a solution, the test must confirm this in a finite number of steps“ (McCarthy, 1956, S. 177). Davon ausgehend ist anzunehmen, dass bei *wohl-definierten/-strukturierten*⁵² Problemen sowohl Anfangs- und Endzustand als auch die Operatoren und Variablen zwischen den Zuständen klar definiert und bekannt sind. Es muss lediglich die Ausführung erfolgen. Anders ist es bei *schlecht-definierten* Problemen. Diese beinhalten nicht-vorhersehbare Aspekte und es gibt keine eindeutige Lösung (Seel, 2012, S. 2691). Auch Jonassen und Hung (2012, S. 2681) definieren *schlecht-definierte* Probleme als solche, bei denen der Zielzustand und der Weg dahin meist unbekannt bzw. nicht-definiert sind. Bei *wohl-definierten* ist sowohl der Weg als auch das Ziel klar definiert. Bei der Zieldefinition ist zu berücksichtigen, dass mehrere Ziele bestrebt werden können – dieser Fall wird als *Polytelie* bezeichnet (Betsch et al., 2011, S. 142–143, 156). Auch sind bei *wohl-strukturierten* Problemen die notwendigen Variablen zur Problemlösung bekannt (Jonassen, 2011, S. 6). Im Schulkontext sind häufig wohl-strukturierte Probleme zu finden (Jonassen, 2000, S. 66-67; Jonassen & Hung, 2012, S. 2681), wobei der Strukturiertegrad nicht als dichotom zu betrachten, sondern auf einem Kontinuum besteht (Jonassen,

⁵² In der Literatur werden sowohl die Begriffe *wohl-* bzw. *schlecht-definiert* als auch *wohl-* bzw. *schlecht-strukturiert* verwendet, was vermutlich mit der unterschiedlichen Übersetzung zusammenhängt. Im Rahmen dieser Arbeit werden die Begrifflichkeiten als Synonym verstanden und genutzt.

2011, S. 6). Auch wird in der Literatur die *Transparenz* angesprochen. Diese ist gering, „wenn viele Merkmale in der Situation vorhanden sind, deren direkte Feststellung schwierig oder ganz unmöglich ist“ (Dörner, 1987, S. 21). Problemlösende sollen sowohl hinsichtlich der Quantität als auch Qualität geeignete Informationen finden und die Limitation einberechnen, möglicherweise nicht alles Relevante berücksichtigt zu haben (Betsch et al., 2011, S. 156). Das Merkmal der Strukturiertheit bzw. Definiiertheit wird für diese Arbeit am Vorgehen von Wuttke et al. (2022) angelehnt, also Strukturiertheit bzw. Definiiertheit hinsichtlich

- des Ist-Zustandes
- des Ziel-Zustands
- der Informationen, die zur Problemlösung gegeben sind (in ihrer Menge und in ihrer Qualität)
- der Offenheit der Lösung

berücksichtigt.

Schlecht-strukturierte Probleme gehen nach Jonassen (2000, S. 68, 2011, S. 9) meist mit erhöhter *Komplexität* einher. Komplexität wiederum bestimmt sich nach Jonassen und Hung (2008, S. 9) nach der Breite des erforderlichen Wissens, der Schwierigkeit des Verständnisses und der Konzeptanwendung, der zur Problemlösung erforderlichen Fähigkeiten und dem Wissensstand sowie dem Grad der Nichtlinearität der Beziehungen zwischen den Variablen innerhalb des Problemraums. Jonassen (2011, S. 7) beschreibt schlecht-strukturierte Probleme als domänenunspezifisch, wobei Wissen aus mehreren Domänen für die Problemlösung nötig ist. Diese Beschreibung hilft zum Verständnis der Beziehung zwischen Strukturierungsgrad und Komplexität.

Unter der *Dynamik* eines Problems ist zu verstehen, dass sich Variablen und Faktoren bzw. die Beziehung dieser über die Zeit verändern (Dörner, 1987, S. 19-20; Jonassen, 2011, S. 10). Jonassen und Hung (2012, S. 2681) beschreiben dies als die schwierigste Dimension im Problemlöseprozess, wobei schlecht-strukturierte Probleme meist dynamischer sind (Jonassen, 2011, S. 10). Dynamische Situationen verlangen von Problemlösenden „unter Zeitdruck zu handeln und ... Entwicklungen abschätzen zu können“ (Dörner, 1987, S. 20).

Gleichzeitig ist auch die *Vernetztheit* eines Problems relevant, zu welcher *Neben- und Folgewirkungen* bzw. *-effekte* gehören (Betsch et al., 2011, S. 155-156; Dörner, 1987, S. 20-21; Funke, 2003, S. 129-130; Wuttke et al., 2022, S. 34). So kann eine Handlung neben dem fokussierten Haupteffekt auch Nebeneffekte hervorrufen (Dörner, 1987, S. 20-21). Letztere können

kurz- oder langfristige Wirkungen haben (Wuttke et al., 2022, S. 34) und positiv oder negativ sein (Wuttke, Seeber, Schumann et al., 2022, S. 28). Auch müssen häufig die *Interessen verschiedener Stakeholder* berücksichtigt werden, welche wiederum die Problemsituation und das Problemlösen beeinflussen (Wuttke et al., 2022, S. 27, 34). Die *Lösung zu reflektieren bzw. zu begründen* ist ebenfalls von Bedeutung (Betsch et al., 2011, S. 150; Gajek et al., 2021, S. 20-21; Sembill, 1992, S. 76; Wuttke et al., 2022, S. 34). In Kapitel 2.5.2 wurde bereits auf das Problemlösen eingegangen, in Kapitel 2.5.4 wird eine weitere Kompetenzfacette des Problemlösens ausgeführt.

2.5.4 Einflussfaktoren auf das Problemlösen

Neben den Merkmalen eines Problems, durch deren Ausprägung dieses variiert (vgl. Kapitel 2.5.3), beeinflussen weitere Faktoren das Problemlösen (vgl. z. B. grafische Darstellung in Jonassen, 2000, S. 66). Nachfolgend wird auf die Merkmale des Problemlösenden und die Problemlösekompetenz eingegangen.

Probleme können nach Jonassen und Hung (2012, S. 2681-2682) in diversen Kontexten vorkommen, sodass eine Person in unterschiedlichen Kontexten unterschiedliche Leistungen im Problemlösen zeigen kann. Auch geht es um die Art, wie Probleme dargestellt werden, z. B. mit Case Studies. Instrumente können die Lernenden dabei unterstützen, Problemräumen zu konstruieren (z.B. Concept Maps) und sind insbesondere wichtig für das Problem-based-Learning.

Laut Jonassen (1997, S. 66, 2000, S. 67) ist es offensichtlich, dass unterschiedliche Problemtypen unterschiedliche Fähigkeiten erfordern. Das Problemlösen ist demnach abhängig von der *Person, die das Problem löst* (Dörner, 1987, S. 14; Jonassen, 2000, S. 66; Jonassen & Hung, 2012, S. 2682). So würden Expert*innen und Laien den Schwierigkeitsgrad derselben Aufgabe aufgrund unterschiedlichen *Wissens* und *Erfahrungen* unterschiedlich bewerten. Dabei geht es u. a. um die *Vertrautheit* der Person mit dem Problemtypen (Jonassen, 2000, S. 69). In diesem Zusammenhang scheint die Erfahrung bzw. die Möglichkeiten zum Problem-based-Learning nicht unbedeutend. Dem Konstruktivismus folgend, ist davon auszugehen, dass die Kompetenz, Probleme zu lösen, nicht durch einen Lehrenden extern erzwungen werden kann, sondern vom Lernenden selbst erlernt werden muss (Büchter & Tramm, 2007, S. 2; Wilbers, 2014a, S. 607-608). Problem-based Learning⁵³ wird beschrieben als fallbasiertes, kooperatives und selbstgesteuertes Lernen (Weber, 2005, S. 96). Es wird davon ausgegangen, dass Lernende durch das

⁵³ Da der Schwerpunkt der Arbeit nicht auf dem Lehr-Lern-Prozess bzw. der Instruktion liegt, sondern auf der Prüfung bzw. dem Assessment, wird Problem-Based-Learning nicht weiter vertieft.

Problemlösen Wissen erwerben und dass das Gelernte nachhaltiger ist, da gleich im Problemkontext gelernt wird (Wijnen, 2008, S. 30). Durch die Anwendung wird eine Trägheit des Wissens vermieden (vgl. Renkl, 1994).

Die Kompetenz, ein Problem zu lösen, wird vom speziellen *Wissen in einer Domäne* beeinflusst (Bransford & Stein, 1993, S. 4). Die Problemlösekompetenz ist domänen- und kontextspezifisch und die Problemlösekompetenz hängt davon ab, über wie viel Wissen eine Person in einer Domäne verfügt (Jonassen, 2000, S. 68-69). Mit Blick auf das Ziel der beruflichen Bildung, also Handlungskompetenz, wird die Notwendigkeit für Fachkompetenz (beinhaltet Fachwissen) deutlich. Denn Fachkompetenz – als Teil der Handlungskompetenz – gilt u. a. als Basis für das Lösen von Problemen (Sekretariat der KMK, 2007, S. 10-11). Auch Sembill (1992, S. 162) geht davon aus, dass Fachwissen Einfluss auf Handlungskompetenz und Problemlösefähigkeit hat.

Wissen kann in *strukturelles* (Jonassen, 2000, S. 69), *deklaratives*, *prozedurales* und *strategisches* Wissen (Wilbers, 2014a, S. 174; Winther, 2010, S. 111) unterteilt werden. Strukturelles Wissen ist das Wissen darüber, wie Konzepte in einer Domäne miteinander verbunden sind (Jonassen, 2000, S. 69). Grotzer (2002, S. 54) greift die Erklärung zum strukturellen Wissen von Jonassen, Beissner und Yacci (1993) auf und erklärt, dass es mehr als das Verstehen von Beziehungen ist. Es geht über die Frage nach dem *Was* und dem *Wie* hinaus und befähigt Menschen, Wissen über das *Warum* zu erwerben. Deklaratives Wissen umfasst das Sachwissen, Wissen über Konzepte und Theorien sowie Fakten (Wilbers, 2014a, S. 174; Winther, 2010, S. 111). Prozedurales Wissen hingegen ist das (implizite) Wissen über Prozesse (Wilbers, 2014a, S. 178-179) sowie „Handlungen, Prozeduren und Arbeitstechniken zur Bearbeitung eines spezifischen Inhaltsbereichs“ (Winther, 2010, S. 111). Das strategische Wissen bezeichnen Achtenhagen und Baethge (2007, S. 56) als „eines der zentralen Ziele der beruflichen Bildung“ und führen weiter aus mit den Worten:

Der Schwerpunkt liegt dabei auf den Bedingungen eines inhaltsbasierten Problemlösens, das auf vertieftes deklaratives und prozedurales Wissen zurückgreift und dabei die möglichen angestrebten Haupteffekte sowie die ungewollten möglichen Nebeneffekte zugleich in Betracht zieht. Entscheidungen werden so in voller Kenntnis möglicher negativer Konsequenzen getroffen. (Achtenhagen & Baethge, 2007, S. 56)

Fachwissen allein ist demnach nicht ausreichend, um Probleme zu lösen (Frackmann & Tärre, 2009, S. 32).

Der Prozess des Problemlösen erfordert nach Jonassen und Hung (2012, S. 2680) eine Vorstellung der Problemsituation, wodurch der Problemlösende das Problem versteht und einem Problemtypen zuordnet. Dies geschieht nicht unabhängig vom Domänenwissen (Jonassen, 2000, S. 69). Zudem ist es erforderlich, bestehende Problemlöseschemata im Gedächtnis des Problemlösenden als Unterstützung im Problemlöseprozess abzurufen. Vorkenntnisse, Erfahrungen, erkenntnistheoretische Überzeugungen, die Denkfähigkeit und weitere individuelle Merkmale sowie der Fokus des Problemlösenden beeinflussen die Problemlösekompetenz (Jonassen & Hung, 2012, S. 2682).

Wuttke et al. (2015, S. 194) und Rausch und Wuttke (2016, S. 177) stellen zum domänenspezifischen Problemlösen ein Kompetenzmodell vor, welches in Abbildung 6 abgebildet wird.

Abbildung 6:
Kompetenzmodell

Kompetenzdimensionen	13 Kompetenzfacetten			
(A) Wissens-anwendung (Kognition)	Handlungsbedarfe und Informationsquellen identifizieren	Informationen verarbeiten	Begründete Entscheidung treffen	Entscheidung angemessen kommunizieren
(B) Handlungsregulation (Metakognition)	Planvolles (strukturiertes) Vorgehen	Persistenz (fokussiertes) Vorgehen	Retrospektive Handlungskontrolle	
(C) Selbstkonzept (Erwartungen)	Angemessenes Vertrauen in die eigene Kompetenz	Unsicherheits-/Unge-wissheitstoleranz	Angemessenes Vertrauen in die eigene Lösung	
(D) Interesse (Valenzen)	Inhaltliches Interesse am Problem	Positive Aktiviertheit	Interesse an Problemfortgang/ an Lernchancen	

(Nachzeichnung nach Wuttke et al., 2015, S. 194, Verweis auf Original auf Englisch in Rausch & Wuttke, 2016, S. 177)

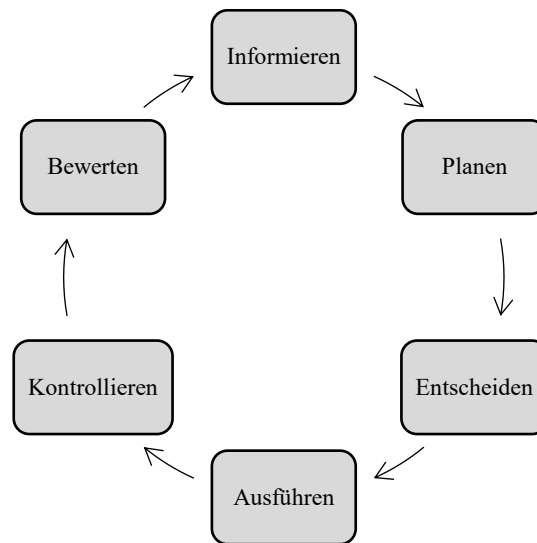
Die Kompetenzen werden in vier Dimensionen und 13 Facetten unterschieden. Im Modell finden sich sowohl einige der vorgestellten Merkmale des Problemlösenden als auch die Schritte des Problemlösens (Kapitel 2.5.2) wieder, weshalb hier auf eine wiederholte Beschreibung verzichtet wird. Stattdessen ist hervorzuheben, dass die Autor*innen in den Dimensionen C und D *Selbstkonzept* und *Interesse* aufnehmen. Die unter der Dimension C gefassten Facetten des (Selbst-)Vertrauens „in die eigene Kompetenz“ und „in die eigene Lösung“ (Wuttke et al., 2015, S. 194) finden sich in ähnlicher Form auch in Jonassen (2000, S. 66) als *self-confidence* (Selbstvertrauen) wieder. Problemlösekompetenten Personen gelingt der Umgang mit Situationen mit unsicherer Entwicklung sowie Lösung besser als weniger problemlösekompetenten

(Rausch & Wuttke, 2016, S. 178). Die *Unsicherheits-/Ungewissheitstoleranz* der Dimension (Wuttke et al., 2015, S. 194) überschneidet sich mit der *perseverence* (Beharrlichkeit) nach Jonassen (2000, S. 66), welche der Autor um Motivation ergänzt. Neben Dörner (1984, S. 14-15) gehen auch Betsch et al. (2011, S. 138) auf die Relevanz der Motivation beim Problemlösen ein und unterscheiden zwischen *high-stake-problems* und *low-stake-problems*. High-stake-problems sind dabei „Probleme, bei denen zur Beseitigung des Hindernisses eine hohe Motivation erforderlich ist“, während bei low-stake-problems „nur wenig Motivation vorliegt, das Hindernis zu beseitigen“ (Betsch et al., 2011, S. 138). Wie auch in anderer Literatur dargestellt (z. B. Klieme, Funke, Leutner, Reimann & Wirth, 2001, S. 185), wird in Dimension D des Modells die Relevanz des Interesses deutlich. Interesse besteht aus den Facetten *Inhaltliches Interesse am Problem*, *Positive Aktiviertheit* sowie *Interesse an Problemfortgang/an Lernchancen* (Wuttke et al., 2015, S. 194).

2.5.5 Problem- vs. Handlungsorientierung

Für die Förderung und Prüfung von Handlungskompetenz⁵⁴ wird der Ansatz der Handlungsorientierung verfolgt (Badura, 2015, S. 13; Bretschneider, Gutschow & Lorig, 2014, S. 27; Sekretariat der KMK, 2021a, S. 12; Wilbers, 2019b, S. 38). In der beruflichen Bildung wird *die vollständige Handlung* bestrebt (Badura, 2015, S. 60-61; Sekretariat der KMK, 2021a, S. 17; Wilbers, 2019b, S. 45-46). Der Ablauf folgt den idealtypischen Schritten: *Informieren*, *Planen*, *Entscheiden*, *Ausführen*, *Kontrollieren* und *Bewerten* (Bundesinstitut für Berufsbildung, 2015, S. 23; vgl. Abbildung 7)

⁵⁴ Auf die Handlungskompetenz als Ziel beruflicher Bildung und dessen Definition wurde bereits in Kapitel 1 eingegangen.

Abbildung 7:*Schritte der vollständigen Handlung*

(Eigene Darstellung nach Bundesinstitut für Berufsbildung, 2015, S. 23)

Die vollständige Handlung erklärt Wilbers (2014a, S. 624-627, 2019b, S. 46) wie folgt: Die Handlung beginnt mit *Informationen*. Hier wird eine (konstruierte)⁵⁵ Situation⁵⁶ eingesetzt, aus der Informationen zur Problemstellung entnommen werden sowie das Ziel der Handlung erkannt wird. Zu der Situation finden die handelnden Personen weitere Materialien, denen sie Informationen entnehmen und filtern können, welche zur Bearbeitung bzw. Lösung der Situation wichtig sind. Durch die Aufforderung zur Handlung in einer Situation soll es dem Handelnden gelingen, sich in dieser Rolle wiederzufinden. Anschließend wird das Vorgehen geplant, um ans Ziel zu gelangen. Zur *Planung* gehört u. a. die Planung der Handlungsschritte und die Identifikation erforderlicher Mittel und Informationen. Je nach Zielgruppe, finden sich hierin ein oder mehrere Handlungsaufträge. Handlungsoptionen werden im nächsten Schritt bewertet und sich für eine *entschieden*. Die geplante Handlung wird nun *ausgeführt*. Anschlie-

⁵⁵ Konstruiert deshalb, da es sich in der Beschreibung um Lernsituationen handeln (vgl. Wilbers, 2014a, S. 624-627). Da die Arbeit sich nicht auf den Lehr-Lern-Kontext fokussiert, wird hier ein weiteres Verständnis verwendet und genutzt, in der solche Situationen auch für Assessmentzwecke genutzt werden.

⁵⁶ Ein typisches Instrument im handlungsorientierten Unterricht bzw. des handlungsorientierten Ansatzes in der beruflichen Bildung sind Lernsituationen (Sekretariat der KMK, 2018, S. 17; Wilbers, 2014a, S. 618) oder auch Problemsituationen (Tramm & Naeve, 2007). Lernsituationen werden als „curriculare Strukturelemente der Lernfeldkonzeption“ (Sekretariat der KMK, 2018, S. 32) beschrieben. Lernfelder sind Formen der Unterrichtsorganisation, die durch die Diskussion der damaligen Ausbildung an Berufsschulen entstand und anstelle von Unterrichtsfächern 1996 eingeführt wurden (Riedl, 2015, S. 128-129, im Online-Dokument S. 2-3). Durch Lernsituationen sollen die Kompetenzziele eines Lernfelds bzw. der Lernfelder abgedeckt werden (Sekretariat der KMK, 2018, S. 12) und es wird bestrebt, dass Lernen durch vollständige Handlung erfolgt (Sekretariat der KMK, 2018, S. 17).

End erfolgt die (*Handlungs-*)*Kontrolle*, in der das Handlungsprodukt bzw. der-prozess kontrolliert wird. Das Resultat der Handlung wird bewertet, indem die handelnde Person das Produkt präsentiert, gefolgt von der eigenen *Reflexion* der Arbeit.

Lernsituationen in der Schule werden i. d. R. von Lehrkräften entwickelt.⁵⁷ Die Qualität bzw. die Güte einer Lernsituation lässt sich anhand mehrerer Kriterien beurteilen⁵⁸, zu welchen auch die Problemstellung bzw. -orientierung gehört (Staatliches Studienseminar für das Lehramt an berufsbildenden Schulen Speyer / Kaiserslautern, 2018, S. 5; Tramm & Naeve, 2007, S. 8; Wilbers, 2014b, S. 145). Für handlungsorientiertes Lernen bzw. für Handlungssituationen inkl. zugehörigen Informationsmaterialien wird gefordert, dass diese realitätsnah bzw. authentisch gestaltet sind (Dubs, 2006, S. 193; Klotz & Winther, 2012, S. 3; Wilbers, 2014a, S. 624). Zudem sollen sie sich an „*komplexen, lebens- und berufsnahen, ganzheitlich betrachtenden Problem-bereichen* [Herv. i. Original]“ (Dubs, 2006, S. 193)⁵⁹ orientieren. Die Komplexität von realitätsnahen Problemen soll daher beibehalten werden.

Zwischen den konstruktivistischen Ansätzen der Handlungs- und Problemorientierung (Wilbers, 2014a, S. 607-608) lassen sich Parallelen identifizieren. So sind sowohl die Handlungs- als auch die Problemorientierung kompetenzorientiert, da die jeweilige Kompetenz – Handlungs- und Problemlösekompetenz – adressiert wird. Handlungs- und Problemorientierung sind also nicht als konkurrierende Konstrukte zu verstehen, sondern als solche, die einen unterschiedlichen Schwerpunkt der (didaktischen) Ausrichtung haben. Es bestehen daher zwischen den Schritten der vollständigen Handlung und des Problemlösens Gemeinsamkeiten. Ebenso finden sich Parallelen in der Gestaltung von handlungs- bzw. problemorientierten Situationen (vgl. Kapitel 2.5.3).

In Anlehnung an Aebli (1981) und Klieme et al. (2001) argumentiert Reetz (2005, S. 3), dass Handeln über die (wissensbasierte) Kompetenz, Probleme zu lösen, entsteht. Folglich schließt Handlungsorientierung Probleme bzw. Problemorientierung ein (vgl. in Bezug auf die Unterrichtsgestaltung bzw. das Lernen z. B. Bremer, 2005, S. 6 im Online-Dokument; Greiner, 2010, S. 58; Riedl & Schelten, 2006, S. 18-19; in Bezug auf Prüfungen z. B. Badura, 2015, S. 61; Ressel & Kiesecker, 2019, S. 22). Im Rahmen dieser Arbeit wird davon ausgegangen, dass

⁵⁷ Auf die Schritte bei der Konstruktion von Lernsituationen wird nicht eingegangen, da der Fokus der Arbeit nicht hierauf liegt. Als Literatur hierfür kann Wilbers (2014b, S. 144) empfohlen werden.

⁵⁸ Auch hierzu sei auf Wilbers (2014b, S. 145) verwiesen, da auf die Darstellung einzelner Kriterien und deren kritischen Reflexion im Rahmen dieser Arbeit verzichtet wird, um den Fokus der Arbeit zu behalten.

⁵⁹ Dubs (2006, S. 193) stellt in seiner Aufforderung mehrere Merkmale auf, die sich auf „*handlungsorientierte[n] Lernen* [Herv. i. Original]“ beziehen. Dieses Verständnis wird hier erweitert, da der Fokus nicht auf die Instruktion gerichtet ist, wohl aber die Inhalte, im Sinne der Curriculum-Instruktion-Assessment-Triade von Pellegrino (2010), sich nicht auf die instruktionale Ebene beschränken, sondern darüber hinaus gehen (sollten).

Problemorientierung ein Konstrukt ist, welches Teil der Handlungsorientierung ist, aber gleichzeitig über die Handlungsorientierung hinausgeht. Feststeht, dass Handlungs- und Problemorientierung nicht synonym verwendet werden. Während bei Problemorientierung das Problem bzw. die Problemlösekompetenz im Vordergrund steht – es gibt keine routinierte Lösung (vgl. Kapitel 2.5.2) – ist die (vollständige) Handlung bzw. Handlungskompetenz bei Handlungsorientierung vordergründig. Bei Handlungsorientierung wird i. d. R. mit Lernsituationen gearbeitet, in denen ein Problem vorgegeben wird. Bei problemorientierten Ansätzen hingegen ist nicht immer eine vollständige Handlung notwendig. So können beispielsweise weniger komplexe Problemsituationen, d. h. dessen Merkmale nicht stark ausgeprägt sind (vgl. Kapitel 2.5.3), auch ohne das Einhalten der Schritte der vollständigen Handlung gelöst werden. Kritisch ist dies, wenn man davon ausgeht, dass die Schritte der vollständigen Handlung eine Routine darstellen.

2.6 Erstellung problemhaltiger Prüfungsaufgaben

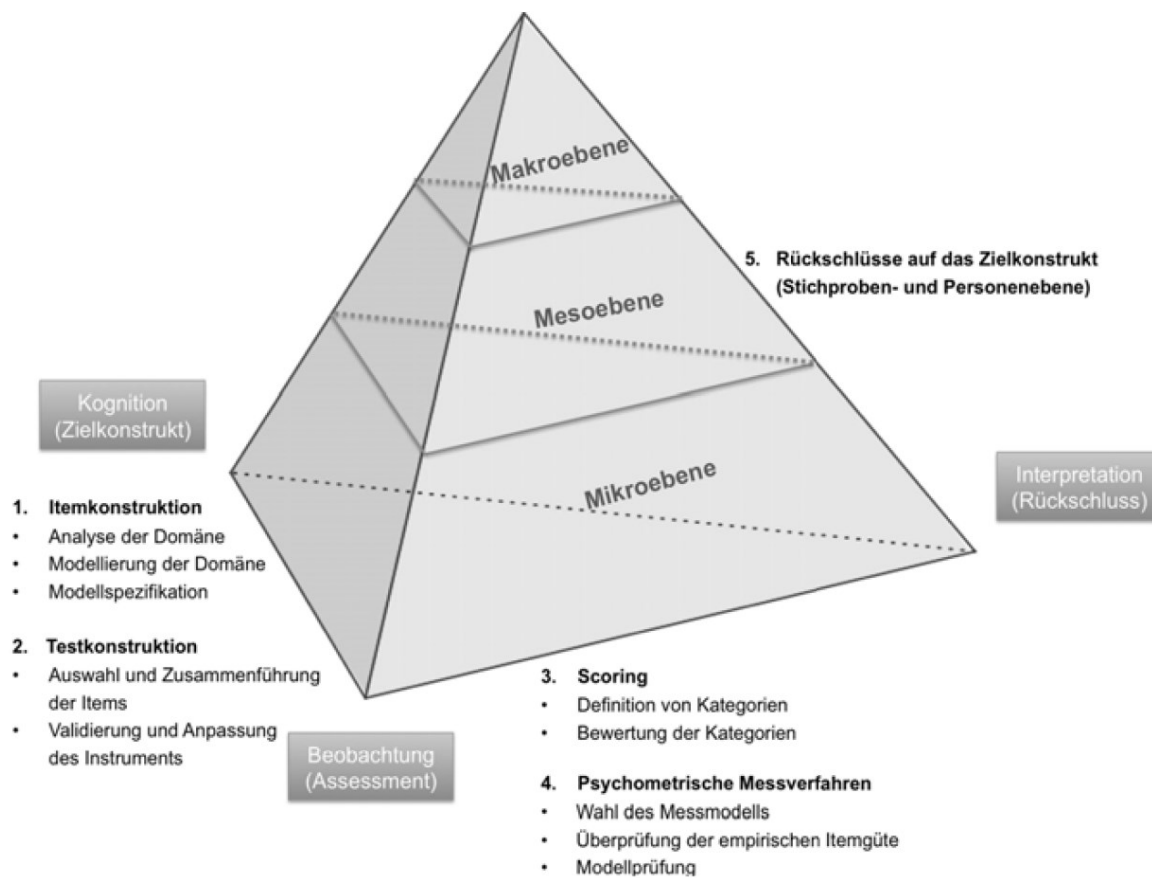
Auch wenn nicht alle Lehrkräfte an der Erstellung von Prüfungen als eine Form des summativen Assessments beteiligt sind (vgl. Kapitel 2.3), gehören dennoch andere Formen pädagogischer Diagnostik zu ihren beruflichen Tätigkeiten, z. B. Erstellen von Klausuren. Durch die Klausur findet eine Kontrolle darüber statt, ob der*die Schüler*in die curricularen Ziele des Unterrichts erreicht hat (Speth, 2007, S. 468-469). Dies ist auch im Sinne der Kompetenzorientierung (vgl. Kapitel 2.5.1) und des von Wilbers (2018, S. 144-145) vorgestellten *Zyklus der Kompetenzorientierung*, welcher Parallelen zur Curriculum-Instruktion-Assessment-Triade von Pellegrino (2010) aufweist. Der Zyklus enthält zum einen die Definition der *Kompetenzerwartung* und zum anderen die Schritte *Kompetenzerwartungen modellieren* und *Learning Outcomes festlegen* (Wilbers, 2018, S. 145), wofür curriculare Vorgaben beachtet werden (Wilbers, 2020), so, wie auch nach der Triade von Pellegrino (2010, S. 4). Diesem folgt die Erfassung der *Kompetenzen als Lernausgangslage*⁶⁰ und anschließend die *Kompetenzentwicklung* (Wilbers, 2018, S. 145), welche im (Lehr-)Lern-Prozess⁶¹ (Wilbers, 2020) bzw. in der Instruktion (Pellegrino, 2010, S. 4-5) geschieht. Im Anschluss daran werden die *Kompetenzen als Lernergebnisse* festgestellt (Wilbers, 2018, S. 145; vgl. hierzu auch Pellegrino, 2010). Im Rahmen dieser Arbeit wird als Ziel (Speth, 2007, S. 468), welches curricular begründet ist (Pellegrino, 2010, S. 4; Wilbers, 2018, S. 145, 2020), vordergründig Problemlösekompetenz betrachtet. Die

⁶⁰ Dieser Punkt ist unter die *Bedingungsanalyse* zu fassen (Wilbers, 2020, o. A.).

⁶¹ Wilbers (2020) schreibt im Leitfaden über *Lernprozesse* (3.0.0) und *Lehr-Lernaktivitäten* (3.1.3) und vom *Unterricht*. Da Lernen und Lehren in Interaktion geschieht, wird die Bezeichnung *Lehr-Lern-Prozess* genutzt (vgl. Wilbers, 2019b, S. 9).

Ebene des *Lehr-Lern-Prozesses* nach Wilbers (2018, S. 145, 2019b, S. 9) bzw. der *Instruktion* nach Pellegrino (2010, S. 4-5) wird nicht fokussiert.

Assessments und die darin gezeigte Performanz werden als Mittel zur Erfassung von Kompetenzen verstanden (Sekretariat der KMK, 2018, S. 32; Wilbers, 2014a, S. 63; vgl. Kapitel 2.4.3 und 2.4.4). Dabei wird die Limitation berücksichtigt, dass Assessments keinen direkten Zugriff auf die im Gehirn des Lernenden stattfindenden Prozessen ermöglichen (Pellegrino, 2010, S. 5). Auch, wenn die Erstellung von Aufgaben zu Assessmentzwecken nicht immer Beachtung in der Definition der *pädagogischen Diagnostik* bzw. *diagnostischen Kompetenz* und in der Beschreibung diagnostischer Tätigkeiten findet, ist diese doch ein wesentlicher Bestandteil pädagogisch-diagnostischer Tätigkeiten (vgl. Kapitel 2.4). Die Erstellung von Aufgaben ist in der *fachdidaktischen Kompetenz* verortet, auch werden hier *diagnostische Potentiale* von Aufgaben thematisiert (Baumert & Kunter, 2011, S. 37; vgl. Kapitel 2.1 und 2.4). Daher wird das Erstellen von problemhaltigen Aufgaben zu summativen Zwecken als eine Anforderung verstanden, für die sowohl diagnostische als auch fachdidaktische Kompetenzen benötigt werden. Zur Gestaltung von problemhaltigen Aufgaben zu diagnostischen Zwecken wird das *Logic-Assessment-Modell* (Abbildung 8) nach Klotz (2015, 72–84) zugrunde gelegt, welches auf bestehenden theoretischen Erkenntnissen aufbaut. Diese werden in der vorliegenden Arbeit nicht wiederholt, vielmehr liegt der Fokus auf der Erklärung des pyramidenförmigen Modells.

Abbildung 8:*Logic-Assessment-Modell*

(Entnommen aus Klotz, 2015, S. 72⁶²)

Angelehnt an Pellegrino, Chudowsky und Glaser (2001, S. 44) werden die Ecken der Grundfläche des Modells als *Kognition*, *Beobachtung* und *Interpretation* bezeichnet (vgl. hierzu auch Pellegrino, 2002, S. 49, 2014). *Kognitionen* sind theoretische Annahmen über Wissensstrukturen in einer bestimmten Domäne, auch der Erwerb dieses Wissens ist hierin enthalten. Sie bilden das Ziel eines Assessments ab. In der *Beobachtung* geht es um theoretische Merkmale von Aufgaben(-situationen), die es in einer Testsituation ermöglichen, Informationen über eine bestimmte Eigenschaft eines*r Lernenden zu gewinnen. Sie beinhaltet „ein inhalts- und erfassungsvalides Itemdesign“ (Klotz, 2015, S. 73). In der *Interpretation* werden Methoden und Verfahren zusammengefasst, „die zum Treffen von Aussagen über die latente Variable Verwendung finden“ (Klotz, 2015, S. 73). Sie „dient als Ausdruck dafür, wie die gewonnenen Beobachtungen aus den konstruierten Aufgabensituationen zur Beweisführung in Bezug auf das theoretische Zielkonstrukt genutzt werden können“ (Klotz, 2015, S. 73 nach Pellegrino et al.,

⁶² Für die Nutzung der Abbildung aus der Primärquelle liegt die freundliche Genehmigung von Prof. Dr. Viola Deutscher vor (V. Deutscher, persönl. Mitteilung, 26.04.2023).

2001, S. 48). In Large-Scale-Assessments werden i. d. R. statistische Modelle zu den Kompetenzniveaus der Prüflinge genutzt (Pellegrino, 2014, S. 69). Im Vergleich dazu nutzen Lehrpersonen in Klausuren (*classroom assessment*) eher intuitive oder qualitative Modelle, seltener formelle. Diese drei Ecken sind als eine Triade zu verstehen, sie sind also zusammenhängend zu betrachten (Klotz, 2015, S. 73-74). Übertragen auf die vorliegende Arbeit bedeutet es, dass berufliche Problemlösekompetenz die Kognition darstellt, die über die Leistungen der Prüflinge in schriftlichen problemhaltigen Aufgaben beobachtet werden soll. Die Interpretation erfolgt über die Beurteilung der Validität, insbesondere in Bezug auf die Problemhaltigkeit der Aufgabe sowie über die Leistung der Prüflinge. Dazu können bestehende Kompetenzmodelle (vgl. Kapitel 2.5.2) genutzt werden. Zur Bestimmung der Niveaustufen werden Scoringverfahren verwendet.

Vertikal werden drei Ebenen diagnostischer Prozesse abgebildet: *Mikro-*, *Meso-* und *Makroebene*. Während in der *Mikroebene* beispielsweise Gespräche zwischen Lehrkraft und Schüler*innen zur Erfassung des Wissensstands zu verstehen sind, sind unter der *Mesoebene* Tests zu fassen, die klassen- bzw. schulbezogen sind. Auf der *Makroebene* hingegen sind Tests zu verorten, die als Large-Scale-Verfahren eingestuft werden (Klotz, 2015, S. 75). Das im Rahmen dieser Arbeit betrachtete Assessment bezieht sich auf die *Makroebene* im Sinne von Prüfungsaufgaben mit der Erweiterung um die *Mesoebene*, da auch diese zu summativen Assessments gezählt werden können und nicht jede (angehende) Lehrkraft bei der Prüfungsaufgabenkonzeption beteiligt ist, wohl aber die Aufgabe hat, Klausuren zu stellen.

Das Modell weist zudem verschiedene nicht voneinander unabhängigen *Schichten* auf, die auf Mislevy, Steinberg und Almond (2003, S. 17) bzw. Mislevy und Riconscente (2005, S. 6) zurückzuführen sind (Klotz, 2015, S. 75). Im Modell werden diese als Schritte aufgenommen, die eine Konsistenz schaffen sollen (Klotz, 2015, S. 75 nach Mislevy, Almond & Lukas, 2003, S. 17). Die Schritte werden nachfolgend aufgeführt und um weitere, in der Literatur zu findenden Informationen, ergänzt:

- **Schritt 1: *Itemkonstruktion*.** Es werden Informationen zu spezifischen Eigenschaften einer jeweiligen Domäne beschafft (Klotz, 2015, S. 76). Die gewonnenen Informationen sind dienlich für die Implikationen des Assessments. Des Weiteren werden Wissensstrukturen, -erwerb, -nutzung und die Kommunikation dessen betrachtet (Mislevy & Haertel, 2006, S. 8, siehe hierzu auch Klotz, 2015, S. 76). Zusammengefasst werden hier die Elemente zu Analyse und Modellierung der Domäne sowie die Modellspezifikation aufgeführt (Klotz, 2015, S. 85). Im *Domain Modeling* (Mislevy & Haertel, 2006, S. 8; Mislevy & Riconscente,

2005, S. 10) folgt die Abbildung der Domäne in seinen charakteristischen Prozessen und Vorstellungen (Klotz, 2015, S. 76 nach Mislevy & Riconscente, 2005, S. 10).

Bei der Erstellung von Prüfungen⁶³ sollen nach La Marca et al. (2000, S. 5-6) zunächst die geprüfte Domäne und zu prüfenden Fähigkeiten definiert werden. Es wird entschieden, was bzw. welchen Bereich der Test prüfen wird.

- *Schritt 2: Testkonstruktion.* Das zentrale Ziel der Testkonstruktion ist „die Umsetzung des Vollständigkeitsanspruchs, der sowohl in kaufmännischen Lehrplänen als auch in den Prüfungsregularien expliziert wird“ (Klotz, 2015, S. 85). Dieser wird, Klotz (2015, S. 85-86) zufolge, etwa durch das Design der Aufgabe erreicht, indem die einzelnen Aufgaben/Items sinnvoll zusammengeführt werden sowie auf Authentizität und einen adäquaten Schwierigkeitsgrad geachtet wird. Gleichzeitig wird die Validität geprüft. Nach Maier (2015, S. 19) ist zu entscheiden, welches Verfahren genutzt wird, um das Diagnoseziel⁶⁴ zu erreichen. Bei der Aufgabenerstellung ist zudem über das Format dieser zu entscheiden und darüber, welche Wissensart gefordert werden soll. Zudem ist die Übereinstimmung der Aufgabe mit den curricularen Anforderungen zu prüfen (La Marca et al., 2000, S. 5).
- *Schritt 3: Scoring.* Hierbei werden Kategorien definiert und diese bewertet (Klotz, 2015, S. 72). Diese dienen der Objektivität, insbesondere der Auswertungsobjektivität (Klotz, 2015, S. 86).
- *Schritt 4: Psychometrische Messverfahren.* Hierin wird ein Messmodell gewählt. Daraufhin folgt die Überprüfung der Itemschwierigkeiten, Itemgüte sowie der Güte des Modells (Klotz, 2015, S. 72, 86).
- *Schritt 5: Rückschlüsse auf das Zielkonstrukt.* Auf der Grundlage „der hergestellten Kausalität zwischen Zielkonstrukt und Datenbasis [werden] Rückschlüsse auf die Ausprägung des Zielkonstrukts gezogen“ (Klotz, 2015, S. 86).

In Kapitel 3.2 wird auf weitere Aspekte der vorgestellten Schritte eingegangen und die Gegebenheiten bestehender schriftlich-kaufmännischer Prüfungsaufgaben kritisch diskutiert.

⁶³ La Marca, Redfield und Winter (2000) schreiben in ihrem Beitrag von Assessment, die für den hier genutzten Kontext als Prüfungen übersetzt werden.

⁶⁴ Maier (2015, S. 19) schreibt hier von *Wissen*, das erfasst werden soll. Dieser Gedanke wird hier erweitert dadurch, dass nicht nur Wissen, sondern Kompetenzen als Diagnoseziel gelten, sprich, durch Verfahren mehr als lediglich Wissen erfasst werden kann bzw. soll.

3 Stand der Forschung zu betrachteten Kompetenzen (angehender) Lehrkräfte

3.1 Diagnostische und fachdidaktische Kompetenz (angehender) Lehrkräfte

Diagnostische Kompetenz wird für eine Lehrkraft bzw. für das Unterrichten als notwendig angesehen (vgl. z. B. Helmke, 2021, S. 121; Helmke, Hosenfeld & Schrader, 2004, S. 119; vgl. auch Kapitel 2). Zur Messung diagnostischer Kompetenz wird häufig die Urteilsgenauigkeit genutzt (Schrader, 2013, S. 161; Spinath, 2005; Südkamp, Möller & Pohlmann, 2008, S. 90-91). So beziehen sich auch die meisten Studien hierzu auf die Urteilskompetenz bzw. Diagnosegenauigkeit, z. T. unter dem Stichwort *Akkuratheit* (z. B. Brunner, Anders, Hochfeld & Krauss, 2011; Ohle, McElvany, Horz & Ullrich, 2015; Schrader, 2013, S. 158-159; Spinath, 2005; van Ophuysen & Behrmann, 2015, S. 82-83), auch in Bezug auf die Korrektur kaufmännischer Prüfungsaufgaben (Winther, 2011, S. 45, 51). Wie im Kapitel 0 beschrieben, geht allerdings diagnostische Kompetenz über Urteilsgenauigkeit hinaus. Die Entwicklung von Instrumenten bzw. die Aufgabenerstellung zählt auch zur diagnostischen Kompetenz einer Lehrkraft. So kritisiert Schrader die Beschränkung auf die Diagnosegenauigkeit als Messwert diagnostischer Kompetenz von Lehrkräften und schreibt, unter Bezugnahme auf Nitko (2004), dass „[z]u einer professionellen Lehrpersonendiagnostik ... zudem die Fähigkeit [gehört], geeignete Strategien und Methoden der Informationsbeschaffung und -verarbeitung einzusetzen, was auch die passgenaue Auswahl und Konstruktion von Tests und anderen Verfahren zur Lernkontrolle einschließt, sowie die Bereitschaft, sich mit den gewonnenen Daten fundiert auseinanderzusetzen und diese sachgerecht zu interpretieren“ (Schrader, 2013, S. 162). Des Weiteren wird die Fähigkeit, Aufgabenschwierigkeiten einzuschätzen bzw. zu bestimmen, als ein weiteres Merkmal diagnostischer Kompetenz beschrieben (Schrader, 2010, S. 104; Schrader & Helmke, 2014, S. 48).

In dieser Arbeit wird die diagnostische Kompetenz nicht in ihrer Gesamtheit betrachtet. Vielmehr liegt der Fokus auf die Erstellung von Aufgaben zu summativen Zwecken, als Kombination diagnostischer und fachdidaktischer Kompetenz. Auch wenn die Aufgabenerstellung nicht unabhängig von anderen Dimensionen diagnostischer Kompetenz ist, wird der Schwerpunkt auf die Aufgabenerstellung beibehalten, sodass die Ergebnisse der oben genannten Studien nicht näher erläutert werden.

Im Vergleich zur Erforschung der Urteilskompetenz sind derzeit kaum explizite Studien zur Erstellung bzw. Nutzung von Instrumenten, mit denen Informationen über den Kompetenz- bzw. Leistungsstand der Schüler*innen gewonnen können, zu finden. Dieser Themenbereich

ist weitestgehend unerforscht und stellt daher ein Forschungsdesiderat dar (Schrader, 2013, S. 162). Zwar wurden beispielsweise in der Studie von Neubrand, Jordan, Krauss, Blum und Löwen (2011) Aufgaben, auch Klassenarbeiten von Lehrkräften⁶⁵ (Neubrand et al., 2011, S. 125), untersucht. Diese bezogen sich allerdings auf den Mathematikunterricht für die 9. und 10. Jahrgangsstufe. Es wurden die *kognitive Aktivierung* (Neubrand et al., 2011, S. 118) sowie die Problemhaltigkeit der Aufgaben (Jordan et al., 2006, S. 31-33; Neubrand et al., 2011, S. 123) betrachtet. Neubrand et al. (2011, S. 126-129) halten fest, dass u. a. die Problemhaltigkeit der Aufgaben gering ist. Aufgaben für die Zielgruppe der Auszubildenden in der (kaufmännisch-)beruflichen Bildung wurden nicht betrachtet. Auch wenn das Problemlösen in mathematischen Kontexten relevante Aspekte für das berufliche Problemlösen beinhaltet, kann damit das berufliche Problemlösen nicht umfänglich abgebildet werden.

In der Evaluationsstudie von Abs (2006) werden neben Ausbilder*innen und Mentor*innen auch LiV⁶⁶ aus Hessen (davon keine*r aus berufsbildenden Schulen) u. a. nach den diagnostischen Standards befragt. Dabei wurden formulierte Aussagen (Items) nach dem Notenschema bewertet. Bei der Betrachtung der beispielhaft aufgeführten Items fällt auf, dass sich einige auch auf die Aufgaben zu diagnostischen Zwecken beziehen, etwa in Bezug auf die Aufgabenstellung nach Leistung der Lernenden oder die Nutzung von Instrumenten (Abs, 2006, S. 222; Abs, Döbrich, Vögele & Eckhard, 2005, S. 60; Items S9 und S18). Andere Items beziehen sich stärker auf diagnostische Tätigkeiten für den Unterricht, wie die Auswahl der Aufgaben bzw. die Entwicklung von Materialien für den Unterricht (Abs et al., 2005, S. 60; Items S10 und S14). Damit sollte, den Angaben des Autors unter Bezug auf Horstkemper (2004) sowie Oser und Spychiger (2005) zufolge, „die Abbildung einer Auffassung von Diagnostik [erfolgen], die im Sinne positiver Fehlerkultur und gerichtetem didaktischen Anschlusshandelns dem Konzept der Förderdiagnostik entspricht“ (Abs, 2006, S. 222). Somit liegt der Fokus nicht auf das summative Assessment. Als Ergebnis der Studie wird festgehalten, dass LiV an Gymnasien ihre diagnostische Kompetenz geringer einschätzen als LiV an Grund-, Haupt-, Real- und Förderschulen (Abs, 2006, S. 225-230). Die Zielgruppe der LiV kaufmännisch-beruflicher Schulen wurde nicht berücksichtigt.

⁶⁵ Neubrand, Jordan, Krauss, Blum und Löwen (2011) beschreiben, dass Lehrkräfte als Proband*innen „gebeten [wurden], eine Zusammenstellung aller bislang im Schuljahr gestellten Klassenarbeiten ... zu festgelegten Themen aus dem Unterricht einzureichen“ (S. 125). Es ist unklar, ob es sich bei den „gestellten“ Arbeiten um von Lehrkräften selbst erstellte Klassenarbeiten oder um adaptiertes oder übernommenes Material aus bestehenden Arbeiten handelt.

⁶⁶ Personen, die sich im Referendariat befinden, werden in der Arbeit mit „LiV“ für „Lehrkräfte im Vorbereitungsdienst“ abgekürzt. Damit sind auch Referendar*innen gemeint.

Hascher (2008, S. 78) nimmt an, dass die diagnostischen Kompetenzen eher im beruflichen Alltag erlernt werden und weniger während der Ausbildungsphase, was die Gefahr der fehlenden Reflexion bergen könne. Helmke (2021, S. 124) bezeichnet es als „massive[s] Versäumnis“, dass kaum schulpraktische, pädagogische Diagnostik in der ersten Phase der Lehrer*innenbildung in Deutschland stattfindet. Dabei ist jedoch zu berücksichtigen, dass Lehrende an Schulen (in der Regel) nach dem Studium ein Referendariat durchlaufen und in ihrer beruflichen Praxis begleitet werden (vgl. Kapitel 2.2). Wird die Unterscheidung nach theoretischem und praktischem Wissen von Fenstermacher (1994) genutzt, ist es nicht unüblich, dass pädagogische Diagnostik durch die Anwendung in der Praxis erlernt wird. Das ggf. zunächst in der Ausbildung erlernte formale Wissen wird durch die praktische Anwendung gefestigt und vertieft, da es konkret auf einen Fall angewandt wird.

Es finden sich keine Studien zur Erstellung von problemhaltigen Prüfungsaufgaben durch (angehende) Lehrkräfte. Dies verdeutlicht nochmals den Forschungsbedarf. Vielmehr finden sich, neben den oben beschriebenen Studien zur diagnostischen Kompetenz, Studien zur fachdidaktischen Kompetenz von (angehenden) Lehrkräften, die sich auf die Aufgabengestaltung/-erstellung beschränken. So berichtet Holtsch (2011) von einer Studie zur *Aufgabenkompetenz*. Dazu gehört auch das *Aufgabewissen*, welches wie folgt beschrieben wird: „Das Aufgabewissen des Lehrenden umfasst die Fähigkeit, problemorientierte Aufgaben nach fachdidaktischen Ansätzen zu konstruieren, auszuwählen und zu modifizieren“ (Holtsch, 2011, S. 26). In ihrer Studie bezieht sie sich nicht auf Test- bzw. Prüfungsaufgaben, sondern auf Lern- und Übungsaufgaben (Holtsch, 2011, S. 27), weshalb ein näheres Betrachten für die vorliegende Arbeit unerheblich ist. Stattdessen werden in Kapitel 3.2 Hinweise aus der Literatur und Studien zur schriftlich-kaufmännischen Prüfungspraxis vorgestellt.

3.2 Hinweise aus der schriftlich-kaufmännischen Prüfungspraxis

Zentrale Abschlussprüfungen in der beruflichen Bildung als summative Assessments (Vogel, 2014, S. 16) haben vielerlei Funktionen (vgl. Kapitel 0, insbesondere Kapitel 2.4.3). Anhand der Prüfungsergebnisse sollen Aussagen darüber getroffen werden können, ob und wie gut der Prüfling das Ausbildungsziel erreicht hat. Schließlich wird auch im BBiG formuliert, dass im

Rahmen der (Abschluss-) ⁶⁷Prüfung die berufliche Handlungskompetenz ⁶⁸ des zu Prüfenden ermittelt werden soll (§ 38 BBiG). Die AkA (2023b) und ZPA Nord-West (2021), die zentral bundesweit (außer in Baden-Württemberg) für die Erstellung der schriftlichen Prüfungsaufgaben für kaufmännisch(-verwandte) Ausbildungsberufe zuständig sind (AkA, 2023b), fordern Validität. Prüfungen als Assessments sollen den curricularen Anforderungen entsprechend gestaltet werden (im Sinne der Curriculum-Instruktion-Assessment-Triade von Pellegrino, 2010; vgl. z. B. auch Winther, Klotz & Sangmeister, 2015, S. 6) und kompetenzorientiert sein (vgl. Wilbers, 2018, S. 145).

Den Prozess, Prüfungsaufgaben basierend auf Inhaltsstandards zu formulieren, beschreiben La Marca et al. (2000, S. 5) als komplex. So besteht die Frage, ob durch die jeweilige Aufgabe der Inhaltsstandard gemessen werden kann. Die Autor*innen plädieren für eine multiprofessionelle Arbeit bei der Erstellung von Prüfungsaufgaben (La Marca et al., 2000, S. 5). Bei der Erstellung und Auswahl von schriftlich-kaufmännischen Prüfungsaufgaben sind mehrere Professionen beteiligt (vgl. AkA, 2022; ZPA Nord-West, 2021), womit dieser Anforderung entsprochen wird. Zudem sehen es La Marca et al. (2000, S. 6) als notwendig, dass die Aufgaben mehrfache Begutachtungen (technische, qualitative sowie statistische) durchlaufen. In der technischen Begutachtung wird überprüft, ob mit der Aufgabe bzw. dem Aufgabentool die Inhaltsstandards gemessen werden können. Bei der qualitativen Begutachtung wird festgestellt, ob es inhaltliche Verzerrungen in der Aufgabe bzw. im Aufgabenpool gibt, z. B. durch Nutzung von Stereotypen. In der statistischen Begutachtung geht es um psychometrische Eigenschaften und die Item-Funktionalität. Statistische und qualitative Beobachtung sind nicht unabhängig voneinander, auch hier führen die Autor*innen als Beispiel die Voreingenommenheit in Bezug auf Ethnizität auf. Der Prozess der Aufgabenerstellung der AkA (2023c) ist mehrstufig, indem mehrere Prüfungen durchlaufen werden.

Neben der Validität bestehen weitere Ansprüche hinsichtlich der *diagnostischen Qualität* von schriftlichen Prüfungsaufgaben – die Objektivität, Effektivität und Ökonomie (AkA, 2023b; ZPA Nord-West, 2021, vgl. z. B. auch Deutscher & Winther, 2019, S. 13-14; Vogel, 2014, S. 16-17 sowie Kapitel 2.3 und 2.4.5). Ingenkamp und Lissmann (2008, S. 143) kritisieren die Gütekriterien bestehender schriftlicher Instrumente, welche die schulische Leistung messen. Auch in Bezug auf die hier betrachteten kaufmännischen Prüfungsaufgaben ist diese Kritik zu

⁶⁷ Auch wenn im § 38 BBiG Bezug auf die Abschlussprüfung genommen wird, wird auch in § 48 BBiG zur Zwischenprüfung u. a. auf § 38 BBiG referiert, sodass der Gegenstand und das Ziel der Prüfungen sich nicht im Wesentlichen ändern, sondern der Zeitpunkt bzw. die Verortung in der Ausbildungsphase.

⁶⁸ Im § 38 BBiG wird von geschrieben, das als *berufliche Handlungskompetenz* verstanden wird, worauf Kapitel 1 eingegangen wurde.

finden. Während die Objektivität der Aufgaben positiv hervorgehoben wird (Deutscher & Winther, 2019, S. 13), werden in Bezug auf die Reliabilität sowohl positive als auch negative Aspekte genannt (Deutscher & Winther, 2019, S. 13-14; Klotz & Winther, 2012, S. 7-11). Die Autorinnen halten zusammenfassend fest, dass zur Reliabilität in Abschlussprüfungen „für den Bereich mittlerer Kompetenzausprägung“ (Klotz & Winther, 2012, S. 11) hohe Werte erreicht werden können, diese ansonsten aber abnehmen.

An Aufgaben besteht der Anspruch, authentisch zu sein und die Problemsituationen nah an der beruflichen Realität zu gestalten (Klotz & Winther, 2012, S. 3; Ressel & Kiesecker, 2019, S. 22-23; Winther et al., 2015, S. 6). Die Authentizität bestehender Aufgaben steht unter Kritik (z. B. Gajek et al., 2021, S. 68; Ressel & Kiesecker, 2019, S. 21; Wuttke et al., 2022, S. 42; siehe auch Weiß, 2011, S. 39 aus der Zusammenfassung nach Grosse-Beck, 1998; Hassler, 2007; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 1998). In ihrer Studie, in der 1.454 schriftliche Prüfungsaufgaben der Ausbildungsberufe der Industriekaufleute und Kaufleute für Büromanagement analysiert wurden, stufen Wuttke et al. (2022, S. 42) lediglich 21,8 % bzw. 14,9 % der Aufgaben als authentisch ein.

Des Weiteren wird gefordert, dass in Prüfungsaufgaben Probleme abgebildet werden bzw. diese problemhaltig sind (Ressel & Kiesecker, 2019, S. 22; Wuttke et al., 2022, S. 26-28), was curricular begründet ist (vgl. z. B. KMK, 2002, S. 4, 2013, S. 4, § 7 Abs. 5 BüroMKfAusbV; § 9 Abs. 3 IndKfmAusbV 2022, vgl. auch Wuttke et al., 2022, S. 27-28). Die *Problemdefinition* ist Bestandteil des *Handlungsmodells für AKA-Prüfungsaufgaben* (Badura, 2015, S. 60). Die Studie von Wuttke et al. (2022) zur Problemhaltigkeit von schriftlich-kaufmännischen Prüfungsaufgaben zeigt, dass diese nicht hinreichend berufliche Probleme abbilden und Problemlösekompetenz erfassen. Den curricularen Anforderungen wird daher nicht hinreichend gerecht. Das damit einhergehende Problem wird von Pellegrino wie folgt beschrieben: „If any of the functions [Curriculum, Instruction and Assessment, ergänzt durch Verf.] is not well synchronized, it will disrupt the balance and skew the educational process. Assessment results will be misleading, or instruction will be ineffective“ (Pellegrino, 2010, S. 5).

In den analysierten schriftlichen Prüfungsaufgaben „überwiegt die Abfrage deklarativen und prozeduralen Wissens“ (Wuttke et al., 2022, S. 44-45). Reine Wissensabfragen in Prüfungen stehen schon lange in der Kritik (vgl. z. B. Labusch & Müller, 2014, o. A.; Lorig, Bretschneider, Gutschow & Mpangara, Mirim, Weber-Höller, Robin, 2014, S. 48 sowie Weiß, 2011, S. 39 nach Grosse-Beck, 1998; Hassler, 2007; Sekretariat der Ständigen Konferenz der

Kultusminister der Länder in der Bundesrepublik Deutschland, 1998; in Bezug auf Klausuren vgl. Fritz & Laueremann, 2019, S. 69), auch wenn es Bestrebungen zur handlungsorientierten Gestaltung von Aufgaben⁶⁹ gibt (vgl. Badura, 2015). Zwar ist Wissen ein wichtiger Bestandteil der Kompetenz (Abele, 2017, S. 62-63; Deutscher & Winther, 2019, S. 12; Nickolaus, 2011, S. 164-165; vgl. auch Siegfried et al., 2019, S. 994-995), doch wird an den Prüfungen kritisiert, dass diese nicht in ein berufliches Problem eingebettet werden (Deutscher & Winther, 2019, S. 12; Wuttke et al., 2022; Zimmer, 2009, S. 61).

Prüfungsaufgaben haben den Anspruch, ein breites Spektrum an Leistungsbereichen abzudecken (AkA, 2023b; ZPA Nord-West, 2021). Problemlösekompetenz wird zur höheren Leistung gezählt (Abele et al., 2021, S. 29; Gajek et al., 2021, S. 20-21). Demnach besteht nicht der Anspruch, ausschließlich Aufgaben in die Prüfung zu stellen, die die Problemlösekompetenz erfassen. Dies würde auch nicht die berufliche Realität widerspiegeln, da Auszubildende im beruflichen Alltag auch Aufgaben bearbeiten, die keine Probleme darstellen (Wuttke et al., 2022, S. 45). Neben Aufgaben, die mit im Vergleich niedrigerer Leistung zu lösen sind, sollen aber auch Aufgaben integriert werden, die berufliche Probleme abbilden und mit denen die Problemlösekompetenz gemessen werden kann.

Trotzdem gebundene Aufgaben in kaufmännischen Prüfungen mit Blick auf die *Objektivität* als „gute Bedingungen für die Reliabilität“ (Wilbers, 2019b, S. 628), *Justiziabilität* sowie *Ökonomie* betrachtet werden (Hewlett & Kahl-Andresen, 2014, S. 8; Wilbers, 2019b, S. 628), stehen sie in der Kritik (Gajek et al., 2021, S. 68-69; Ressel & Kiesecker, 2019, S. 4; Wilbers, 2019b, S. 628; zusammengefasst von Weiß, 2011, S. 39 nach Grosse-Beck, 1998; Hassler, 2007; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 1998). Begründet wird dies damit, dass kompetenzmessende Aufgaben den Prüflingen die Möglichkeit geben müssen, individuelle Lösungen zu finden. Geschlossene Aufgaben können dies aufgrund ihrer Eigenschaften nicht bieten (Gajek et al., 2021, S. 84; Körber, 2010, S. 11). Ferner wird argumentiert, dass für die Prüfung der *beruflichen Handlungskompetenz* komplexe Aufgaben notwendig sind (Gajek et al., 2021, S. 84), was darauf schließen lässt, dass gebundene Aufgaben als weniger komplex eingestuft werden. In der Literatur finden sich jedoch Hinweise darauf, dass auch (Prüfungs-)Aufgaben im geschlossenen Format in beruflichen Situationen eingebettet und überzeugend sein können (Deutscher & Winther, 2019, S. 12,

⁶⁹ Unter Berücksichtigung dessen, dass zwischen Handlungs- und Problemorientierung bzw. Aufgaben hierzu Parallelen bestehen (vgl. Kapitel 2.5) ist dies besonders interessant.

zur Gestaltung von Disktraktoren siehe z. B. Wilbers, 2019b, S. 624-625). Deutscher und Wintner (2019, S. 12) ergänzen dazu, dass offene Aufgaben dann präferiert werden sollten, wenn für die Handlung *argumentativ-sprachliche Kompetenzen* notwendig sind. Auch ist anzunehmen, dass geschlossene Aufgabenformate einen gewissen Grad an Komplexität aufweisen können, die erforderlich ist, um (Problemlöse-)Kompetenz zu messen. Dies ist damit zu begründen, dass zwar die Offenheit der Lösungen eine Dimension ist, die die Gestaltung von problemhaltigen Aufgaben beeinflusst, aber nicht alleine darüber entscheidet, ob mit dieser Aufgabe Problemlösekompetenz gemessen werden kann (vgl. hierzu Kapitel 2.5.3).

Kritik findet sich auch in Bezug auf die Organisation von Prüfungen. Gründe hierfür sind, dass zu bestimmten Zeitpunkten geprüft wird und erbrachte (Vor-)Leistungen in den Lernorten (Betrieb und Berufsschule) keine Beachtung finden (Weiß, 2011, S. 39 nach Grosse-Beck, 1998; Hassler, 2007; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 1998; vgl. auch Ressel & Kiesecker, 2019, S. 8-9; Sekretariat der KMK, 2021b, S. 2-4; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2009, S. 2). Da der Fokus der Arbeit nicht auf der Prüfungsorganisation liegt, wird diese nicht weiter vertieft.

4 Studien zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte

4.1 Überblick über die Fragestellungen und das Design der Studien

An der Erstellung von schriftlichen, kaufmännischen Prüfungsaufgaben sind u. a. Lehrkräfte beteiligt (vgl. Kapitel 2.3). Aufgrund der Kritik an der (kaum problemhaltigen) Gestaltung der Aufgaben sowie an der diagnostischen Qualität (vgl. Kapitel 3.2) stellt sich die Frage, wie Lehrkräfte in ihrer Ausbildung (erste und zweite Phase der Lehrer*innenbildung) auf diese Tätigkeit vorbereitet werden.

Um an kaufmännisch-beruflichen Schulen als Lehrkraft tätig zu sein, gibt es in der Bundesrepublik Deutschland verschiedene Wege⁷⁰, etwa das Studium *Wirtschaftspädagogik* oder *Lehramt an Berufskollegs* oder aber über einen Quereinstieg. Entsprechend gestalten sich die Qualifikationswege unterschiedlich. Da die Betrachtung aller Möglichkeiten nicht das Ziel der vorliegenden Arbeit ist, wurde auf Wirtschaftspädagog*innen fokussiert und folgende Forschungsfragen untersucht:

*F₁: Wie werden angehende Lehrkräfte an kaufmännisch-beruflichen Schulen/Wirtschaftspädagog*innen auf die Erstellung von problemhaltigen (Prüfungs-)Aufgaben vorbereitet?*

*F₁₋₁: Wie werden angehende Lehrkräfte an kaufmännisch-beruflichen Schulen/Wirtschaftspädagog*innen im **Studium** auf die Erstellung problemhaltigen (Prüfungs-)Aufgaben vorbereitet?*

*F₁₋₂: Wie werden angehende Lehrkräfte an kaufmännisch-beruflichen Schulen/Wirtschaftspädagog*innen im **Referendariat** auf die Erstellung von problemhaltigen (Prüfungs-)Aufgaben vorbereitet?*

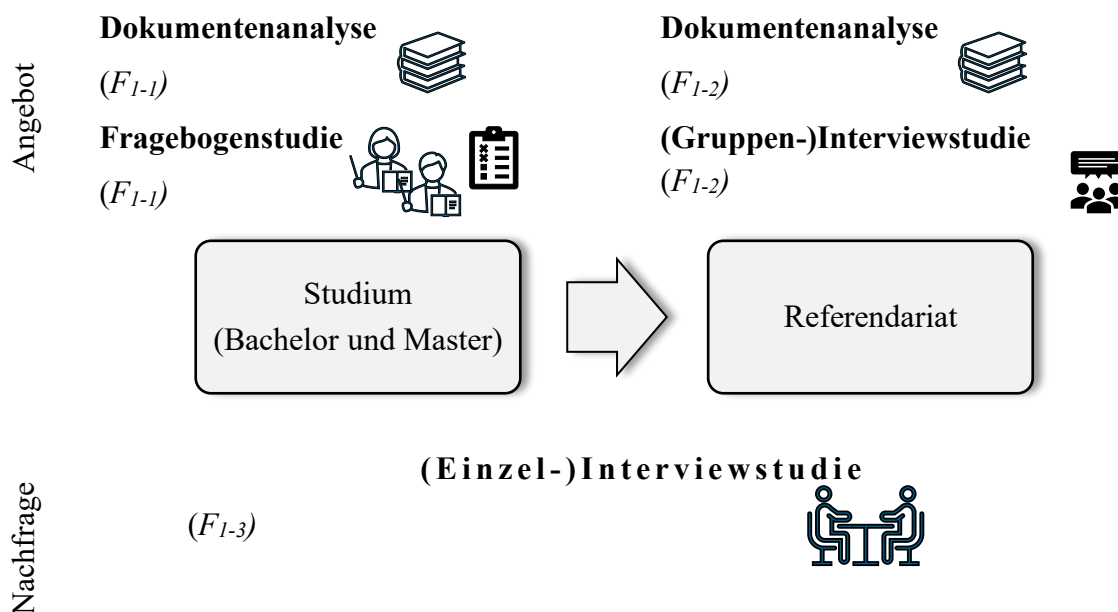
*F₁₋₃: Wie nehmen angehende Lehrkräfte an kaufmännisch-beruflichen Schulen/Wirtschaftspädagog*innen ihre Vorbereitung auf die Erstellung von problemhaltigen (Prüfungs-)Aufgaben wahr?*

Zur Beantwortung der Forschungsfragen wurden mehrere Studien durchgeführt. Abbildung 9 dient der Übersicht über die Studien zur Erfassung der Lehrenden- bzw. Angebots- (F₁₋₁ und F₁₋₂) und Lernenden- bzw. Nachfrageperspektive (F₁₋₃).

⁷⁰ Diese Wege unterscheiden sich z. T. auch unter den Bundesländern.

Abbildung 9:

Durchgeführte Studien zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte



(Eigene Darstellung)

Der Studiengang *Wirtschaftspädagogik* (erste Phase der Lehrer*innenbildung) wird fast im gesamten Bundesgebiet angeboten (vgl. Tabelle 3), jedoch sind die Studiengänge inhaltlich nicht einheitlich gestaltet. Gleiches gilt auch für das *Referendariat* (zweite Phase der Lehrer*innenbildung). Aufgrund dessen ist es notwendig, die Inhalte der Ausbildungsphasen systematisch zu untersuchen. Hierzu wurde zunächst für beide Phasen der Lehrer*innenbildung (F_{1-1} und F_{1-2}) eine *genuine Dokumentenanalyse* durchgeführt, in der bestehende Dokumente (Döring & Bortz, 2016, S. 533) berücksichtigt wurden (vgl. Kapitel 4.2.1). Unter Berücksichtigung, dass in Curricula lehramtsbezogener Studiengänge insbesondere Kompetenzen zum Professionswissen vorhanden sind (Artelt & Kunter, 2019, S. 399) und in der Arbeit das Professionswissen (vgl. Kapitel 2.1 und 2.4) fokussiert wird, wurde erwartet, mittels der Dokumentenanalyse die Forschungsfragen beantworten zu können. Bei den Dokumentenanalysen zeigten sich verschiedene Einschränkungen. Zum einen konnten keine abschließenden Ergebnisse für alle Standorte berichtet werden, da nicht alle Informationen öffentlich zugänglich waren. Zum anderen wurden einige Module in den Dokumenten nicht (ausführlich) beschrieben oder benötigen eine eigene Interpretation einzelner Stichworte (z. B. „Handlungsorientierte Prüfungen“ in der Master-Veranstaltung *Grund- und Erstausbildung* der FAU [2019a, S. 84]). Aus diesem Grund wurden weitere Studien durchgeführt.

Für die erste Phase der Lehrer*innenbildung wurde eine Fragebogenstudie mit (vertretenden) Studiengangsleiter*innen der Wirtschaftspädagogik an deutschen staatlichen Universitäten durchgeführt (vgl. Kapitel 4.2.3) und damit die Forschungsfrage F_{1-1} fokussiert. Mit der Befragung mittels eines Fragebogens sollte einerseits ermittelt werden, welche Module im wirtschaftspädagogischen Bachelor- und Masterstudium angeboten werden und ob es sich um Pflicht- oder Wahlmodule handelt. Andererseits sollte herausgefunden werden, ob und wenn ja, welche Inhalte zur Erstellung problemhaltiger (Prüfungs-)Aufgaben behandelt werden.

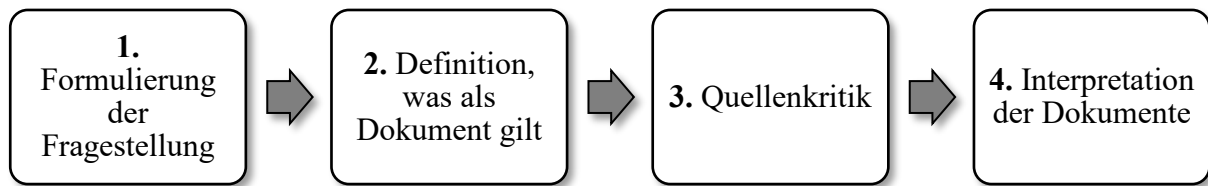
Begründet durch die wenig ausführlichen Antworten der (vertretenden) Studiengangsleitungen zu den offenen Fragen der Fragebogenstudie (vgl. Kapitel 5.1.4), wurde die geplante Fragebogenstudie mit Fachleitungen der Studienseminare verworfen. Stattdessen wurde eine Interviewstudie entwickelt und durchgeführt (vgl. Kapitel 4.2.5). Die gewählte Methode bietet – im Unterschied zur Fragebogenstudie – die Möglichkeit, Rückfragen bei Unklarheiten zu stellen (vgl. Lamnek & Krell, 2016, S. 318). So konnten Informationen über die Vorgehensweisen und Lerngelegenheiten zu pädagogisch-diagnostischen Zwecken und zur Erstellung von problemhaltigen (Prüfungs-)Aufgaben für die zweite Phase der Lehrer*innenbildung (F_{1-2}) gesammelt werden. Die ergänzenden Studien zur Dokumentenanalyse sollten dazu beitragen, belastbare Antworten zu den Forschungsfragen F_{1-1} bzw. F_{1-2} zu gewinnen.

Um neben der Angebotsperspektive zur Ausbildung von (angehenden) kaufmännisch-beruflichen Lehrkräften auch die Perspektive der Nachfrage zu beleuchten (F_{1-3}), wurde eine Interviewstudie mit Masterstudierenden der Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen durchgeführt (vgl. Kapitel 4.3). Das Studiendesign und die Auswertung der gewonnenen Daten werden in Kapitel 4.3 dargestellt. Die Ergebnisse und Implikationen finden sich in Kapitel 5.

4.2 Studien zur Angebotsperspektive

*4.2.1 Methodisches Vorgehen in den Dokumentenanalysen zur ersten und zweiten Phase der Lehrer*innenbildung*

Ein Vorteil der Dokumentenanalyse ist, dass mit bestehenden Informationen gearbeitet wird und keine neuen Daten erhoben werden müssen. Die bestehenden Informationen sind dabei „völlig unabhängig vom Forschungsprozess produziert“ (Döring & Bortz, 2016, S. 533), sodass mögliche Fehlerquellen im Rahmen einer Erhebung vermieden werden. Auch die Gefahr des subjektiven Vorgehens wird minimiert, sodass von einer besseren Durchführungsobjektivität auszugehen ist (Döring & Bortz, 2016, S. 533; Mayring, 2016, S. 47). Die Schritte der Dokumentenanalyse nach Mayring (2016, S. 48-49) sind in Abbildung 10 dargestellt.

Abbildung 10:*Ablauf der Dokumentenanalyse*

(Eigene Darstellung in Anlehnung an Mayring, 2016, S. 48-49)

Nachdem zunächst die Forschungsfragen (**Schritt 1**) festgehalten wurden (vgl. Kapitel 4.1), wurden die Dokumente definiert (**Schritt 2**), kritisch betrachtet (**Schritt 3**) und interpretiert (**Schritt 4**).

Definition der Dokumente (Schritt 2)*Zum wirtschaftspädagogischen Studium*

Bei der Analyse der Dokumente zum wirtschaftspädagogischen Studium wurden als Dokumente alle aktuellen, offiziellen und öffentlich zugänglichen Dokumente (Döring & Bortz, 2016, S. 534-535) der staatlichen Universitäten betrachtet, in denen die Module aus der ersten Phase der Lehrer*innenbildung der Wirtschaftspädagog*innen beschrieben werden. Diese finden sich in Form von Modulhandbüchern, Studiengangsbeschreibungen sowie Studien- und/oder Prüfungsordnungen wieder. Wirtschaftspädagog*innen werden in der Bundesrepublik an 25 Standorten im Rahmen von Bachelor- bzw. Masterstudiengängen ausgebildet (vgl. Tabelle 3). Da es keine bundesweit einheitliche Regelung dazu gibt, dass nur eine Studienrichtung der Wirtschaftspädagogik zum Referendariat berechtigt, wurden sowohl die Dokumente zu Studienrichtung I (ohne Zweifach) als auch zu Studienrichtung II (mit Zweifach) betrachtet. Die meisten Universitäten bieten Wirtschaftspädagogik als einen eigenständigen Studiengang an, an acht von 25 Studienstandorten ist Wirtschaftspädagogik im Bachelorstudium lediglich ein Studienschwerpunkt. Die Bachelorabschlüsse an diesen Universitäten berechtigen jedoch auch dazu, ein wirtschaftspädagogisches Masterstudium anzuschließen, was wiederum ermöglicht, mit dem Referendariat zu beginnen. Aus diesem Grund wurden diese Studiengänge in die Analyse einbezogen.

Zur Beantwortung der Forschungsfrage wurden zum Zeitpunkt der Analyse vorhandene, aktuelle Dokumente betrachtet. Aufgrund des Dokumententyps⁷¹, der Ausrichtung der Studiengänge und der Tatsache, dass nicht alle relevanten Dokumente an allen Standorten öffentlich zugänglich waren (auch auf Nachfragen war es nicht möglich, benötigte Dokumente zu erhalten), konnten keine allumfassende Analyse und Vergleich aller Standorte durchgeführt werden. Die in der Analyse einbezogenen Universitäten sind in Tabelle 3 **fett** markiert.

Tabelle 3:

*Universitäten in der BRD, an denen Wirtschaftspädagog*innen ausgebildet werden, inkl. Markierung der Standorte*

Bundesland	Universität	Bachelor	Master
Berlin	Humboldt-Universität zu Berlin	WiPäd, B. Sc.	WiPäd, M. Ed.
Mecklenburg-Vorpommern	Universität Rostock	WiPäd, B. A.	WiPäd, M. A.
Niedersachsen	Carl von Ossietzky Universität Oldenburg	WiWi – WiPäd, B. A./ B. Sc.	WiPäd, M. Ed.
	Universität Göttingen	WiPäd, B. A.	WiPäd und PE, M. Sc.
	Leuphana Universität Lüneburg	WiPäd, B. A.	Lehramt an berufsbildenden Schulen - WiWi, M. Ed.
Sachsen-Anhalt	Otto-von-Guericke Universität Magdeburg	Beruf und Bildung – WiPäd, B. Sc.	Lehramt an berufsbildenden Schulen, M. Ed.
Sachsen	Universität Leipzig	WiPäd, B. Sc.	WiPäd, M. Sc.
	Technische Universität Dresden	WiPäd, B. Sc.	WiPäd, M. Sc.

⁷¹ Betrifft die Universität Paderborn und die Technische Universität München, die kein Modulhandbuch veröffentlichten, sondern eine Online-Datenbank zu den im ausgewählten Semester angebotenen Modulen bzw. ein verlinktes Dokument zur Verfügung stellen. Diese konnten zwar möglich (für ein bestimmtes Semester) als Dokument heruntergeladen werden, allerdings würde das nicht der hier genutzten Definition von Dokumenten entsprechen und wäre im Vergleich zu den Dokumenten der anderen Standorte viel dynamischer, weshalb diese keine Beachtung fanden.

	Technische Universität Chemnitz	WiWi⁷² – Wirtschaftstraining und Bildungsmanagement⁷³, B. Sc.	Berufs- und Wirtschaftspädagogik, M.Sc.
Schleswig-Holstein	Christian-Albrechts-Universität zu Kiel	WiWi – WiPäd, B. A./ B. Sc.	WiWi – WiPäd, M. A./ M. Sc.
Nordrhein-Westfalen	Universität Duisburg-Essen	WiPäd, B. Sc.	WiPäd, M. Ed.
	Universität Paderborn	WiWi – WiPäd, B. Sc.	WiPäd, M. Sc., WiPäd, M. Ed.
	Universität zu Köln	WiWi – berufl. Fachrichtung, B. A.	WiPäd, M. Sc., WiPäd, M. Ed.
Saarland	Universität des Saarlands	WiPäd, B. Sc.	WiPäd, M. Sc.
Rheinland-Pfalz	Johannes-Gutenberg-Universität Mainz	WiPäd, B. Sc.	WiPäd I, M. Sc. WiPäd II, M. Ed.
Hessen	Universität Kassel	WiPäd, B. Ed.	WiPäd, M. Ed.
	Goethe-Universität Frankfurt	WiPäd, B. Sc.	WiPäd, M. Sc.
Thüringen	Friedrich-Schiller-Universität Jena	WiWi – WiPäd, B. A.	WiPäd I, M. Sc. WiPäd II, M. Ed.
Baden-Württemberg	Universität Hohenheim	WiPäd, B. Sc.	WiPäd, M. Sc.
	Universität Konstanz	WiWi – WiPäd, B. Sc.	WiPäd, M. Sc.
	Universität Mannheim	WiPäd, B. Sc.	WiPäd, M. Sc.
Bayern	Otto-Friedrich-Universität Bamberg	BWL/WI – WiPäd, B. Sc.	WiPäd, M. Sc.
	Friedrich-Alexander-Universität Erlangen-Nürnberg	WiWi – WiPäd, B. A.	WiPäd, M. Sc.
	Technische Universität München	--	WiPäd, M. Sc.
	Ludwig-Maximilians-Universität München	WiPäd, B. Sc.	WiPäd, M. Sc.

(Eigene Darstellung)

⁷² Das wirtschaftswissenschaftliche Studium an der Technischen Universität Chemnitz bietet zwar keinen wirtschaftspädagogischen Schwerpunkt an, der Abschluss berechtigt allerdings auch zur Aufnahme eines Master-Studiums in der Berufs- und Wirtschaftspädagogik, weshalb der Bachelor-Studiengang in die Analyse eingeschlossen wurde.

⁷³ Aus der Prüfungsordnung des Masterstudiengangs Berufs- und Wirtschaftspädagogik der Technische Universität Chemnitz (2018, S. 2444) ist zu schließen, dass der Schwerpunkt *Wirtschaftstraining und Bildungsmanagement* der entsprechende Schwerpunkt für die (Berufs- und) Wirtschaftspädagogik ist.

Zum Referendariat

Auch bei der Analyse der Dokumente zum Referendariat an kaufmännisch-beruflichen Schulen wurden jene einbezogen, die seitens der zuständigen Stellen veröffentlicht und damit verfügbar waren (vgl. Döring & Bortz, 2016, S. 534-535) und die Ausbildungsinhalte widerspiegeln. Allerdings beschränkte sich die Analyse der Dokumente auf die Bundesländer Bayern, Baden-Württemberg, Hessen, Niedersachsen, Nordrhein-Westfalen und Rheinland-Pfalz. Die Beschränkung ist damit zu erklären, dass teilnehmende LiV der Interventions- und Kontrollgruppe aus diesen Bundesländern erwartet wurden (vgl. Kapitel 7.3.2 und 7.4.1 mit Verweis auf Kapitel 4.3.3). Auf die Unterschiede, den Aufbau und den Quellen der Dokumente wird in der Quellenkritik (Schritt 3) eingegangen.

Quellenkritik (Schritt 3)

Die definierten Dokumente wurden anhand der Kriterien nach Mayring (2016, S. 48) zur *Dokumentart*, den *äußeren Merkmalen*, der *Intendiertheit*, der *Dokumentnähe zum Gegenstand* sowie der *Dokumentherkunft* kritisch betrachtet. Alle in die Untersuchung einbezogenen Dokumente sind von Institutionen veröffentlichte (oder auf Nachfrage zur Verfügung gestellte) Unterlagen, die die Ausbildungsinhalte angehender Lehrkräfte in der entsprechenden Phase berücksichtigen. Dies entspricht der Intendiertheit, da durch die Veröffentlichung bzw. Verfügbarmachung eine Transparenz für Interessierte gewährleistet wird. Die meisten Dokumente wurden bereits vor der Recherche hochgeladen, wodurch eine Stabilität angenommen wird. Für die Dokumente beider Phasen kann in Bezug auf den äußeren Zustand festgehalten werden, dass alle Dokumente digital zur Verfügung standen und einen gut lesbaren Zustand aufwiesen. Zudem wurden sowohl Dokumente zum Studium als auch zum Referendariat aus verlässlichen Quellen bezogen, nämlich aus den Homepages der jeweiligen Institutionen. Die Dokumente werden als nah am Forschungsgegenstand und dadurch als valide eingeschätzt. Im Nachfolgenden wird näher auf das genaue Vorgehen beim Auswählen und Filtern der Materialien eingegangen. Dabei wird auch das Material beschrieben, das final in die Analyse einbezogen wurde.

Zum wirtschaftspädagogischen Studium

Die definierten Dokumente werden von Studiengangsverantwortlichen (mit-)gestaltet und wurden einer Prüfung unterzogen. Sie sind rechtsverbindlich, stellen die Aktualität sicher, sind nah am Forschungsgegenstand und valide. In der Analyse zur ersten Phase der Lehrer*innenbildung wurden alle (berufs- und) wirtschaftspädagogischen, allgemeinpädagogischen sowie bildungs-

und erziehungswissenschaftlichen⁷⁴ und fachdidaktischen (Wirtschaft und Verwaltung/Wirtschaftswissenschaften)⁷⁵ Pflicht⁷⁶- und Wahlmodule betrachtet und auf deren Inhalt analysiert. Module zum Zwecke von Qualifikationsarbeiten (Bachelor- und Masterarbeiten) wurden nicht beachtet, da grundsätzlich davon auszugehen ist, dass diese auch an anderen Lehrstühlen bzw. Abteilungen außerhalb der Wirtschaftspädagogik geschrieben werden können. Auch können die Themen und die Art der Bearbeitung dieser grundsätzlich unterschiedlich gestaltet sein. Zudem konnten nur die Module ausgewertet werden, zu denen es Informationen in den analysierten Dokumenten gab. Beispielsweise werden im Modulhandbuch des Bachelorstudiums Wirtschaftspädagogik der Universität zu Köln bildungswissenschaftliche Module nicht aufgeführt. Diese werden von „weiteren lehrerbildenden Fakultäten oder den kooperierenden Hochschulen angeboten“ (Universität zu Köln, 2019a, S. 2). Insgesamt wurden Dokumente von 22 Standorten analysiert, die sich wie folgt zusammensetzen (vgl. Tabelle 4):

Tabelle 4:*Anzahl der Module zum wirtschaftspädagogischen Studium*

Bundesland	Universität	Abschluss	Anzahl Pflichtmodule	Anzahl Wahlmodule
Berlin	Humboldt-Universität zu Berlin	B. Sc.	2	--
		M. Ed.	10	1
Mecklenburg-Vorpommern	Universität Rostock	B. A.	5	--
		M. A.	--	--
Niedersachsen	Universität Göttingen	B. A.	6	--
		M. Sc.	4	2
	Leuphana Universität Lüneburg	B. A.	4	--
		M. Ed.	--	--
Sachsen-Anhalt	Otto-von-Guericke Universität Magdeburg	B. Sc.	14	3
		M. Ed.	2	--
Sachsen	Universität Leipzig	B. Sc.	5	--
		M. Sc.	3	--
	Technische Universität Dresden	B. Sc.	5	--
		M. Sc.	6	12
	TU Chemnitz	B. Sc.	2	--
		M. Sc.	7	--

⁷⁴ Der Erläuterung von Kunter, Pohlmann und Decker (2021, S. 282) folgend, sind „Erziehungswissenschaft, Psychologie und Soziologie“ unter der Bildungswissenschaft eingegliedert. Da allerdings diese teilweise getrennt berichtet werden, wurden diese, zu Zwecken der Transparenz, zunächst separat aufgeführt.

⁷⁵ Zur besseren Lesbarkeit wird im Folgenden nur noch von *(wirtschafts-)pädagogischen Modulen* gesprochen.

⁷⁶ Die Zuordnung der Wahlpflichtmodule wird zu Tabelle 6 erläutert.

Schleswig-Holstein	Universität Kiel	B. A./ B. Sc.	4	--
		M. A./ M. Sc.	6	--
Nordrhein-Westfalen	Universität Duisburg-Essen	B. Sc.	5	--
		M. Ed.	7	--
	Universität zu Köln ⁷⁷	B. A.	1	--
		M. Ed. I	8	2
		M. Ed. II	10	2
		M. Sc.	9	2
Saarland	Universität des Saarlands	B. Sc.	2	--
		M. Sc.	2	1
Rheinland-Pfalz	Johannes-Gutenberg-Universität Mainz ⁷⁸	B. Sc.	5	--
		M. Ed.	5	1
		M. Sc.	4	--
Hessen	Universität Kassel	B. Ed.	7	--
		M. Ed.	3	--
	Goethe-Universität Frankfurt	B. Sc.	5	--
		M. Sc.	6	--
Thüringen	Friedrich-Schiller-Universität Jena ⁷⁹	B. A.	7	--
		M. Ed.	7	--
		M. Sc.	7	--
Baden-Württemberg	Universität Hohenheim	B. Sc.	4	--
		M. Sc.	8	--
	Universität Konstanz ⁸⁰	B. Sc. (B)	4	--
		B. Sc. (C)	5	--
	Universität Mannheim	M. Sc.	4	--
		B. Sc.	6	2
Bayern	Otto-Friedrich-Universität Bamberg ⁸¹	M. Sc.	3	--
		M. Sc. (WiPäd I)	6	--
		M. Sc. (WiPäd II)	6	--
	Friedrich-Alexander-Universität Erlangen-Nürnberg	B. A.	6	14
		M. Sc.	5	1
		B. Sc.	2	--
Ludwig-Maximilians-Universität München	B. Sc.	2	--	
	M. Sc.	4	6	

(Eigene Darstellung)

⁷⁷ Die drei Master-Studienrichtungen überschneiden sich in sechs Pflichtmodulen. Die Wahlmodule sind dieselben.

⁷⁸ Die beiden Master-Studienrichtungen überschneiden sich in zwei Pflichtmodulen.

⁷⁹ Die beiden Master-Studienrichtungen überschneiden sich in sechs Pflichtmodulen.

⁸⁰ Da sowohl die Vertiefungsrichtung B als auch die Vertiefungsrichtung C des genannten Bachelorstudiengangs der Universität Konstanz grundsätzlich für den Master Wirtschaftspädagogik berechtigen, was wiederum den Eintritt ins Referendariat ermöglicht (bestätigt über Burkhart, 2023), wurden relevante Module beider Richtungen betrachtet.

⁸¹ Die beiden Master-Studienrichtungen überschneiden sich in vier Pflichtmodulen.

An einigen Standorten unterschied sich der Modulplan zwischen den Studienrichtungen. Gleichzeitig gab es Module, die in allen Studienrichtungen eines Standorts angeboten werden. Eine separate Betrachtung nach Studiengängen würde das Ergebnis verzerren, da zum einen einige Module doppelt gewichtet werden würden und zum anderen es auch Standorte gab, die zwar auch mindestens zwei Studienrichtungen in der Wirtschaftspädagogik anbieten, sich die Module zur (Berufs- und Wirtschafts-)Pädagogik jedoch nicht unterscheiden. So werden in der Darstellung der Ergebnisse die Module nicht nach Studienrichtung unterteilt, sondern danach, ob sie im Bachelor- oder Masterstudium angeboten werden. Die Ergebnisse werden auch nicht nach Standort getrennt dargestellt, da es sich nicht um eine Standortevaluation handelt. Informationen zur Anzahl der analysierten Module finden sich in Tabelle 5.

Tabelle 5:

Anzahl der analysierten Module zum wirtschaftspädagogischen Studium

Bundesland	Universität	Abschluss	Anzahl Pflichtmodule	Anzahl Wahlmodule
Berlin	Humboldt-Universität zu Berlin	B. Sc.	2	--
		M. Ed.	10	1
Mecklenburg-Vorpommern	Universität Rostock	B. A.	5	--
		M. A.	--	--
Niedersachsen	Universität Göttingen	B. A.	6	--
		M. Sc.	4	2
	Leuphana Universität Lüneburg	B. A.	4	--
Sachsen-Anhalt	Otto-von-Guericke Universität Magdeburg	M. Ed.	--	--
		B. Sc.	14	3
Sachsen	Universität Leipzig	M. Ed.	2	--
		B. Sc.	5	--
	Technische Universität Dresden	M. Sc.	3	--
		B. Sc.	5	--
		M. Sc.	6	12
TU Chemnitz	B. Sc.	2	--	
M. Sc.	7	--		
Schleswig-Holstein	Universität Kiel	B. A./ B. Sc.	4	--
		M. A./ M. Sc.	6	--
Nordrhein-Westfalen	Universität Duisburg-Essen	B. Sc.	5	--
		M. Ed.	7	--
	Universität zu Köln	B. A.	1	--
		M. Ed. I, M.Ed.II, M. Sc	14	2

Saarland	Universität des Saarlands	B. Sc.	2	--
		M. Sc.	2	1
Rheinland-Pfalz	Johannes-Gutenberg-Universität Mainz	B. Sc.	5	--
		M. Ed., M. Sc.	8	1
Hessen	Universität Kassel	B. Ed.	7	--
		M. Ed.	3	--
	Goethe-Universität Frankfurt	B. Sc.	5	--
		M. Sc.	6	--
Thüringen	Friedrich-Schiller-Universität Jena	B. A.	7	--
		M. Ed., M. Sc.	8	--
Baden-Württemberg	Universität Hohenheim	B. Sc.	4	--
		M. Sc.	8	--
	Universität Konstanz	B. Sc. (B, C)	7	--
		M. Sc.	4	--
	Universität Mannheim	B. Sc.	6	2
		M. Sc.	4	--
Bayern	Otto-Friedrich-Universität Bamberg	B. Sc.	3	--
		M. Sc. (WiPäd I, WiPäd II)	8	--
	Friedrich-Alexander-Universität Erlangen-Nürnberg	B. A.	6	14
		M. Sc.	5	1
	Ludwig-Maximilians-Universität München	B. Sc.	2	--
		M. Sc.	4	6

(Eigene Darstellung)

Bei der Interpretation der in Tabelle 5 angegebenen Zahlen in den Spalten *Pflichtmodule* und *Wahlmodule* ist zu berücksichtigen, dass es sich um die Anzahl der analysierten Module handelt, nicht um die tatsächlich in dem Studiengang angebotenen. Dies liegt daran, dass in den Dokumenten einerseits nicht alle relevanten Module beschrieben wurden und andererseits nicht alle Module enthalten waren. Zudem ist zu beachten, dass an den Standorten, an denen Wahlmodule angeboten werden, auch einige dabei waren, die mehr Module im Wahlbereich anboten als in Tabelle 5 abgebildet. In den Dokumenten zu den Wahlmodulen wurde teilweise der Block der Wahlmodule als ein Modul vorgestellt (z. B. Mastermodul *Berufs- und wirtschaftspädagogischer Wahlbereich* an der FAU, 2019a, S. 49-50). Weiterhin ist wichtig festzuhalten, dass es mehr Wahlmodule geben kann, die jedoch nicht alle belegt werden müssen, sondern mindestens eines zur Wahl steht (z. B. wirtschaftspädagogischer Wahlpflichtbereich im wirtschaftspädagogischen Masterstudium an der Technischen Universität Dresden, 2020). Insgesamt lässt sich das Datenmaterial wie in Tabelle 6 dargestellt beschreiben.

Tabelle 6:

Beschreibung des analysierten Materials zum Bachelor- und Masterstudium Wirtschaftspädagogik

	Bachelor			Master		
	Pflichtbereich	Wahlbereich	Gesamt	Pflichtbereich	Wahlbereich	Gesamt
$N_{\text{Universitäten}}$	22	3	22	20	8	21 ⁸²
N_{Module}	107	19	126	119	28	147
$\text{Min}_{\text{Module}}$	1	2	1	1	1	1
$\text{Max}_{\text{Module}}$	14	14	14	10	12	12
M_{Module} ⁸³	4.86	6.33	5.73	5.95	2.80	7.00
SD_{Module}	2.64	6.66	3.16	2.39	3.96	2.75

(Eigene Darstellung)

Insgesamt wurden Dokumente von 22 Standorten zum Bachelor- und 21 Standorten zum Masterstudium Wirtschaftspädagogik analysiert, wobei mindestens ein (Bachelor- und Masterstudium) und maximal 14 (Bachelorstudium) bzw. 12 (Masterstudium) Module betrachtet wurden. Der in der Tabelle 6 dargestellte Pflichtbereich enthält auch die Module, die als Wahlpflicht für Studierende der Wirtschaftspädagogik festgelegt waren. Auch im Wahlbereich finden sich die Wahlpflichtmodule, bei denen die Studierenden die Möglichkeit hatten, im pädagogischen Bereich aus diversen Modulen mindestens eines auszuwählen. Hier fällt die mangelnde Stringenz in der Organisation der Wahlpflichtmodule und deren Verständnis auf.

An drei Standorten wurden im Wahlbereich des Bachelorstudiengangs zwei bis 14 Module angeboten, aus denen Studierende eines bis vier auswählen. Bei den Masterstudiengängen boten acht Standorte Module im Wahlbereich an. Hier wurden ein bis zwölf Module angeboten. Bei der Interpretation dieser Informationen ist zu berücksichtigen, dass an manchen Standorten mehr Wahlmodule angeboten wurden als an anderen, und dadurch mehr Kodierungen entstanden. Dies bedeutet jedoch nicht, dass an diesen Standorten die Studierenden mehr Module besuchen als an anderen. Studierende wählen und belegen in der Regel eines der Module aus einem Pool an Wahlmodulen der Wirtschaftspädagogik aus.

Zum Referendariat

Die Dokumente der Bundesländer zum Referendariat wiesen ebenfalls Unterschiede auf. Beginnend mit der Bezeichnung der Veranstaltungseinheiten, heißen diese in Bayern, Hessen und

⁸² An einem Standort (Universität Rostock) konnten die Pflichtmodule des Masterstudiums Wirtschaftspädagogik nicht analysiert werden, jedoch die Wahlmodule. Insgesamt wurden also Dokumente von 21 Standorten zum Masterstudium Wirtschaftspädagogik analysiert.

⁸³ Der Mittelwert der Module im Pflicht- und Wahlbereich wurde unter Berücksichtigung der Studienrichtungen berechnet.

Rheinland-Pfalz *Module*, in Baden-Württemberg *Fachdidaktikstandards*, in Niedersachsen *Lernfelder* und in Nordrhein-Westfalen *Handlungsfelder*. Zur einheitlichen Bezeichnung werden diese im Folgenden als *Module* bezeichnet.

Während es für das Referendariat für die verschiedenen Lehrämter in Hessen (Grund-, Haupt-, Real- und Förderschulen, Gymnasien und berufliche Schulen) und Rheinlandpfalz (Grund- und Förderschulen, Realschulen plus, Gymnasien und berufsbildende Schulen) je ein gemeinsames Dokument gibt, wird die zweite Phase der Lehrer*innenbildung an beruflichen Schulen in Bayern, Baden-Württemberg, Niedersachsen und Nordrhein-Westfalen in einem separaten Dokument beschrieben. In Baden-Württemberg wird nach Technischen und Wissenschaftlichen Lehrkräften unterschieden. Aufgrund der Unterschiede in Bezug auf den Einstieg in das Referendariat und die Laufbahnen der beiden Arten von Lehrkräften, wurde das Dokument für die technischen Lehrkräfte (vgl. Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2016) nicht analysiert. Das Dokument für Bayern ist in zwei Teilen geteilt, wobei sich das eine auf das erste und das andere auf das zweite Ausbildungsjahr bezieht. In den analysierten Dokumenten für Bayern werden keine Module zur Fachdidaktik der Wirtschaftswissenschaften/Wirtschaft und Verwaltung aufgeführt, weshalb zunächst bei der zuständigen Stelle Nachforschungen angestellt wurden. Diese teilte mit, dass die genannten fachdidaktischen Module „von ... Seminarlehrkräften an den Seminarschulen vor Ort umgesetzt und laufend an aktuelle Gegebenheiten angepasst“ (M. Amend, persönl. Mitteilung, 29.10.2021) und diese daher nicht veröffentlicht werden. Aus diesem Grund konnten diese Module nicht berücksichtigt werden.

Im Dokument für Hessen, in dem die Module mehrerer Lehrämter aufgeführt werden, wurden nur die analysiert, die für das Lehramt an beruflichen Schulen relevant sind. Eine weitere Einschränkung wurde vorgenommen, indem nur diejenigen berücksichtigt wurden, die zur beruflichen Fachrichtung *Wirtschaft und Verwaltung* sowie zu *Politik und Wirtschaft* angeboten wurden. Für die Analyse der Ausbildung der Lehrkräfte an kaufmännisch-beruflichen Schulen an niedersächsischen Studienseminaren wurden zwei Dokumente analysiert, wobei diese nicht unabhängig voneinander sind und als *Landesweit abgestimmte Seminarlehrpläne für die Studienseminare in Niedersachsen* (2013a, 2013b) in Teil I und Teil II aufgeteilt werden. Diese und weitere Dokumente zu rechtlichen Grundlagen wurden auf Nachfrage beim Niedersächsischen Kultusministerium zur Verfügung gestellt (vgl. J. Leonard, persönl. Mitteilung, 30.06.2021). Laut Auskunft des Ministeriums für Schule und Bildung Nordrhein-Westfalen, bilden das *Lehrerausbildungsgesetz*, die *Ordnung des Referendariats und der Staatsprüfung*, das *Kerncurriculum*, der *Referenzrahmen Schulqualität* und der *Orientierungsrahmen* die Grundlage für das

Referendariat. Unter Berücksichtigung dieser seien die jeweiligen Zentren für Lehrerbildung in Nordrhein-Westfalen für den Aufbau des Ausbildungsprogramms zuständig. Zudem wurde mitgeteilt, dass basierend darauf die Zentren für schulpraktische Lehrerbildung die fachlichen Ausbildungsmodule festlegen (N.-H. Redenius, persönl. Mitteilung, 30.09.2021). Aufgrund dessen wurde auch das Ausbildungsprogramm eines Zentrums für Lehrerbildung (Köln) zunächst in die Analyse einbezogen. Die Auswahl ist in der Größe des Bezirkes (Köln ist der zweitgrößte in Nordrhein-Westfalen; vgl. Bezirksregierung Köln, 2021) und der unproblematischen Beschaffung des Dokumentes (Download über die Homepage des Zentrums) begründet. Da der Aufbau der Ausbildung in allen Zentren auf derselben Grundlage beruht, wurde davon ausgegangen, dass sich die Programme nicht wesentlich voneinander unterscheiden. Daher wurde auf eine Analyse aller Programme der einzelnen Zentren verzichtet. Die Analyse des Ausbildungsprogramms in Köln zeigte jedoch, dass darin nicht die Handlungsfelder differenzierter betrachtet, sondern vielmehr das Leit- bzw. Selbstbild, die Organisation und die Orientierung des Seminarstandorts beschrieben werden (vgl. Zentrum für schulpraktische Lehrerbildung Köln, 2020). Es waren demnach keine Module zur Fachdidaktik der Wirtschaftswissenschaften/Wirtschaft und Verwaltung beachtet, sodass das Ergebnis der Analyse dieses Dokuments nicht in die Ergebnisdarstellung einbezogen wurde. Auch konnten auf den Homepages der anderen Zentren keine weiteren Dokumente zur Beschreibung der fachdidaktischen Lerneinheiten gefunden werden (vgl. Zentrum für schulpraktische Lehrerbildung Aachen, 2021; Zentrum für schulpraktische Lehrerbildung Bielefeld, 2021; Zentrum für schulpraktische Lehrerbildung Dortmund, 2021; Zentrum für schulpraktische Lehrerbildung Duisburg, 2021; Zentrum für schulpraktische Lehrerbildung Düsseldorf, 2021; Zentrum für schulpraktische Lehrerbildung Gelsenkirchen, 2021; Zentrum für schulpraktische Lehrerbildung Hagen, 2021; Zentrum für schulpraktische Lehrerbildung Köln, 2020; Zentrum für schulpraktische Lehrerbildung Münster, 2021; Zentrum für schulpraktische Lehrerbildung Paderborn, 2021; Zentrum für schulpraktische Lehrerbildung Solingen, 2021). Lediglich auf den Homepages der Standorte Krefeld und Leverkusen wurde jeweils ein Dokument zur fachdidaktischen Ausbildung gefunden (vgl. Zentrum für schulpraktische Lehrerbildung Krefeld, 2021; Zentrum für schulpraktische Lehrerbildung Leverkusen, 2021). Da die Dokumente der Homepage des Zentrums in Krefeld als *Entwurf* bezeichnet wurden (vgl. z. B. Zentrum für schulpraktische Lehrerbildung Krefeld, o. A.), wurden lediglich die Dokumente aus Leverkusen zu den Fachseminaren *Wirtschaftslehre/Politik* und *Wirtschaftswissenschaften* genutzt und die fachdidaktischen Module hieraus in der Analyse berücksichtigt.

Auch die Struktur der Dokumente war unterschiedlich. So wurde im Dokument des Bundeslandes Bayern nach Modulen unterschieden, während beispielsweise in Baden-Württemberg in den Fachdidaktikstandards nach entsprechenden Fachdidaktiken unterteilt wurde. Bezüglich der Module des Referendariats in Rheinland-Pfalz fand sich der Hinweis, dass im Dokument berufspraktische und fachdidaktische Module zu finden sind und zu den Themen darin eine Konkretisierung nach Lehramt stattfindet (Ministerium der Justiz Rheinland-Pfalz, 2012, o.A.). Für Bayern bestand die Möglichkeit, die Module den Ausbildungsjahren zuzuordnen. Für den Standort Hessen mussten zwei Dokumente analysiert werden. Neben den Beschreibungen der Module für alle Lehrämter, konnte aus dem Strukturmodell für das Lehramt an beruflichen Schulen entnommen werden, welches Modul in welcher Phase zum Einsatz kommt.

In den Dokumenten fanden sich in der Regel Vorworte, in denen die Inhalte des Referendariats kurz vorgestellt werden. Diese wurden nicht kodiert, um Übergewichtungen der kodierten Module zu vermeiden – entsprechend Regel 3 (vgl. Kapitel 4.2.2) zur Vorbereitung des zu analysierenden Materials. Gleiches galt für Teil I der Seminarpläne des Niedersächsischen Kultusministeriums (2013a), welches einen Überblick über die Lernfelder und deren Organisation gibt. Da die in der Analyse betrachteten Lernfelder in Teil II der Seminarpläne des o. g. Bundeslandes näher erläutert werden, wurde Teil I nicht berücksichtigt. Mit derselben Begründung wurden in den Dokumenten zum Referendariat in Hessen (vgl. Amt für Lehrerbildung Hessen, 2012) und Baden-Württemberg (vgl. Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 4) lediglich die Module kodiert. Gleiches Vorgehen galt auch für die Dokumente in Bayern, auch hier wurden nur die Modulbeschreibungen und nicht die Modulübersichten zu Beginn des Dokuments (vgl. Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, 2019) kodiert⁸⁴. Auch das *Vorwort*, die *Informationen zum Kerncurriculum* und die *Leitgedanken der schulpraktischen Lehrerausbildung* des Landes Nordrhein-Westfalen (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021, S. 2-6) wurden nicht kodiert. Tabelle 7 beschreibt das analysierte Material⁸⁵ zum Referendariat.

⁸⁴ Lediglich in der Übersicht „Module zu ausgewählten Themen“ (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2019, o.A.) wurde in der Tabelle kodiert, da die hier benannten Module im Verlauf des Dokuments nicht konkreter aufgeführt werden.

⁸⁵ Das Material wurde unter Berücksichtigung der Regeln aus Kapitel 4.2.2 aufbereitet und kodiert.

Tabelle 7:*Beschreibung des analysierten Materials zum Referendariat*

	Module
$N_{\text{Bundesländer}}$	6
N_{Module}	108
$\text{Min}_{\text{Module Bundesland}}$	5
$\text{Max}_{\text{Module Bundesland}}$	45
$M_{\text{Module Bundesland}}$	12.00
$\text{SD}_{\text{Module Bundesland}}$	13.09

(Eigene Darstellung)

Aus den Dokumenten der sechs betrachteten Bundesländer wurden insgesamt 108 Module analysiert, wobei pro Bundesland mindestens fünf und maximal 45 Module in die Analyse eingeflossen sind. Die Interpretationen der analysierten Dokumente (**Schritt 4** nach Mayring, 2016, S. 49) folgen in Kapitel 5.1.1 zum Studium und in Kapitel 5.1.3 zum Referendariat.

4.2.2 Auswertungsverfahren in den Dokumentenanalysen

An (angehenden) kaufmännisch-berufliche Lehrkräfte werden Anforderungen und Kompetenzziele in Bezug auf ihre Ausbildung gestellt. Diese umfassen neben der Kompetenz in der Fachdidaktik auch die diagnostische Kompetenz (vgl. Kapitel 0). Beide sind erforderlich, um kompetenzorientierte Aufgaben unter Berücksichtigung diagnostischer Gütekriterien zu pädagogischen Zwecken zu erstellen. Herkunftsbegründet ist davon auszugehen, dass sich diese Anforderungen sowohl im Studium der Wirtschaftspädagogik als auch im Referendariat an kaufmännisch-beruflichen Schulen wiederfinden. In der Dokumentenanalyse wurde das Vorhandensein der geforderten Inhalte in den Modulbeschreibungen der Bachelor- und Masterstudiengänge Wirtschaftspädagogik und im Referendariat für das Lehramt an kaufmännisch-beruflichen Schulen in der Bundesrepublik Deutschland bzw. in einzelnen Bundesländern geprüft. Der Fokus lag dabei auf den (berufs- und) wirtschaftspädagogischen⁸⁶ Bereich (vgl. Kapitel 4.2.1).

Für die Inhaltsanalyse wurde das Ablaufmodell nach Mayring (2015, S. 62, zur ausführlichen Beschreibung siehe S. 50 ff.) zugrunde gelegt⁸⁷. Zunächst wurde das zu analysierende Material

⁸⁶ Die anderen genannten Bereiche in Kapitel 4.2.1 sind auch hier inbegriffen.

⁸⁷ Im Modell Mayrings (2015, S. 62) sind die Ablaufschritte z. T. anders als nachfolgend beschrieben. Doch auch Mayring (2015) selbst schreibt, dass „[d]as Ablaufmodell der Analyse ... im konkreten Fall an das jeweilige Material und die jeweilige Forschungsfrage angepasst werden [muss]“ (S. 61).

bestimmt und beschrieben. Der Kontext, in dem das Material entstand, wurde ebenfalls wiedergegeben (vgl. Kapitel 4.2.1). Anschließend wurden zur formalen Charakterisierung des Materials *Kodierregeln* aufgestellt sowie die *Kodiereinheit* bestimmt (Mayring, 2015, S. 61).

Aufbereitung der Dokumente zum Studium und Kodierregeln

In Vorbereitung auf die Kodierung mit der Auswertungssoftware MAXQDA wurden die Dokumente der Universitäten auf die Seiten gekürzt, die die relevanten Module beschreiben. Dabei wurden pro Dokument jeweils zwei Dateien angelegt; in einer wurden die Module aus dem Pflichtbereich dokumentiert, wobei diese auch den Wahlpflichtbereich enthalten. In der anderen wurden die Module aus dem Wahlbereich aufgeführt. Zweitfächer wurden nicht kodiert, da zum einen Studierende der Studienrichtung I kein Zweitfach haben und zum anderen nicht in jedem Bundesland ein Zweitfach für die Zulassung zum Vorbereitungsdienst erforderlich ist.

Auch wenn die Dokumente (für das Studium bzw. Referendariat) dieselbe Funktion haben, so unterschieden sie sich zwischen den Standorten z. T. in ihrem Aufbau. Daher wurde in den Regeln zur Analyse auch auf die formalen Aspekte des Materials eingegangen. Um eine unterschiedliche Zusammenstellung der Dokumente zu vermeiden, wurden folgende Regeln für die Vor- und Aufbereitung des Materials sowie zum Kodieren aufgestellt:

- Regel 1: Wenn im Pflichtbereich Wahlmodule zu finden sind, werden diese Module nur im Dokument zum Wahlbereich aufgenommen. Die Begründung ist, dass die Studierenden die Möglichkeit haben, ein Modul auszuwählen.
- Regel 2: Wahlpflichtmodule werden im Pflichtbereich aufgenommen, sofern diese keine Wahlmodule darstellen. Grund hierfür ist, dass die im Wahlpflichtbereich genannten Module i. d. R. nicht variieren und Studierende diese absolvieren müssen.
- Regel 3: Jedes Modul wird nur einmal analysiert. Wenn beispielsweise ein Modul in zwei Studienrichtungen aufgeführt wird, die Inhalte und die Bezeichnung des Moduls sich aber nicht für die jeweilige Studienrichtung unterscheiden, wird das Modul nur einmal kodiert. Dies dient der Vermeidung von Übergewichtungen von Modulen.
- Regel 4: Wenn im Modulhandbuch auf andere Modulhandbücher oder auf Meldungen des Prüfungsamts verwiesen wird, werden diese nicht kodiert.
- Regel 5: Im Winter 2017/18 ausgelaufene Module werden nicht kodiert.

Regel 6: Es werden nur Module betrachtet, die in den berufs- und wirtschaftspädagogischen, allgemeinpädagogischen, psychologischen, bildungs- und/oder erziehungswissenschaftlichen Bereichen und in der Fachdidaktik Wirtschaftswissenschaften/Wirtschaft und Verwaltung liegen. Module zu beispielsweise den Fachwissenschaften Wirtschaft und zu Zweitfächern werden nicht kodiert.

Regel 7: Modulinhalt zum Thema Personalmanagement werden nicht kodiert. Grund hierfür ist, dass sie einen anderen Anwendungsbereich haben als hier fokussiert. Beispiel: „Dies umfasst u. a. die gezielte Förderung diagnostischer Kompetenz (insb. für den Assessment-Bereich) und kommunikativer Kompetenz als einer Dimension von Sozialkompetenz, die im Bereich der Führung von profit- und non-profit-organisationen von hoher Bedeutung ist (z. B. in Form von Personalgesprächen oder im Konfliktmanagement)“ (Der Präsident der Humboldt Universität zu Berlin, 2008, S. 7).

Regel 8: Als Kodiereinheit wird das jeweilige Modul bestimmt. Das heißt, dass das ganze Modul – beginnend mit dem Modulnamen, endend mit dem letzten Wort des Inhalts bzw. des Ziels des Moduls – markiert und kodiert wird. Bei der Auswertung geht es schließlich nicht darum, die Häufigkeit einer Kategorie in einem Modul zu überprüfen, sondern ob diese in dem Modul vorkommt.

Regel 9: Mehrfachkodierungen sind möglich.

Regel 10: Es wird immer auf der untersten (Gliederungs-)Ebene der Kategorie kodiert.

Regel 11: Im Kodierleitfaden wird jede (Sub-)Kategorie benannt, beschrieben und die dazugehörigen Kodierregeln aufgestellt. Zudem sind aus dem Material Ankerbeispiele zu den jeweiligen Kategorien eingefügt. In Kategorien, für die es kein Beispiel aus dem Material gibt, wird ein fiktives eingefügt und dieses als solches gekennzeichnet.

Regel 12: Die Kategorien sind in Phasen (Eingang, Prozess und Output) und phasenübergreifenden Kategorien aufgeteilt. Ist aus einer Passage bzw. Modulbeschreibung ersichtlich, auf welche Phase sich die Inhalte beziehen, wird die nachfolgende Aufzählung / der nachfolgende (Neben-)Satz zu dieser Phase kodiert, auch wenn die konkrete Passage die Phasenbezeichnung nicht enthält.

Aufbereitung der Dokumente zum Referendariat und Kodierregeln

Im Analyseschritt der Ausbildungsinhalte der zweiten Phase der Lehrer*innenbildung wurden in den Dokumenten die Module *allgemein-pädagogisches Seminar* sowie Module zur *Fachdidaktik Wirtschaftswissenschaften/ Wirtschaft und Verwaltung* betrachtet. Diese Regel wurde

erweitert für den Fall, dass Module zur pädagogischen Diagnostik nicht im genannten Seminar, sondern separat aufgeführt werden. Die zum Studium aufgeführten Kodierregeln wurden auch hier angewendet, sofern sie für die Dokumente der zweiten Phase zutreffend waren.

Analyserichtung und Auswertungsmethode

Die Analyserichtung nach Mayring (2015, S. 58-60) liegt auf den betrachteten *Gegenstand* (Mayring, 2015, S. 58). Zur Beantwortung der im Vorfeld definierten Forschungsfrage wurden eingangs die theoretischen Grundlagen gelegt (vgl. insbesondere Kapitel 2), die auch im Rahmen der Analysearbeit berücksichtigt wurden. Als Auswertungsmethode wurde die strukturierte Inhaltsanalyse nach Mayring (2015, S. 98; 2016, S. 120) angewandt. Die Methodenwahl wurde als geeignet eingeschätzt, da es bei der Analyse des Materials nicht darum ging, einen Überblick über alle angebotenen Inhalte zu geben (in Anlehnung an die Beschreibung von Mayring, 2015, S. 67-68, 2016, S. 118), was beispielsweise mittels der zusammenfassenden qualitativen Inhaltsanalyse möglich ist (vgl. z. B. Mayring, 2016, S. 115). Vielmehr ging es darum herauszufinden, ob (und wie) an den Standorten zu dem hier untersuchten Themenfeld gelehrt wird. Bei diesem Verfahren ist es möglich, die Materialien nach demselben Schema zu kodieren. Dadurch wurde gewährleistet, dass in allen Materialien nach den gleichen Kriterien kodiert wird. So war eine Vergleichbarkeit der Ergebnisse unter den Dokumenten der jeweiligen Standorte möglich (vgl. hierzu auch Lamnek, 2010, S. 456).

In Vorbereitung der inhaltsanalytischen Auswertung konnten keine bestehenden Forschungsergebnisse zum Thema gefunden werden, anhand welcher ein deduktives Vorgehen bei der Entwicklung des Kategoriensystems möglich gewesen wäre. Jedoch konnten anhand der Ergebnisse der Literaturrecherche (vgl. Kapitel 0 und Kapitel 3) Erwartungen an das Material formuliert und für das Kategoriensystem zugrunde gelegt werden, um die Forschungsfragen F_{1-1} und F_{1-2} zu beantworten. Beim Kodieren des Materials stellte sich heraus, dass auch relevante Themen behandelt werden, die keiner im Voraus definierten Kategorie zugeordnet werden konnten. Aus diesem Grund wurden, neben den deduktiv formulierten Kategorien, auch induktive Kategorien gebildet. Diese werden im Verlauf der Arbeit vorgestellt. Die induktiv gebildeten Kategorien stellen keine neue theoretische Erkenntnis dar. Sie sind vielmehr als Kategorien zu verstehen, die vage formulierte Aussagen beinhalten und als sinnvoll für die Beantwortung der Forschungsfragen eingeschätzt wurden. Da zu Beginn der Analysen solch unkonkreten Passagen nicht erwartet wurden, wurden zunächst keine separaten Kategorien abgeleitet, sondern erst induktiv während der Sichtung des Materials. Hierbei wurde die angewandte Methode erweitert um die Empfehlung von Kuckartz (2016, S. 101-102), der auch induktive Kategorienbildung bei der inhaltlich-strukturierenden qualitativen Inhaltsanalyse zulässt.

Zu den deduktiv gebildeten Kategorien

Zwar liegt der Fokus der Arbeit auf der Outputdiagnostik (und Erstellung von Aufgaben für das Assessment), dennoch wurden mehr als nur diese eine Kategorie gebildet. Grund hierfür ist, dass (a) im Material nicht viele Passagen zur Aufgabenerstellung zur Outputdiagnostik zu finden waren, (b) auch weitere Inhalte zur pädagogischen Diagnostik, wie etwa Gütekriterien, von Bedeutung für die hier aufgestellten Forschungsfragen sind und (c) auch das Wissen bzw. die Kompetenzen, die in der Eingangs- und Prozessdiagnostik genutzt werden, relevant für die Outputdiagnostik sind. Es ist anzunehmen, dass diagnostische Verfahren in der Prozessdiagnostik sich im Aufbau und in der Durchführung nicht wesentlich von der Outputdiagnostik unterscheiden. Ebenso wurden unter (c) ausgewählte Inputvariablen analysiert, die für die diagnostische Tätigkeit bedeutend sind: Vorgaben, Richtlinien und Regelungen sowie Voraussetzungen der Lehrenden. Des Weiteren wurde danach untersucht, ob (d) handlungs- und problemorientierte Ansätze in den Phasen behandelt werden, weshalb auch hierzu Kategorien entwickelt wurden.

Es ist schwierig, formative und summative Assessments voneinander zu trennen. Für summative Assessments konzipierte Instrumente bzw. Verfahren und Materialien können auch in formativen Assessments genutzt werden (z. B. IHK-Prüfungsaufgaben vergangenen Jahren als Übungsaufgaben im Unterricht, vgl. Kapitel 2.4.3 und 2.4.4). Aus diesem Grund wurde im Kategoriensystem nicht danach unterschieden. Stattdessen wurde eine deduktive Unterscheidung zwischen *Input/Eingang*, *Prozess* und *Output* vorgenommen (in Anlehnung an z. B. Ditton, 2000; Frommberger, 2013; Halbheer & Reusser, 2008; Jürgens & Lissmann, 2015; Klieme et al., 2003; Schrader, 2013; Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2003, 2014; Wild & Krapp, 2006; Wilhelm & Kunina-Habenicht, 2021; Winther, 2006; Wottawa, 2006). Bei den Kategorien besteht kein Anspruch auf Vollständigkeit. Vielmehr wurden aus jeder Phase ausgewählte Inhalte beachtet, die zur Beantwortung der Forschungsfragen als hilfreich eingestuft wurden.

Die Hauptkategorie *Input* gliedert sich in zwei Subkategorien. In der ersten Subkategorie *Vorgaben, Richtlinien und Regelungen*⁸⁸ wurden Passagen kodiert, die organisatorische und/oder rechtliche Ordnungen enthalten bzw. auf diese verweisen, wie etwa Bildungsstandards, Berufsbildungs- aber auch Prüfungsrecht⁸⁹. Diese haben eine steuernde Funktion⁹⁰ für den Lehr-Lern-Prozess und Prüfungen. Die Intention war herauszufinden, inwiefern angehenden Lehrkräften

⁸⁸ Diese sind als Inputvariablen in der Bildung zu verstehen, vgl. hierzu z. B. Klieme et al. (2003, S. 11, 32).

⁸⁹ Die Subkategorie wurde nicht weiter unterteilt, da die Kodierung nach speziellen Regelungen, Richtlinien und Vorgaben nicht trennscharf war.

die Möglichkeit geboten wird, sich mit diesen auseinanderzusetzen, die relevante Inhalte für die Prüfungsaufgabenerstellung im Speziellen und zum Diagnostizieren im Allgemeinen enthalten. Nicht nur konkrete Abschnitte der Regelwerke zur Prüfung sind relevant, sondern auch beispielsweise der Rahmenlehrplan, in dem verankert ist, welche Themen Bestandteil der Ausbildung sind. Leistungserhebungen bzw. Prüfungen sollten sich, entsprechend auch der Curriculum-Instruktion-Assessment-Triade (Pellegrino, 2010), nach curricularen Vorgaben⁹⁰ richten. Das Wissen und der Umgang mit den curricularen Vorgaben sind demnach bei der Erstellung von Prüfungsaufgaben relevant. Ebenso können in den allgemeinen Formulierungen in den Curricula Informationen zu Prüfungen zu finden sein, die nicht explizit ausgewiesen sind. Gleichzeitig sind (Ausbildungs-)Standards nicht unbedeutend für die Erstellung von Prüfungen. Ausbildungsstandards enthalten u. a. Kompetenzziele, die verbindlich sind (Bundesinstitut für Berufsbildung, 2015, S. 11). Diese sind Ziel der beruflichen Bildung (§ 1 Abs. 3 BBiG), welche zum Abschluss einer Ausbildung geprüft werden sollen (§ 38 BBiG). In dieser Subkategorie sind demnach unspezifische Passagen zu finden, wie z. B. im Bachelor-Wahlmodul *Berufs- und wirtschaftspädagogische Vertiefung: Transferseminar Grundlagen der Berufsausbildung* der Universität Erlangen-Nürnberg: „Schwerpunkt des Seminars sind die rechtlichen Grundlagen der Berufsbildung in Deutschland mit dem Schwerpunkt auf der Beantwortung organisatorischer und didaktischer Fragestellungen zur Gestaltung von betrieblicher Aus- und Weiterbildung unter besonderer Berücksichtigung der Ausbildereignung“ (FAU, 2019b, S. 136). Daneben sind hierin spezifische Passagen zu finden, die Regelungen zum Assessment enthalten, wie etwa eine Passage aus der Beschreibung des Mastermoduls *Lern- und Leistungsdiagnostik in der beruflichen Bildung* der Universität Göttingen: „Spezielle rechtliche ... Aspekte der Diagnostik im betrieblichen Kontext“ (Georg-August-Universität Göttingen, 2019, S. 150). Des Weiteren finden sich in dieser Kategorie Formulierungen zu Standards, wie z. B. die Passage aus der Beschreibung des Mastermoduls *Ausbildungsstandards und Kompetenzerwerb im Lernfeld Betriebliches Rechnungswesen* der Universität Berlin, in dem „(beruflichen) Bildungs- und Ausbildungsstandards, Ausbildungsreife, Berufsfähigkeit, Beschäftigungsfähigkeit und den Kompetenzerwerb im Verlauf der beruflichen Lernprozesse der Jugendlichen“ (Der Präsident der Humboldt Universität zu Berlin, 2008, S. 11) thematisiert werden. In der zweiten Subkategorie der Hauptkategorie *Input* wurden Passagen zu *Voraussetzungen der Lehrenden* kodiert, in denen Anforderungen an die Bildungsarbeit/das Bildungspersonal genannt werden.

⁹⁰ Auf die Kritik an die kaufmännisch-berufliche Prüfungspraxis wurde in Kapitel 3.2 eingegangen.

In den Hauptkategorien *Eingangsdiagnostik*, *Prozessdiagnostik* und *Outputdiagnostik* wurden hierfür relevante Inhalte kodiert. Während in der *Eingangsdiagnostik* der Fokus auf die Lern- und Leistungsvoraussetzungen der Lernenden lag, richtete sich der Blick in der *Prozessdiagnostik* auf den Lehr-Lern-Prozess (als formative Diagnostik, vgl. Kapitel 2.4.3). Zwar gehört zur pädagogischen Diagnostik auch die Unterrichtsplanung (Jürgens & Lissmann, 2015, S. 58), doch wurde dieser in der vorliegenden Arbeit aufgrund des Schwerpunkts nicht fokussiert. Eine Zuordnung zur Hauptkategorie *Outputdiagnostik* erfolgte dann, wenn aus einer Passage hervorging, dass es sich um eine Diagnostik des Outputs handelt, d. h., wenn beispielsweise die Wörter *Prüfung*, *Klassenarbeit*, *Klausur* oder *Outputanalyse* vorzufinden waren. Bezeichnungen wie etwa Leistungs- und Kompetenzmessung wurden nicht unter die Outputdiagnostik kodiert, wohlwissend, dass diese *auch* zur Diagnostik des Outputs eingesetzt werden, aber nicht nur (vgl. Kapitel 2.4.4). Um in den Kategorien *Eingangs-*, *Prozess-* und *Outputdiagnostik* dieselbe Basis zu schaffen, wurden die nachfolgenden fünf Subkategorien in jeder dieser Hauptkategorien eingesetzt:

- *Aufgabenerstellung*: Zur Erhebung der Leistungen bzw. Kompetenzen der Lernenden ist es erforderlich, zunächst eine Situation zu entwickeln, die es ermöglicht, die Erhebung durchzuführen. Diese Kategorie wurde gebildet, da in der vorliegenden Arbeit auf schriftliche (Prüfungs-)Aufgaben fokussiert wird (in Anlehnung an z. B. La Marca et al., 2000, S. 5-6; Maier, 2015, S. 19; vgl. Kapitel 0 und 3).
- *Bewertung und Beurteilung*: Diese als Synonyme verstandenen Begriffe (z. B. Jürgens & Lissmann, 2015, S. 69; Lissmann, 2010a, S. 7) sind ein bedeutender Schritt in der pädagogischen Diagnostik und daher Teil des Kategoriensystems (in Anlehnung an z. B. Jürgens & Lissmann, 2015, S. 68-75; Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2003, S. 17; vgl. Kapitel 0 und 3).
- *Gütekriterien* werden als Qualitätsmerkmale verstanden, welche in der pädagogischen Diagnostik unterstrichen werden (in Anlehnung an z. B. Klein, 2009, S. 25-27; Knight & Yorke, 2003, S. 17-23; La Marca et al., 2000, S. 6; Wilhelm & Kunina-Habenicht, 2021, S. 319-322; vgl. Kapitel 0 und 3).
- *Handlungsorientierung* umfasst Problemorientierung, weshalb auch diese Kategorie in den Kodierleitfaden aufgenommen wurde (in Anlehnung an z. B. Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2003, S. 17; vgl. Kapitel 0 und 3).

- *Problemorientierung* wird als Kategorie aufgenommen, da in der vorliegenden Forschungsarbeit die Frage zugrunde gelegt wird, ob in der Ausbildung der Wirtschaftspädagog*innen Problemorientierung als Ansatz zur Gestaltung von Aufgaben zu diagnostischen Zwecken behandelt wird (in Anlehnung an z. B. Wuttke et al., 2022, vgl. Kapitel 0 und 3).

Wohlwissend, dass diagnostische Tätigkeiten mehr umfassen als in den Subkategorien abgebildet (vgl. Kapitel 2), wurden nur jene Aspekte beleuchtet, die dem Fokus der Forschungsarbeit entsprechen. Es besteht demnach kein Anspruch auf Vollständigkeit. Wie im Verlauf der Arbeit aufgezeigt wird, konnte nicht zu jeder dieser Subkategorien ein Ankerbeispiel aus dem Material gefunden werden, weshalb hier fiktive Beispiele eingefügt wurden. Doch auch die Erkenntnis, dass zu der jeweiligen Subkategorie keine Passage anhand des Materials kodiert werden konnte, ist für das Ergebnis relevant (siehe Kapitel 5.1).

In der Hauptkategorie *Handlungs- und Problemorientierung* wurde überprüft, ob diese Themen in den Modulen behandelt werden. Zur Handlungsorientierung wurden Passagen kodiert, die sich auf den Unterricht und/oder die Gestaltung diagnostischer Verfahren beziehen. Analog hierzu wurden zur Problemorientierung die Passagen kodiert, in denen die Problemorientierung bzw. Problemlösekompetenz angesprochen wurde(n), die sich auf den Unterricht und/oder die Gestaltung diagnostischer Verfahren beziehen. An einigen Standorten wird zwar auf die didaktische Gestaltung des Unterrichts und/oder auf das Kompetenzmanagement eingegangen, jedoch nicht weiter konkretisiert, sodass eine Kodierung hier nicht möglich war. Beispielhaft wird im Folgenden ein Lernziel des Bachelormoduls *Betriebliche Aus- und Weiterbildung* der FAU (2019b) vorgestellt: „Die Studierenden ... verstehen die Systematik sowie eingesetzte Instrumente eines betrieblichen Kompetenzmanagements“ (S. 142). Zwar kann dadurch auch auf das Assessment geschlossen werden, allerdings ist eine eindeutige Aussage nicht möglich. Bei der Kodierung wurde darauf geachtet, eine Überinterpretation der Inhalte bzw. der Ziele zu vermeiden. Das schließt allerdings nicht aus, dass in Modulen auch Ansätze zur problem- und/oder handlungsorientierten Gestaltung beinhaltet sind.

Anhand der deduktiv gebildeten Kategorien wurden die ersten Dokumente zur Beschreibung der Module im Bachelor- und Masterstudium Wirtschaftspädagogik von sechs Universitäten kodiert. Es wird empfohlen, „einen ersten, zumindest ausschnittweisen, Materialdurchgang“ (Mayring, 2015, S. 97) vorzunehmen, was ziemlich ungenau ist. Auch, wenn in dieser Arbeit insbesondere die Methodenbeschreibung von Mayring (2015, 2016) zugrunde gelegt wurde, wurden auch Empfehlungen von Kuckartz (2016, S. 102) berücksichtigt. Kuckartz (2016,

S. 102) empfiehlt für die Erstkodierung die Kodierung von 10 bis 25 % des Materials. Der prozentuale Anteil des in der Arbeit kodierten Materials liegt mit ca. 26 % nur knapp über dem Richtwert.

Zu den induktiv gebildeten Kategorien

Zu Beginn der Kodierarbeit wurde, der Empfehlung Mayrings (2015, S. 61) folgend, die Kodiereinheit bestimmt (vgl. Regel 8). Während der ersten Kodierarbeit wurden weitere, relevant erscheinende Kategorien induktiv entwickelt, die parallel dokumentiert wurden. Entsprechend dem Ablaufplan nach Mayring (2016, S. 116) wurde zunächst überprüft, ob diese Kategorien zur Beantwortung der Fragestellung beitragen können. Wurde die Frage bejaht, wurden diese Kategorien mit (Sub-)Kategoriennamen, Beschreibung und Ankerbeispiel in den Kodierleitfäden aufgenommen. Anschließend wurde das Material ein weiteres Mal gesichtet und die Kategorien kritisch überprüft, präzisiert, optimiert und ggf. die (Sub-)Kategorien neu geordnet. Anschließend fand die *Revision* nach Mayrings (2016, S. 116) statt, indem die Kodierarbeit kritisch mit Blick auf die Fragestellung und die Kategorien bewertet wurde. Kategorien, die induktiv gebildet worden sind, sich aber beim Sichten des Materials bzw. im Revisionsprozess nicht bewährt haben, wurden aus dem Kategoriensystem wieder entfernt.

Die induktiven Kategorien dienen dazu, die deduktiven Kategorien differenzierter zu betrachten. So etwa die Kategorie *1.2 Voraussetzungen der Lehrenden*, welche unterteilt wurde in *1.2.1 Unspezifische Formulierungen* und *1.2.2 Diagnostische Kompetenz*. Während in *1.2.2* Passagen kodiert wurden, bei denen diagnostische Kompetenz als konkrete bzw. notwendige Kompetenz an das Bildungspersonal formuliert wurde, wurden unter Kategorie *1.2.1* unspezifische Formulierungen zu den Anforderungen an die Bildungsarbeit kodiert.

Diagnostische Kompetenz scheint zur Anwendung diagnostischer Verfahren, u. a. auch die Erstellung von Aufgaben, selbstverständlich. Ein Beispiel für die Kategorie zeigt die Passage zur Beschreibung der Masterveranstaltung *Pädagogische Diagnostik und Evaluation*: „Die Studierenden besitzen die Kompetenz ... ihre Kenntnisse in empirischen Forschungsmethoden auf die besonderen Probleme pädagogischer Diagnostik anzuwenden“ (Universität Duisburg-Essen, 2019b, S. 24). Da allerdings zu vermuten ist, dass in den Modulbeschreibungen nicht immer einzelne Kompetenzdimensionen beschrieben werden, kann davon ausgegangen werden, dass auch in ungenauen Passagen diagnostische Kompetenzen als Anforderung enthalten sind. So beispielsweise im Bachelormodul *Schulpraktische Studien I* der Universität Mannheim (2019): „Die Studierenden ... kennen den Arbeitsplatz ‚Schule‘ sowie die damit verbundenen Anforder-

derungen“ (S. 37). Anhand dieses Beispiels wird deutlich, dass die Diagnostik auch eine Anforderung des Arbeitsplatzes ist. In den *Standards für die Lehrerbildung*, in denen die KMK die Anforderungen an die Lehrer*innen definiert, ist auch folgende Anforderung enthalten: „Lehrerinnen und Lehrer üben ihre Beurteilungs- und Beratungsaufgabe im Unterricht und bei der Vergabe von Berechtigungen für Ausbildungs- und Berufswege kompetent, gerecht und verantwortungsbewusst aus. Dafür sind hohe pädagogisch-psychologische und diagnostische Kompetenzen von Lehrkräften erforderlich“ (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004, S. 3). Ohne die Kodierung dieser und ähnlicher Passagen, bestünde die Gefahr, dass relevante Stellen unbeachtet bleiben.

In der Analyse fiel auf, dass häufig das Thema der Diagnostik im Allgemeinen in den einzelnen Phasen benannt wird. Aus diesem Grund wurden die Subkategorien *2.1 Lern- und Leistungsvoraussetzungen diagnostizieren*, *3.1 Diagnostik im Lernprozess der Lernenden* und *4.1 Diagnostik des Outputs* gebildet. So wurde beispielsweise die Passage „Diagnostik: ... Vorwissen und Kompetenzen“ (Goethe-Universität Frankfurt, 2015b, S. 41) der Kategorie 2.1 zugeordnet und die Passage „Lernprozesse ... von Schüler/-innen in ihrer Unterschiedlichkeit erkennen und diagnostizieren“ (Universität Kassel, 2014, S. 1024) der Kategorie 3.1. Die Passage "Evaluation, dabei insbesondere: Outputanalysen in Form von Prüfungen“ (Goethe-Universität Frankfurt, 2015b, S. 41) wurde der Kategorie 4.1 zugeordnet. Passagen, die mehrere Phasen ansprechen, wurden mehrfach kodiert (vgl. Regel 9). So etwa das Modul *Lern- und Leistungsdiagnostik in der beruflichen Bildung* der Georg-August-Universität Göttingen (2019) mit folgender Passage: „Mit Abschluss des Moduls sind die Studierenden in der Lage, Verfahren und Instrumente der Diagnostik von Lernvoraussetzungen, Lernprozessen und Lernergebnissen zu erläutern sowie für ausgewählte diagnostische Anlässe in einem Bereich der beruflichen Bildung kritisch zu bewerten“ (S. 149).

Des Weiteren wurde festgestellt, dass es auch Passagen zur pädagogischen Diagnostik gibt, die keine der Phasen *Eingangs-*, *Prozess-* und *Outputdiagnostik* zugeordnet werden konnten. Deshalb wurde die Hauptkategorie *5. Diagnostik in der Pädagogik* gebildet, die wiederum in folgenden Subkategorien unterteilt wurde: *5.1 Diagnostik allgemein*, *5.2 Aufgabenerstellung zur Diagnostik im Allgemeinen*, *5.3 Bewertung und Beurteilung*, *5.4 Gütekriterien* und *5.5 Diagnoseinstrumente und -verfahren*. Die induktiv gebildete Kategorie 5.5 wurde zudem weiter unterteilt. Wie bereits in Kapitel 2.4.4 erläutert, ist es bei Diagnoseinstrumenten und -verfahren nicht immer möglich zu sagen, ob sie für summative und/oder formative Assessments bzw. in welcher Phase diese genutzt werden (z. B. Schütze et al., 2008, S. 700; Seeber & Nickolaus,

2010, S. 11; Tribelhorn, 2014). Insofern würde es das Ergebnis verfälschen, wenn sie zu den Kategorien der Phasen (Hauptkategorien 2 bis 4) zugeordnet werden würden. Stattdessen fand eine Differenzierung in *5.5.1 Diagnoseinstrumente und -verfahren allgemein*, *5.5.2 Kompetenzmessung und -diagnostik*, *5.5.3 Leistungsmessung und -diagnostik* und *5.5.4 Lernerfolgskontrollen und -diagnostik* statt. In die Kategorie 5.5.1 wurden Passagen kodiert, in denen allgemeine und unspezifische Angaben zu Diagnoseinstrumenten bzw. -verfahren zu finden sind, z. B. das Mastermodul *Pädagogische Diagnostik* der Universität Duisburg-Essen (2019b), das darauf abzielt, „diagnostische Instrumente zu bewerten und selbst zu entwickeln“ (S. 24). In die folgenden drei Kategorien (5.5.2 bis 5.5.4) wurden Passagen kodiert, in denen Verfahren und Instrumente mit der jeweiligen oder ähnlichen Bezeichnung zu finden waren. Diese wurden bewusst separat betrachtet, da die Begriffe *Kompetenz*, *Leistung* und *Lernergebnis* nicht gleich zu setzen sind, auch wenn sie sich überschneiden (vgl. Kapitel 2.4.4). Bei der Analyse der Dokumente stellte sich schließlich heraus, dass nicht alle Universitäten dieselben Bezeichnungen nutzen bzw. dieselben Instrumente und Verfahren zur Diagnostik behandeln.

Im ersten Materialdurchgang zeigte sich zudem, dass im Hinblick auf Handlungs- und Problemorientierung nicht immer konkret angegeben wurde, in welcher Phase diese behandelt werden. Daher wurden mit einer entsprechenden Beschreibung und einem Ankerbeispiel die Subkategorien *6.1 Handlungsorientierung* und *6.2 Problemorientierung* unter die Hauptkategorie *6. Handlungs- und Problemorientierung* gefasst.

Häufig waren auch Passagen zu finden, in denen aktuelle Fragen der (Berufs- und) Wirtschaftspädagogik thematisiert werden. Da das Thema der Kompetenzmessung auch eine aktuelle Fragestellung der Forschung ist (z. B. Klotz, 2015; Lohmeyer & Velten, 2019; Seifried, Wuttke, Kögler et al., 2015; Winther, 2010; Wuttke et al., 2022; vgl. auch Kapitel 1 und 3.2), wurden auch diese Passagen erfasst (*7. Aktuelle Fragestellungen*). Als Beispiel soll das Modul *Aktuelle Fragen der Wirtschaftspädagogik* der Friedrich-Schiller-Universität Jena (2019) dienen: „Das Modul greift aktuelle Diskussionslinien der Wirtschaftspädagogik auf, wobei wissenschaftliche Bedeutsamkeit mit bildungspraktischer und bildungspolitischer Relevanz verknüpft werden“ (S. 224). Bei der Interpretation dieser Kategorie ist zu beachten, dass es sein kann, dass auch für die Forschungsfrage irrelevante Themen behandelt werden. Während also in den Kategorien 1 bis 5 Passagen kodiert wurden, die sich auf die pädagogische Diagnostik beziehen, ist dies in den Kategorien 6 und 7 nicht zwingend der Fall.

Die Kategorien wurden schließlich auf ihre Reliabilität geprüft. Hierfür wurden zunächst die Modulhandbücher eines Standorts zur Kodierer*innenschulung genutzt. Danach wurde der Kodierleitfaden validiert, unkonkrete Kodierregeln angepasst und entsprechende Ankerbeispiele aus dem Material hinzugefügt – oder aufgrund von fehlenden echten Beispielen, fiktive gewählt. Schließlich wurden ca. 20 % des Materials (dies entspricht den Modulhandbüchern zum Bachelor- und Masterstudium von vier Standorten) doppelt kodiert und die Intercoderrealibali-tät (Kappa-Werte, κ) einzelner Kategorien berechnet. Die Berechnung erfolgte nach Brennan und Prediger (1981) über MAXQDA. Die Werte für die (Sub-)Kategorien bis zur zweiten Ebene werden in Tabelle 8 dargestellt, die Werte für die tieferen Kategorienebenen finden sich in Anhang A.

Tabelle 8:*Reliabilitäten der Kategorien in der Dokumentenanalyse*

(Sub-)Kategorien		κ	
1. Input	1.1 Vorgaben, Richtlinien und Regelungen	.75	.79
	1.2 Voraussetzungen der Lehrenden	.90	
2. Eingangsdiagnostik	2.1 Diagnostik der Lern- und Leistungsvoraussetzungen	.81	.91
	2.2 Aufgabenerstellung zur Eingangsdiagnostik	k. K.	
	2.3 Bewertung und Beurteilung der Lern- und Leistungsvoraussetzungen	1.00	
	2.4 Gütekriterien der Eingangsdiagnostik	1.00	
	2.5 Handlungsorientierung in der Eingangsdiagnostik	k. K.	
	2.6 Problemorientierung in der Eingangsdiagnostik	k. K.	
3. Prozessdiagnostik	3.1 Diagnostik des Lernprozesses der Lernenden	.73	.85
	3.2 Aufgabenerstellung zur Prozessdiagnostik	1.00	
	3.3 Bewertung und Beurteilung der Lernenden im Lehr-Lern-Prozess	1.00	
	3.4 Gütekriterien in der Prozessdiagnostik	k. K. ^{a)}	
	3.5 Handlungsorientierung im Prozess	.73	
	3.6 Problemorientierung im Prozess	.81	
4. Outputdiagnostik	4.1 Diagnostik des Outputs	1.00	1.00
	4.2 Aufgabenerstellung zur Outputdiagnostik	1.00	
	4.3 Bewertung und Beurteilung des Outputs	k. K. ^{a)}	
	4.4 Gütekriterien in der Outputdiagnostik	1.00	
	4.5 Handlungsorientierung in der Outputdiagnostik	k. K. ^{a)}	
	4.6 Problemorientierung in der Outputdiagnostik	k. K.	
5. Diagnostik in der Pädagogik	5.1 Diagnostik	.88	.83
	5.2 Aufgabenerstellung zur Diagnostik im Allgemeinen	1.00	
	5.3 Bewertung und Beurteilung	.73	
	5.4 Gütekriterien	1.00	
	5.5 Diagnoseinstrumente und -verfahren	.78	
6. Handlungs- und Problemorientierung	6.1 Handlungsorientierung	1.00	1.00
	6.2 Problemorientierung	1.00	
7. Aktuelle Fragestellungen		1.00	

a) Zu diesen Kategorien konnten im doppelt kodierten Material keine Passagen gefunden und kodiert werden. (Eigene Darstellung)

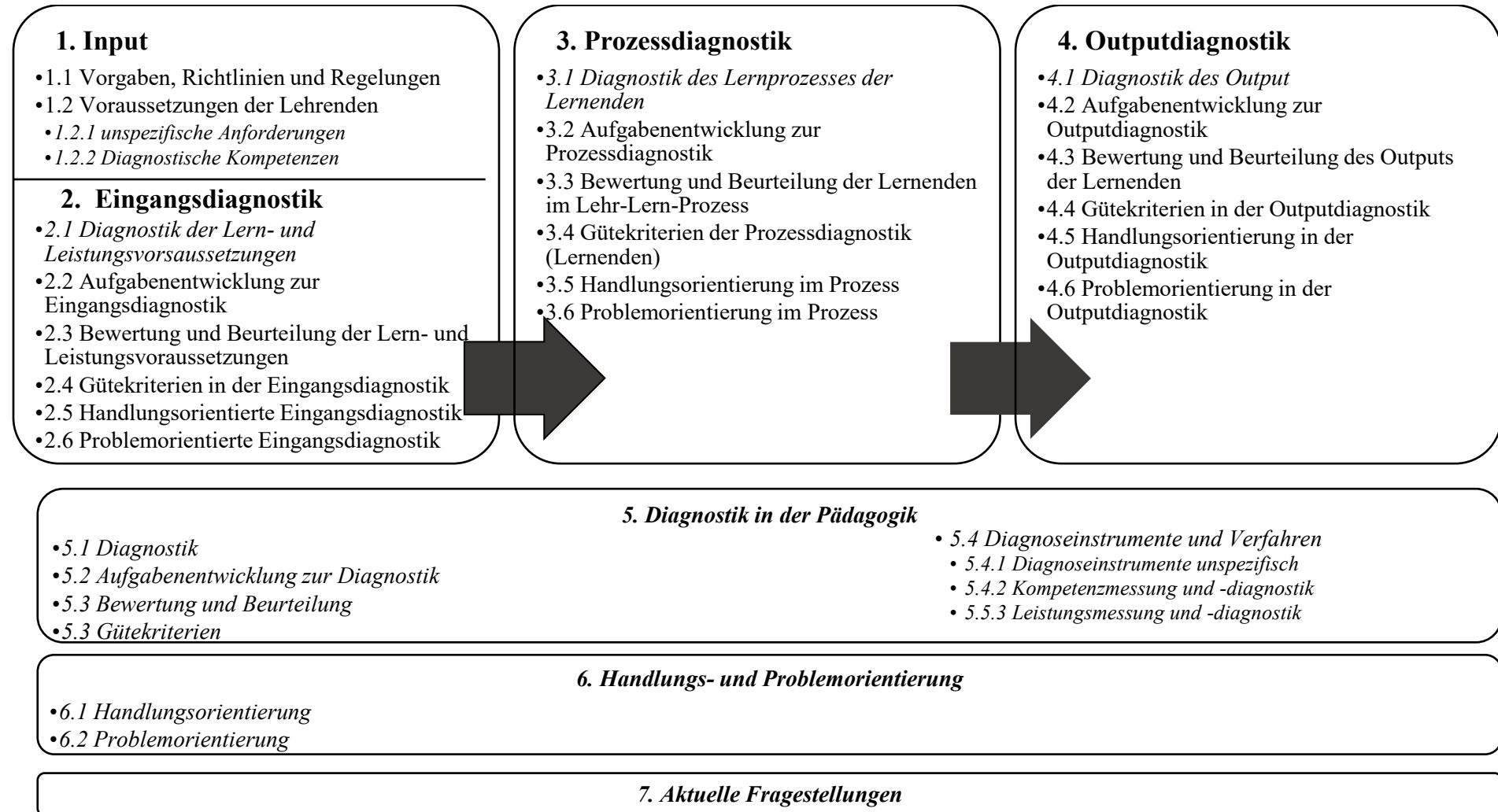
Während die geringste Übereinstimmung mit einem Wert von .67 in der Subkategorie 5.5.3 *Kompetenzmessung und -diagnostik* gemessen wurde – nach Landis und Koch (1977, S. 165) ist dies als *substanzielles*⁹¹ Ergebnis zu bezeichnen – wurde bei elf Kategorien der

⁹¹ Die Bezeichnung des Bereichs ist aus Landis und Koch (1977, S. 165) und wurde aus dem Englischen ins Deutsche übersetzt.

Höchstwert vom 1.0 erreicht. Die durchschnittliche Reliabilität über alle Kategorien liegt bei .79 und kann ebenfalls nach Landis und Koch (1977, S. 165) als *substanziell* eingestuft werden. Nach der Revision und der Doppelkodierung sowie der Optimierung des Kodierleitfadens entstand schließlich der finale Kodierleitfaden. Dieser fand bei der Analyse der Dokumente sowohl zum wirtschaftspädagogischen Studium als auch zum Referendariat Anwendung. Jedoch wurde dieser um passende Ankerbeispiele zum Referendariat ergänzt. Der finale Kodierleitfaden (Anhang A) besteht aus den in Abbildung 11 dargestellten Kategorien, wobei die kursiv markierten Kategorien diejenigen darstellen, die anhand des Materials induktiv gebildet wurden. Die Grafik dient zum besseren Verständnis.

Abbildung 11:

Grafische Darstellung der Kategorien zu den Ausbildungsinhalten



(Eigene Darstellung)

4.2.3 Methodisches Vorgehen in der Fragebogenstudie zur ersten Phase der Lehrer*innenbildung und Auswertung

Die Datenerhebung zu den Ausbildungsinhalten im wirtschaftspädagogischen Studium im Rahmen der Fragebogenstudie bei (vertretenden) Studiengangsleiter*innen erfolgte mittels eines Online-Fragebogens über *Questionstar*. Das Tool wurde ausgewählt, da es neben einer *Filter*-Funktion auch eine *Piping*-Funktion hat. Im Fragebogen fanden sich zunächst allgemeine Fragen zur wirtschaftspädagogischen Abteilung und den Studierenden am Standort. Anschließend wurde nach den Inhalten und der Methode in dem jeweiligen Modul gefragt. Hierbei wurden die Teilnehmenden gebeten, die pädagogischen Module im wirtschaftspädagogischen Bachelor- und Masterstudium ihres Standorts zu nennen. Durch die *Piping*-Funktion konnten so die genannten Module auf die nächste(n) Frage(n) übertragen werden, wodurch eine Spezifikation der Fragen zu den jeweiligen Modulen ermöglicht wurde. Die *Filter*-Funktion kam zum Einsatz, wenn in einer vorgehenden Frage festgestellt wurde, dass kein für die Forschungsfrage relevantes Thema in einem Modul behandelt wird. Dieses Modul wurde bei den folgenden Fragen eliminiert und nur noch Module angezeigt, die mindestens eines der Kriterien erfüllten. Der Fragebogen wurde im Zeitraum von 9. April 2020 bis 14. April 2020 von drei Mitarbeitenden des Lehrstuhls für Wirtschaftspädagogik der Universität Frankfurt auf Verständlichkeit überprüft, bevor die Studiengangsleiter*innen zur Teilnahme an der Studie eingeladen wurden. Insgesamt wurden 25 Leiter*innen des Studiengangs Wirtschaftspädagogik an staatlichen Universitäten der Bundesrepublik Deutschland kontaktiert. Die Einladungs-E-Mails mit dem Link zum Fragebogen wurden am 27.04.2020⁹², die Erinnerungs-E-Mails größtenteils am 04.06.2020 verschickt.

Es wäre möglich gewesen, nur den Universitäten den Fragebogen zu senden, deren Dokumente mindestens eine Einschränkung aus Kapitel 4.1 bzw. Kapitel 5.1.4 aufwiesen. Allerdings bestünde hier die Gefahr, dass die anderen Universitäten nicht die Möglichkeit gehabt hätten, Ergänzungen zu den Modulbeschreibungen aus den Dokumenten zu geben. Daher wurde die Entscheidung getroffen, alle Studiengangsleiter*innen wirtschaftspädagogischer Studiengänge in die „gezielte Stichprobe“ (Rost, 2007, S. 92) aufzunehmen und diesen die Einladung zur Teilnahme am Fragebogen zukommen zu lassen. Es wurde erwartet, dass die befragten Studiengangsleiter*innen selbst nicht immer in den Modulen lehren, weshalb sie nur bedingt Aussagen über die Gestaltung machen können. Gleichzeitig wurde jedoch angenommen, dass sie ei-

⁹² Der Studiengangsleiter der TU Chemnitz erhielt die Mail erst am 27.05.2020.

nen Überblick über die Veranstaltungen haben. Eine Kontaktierung aller Lehrenden war aufgrund des großen Umfangs nicht umsetzbar. Um negative Effekte (z. B. Ermüdung) bei der Bearbeitung des Fragebogens aufgrund von Pflichtfeldern zu vermeiden, wurden diese so gering wie möglich gehalten. Die Teilnahme am Fragebogen war freiwillig.

Die Auswertung der Fragebogenstudie erfolgte deskriptiv. Antworten auf die offenen Fragen wurden anhand des Kodierleitfadens kodiert, der für die Dokumentenanalyse entwickelt und genutzt wurde (vgl. Anhang A und Kapitel 4.2.2). Dabei wurde als Kodiereinheit auch hier das gesamte Modul bestimmt, sodass die gesamte Zelle kodiert wurde. Eine Zelle konnte dabei auch mehrfach kodiert werden. Aussagen, die nicht zur Beantwortung der Fragestellung dienten oder nur Hinweise wie „s.o.“ enthielten, wurden nicht kodiert.

4.2.4 Beschreibung der Stichprobe der Teilnehmenden an der Fragebogenstudie und deren Studienangebot

14 von 25 Studiengangsleiter*innen (kontaktierte Personen⁹³) sind der Einladung gefolgt. Dies entspricht einer Rücklaufquote von 56 %. Aus Datenschutzgründen werden die Teilnehmenden nicht nach Bundesland vorgestellt. Bei der Aufbereitung der Daten zur Auswertung fiel auf, dass teilweise Fragebögen doppelt ausgefüllt wurden, wobei nur einmalig vollständig ausgefüllte Fragebögen berücksichtigt wurden. Die Stichprobe lässt sich wie folgt beschreiben (vgl. Tabelle 9):

⁹³ An einem Standort war die Stelle bzw. die Funktion der Studiengangsleitung nicht besetzt.

Tabelle 9:*Beschreibung der Stichprobe und allgemeine Angaben zu den Standorten*

	N	Min	Max	M	SD
Lebenszeitprofessor*innen	13	0	5	2.00	1.53
Seniorprofessor*innen ⁹⁴	4	0	1	0.25	0.50
Juniorprofessor*innen	5	0	1	0.40	0.55
Sonstige	3	0	1	0.33	0.58
Post-Doc	13	0	5	2.00	1.96
Masterabsolvent*innen	13	20	85	53.08	21.07
Nach dem Studium ins Referendariat ⁹⁵	12	15	99	69.83	24.83

(Eigene Darstellung)

In den Abteilungen der Wirtschaftspädagogik der Universitäten finden sich im Durchschnitt zwei Lebenszeitprofessor*innen (SD = 1.53; Min = 0; Max = 5). Seniorprofessor*innen sind im Vergleich im Durchschnitt weniger häufig vorhanden – lediglich ein Standort, dessen Studiengangsleiter*in an der Umfrage teilnahm, verfügt über eine Seniorprofessur (M = 0.25; SD = 0.50; Min = 0; Max = 1). Ebenso finden sich wenige Standorte mit Juniorprofessor*innen. Der Mittelwert hierzu liegt bei M = 0.40 (SD = 0.55; Min = 0; Max = 1). Im Vergleich sind an den Standorten mit einem Mittelwert von M = 2.00 (Min = 0; Max = 5) relativ häufig Post-Doktorand*innen vertreten, wobei die Standardabweichung von SD = 1.96 eine große Streuung zwischen den Standorten aufzeigt. Diese Zahlen verdeutlichen die unterschiedliche Besetzung der Abteilungen an den Standorten, was sich auf das Angebot an Modulen auswirken könnte.

An den Standorten gibt es laut Studiengangsleiter*innen im Durchschnitt mindestens 20 und maximal 85 Masterabsolvent*innen jährlich (M = 53.08). Die Standardabweichung von SD = 21.07 zeigt jedoch eine breite Streuung. Zudem wurde angegeben, dass im Durchschnitt etwa 70 % (M = 69.83) der Studierenden nach dem Masterabschluss den Vorbereitungsdienst wählen (SD = 24.83; Min = 15; Max = 99). Im Durchschnitt entscheiden sich also mehr als die Hälfte der Studierenden für das Referendariat, wobei es auch hier Unterschiede zwischen den Standorten gibt.

⁹⁴ Es gab zum Teil leere Felder und Felder, in denen der Wert 0 eingetragen wurde. Diese wurden nicht geändert und in die leeren Felder wurde nicht 0 eingetragen. Leer gelassene Felder gelten somit als fehlende Werte. Das gilt für die Frage nach der Senior- und Juniorprofessor*innen und Post-Doktorand*innen.

⁹⁵ Zu dieser Frage gaben zwei Proband*innen einen Wertebereich ein. Diese wurden mit dem Mittelwert in die Berechnung einbezogen.

Aus den Antworten auf die Fragen, wie viele Module im Pflicht- und Wahlbereich des Bachelor- und Masterstudium erreicht werden können, ergibt sich folgendes Bild (vgl. Tabelle 10):

Tabelle 10:

Verteilung der Module auf den Pflicht- und Wahlbereich des Bachelor- und Masterstudiums

Module im	Bachelor		Master	
	Pflicht	Wahl	Pflicht	Wahl
Anzahl Antwortende	10			
Anzahl Module gesamt	45	6	40	11
Min _{Module}	1	0	0	0
Max _{Module}	8	2	7	7
M _{Module}	4.50	0.60	4.00	1.10
SD _{Module}	2.32	0.97	2.11	2.18

(Eigene Darstellung)

Die Anzahl der Module berechnet sich aus den Antworten zur Frage nach den Modulen im Bachelor- und Masterstudium. Die Frage nach dem Modulnamen wurde von zehn Teilnehmenden beantwortet. Daraus ergibt sich, dass an den Standorten im Bachelorstudium im Durchschnitt 4,5 Module im Pflichtbereich (SD = 2.32; Min = 1; Max = 8) und 0,6 Module im Wahlbereich (SD = 0.97; Min = 0; Max = 2) angeboten werden. Im Masterstudium liegt die Anzahl der Module im Pflichtbereich bei durchschnittlich vier Modulen (SD = 2.11; Min = 0; Max = 7) und 1,1 Modulen im Wahlbereich (SD = 2.18; Min = 0; Max = 7). Aus der Dokumentenanalyse zeigte sich, dass es mehr (wirtschafts-)pädagogische Module im Master- als im Bachelorstudium gibt (vgl. Tabelle 6). Anhand der Nennungen in den Fragebögen ergibt sich, dass es im Pflichtbereich des Bachelors (45 Module) mehr Module als im Master (40 Module) gibt – auch wenn zu beiden Studienabschnitten insgesamt jeweils 51 Module genannt wurden. Diese Diskrepanz kann etwa dadurch erklärt werden, dass nicht alle Studiengangsleiter*innen der wirtschaftspädagogischen Standorte an der Umfrage teilnahmen.

*4.2.5 Methodisches Vorgehen bei der Interviewstudie zur zweiten Phase der Lehrer*innenbildung und Auswertung*

Der Leitfaden zum Interview mit Fachleitungen der Studienseminare zur zweiten Phase der Lehrer*innenbildung ist teilstandardisiert aufgebaut (vgl. z. B. Hopf, 1995, S. 177). Mittels des Leitfadens konnten allen Teilnehmer*innen die gleichen Fragen gestellt und gleichzeitig eine gewisse Flexibilität in der Reihenfolge der Fragen und den Nachfragen geboten werden. So wurde auch eine Anpassung der Gesprächssituation ermöglicht (vgl. Berge, 2020, S. 279, 287;

Döring & Bortz, 2016, S. 322; Gläser & Laudel, 2009, S. 42; Meuser & Nagel, 2002, S. 78). Der Leitfaden gliedert sich in die (1) *Informationsphase*, (2) *Aufwärm- und Einstiegsphase*, (3) *Hauptphase* und (4) *Ausklang- und Abschlussphase* (Misoch, 2019, S. 68). Mit 14 Fragen (inkl. organisatorischer Fragen) liegt die Anzahl der Fragen im von Gläser und Laudel (2009, S. 144) empfohlenen Rahmen. Das Interview begann mit einer Erläuterung der Forschungsfrage und einer Erklärung des Ziels des Gesprächs. Es wurde erklärt, wie die Datenaufzeichnung funktioniert und dass bestimmte Datenschutzrichtlinien beachtet werden⁹⁶ sowie einem Dank für die Teilnahme (vgl. Gläser & Laudel, 2009, S. 144; Misoch, 2019, S. 68). Beendet wurde das Gespräch in der letzten Phase mit einem Dank und der Frage, ob aus Sicht der Interviewten noch weitere Themen angesprochen werden sollten und ob weitere Fragen bestehen (vgl. Misoch, 2019, S. 68-69). Der vollständige Leitfaden findet sich in Anhang B.

Den Fachleitungen der Studienseminare wurde die Möglichkeit gegeben, alleine oder mit bis zu zwei Kolleg*innen am Interview teilzunehmen. Zwar unterscheidete sich dadurch die Methode (Einzel- oder Gruppeninterview), doch wurde dies gewählt, um die Bereitschaft zur Teilnahme zu erhöhen und ergänzende Aussagen von Kolleg*innen zuzulassen. Das gewählte Verfahren zum Sampling entspricht dem *qualitativen Stichprobenplan*⁹⁷, indem darauf geachtet wurde, dass die Teilnehmenden des Interviews

- eine Position als Fachleiter*in für Wirtschaft und Verwaltung/Wirtschaftswissenschaften oder (allgemeiner) Pädagogik oder pädagogischer Psychologie⁹⁸ haben,
- aktuell in der Lehrer*innenbildung beteiligt und
- an einem Studienseminar aus den untersuchten Bundesländern tätig sind.

Die gewählte Methode und die Auswahl der Interviewbeteiligten sollte zur Validität der Studie beitragen (vgl. Mayring, 2012, S. 297). Um eine Interviewstudie mit Mitarbeitenden von Studienseminaren durchführen zu können, war es erforderlich, in den Bundesländern Bayern, Hessen, Baden-Württemberg und Nordrhein-Westfalen einen Antrag bei den zuständigen Behörden zu stellen. In den Bundesländern Niedersachsen und Rheinland-Pfalz gab es keine solche Vor-

⁹⁶ Teilnehmende erhielten bereits vor dem Interview eine Datenschutzerklärung, dem sie in Vorfeld schriftlich zugestimmt haben. Zusätzlich dazu wurde vor Beginn der Tonaufnahme erneut die Frage nach der Zustimmung gestellt und für das Band wiederholt.

⁹⁷ „Der qualitative Stichprobenplan wird bewusst so zusammengestellt, dass möglichst alle für den untersuchten Sachverhalt besonders wichtigen Merkmale und Merkmalskombinationen im Sample vorkommen“ (Döring & Bortz, 2016, S. 303).

⁹⁸ Die Bezeichnungen der Fächer haben sich von Bundesland zu Bundesland unterschieden. Bei den Einladungsschreiben wurde auf die entsprechende Bezeichnung geachtet und die Personen angesprochen, die dem untersuchten Fach entsprechend Aufgaben haben.

gabe. Die Erhebung in Bayern war nicht möglich, da die Studie nicht genehmigt wurde. Genehmigungen für die Bundesländer Baden-Württemberg, Nordrhein-Westfalen und Hessen lagen vor.

Die Durchführungszeiträume der Interviews unterscheiden sich. Dies lag vor allem daran, dass die Bearbeitungszeiten für die Genehmigungen aufseiten der Behörden der Bundesländer unterschiedlich lang waren, auch wenn die erforderlichen Anträge nahezu zeitgleich in den KW 50-51 des Jahres 2021 eingereicht wurden. Die Interviews mit den Teilnehmenden aus Bundesländern, in denen kein Antrag notwendig war, wurden in den KW 4 und 5 des Jahres 2022 durchgeführt. Das letzte Gespräch erfolgte in KW 29 des Jahres 2022. Die Interviews dauerten im Durchschnitt 43:22 Minuten (Min = 27:34 Minuten; Max = 56:10 Minuten). Die Audioaufnahmen und das jeweilige transkribierte Interview nach Kuckartz (2016, S. 164-172) wurden mit einem Vornamen und dem Buchstaben **L** für Fachleiter*innen benannt (z. B. Lars). Bei den Namen handelt es sich aus Datenschutzgründen um Pseudonyme.

Zur inhaltsanalytischen Auswertung des Interviews mit Fachleiter*innen der Studienseminare wurde der Kodierleitfaden der Dokumentenanalyse verwendet (vgl. Anhang A). Dieser war geeignet, da es sich hierbei um einen bereits erprobten und funktionierenden Leitfaden handelt (vgl. Kapitel 4.2.2). Ein weiterer Grund war, dass derselben Forschungsfrage (F_{1-2}) nachgegangen wurde. Dies ermöglichte eine Vergleichbarkeit der erzielten Ergebnisse der Studien.

Zur Auswertung wurde das Programm MAXQDA eingesetzt. Als Kodiereinheit wurden Sineseeinheiten (mindestens ein Satz) festgelegt. Beim Berichten der Ergebnisse wird auf eine Standortnennung verzichtet, was zum einen datenschutzrechtlich begründet ist. Zum anderen ist es nicht Ziel der Studie, eine Standortevaluation vorzunehmen, sondern einen Einblick in die Ausbildungsinhalte an Studienseminaren für kaufmännisch-berufliche Schulen zu gewinnen.

*4.2.6 Beschreibung der Stichprobe der Teilnehmenden der Interviewstudie zur zweiten Phase der Lehrer*innenbildung*

Der Einladung zur Teilnahme an der Studie folgten insgesamt zehn Personen der Studienseminare der Bundesländer Niedersachsen, Rheinland-Pfalz, Baden-Württemberg und Hessen (insgesamt fünf Standorte). Die Teilnehmenden waren Fachleitungen und/oder Lehrende im Bereich Wirtschaftswissenschaften⁹⁹ (acht Teilnehmende) und/oder des allgemein-pädagogischen

⁹⁹ Die Teilnehmende nannten den Bereich z. T. nur „Wirtschaft“, „Wirtschaft und Verwaltung“, „BWL mit Rechnungswesen“ o. Ä. Diese werden in der Arbeit zusammengefasst unter *Wirtschaftswissenschaften*. Welche der

Seminars¹⁰⁰ (sechs Teilnehmende). Vier Teilnehmende gaben an, für beide Bereiche zuständig zu sein bzw. darin zu lehren. Dass die Teilnehmenden zum Großteil aus dem Bereich Fachdidaktik stammten, war förderlich für die Beantwortung der Forschungsfrage, da hier der Hauptschwerpunkt zur Aufgabenerstellung liegt (Laurentia, Pos. 8; ähnliche Aussagen kamen auch von anderen Interviewpartner*innen z. B. in Lars, Lara und Lamia, Pos. 19). Auch die Klausurerstellung wurde in diesem Zusammenhang genannt (Laurentia, Pos. 12). Es zeigte sich ebenfalls als zielführend, Fachleitungen des allgemein-pädagogischen Seminars in die Stichprobe einzubeziehen, da Grundlagen pädagogischer Diagnostik hierin gelegt werden (Lars, Lara und Lamia, Pos. 19).

Fünf Teilnehmende berichteten über ihre Erfahrungsjahre am Studienseminar, der im Durchschnitt bei 7,4 Jahren lag (SD = 5.55; Min = 2; Max = 14). Den Angaben der Teilnehmenden zufolge unterscheiden sich die Standorte wesentlich in der durchschnittlichen Anzahl ihrer Absolvent*innen pro Jahr¹⁰¹. Dieser liegt im Durchschnitt bei knapp 45 Referendar*innen (M = 44.83; SD = 45.38; Min = 6; Max = 130).

4.3 Studie zur Nachfrageperspektive

4.3.1 Methodisches Vorgehen in der Interviewstudie mit angehenden Lehrkräften

Um zu den Ausbildungsinhalten angehender Lehrkräfte die Perspektive um die Nachfrageseite zu erweitern, wurden Interviews mit angehenden Lehrkräften durchgeführt. Es wurde die erste und zweite Phase der Lehrer*innenbildung betrachtet, indem Masterstudierende der Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen zum Interview eingeladen wurden.

Zur Erreichung des Ziels der Teilstudie wurden die Interviews basierend auf einem teilstandardisierten Leitfaden geführt (Anhang C). Auch hier wurde ein teilstandardisierter Leitfaden eingesetzt, um eine gewisse Flexibilität im Hinblick auf die Nachfragen zu ermöglichen (vgl. Kapitel 4.2.5). Die Phasen des Interviews sowie die Erläuterungen zu Beginn entsprechen den

Teilnehmenden welche Bezeichnung nutze, wird aus Gründen des Datenschutzes nicht angegeben, da die Bezeichnungen teilweise bundeslandspezifisch sind.

¹⁰⁰ Auch hier wurden teilweise unterschiedliche Bezeichnungen genutzt, z. B. „Pädagogik und pädagogische Psychologie“, die unter dem *allgemein-pädagogischen Seminar* zusammengefasst wurden. Aus oben genannten Gründen wird nicht angegeben, welche Person welche Bezeichnung nutzte.

¹⁰¹ Teilnehmende gaben z. T. Wertebereiche zu der Anzahl ihrer Absolvent*innen an. Hier wurde zur Berechnung der berichteten Werte der jeweilige Höchstwert des Bereichs genutzt.

Erläuterungen in Kapitel 4.2.5.¹⁰² Zusätzlich wurden die Teilnehmenden dieser Teilstudie über die Incentives informiert.

Im Interview wurden die angehenden Lehrkräfte nach ihrer Erfahrung bzw. dem hypothetischen Vorgehen beim Klausurerstellen befragt. Angehende Lehrkräfte haben i. d. R. wenig bis keine Erfahrung im Prüfungsaufgabenerstellen, auch wenn sie im Rahmen des Trainings (vgl. Kapitel 7) darauf fokussieren. Da das Klausurerstellen berufsalltagsnah ist und auch als summatives Assessment gilt (vgl. Kapitel 2.4.3 und 2.6), fiel die Entscheidung darauf. In der Abschlussphase wurde den Teilnehmenden gedankt und gefragt, ob sie weitere Themen ansprechen möchten und ihrerseits Fragen bestehen (vgl. Misoch, 2019, S. 68-69).

Auch wenn zur Beantwortung der Forschungsfrage F_{1-3} nicht systematisch nach explizit geforderten Standards gefragt wurde, ist dieses Vorgehen als Selbsteinschätzung im Sinne der Beschreibung von Terhart (2002, S. 37) zu verstehen. Das methodische Vorgehen zur Datengewinnung entspricht am ehesten einem episodischen Interview¹⁰³ (vgl. z. B. Flick, 2010, S. 281-282, 2011; Lamnek & Krell, 2016, S. 343-344), in welchem „Erzählung und Befragung – miteinander kombiniert“ (Flick, 2006, S. 222) und u. a. „Subjektive Erfahrungen“ (Misoch, 2019, S. 280) fokussiert werden. In dem durchgeführten Interview wurden die Teilnehmenden zum einen zur Erzählung ihrer bisherigen bzw. hypothetischen Vorgehensweisen zur Klausuraufgabenerstellung aufgefordert und zum anderen nach ihrem Wissen bzw. ihren Erfahrungen in der Ausbildung befragt.

Die Stichprobenauswahl erfolgte gezielt (vgl. Misoch, 2019, S. 208-209; Rost, 2007, S. 92); es wurden Masterstudierende der Wirtschaftspädagogik sowie LiV kaufmännisch-beruflicher Schulen unterschiedlicher Standorte zur Teilnahme eingeladen. Die Teilnehmenden der Interviews hatten vorab an einer Trainingsstudie teilgenommen (das Training wird in Kapitel 7 beschrieben). Die angehenden Lehrkräfte wurden am Ende des Trainings eingeladen, an einem ca. einstündigen Interview teilzunehmen. Der Einladung folgten 16 Personen – die Stichprobenbeschreibung erfolgt in Kapitel 4.3.3. Die Stichprobe war ungleich verteilt im Hinblick auf Masterstudierende und LiV, was sich aus der Ungleichverteilung der Teilnehmenden an den

¹⁰² Die Teilnehmenden erhielten bereits vor dem Interview eine Datenschutzerklärung, dem sie in Vorfeld schriftlich zugestimmt haben. Zusätzlich dazu wurde vor Beginn der Tonaufnahme erneut die Frage nach der Zustimmung gestellt und für das Band wiederholt.

¹⁰³ Den Beschreibungen von Flick (2010, S. 281-282) zufolge ist es anhand der Methode möglich, narrativ-episodisches und begrifflich semantisches Wissen zu erfassen. In den vorgestellten Interviewstudien (sowohl zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte als auch zur Trainingsevaluation, vgl. Kapitel 7.4) wird der Fokus auf das Narrativ-Episodische gelegt.

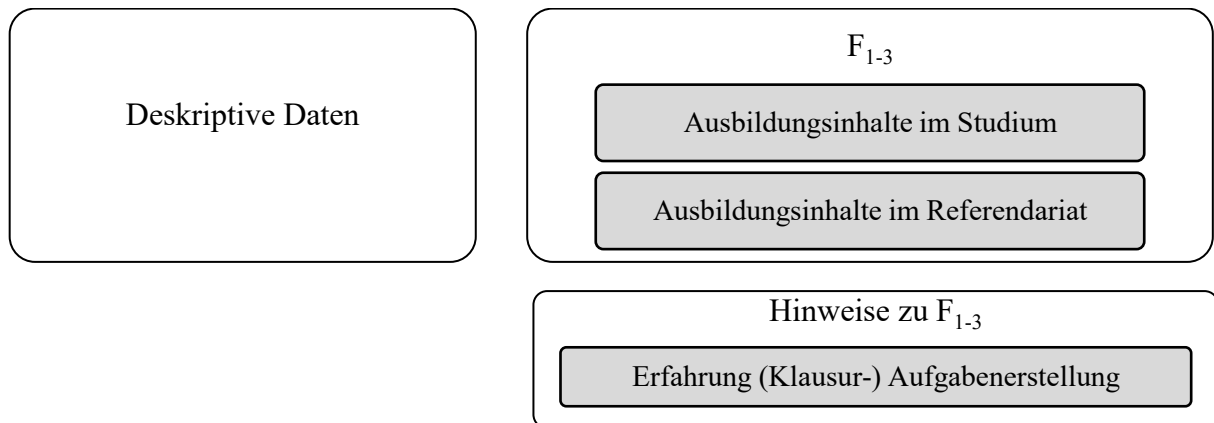
Trainings ergab (vgl. Kapitel 7.3.3). Die Stichprobe wurde nicht erweitert, da die Merkmale der Befragten, nach denen selektiert wurde, sonst nicht hätten eingehalten werden können.

Im Interview wurden die Teilnehmenden sowohl zu den Ausbildungsinhalten aus der Nachfragerperspektive befragt als auch zur Trainingsevaluation (vgl. Kapitel 7.4). Die Fragen des Interviews wurden thematisch sortiert, sodass die Fragen den jeweiligen Forschungsfragen zugeordnet werden konnten (vgl. Gläser & Laudel, 2009, S. 146; vgl. Anhang C). Die Teilnehmenden erhielten als Dank für die Teilnahme 30 Euro, was im Vorfeld angekündigt wurde.

Die online durchgeführten Interviews, welche nach Kuckartz (2016, S. 164-172) transkribiert wurden, dauerten im Durchschnitt 39:09 Minuten (Min = 27:30 Minuten; Max = 49:49 Minuten), wobei in dieser Zeit auch Daten zur qualitativen Trainingsevaluation (Kapitel 7.4) erhoben wurden. Das Interviewmaterial wurde mit einem Vornamen und dem Buchstaben **R** für LiV bzw. **Referendar*innen** (z. B. **Rebecca**) oder dem Buchstaben **S** für Studierende des Masterstudiengangs Wirtschaftspädagogik (z. B. **Sabrina**) benannt. Bei den Namen handelt es sich aus Datenschutzgründen um Pseudonyme.

4.3.2 Auswertung der Interviewstudie mit angehenden Lehrkräften

Die Daten aus dieser Interviewstudie wurden in MAXQDA ausgewertet. Speziell wurde die Methode der zusammenfassenden qualitativen Inhaltsanalyse nach Mayring (2016, S. 96) angewendet. Die Kategorien des Kodierleitfadens wurden induktiv aus dem vorliegenden Material der Interviewstudie entwickelt, wodurch von der Offenheit als Vorteil der gewählten Auswertungsmethode (Gläser-Zikuda, 2013, S. 137) gebraucht wurde. Durch die Bildung der induktiven Kategorien bei der zusammenfassenden qualitativen Inhaltsanalyse (vgl. Mayring, 2016, S. 115-117) ist es möglich, neue Erkenntnisse zu gewinnen, in dem auch „nicht erwartete Informationen aufgenommen werden können“ (Gläser-Zikuda, 2013, S. 137). Als *Selektionskriterium* für die Bildung der Kategorien muss allerdings ein *deduktives Element* gewählt werden, welches „mit theoretischen Erwägungen über Gegenstand und Ziel der Analyse begründet werden [muss]“ (Mayring, 2016, S. 115-116). In der Teilstudie der vorliegenden Arbeit wurden zunächst der *Gegenstand der Analyse* sowie die *Theorie* (Mayring, 2016, S. 116) erarbeitet und die Hauptkategorien deduktiv aus dem Leitfaden abgeleitet. Diese werden in Abbildung 12 vorgestellt.

Abbildung 12:*Hauptkategorien nach Forschungsfragen*

(Eigene Darstellung)

Die Zuordnung der Leitfragen des Interviewleitfadens zu den Forschungsfragen und den Kategorien findet sich in Anhang C. Für das Kodieren des Materials wurden im Vorfeld folgende Regeln aufgestellt:

Regel 1: Als Kodiereinheit gilt die jeweilige Zeile im Transkript (in MAXQDA), d. h. es wird die komplette Zeile markiert und kodiert, in der die passende Passage zur (Sub-)Kategorie enthalten ist.

Regel 2: Kodiert wird immer in die unterste Ebene einer (Sub-)Kategorie.

Regel 3: Es werden nur Aussagen von Proband*innen kodiert, nicht von der Interviewerin.

Regel 4: Soweit nicht anders vermerkt, werden alle (Sub-)Kategorien sowohl für Transkripte von Interviews mit Studierenden als auch mit LiV genutzt.

Regel 5: Es ist zu beachten, dass, wenn Studierende von *DBWA* bzw. *Didaktik beruflicher Aus- und Weiterbildung* oder *VFD* bzw. *Vertiefende Fachdidaktik* oder LiV vom *Training* sprechen, sie die Intervention(-sstudie) meinen.

Im Anschluss wurde das Material nochmals gesichtet, Kategorien induktiv aus dem Material abgeleitet, neue Kategorien gebildet und diese stellenweise zusammengefasst. Hierfür wurden sechs Interviews zugrunde gelegt (vier Transkripte zu Interviews mit Studierenden und zwei Transkripte zu Interviews mit LiV) und diese doppelt kodiert. Dies entspricht einem prozentualen Anteil des Gesamtmaterials von 37,5 %, was im Rahmen des von Mayring (2016, S. 117) empfohlenen Anteils liegt. Im Kodierprozess wurden sodann teilweise einzelne Definitionen überarbeitet und Regeln genauer beschrieben. Im Rahmen dessen wurden Kategorien kritisch hinterfragt und revidiert. So wurde etwa von der Frage, wo Teilnehmende das Vorgehen beim Klausurerstellen gelernt haben, auf die vorhandenen und fehlenden Inhalte im Studium bzw.

Referendariat übergeleitet. Anschließend wurden die Proband*innen nach ihren Verbesserungswünschen für die Ausbildung befragt. Die Wünsche überschneiden sich stark mit den fehlenden Inhalten, sodass diese zunächst getrennten zwei Kategorien zu einer Kategorie mit entsprechenden Subkategorien für beide Ausbildungsphasen zusammengefasst wurden. Der finale Kodierleitfaden besteht aus den folgenden vier Hauptkategorien und findet sich vollständig in Anhang D:

- *D – Deskriptive Daten*
- *E – Erfahrung problemhaltiger (Klausur-)Aufgabenerstellung*
- *S – Ausbildungsinhalte im Studium*
- *REF – Ausbildungsinhalte im Referendariat*

Eine Sättigung (vgl. Akremi, 2014, S. 277; Döring & Bortz, 2016, S. 26; Küsters, 2009, S. 48; Strübing, Hirschauer, Ayaß, Krähnke & Scheffer, 2018, S. 88-90) konnte während des Kodierens erreicht werden. Diese zeigte sich am deutlichsten in der Zielgruppe der Studierenden, und insbesondere bei den studiumsbezogenen Kategorien. Auf die referendariatsbezogenen Kategorien wird in der Limitation (vgl. Kapitel 5.2.2) explizit eingegangen. Zum Abschluss der Doppelkodierung wurden die in Tabelle 11 genannten Reliabilitätswerte erreicht, die im von Landis und Koch (1977, S. 165) definierten guten bis sehr guten Bereich liegen. Die Kappawerte (κ) zu den tieferen Ebenen der Kategorien können dem Anhang D entnommen werden.

Tabelle 11:

Kappa-Werte (κ) zu den Kategorien des Interviews zu den Ausbildungsinhalten aus Nachfragerperspektive

Kürzel	(Sub-)Kategorien	κ	
D	D1 – Berufsausbildung absolviert	.88	.83
	D2 – Wunsch ins Referendariat (nur Studierende)	.73	
	D3 – Erfahrung in Unterricht/Schule (nur Studierende)	.86	
	D4 – Ausbildungsfortschritt	.67	
E	E1 Bisheriges Vorgehen Klausurerstellung	.77	.77
S	S1 – Vorhandene Inhalte im Studium	.75	.75
	S2 – Fehlende Inhalte im Studium & Wünsche zur Verbesserung	.72	
REF	REF1 – Vorhandene Inhalte im Referendariat	.88	.85
	REF2 – Fehlende Inhalte im Referendariat & Wünsche zur Verbesserung	.78	

(Eigene Darstellung)

Mit finalen Kodierleitfaden wurden anschließend die bereits kodierten Dokumente erneut durchlaufen, um zu bestätigen, dass keine relevante Stelle übersprungen wurde. Danach wurden die verbleibenden Materialien kodiert, bevor die Ergebnisse aufbereitet und vorgestellt wurden.

4.3.3 Beschreibung der Stichprobe

An der Interviewstudie haben insgesamt 16 Personen teilgenommen, wobei drei LiV an kaufmännisch-beruflichen Schulen ($m_{\text{LiV}|\text{Interview}} = 2$; $w_{\text{LiV}|\text{Interview}} = 1$) waren und 13 Studierende des Masterstudiengangs Wirtschaftspädagogik ($m_{\text{Stud}|\text{Interview}} = 4$; $w_{\text{Stud}|\text{Interview}} = 9$). Von den LiV waren zwei aus Hessen und eine aus Rheinland-Pfalz. Sieben der 13 Studierenden waren Studierende der Universität Frankfurt, sechs der Universität Göttingen. Alle äußerten den Wunsch, nach dem Studium mit dem Referendariat zu beginnen. Von den Teilnehmenden, die bereits im Referendariat sind, waren zwei Personen im ersten Ausbildungsjahr und eine im zweiten ($M_{\text{LiV}|\text{Interview}} = 1.33$; $SD_{\text{LiV}|\text{Interview}} = 0.58$). Die studentischen Teilnehmenden waren im Durchschnitt im 2,5 Semester ($SD_{\text{Stud}|\text{Interview}} = 1.00$). Während alle teilnehmenden LiV vor dem Studium bzw. dem Referendariat mindestens eine kaufmännische Ausbildung absolviert haben, waren es unter den Studierenden elf. Eine*r studentische*r Teilnehmer*in gab an, keine Ausbildung absolviert zu haben, von einer weiteren fehlt die Angabe hierzu. Alle LiV hatten bereits Lehrerfahrung; zehn studentische Teilnehmende hatten bereits außerhalb der universitären Veranstaltungen (z. B. Praktika) Lehrerfahrung gesammelt – etwa als Vertretungslehrkraft an beruflichen Schulen. Zwei LiV konnten von ihrer Erfahrung in Bezug auf das Erstellen von Klausuren berichten. Unter den Studierenden waren es vier Personen, die Erfahrung beim Klausurerstellen hatten. Alle Interviewteilnehmer*innen haben an der Interventionsstudie (vgl. Kapitel 7) teilgenommen und diese abgeschlossen.

5 Ergebnisse, Limitationen und Implikationen der Studien zur Ausbildung angehender kaufmännisch-beruflicher Lehrkräfte

5.1 Ergebnisse der Studien zur Angebotsperspektive

5.1.1 *Vorwort zur Ergebnisdarstellung der Studien zur Angebotsperspektive*

In Kapitel 4.1 wurde dargestellt, welche Teilstudien zur Beantwortung der Forschungsfragen durchgeführt wurden. Eine separate Darstellung der Ergebnisse nach Teilstudien würde zu Redundanzen führen. Stattdessen werden durch die Darstellung des Gesamtergebnisses gleichzeitig die Vorzüge der methodischen Vielfalt genutzt, durch welche ein tiefer Einblick in das Forschungsfeld ermöglicht wird. Die Ergebnisse werden nach der Angebotsperspektive in der ersten Phase der Lehrer*innenbildung (Kapitel 5.1.2) und der zweiten Phase der Lehrer*innenbildung (Kapitel 5.1.3) vorge stellt. Da die Intention dieser Forschungsarbeit nicht ist, einzelne Standortangebote zu betrachten und zu evaluieren, sondern ein Gesamtbild für die ersten beiden Phasen der Lehrer*innenbildung an kaufmännisch-beruflichen Schulen, insb. Wirtschaftspädagog*innen, in Deutschland zu schaffen, werden im Folgenden die Ergebnisse als Gesamtbild dargestellt, wenngleich Beispiele aus den Modulen einzelner Standorte gegeben werden. Bei der Ergebnisdarstellung zur ersten Phase der Lehrer*innenbildung werden die Ergebnisse der Dokumentenanalyse und der Fragebogenstudie gemeinsam präsentiert, wobei jeweils Bezug auf die analysierten Module der Dokumente genommen wird (vgl. Kapitel 4.2.1) bzw. zu den Modulen, zu denen die Studiengangsleiter*innen Angaben machten (vgl. Kapitel 4.2.4). Zur zweiten Phase der Lehrer*innenbildung werden die Ergebnisse der Analyse entsprechender Dokumente (vgl. Kapitel 4.2.1) sowie die Ergebnisse der Interviewstudie mit den Fachleitungen der Seminare (vgl. Kapitel 4.2.6) vorgestellt. Abschließend wird in Kapitel 5.1.4 auf die Limitationen eingegangen. Da in den Studien Parallelen bestehen und die Limitationen sich nicht immer auf eine Teilstudie beschränken, werden diese nicht nach Teilstudie getrennt. Limitationen, die sich auf eine Teilstudie beschränken, werden als solche kenntlich gemacht.

5.1.2 *Ergebnisse zum wirtschaftspädagogischen Studium*

Bei der Interpretation der in Tabelle 12 vorgestellten Werte zu den Kodierungen (Codes) ist zu beachten, dass zwar jeder Code für ein Modul steht, es allerdings nicht immer entsprechend viele Module gibt, die die Themen der jeweiligen Kategorie behandeln. Das liegt daran, dass Module mehrfach kodiert werden konnten. Während der niedrigste Wert für die Summe der Kodierungen über alle Kategorien in einem Dokument bei 0 liegt, liegt der Höchstwert bei

48¹⁰⁴. Wird die Verteilung in den einzelnen Kategorien untersucht, können auch hier Unterschiede festgestellt werden. In Tabelle 12 sind diese und weitere Werte aufgeführt, wobei stets zwischen den Modulen im Bachelor- und Masterstudium sowie nach Pflicht- und Wahlmodulen unterschieden wird. Zudem wird ein Gesamtbild der Pflicht- und Wahlmodule aufgezeigt.

Mit insgesamt 36 Kodierungen im Pflichtbereich (SD = 0.93; Min = 0; Max = 3; M = 1.04) bzw. 37 Kodierungen insgesamt (SD = 1.33; Min = 0; Max = 5; M = 1.42) finden sich die meisten Kodierungen bei den untersuchten Bachelormodulen in der Kategorie *Input*. Von den analysierten 107 Pflicht- und 19 Wahlmodulen zum Bachelorstudium (vgl. Tabelle 6) behandeln demnach 37 Module Themen hierzu. Betrachtet man den höchsten Wert der Kodierungen im Masterstudium, liegt dieser in der Kategorie *Diagnostik in der Pädagogik* mit insgesamt 66 Kodierungen (SD = 2.70; Min = 0; Max = 9; M = 1.94), die meisten davon im Pflichtbereich (58 Kodierungen; M = 2.23; SD = 2.75; Min = 0; Max = 9).

¹⁰⁴ Der Wert gilt für die Dokumente zum Masterstudium an der Universität zu Köln und umfasst sowohl den Pflicht- als auch den Wahlbereich aller angebotenen Studienrichtungen zur Wirtschaftspädagogik des Standorts.

Tabelle 12:*Übersicht der deskriptiven Werte der Kodierungen in der Dokumentenanalyse zum Studium*

	Module im	Bachelorstudium			Masterstudium ¹⁰⁵				
		Pflicht	Wahl	Gesamt	Standorte gesamt	Pflicht	Wahl	Gesamt	Standorte gesamt
1. Input	NCodings;Input	36	1	37	16	24	7	31	17
	MinCodings;Input	0	0	0	22	0	0	0	21
	MaxCodings;Input	3	1	5		3	2	3	
	MCodings;Input	1.09	0.33	1.42		0.92	0.88	0.91	
	SDCodings;Input	0.95	0.58	1.33		0.94	0.99	0.93	
2. Eingangsdagnostik	NCodings;StatusD	4	0	4	2	13	2	15	8
	MinCodings;StatusD	0	--	0	22	0	0	0	21
	MaxCodings;StatusD	2	--	2		3	2	3	
	MCodings;StatusD	0.17	--	0.15		0.50	0.25	0.44	
	SDCodings;StatusD	0.58	--	0.54		0.76	0.71	0.75	
3. Prozessdiagnostik	NCodings;ProzessD	15	3	18	10	33	1	34	12
	MinCodings;ProzessD	0	0	0	22	0	0	0	21
	MaxCodings;ProzessD	4	2	4		5	1	5	
	MCodings;ProzessD	0.65	1.00	0.69		1.27	0.13	1.00	
	SDCodings;ProzessD	1.07	1.00	1.05		1.66	0.35	1.54	
4. Outputdiagnostik	NCodings;OutputD	8	0	8	4	13	1	14	4
	MinCodings;OutputD	0	0	0	22	0	0	0	21
	MaxCodings;OutputD	4	0	4		3	1	3	
	MCodings;OutputD	0.35	0.00	0.31		0.50	0.13	0.41	
	SDCodings;OutputD	0.94	0.00	0.88		1.11	0.35	0.99	
5. Diagnostik in der Pädagogik	NCodings;Diagnostik	18	4	22	9	58	8	66	14
	MinCodings;Diagnostik	0	0	0	22	0	0	0	21
	MaxCodings;Diagnostik	4	4	4		9	7	9	
	MCodings;Diagnostik	0.78	1.33	0.85		2.23	1.00	1.94	
	SDCodings;Diagnostik	1.24	2.31	1.35		2.75	2.45	2.70	
6. Handlungs- und Problemlösungen	NCodings;HPO	7	0	7	3	1	0	1	1
	MinCodings;HPO	0	--	0	22	0	--	0	21
	MaxCodings;HPO	4	--	4		1	--	1	
	MCodings;HPO	0.30	--	0.27		0.04	--	0.03	
	SDCodings;HPO	0.88	--	0.83		0.20	--	0.17	
7. Aktuelle Fragestellungen	NCodings;Aktuell	11	1	12	8	21	12	33	16
	MinCodings;Aktuell	0	0	0	22	0	0	0	21
	MaxCodings;Aktuell	2	1	2		3	3	3	
	MCodings;Aktuell	0.48	0.33	0.46		0.81	1.50	0.97	
	SDCodings;Aktuell	0.67	0.58	0.65		0.90	1.07	0.97	
Gesamt	NCodings;GESAMT	99	9	108	20	163	31	194	20
	MinCodings;GESAMT	0	1	0	22	0	0	0	21
	MaxCodings;GESAMT	18	5	19		21	13	48	
	MCodings;GESAMT	4.30	3.00	5.14		6.27	3.88	9.24	
	SDCodings;GESAMT	3.77	2.00	4.85		5.26	3.94	10.06	

(Tabelle 12: Eigene Darstellung)

In den 20 von 22 analysierten Dokumenten wurden insgesamt 108 Kodierungen zum Bachelorstudium vorgenommen (99 Kodierungen im Pflicht- und neun im Wahlbereich). Im Masterstudium wurden in 20 von 21 analysierten Dokumenten insgesamt 194 Kodierungen vorgenommen (163 Kodierungen im Pflicht- und 31 im Wahlbereich). Die deskriptiven Werte zeigen in der Gesamtbetrachtung Unterschiede in der Verteilung der Kodierungen. Während im Masterstudium der Mittelwert bei $M = 9.24$ ($SD = 10.06$) liegt, ist dieser im Bachelorstudium mit $M = 5.14$ ($SD = 4.85$) geringer. Die Standardabweichung weist in den Masterstudiengängen eine stärkere Streuung auf, werden die Bachelor- und Masterstudiengänge der Standorte verglichen. Das heißt, dass sich die Masterstudiengänge der Standorte stärker in den beobachteten Themen unterscheiden als die Bachelorstudiengänge.

Die Ergebnisse zeigen, dass Themen der pädagogischen Diagnostik im wirtschaftspädagogischen Studium verankert sind, wobei die Analyseergebnisse auch zeigen, dass diese häufiger im Master- als im Bachelorstudium thematisiert werden. Dies kann damit erklärt werden, dass im Masterstudium im Durchschnitt mehr (wirtschafts-)pädagogische Module angeboten werden ($M = 7.00$; $SD = 2.75$) als im Bachelorstudium ($M = 5.73$; $SD = 3.16$), und demzufolge auch mehr Mastermodule analysiert wurden als Bachelormodule (126 Bachelor- und 147 Mastermodule, vgl. Tabelle 6). Insgesamt – und im Vergleich zu den phasenorientierten Kategorien (Kategorien 2 bis 4) – wurden die meisten Kodierungen in der Kategorie 5. *Diagnostik in der Pädagogik* vorgenommen (22 Kodierungen im Bachelor- und 66 Kodierungen im Masterstudium). Themen der Prozessdiagnostik werden häufiger thematisiert als Themen der Eingangs- und Outputdiagnostik (vgl. Tabelle 12). Das Erstellen von Assessmentaufgaben wird selten thematisiert. Vielmehr erwecken die Beschreibungen der Module den Eindruck, dass es um das Kennen und Verstehen sowie die kritische Bewertung bestehender Instrumente geht. Dass insbesondere Themen zur Prozessdiagnostik stärkere im Fokus des Studiums stehen (gemessen an den Kodierungen), zeigt sich deutlich an den häufigen Kodierungen in den Subkategorien Handlungs- und Problemorientierung im Prozess (vgl. Tabelle 13). Bei der Interpretation ist zu beachten, dass unklar ist, ob diese Methoden tatsächlich (auch) in der Diagnostik im Prozess eine Rolle spielen oder sich (ausschließlich) auf die Unterrichtsgestaltung beziehen. Um eine Unterschätzung zu vermeiden, wurden diese dennoch kodiert. So fanden sich relativ häufig Kodierungen zu den systematisch angelegten Subkategorien *Diagnostik* (für unkonkrete Passagen), *Aufgabenerstellung*, *Bewertung und Beurteilung* sowie *Gütekriterien* (vgl. Kapitel 4.2.2).

Tabelle 13:

Anzahl der in der Dokumentenanalyse kodierten Module im Studium zu den systematisch angelegten Kategorien der Eingangs-, Prozess- und Outputdiagnostik

Phase Kategorie	Anzahl	Eingangs- diagnostik (Kategorie 2)	Prozess- diagnostik (Kategorie 3)	Output- diagnostik (Kategorie 4)
Diagnostik (Subkategorie 1)	Module	14, 2 BM, 12 MM (2)	16, 4 BM, 12 MM (1)	13, 6 BM, 7 MM (1)
	Standorte	8	11	8
Aufgaben- erstellung (Subkategorie 2)	Module	0	1, 1 MM	2, 2 MM
	Standorte	--	1	2
Bewertung und Beurteilung (Subkategorie 3)	Module	4, 2 BA, 2 MA	3, 1 BA, 2 MA	2, 1 BM, 1 MM
	Standorte	4	3	2
Gütekriterien (Subkategorie 4)	Module	1, 1 MA	0	4, 1 BA, 3 MA
	Standorte	1	--	4
Handlungs- orientierung (Subkategorie 5)	Module	0	20, 9 BM (2), 11 MM	1, 1 MM
	Standorte	--	11	1
Problem- orientierung (Subkategorie 6)	Module	0	12, 4 BM (1), 8 MM	0
	Standorte	--	9	--

Abkürzungen: BM = Bachelormodul; MM = Mastermodul; (#) = Anzahl der Wahlmodule

z. B. „14, 2 BM, 12 MM (2)“ bedeutet, dass zu der Kategorie insgesamt 14 Module kodiert wurden, die sich auf 12 Bachelor- und 12 Mastermodule verteilen. Zwei der Mastermodule sind Wahlmodule.

(Eigene Darstellung)

Über alle drei Phasen ist die diagnostische Tätigkeit in der jeweiligen Phase häufig unkonkret angegeben. So ist aus den Passagen meist nicht ersichtlich, welche Themen in der Diagnostik behandelt werden. Beispielsweise kann „Prüfungen in der beruflichen Bildung“ (Otto von Guericke Universität Magdeburg, 2019, S. 10, Bachelormodul *Grundlagen der beruflichen Fachdidaktiken*) auch das Erstellen der genannten Prüfung beinhalten. Dies ist allerdings unklar. Vergleicht man die Kodierungen in den Kategorien *Aufgabenerstellung*, *Bewertung und Beurteilung* sowie *Gütekriterien* der jeweiligen Phasen, sind die meisten Kodierungen in *Bewertung und Beurteilung* zu finden. *Bewertung und Beurteilung* wird in der *Eingangsdiagnostik* in vier Modulen und in der *Prozessdiagnostik* in drei Modulen genannt und findet sich damit häufiger wieder als in der *Outputdiagnostik* (ein Bachelor- und ein Mastermodul).

Zwar finden sich zur *Prozessdiagnostik* Passagen, aus denen hervorgeht, dass Problemorientierung als Methode zur Gestaltung des Lehr-Lern-Prozesses thematisiert wird. Allerdings ist daraus nicht konkret erkennbar, ob die Aufgaben im Lehr-Lern-Prozess problemhaltig gestaltet und/oder im Sinne des formativen Assessments zur Diagnostik eingesetzt werden. Gleiches gilt

mit 20 kodierten Modulen für die Kategorie *3.5 Handlungsorientierung im Prozess*. Die Aufgabenerstellung in der Prozessdiagnostik konnte nur für ein Mastermodul kodiert werden, wobei daraus nicht erkennbar ist, wie diese Aufgaben gestaltet werden sollen (vgl. Georg-August-Universität Göttingen, 2019, S. 149).

Das Thema der Aufgabenerstellung in der *Outputdiagnostik* wird den Analyseergebnissen der Dokumente zufolge in keinem Bachelor-, jedoch in zwei Pflichtmodulen im Masterstudium zweier Standorte thematisiert. So wird beispielsweise das Erstellen von Aufgaben zur *Outputdiagnostik* Inhalt des Mastermoduls *Ergänzungsmodul Bildungsgangdidaktik*: „Die Studierenden ... entwickeln Aufgabendesigns für spezifische Kompetenzdimensionen und vergleichen diese mit Ansprüchen typischen Aufgaben abschlussbezogener Prüfungen“ (Universität zu Köln, 2019b, S. 52). Im gesamten Material findet sich lediglich ein Mastermodul, das *Handlungsorientierung in der Outputdiagnostik* (Kategorie 4.5) als handlungsorientierte Prüfungen thematisiert (vgl. FAU, 2019a, S. 84, Mastermodul *Grund- und Erstausbildung*). Zur *4.6 Problemorientierung in der Outputdiagnostik* konnte in den Materialien keine Passage identifiziert werden. Dieses Ergebnis ist insofern wichtig, als dass zwar aus der Literatur ersichtlich ist, dass die Handlungsorientierung die Problemorientierung einschließt (vgl. Kapitel 2.5), die explizite Gestaltung und der Einsatz problemhaltiger (Prüfungs-)Aufgaben jedoch Optimierungspotential aufweist (vgl. auch Kapitel 3).

In Bezug auf die *Problemorientierung* liefern die Ergebnisse der Fragebogenstudie mit Studiengangsleiter*innen ein optimistischeres Ergebnis als die Dokumentenanalyse. So wurden 21 Bachelor- und 31 Mastermodule genannt, in welchen Problemorientierung beinhaltet ist – davon zwei Wahlmodule im Bachelor und acht Wahlmodule im Master. Den Daten zufolge wird durchschnittlich in 2,1 Bachelor- (SD = 1.80; Min = 0; Max = 6) und 3,1 Mastermodulen (SD = 2.89; Min = 0; Max = 10) das Thema *Problemorientierung* behandelt. Die meisten offenen Rückmeldungen in der Fragebogenstudie betrafen ebenfalls diesen Inhaltsbereich. Insgesamt wurden hierzu 26 Kodierungen vorgenommen. Die meisten der Kodierungen (sechs von 26) bezogen sich auf die *Problemorientierung im Prozess* (Kategorie 5.6), was das Ergebnis der Dokumentenanalyse bestätigt. Fünf weitere Rückmeldungen konnten keiner Phase zugeordnet werden, sodass sie zur allgemeinen Kategorie *6.2 Problemorientierung* kodiert wurden. Jeweils vier Kodierungen wurden für die Kategorien *1.1 Vorgaben, Richtlinien, Regelungen*, *3.5 Handlungsorientierung im Prozess* und *4.6 Problemorientierung in der Outputdiagnostik* vorgenommen. Zudem wurde eine Kodierung in der Kategorie *5.1 Diagnostik allgemein* vorgenommen und zwei weitere in der Kategorie *6.1 Handlungsorientierung*.

Auch in Bezug auf die *Outputdiagnostik* liefert die Fragebogenstudie bessere Ergebnisse als die Dokumentenanalyse. Die Studiengangsleiter*innen wurden danach gefragt, ob in den genannten Modulen des Bachelor- und Masterstudiums Inhalte zur *(Prüfungs-)Aufgabenstellung und -auswertung* sowie *Problemorientierung* behandelt werden. Dabei wurde bewusst nicht danach gefragt, ob allgemein Inhalte zur pädagogischen Diagnostik behandelt werden, um konkretere Aussagen zu erhalten. Da beispielsweise bei der Leistungs- und Kompetenzmessung bzw. -diagnostik häufig unklar ist, in welcher Phase (Eingang, Prozess und/oder Output) diese zum Einsatz kommen (vgl. Kapitel 2.4.4), ist es im Bereich der *Prüfungen* eindeutiger. Daher wurde das Wort *Prüfung* bei den Fragen jeweils in Klammern gesetzt. In der Analyse wurden die Daten der zehn Standorte zugrunde gelegt, in denen die Modulnamen genannt wurden, da ansonsten die Piping-Funktion nicht hätte genutzt werden können. Insgesamt wurden 51 Bachelor- und 51 Mastermodule genannt, davon sechs bzw. elf im Wahlbereich. In Tabelle 14 werden die Ergebnisse zusammengefasst dargestellt und anschließend erläutert.

Tabelle 14:

Inhalte zur Prüfungsaufgabenerstellung und -bewertung in den Bachelor- und Mastermodulen

Inhalte zur Gesamt	(Prüfungs-) Aufgabenerstellung 35		(Prüfungs-) Aufgabenbewertung 34	
	Bachelor	Master	Bachelor	Master
Anzahl Module gesamt	13	22	13	21
davon Wahlmodule	1	5	1	5
Min	0	0	0	0
Max	3	6	3	6
M	1.30	2.20	1.30	2.10
SD	0.82	1.81	0.95	1.73

(Eigene Darstellung)

Basierend auf den betrachteten zehn Standorten, zeigt sich, dass im Studium in insgesamt 13 Bachelor- und 22 Mastermodulen Themen zur *(Prüfungs-)Aufgabenerstellung* behandelt werden – ein Bachelor- und fünf Mastermodule sind Wahlmodule. Das entspricht für die analysierten 51 Bachelor- und 51 Mastermodule einem Anteil von 25,49 % im Bachelor- und 43,14 % im Masterstudium. Demnach werden im Masterstudium häufiger Themen zur *(Prüfungs-)Aufgabenerstellung* als im Bachelorstudium. Im Durchschnitt wird dieses Thema in 1,3 Modulen im Bachelor- (SD = 0.82; Min = 0; Max = 3) und in 2,2 Modulen im Masterstudium (SD = 1.81; Min = 0; Max = 6) angesprochen.

Ähnlich verhält es sich bei der Bewertung von (Prüfungs-)Aufgaben. Die Studiengangsleiter*innen nannten hierzu 13 Bachelor- und 21 Mastermodule. Im Durchschnitt entspricht dies 1,3 Modulen im Bachelor- (SD = 0.95; Min = 0; Max = 3) und 2,1 Modulen im Masterstudium (SD = 1.80; Min = 0; Max = 6). Somit wird das Thema in 25,49 % der Bachelor- und 42,18 % der Mastermodule behandelt – ein Bachelor- und fünf Mastermodule sind Wahlmodule.

Anschließend wurden die Studiengangsleiter*innen nach der inhaltlichen Gestaltung der Module befragt. Die Rückmeldungen wurden in die drei Kategorien *(Prüfungs-)Aufgabenerstellung*, *(Prüfungs-)Aufgabenbewertung* und *Problemorientierung* aufgeteilt. Rückmeldungen ohne konkrete Antwort auf die Frage wurden nicht kodiert. Sowohl die Antworten als auch die Dokumentenanalyse zeigten, dass die drei genannten Themen nicht unabhängig voneinander behandelt werden, sondern sich häufig überschneiden. Die häufigste Kodierung zur *(Prüfungs-)Aufgabenerstellung* bezieht sich mit fünf Kodierungen auf die Kategorie der *Aufgabenerstellung zur Prozessdiagnostik* (Kategorie 3.2). Hierbei ist anzumerken, dass die Erstellung von Arbeitsblättern dieser Kategorie zugeordnet wurde. Zu einem Modul wurde angegeben, dass die *Aufgabenerstellung* in Verbindung mit der *Outputdiagnostik* (Kategorie 4.2) thematisiert wird. In zwei Fällen wurde die *Aufgabenerstellung* der Diagnostik allgemein (Kategorie 5.2) zugeordnet und konnte keiner Phasen (Eingang, Prozess und/oder Output) zugeordnet werden. Auch das Thema der *Kompetenzmessung und -diagnostik* (Kategorie 5.5.2) wird in Bezug auf die *(Prüfungs-)Aufgabenerstellung* in drei Modulen thematisiert. Die Ergebnisse überschneiden sich stark mit den Ergebnissen der Dokumentenanalyse.

Weiter wurde das Thema der *Bewertung und Beurteilung* in drei Modulen und die *Diagnostik im Allgemeinen* in zwei Modulen genannt, wobei nicht zugeordnet werden konnte, auf welche Phase (Eingang, Prozess und/oder Output) sich die *Bewertung und Beurteilung* beziehen. Die Daten zeigen auch, dass bei der Erstellung von (Prüfungs-)Aufgaben Themen zu *Vorgaben* sowie *Richtlinien und Regelungen* thematisiert werden (zwei Kodierungen). Des Weiteren finden sich einzelne Kodierungen in den Kategorien *Diagnostik des Outputs* (Kategorie 4.1) und *Gütekriterien* (Kategorie 5.4).

Zum Inhaltsbereich der *Bewertung von (Prüfungs-)Aufgaben* wurden 17 Rückmeldungen gezählt, davon wurden neun kodiert. Die Differenz ist dadurch zu erklären, dass bei acht von 17 Rückmeldungen ein Verweis auf vorher genanntes („s.o.“) gemacht wurde. Es konnte nicht zweifelsfrei erkannt werden, auf welche Stelle sich die Verweise bezogen. Bei zwei Rückmeldungen wurde angegeben, dass keine Angaben gemacht werden können, da externe

Lehrbeauftragte für das Modul zuständig sind. Drei der neun Kodierungen wurden in der Kategorie 5.5.3 *Leistungsmessung und -diagnostik* vorgenommen. Daraus ist zu schließen, dass die Bewertung sich meist auf die Leistungsmessung bzw. -diagnostik bezieht. Hierbei ist jedoch unklar, in welcher Phase (Eingang, Prozess und/oder Output) diese zum Einsatz kommt. Zwei Rückmeldungen wurden in der Kategorie 5.3 *Bewertung und Beurteilung* in der pädagogischen Diagnostik kodiert, wobei auch hier nicht klar war, auf welche Phase sich diese beziehen – es wurden lediglich Stichworte genannt. Vereinzelt Rückmeldungen zur Bewertung von (Prüfungs-)Aufgaben wurden den Kategorien 5.1 *Diagnostik*, 5.2 *Aufgabenerstellung zur Diagnostik*, 5.4 *Gütekriterien* und 5.5.3 *Leistungsmessung und -diagnostik* zugeordnet. Im Unterschied zu den Inhaltsbereichen *Aufgabenerstellung* und *Problemorientierung*, fanden sich *Bewertung von (Prüfungs-)Aufgaben* ausschließlich Passagen, die der (grafisch betrachtet, vgl. Abbildung 11) querliegenden Kategorie 5. *Diagnostik in der Pädagogik* zugeordnet werden. Das heißt, die Angaben wurden allgemein formuliert, sodass daraus nicht ersichtlich war, ob diese sich auf eine bestimmte Phase (Eingang, Prozess und Output) beziehen. Dieses Ergebnis überschneidet sich mit dem der Dokumentenanalyse. In der Analyse der Dokumente zeigte sich ebenfalls, dass häufig Themen zur *pädagogischen Diagnostik* in den Beschreibungen der Module zu finden sind, aus welchen jedoch die Phase nicht hervorgeht (vgl. Tabelle 12).

Weiter zeigt die Analyse der Dokumente, dass im Bachelor- und Masterstudium an etwa der Hälfte der berücksichtigten Standorte Themen zu *Vorgaben, Richtlinien und Regelungen* (Kategorie 1.1) behandelt werden (elf von 22 Standorten im Bachelor-, elf von 21 Standorten im Masterstudium). Das ist insofern relevant, als dass diese Themen wichtig für die Gestaltung von Assessments sind (vgl. insbesondere Kapitel 2.6 und 3.2). Die Notwendigkeit besteht nicht nur für die pädagogische Diagnostik im Sinne eines formalen bzw. summativen Assessments, sondern u. a. auch für die Gestaltung von Lehr-Lern-Prozessen. Die Elemente schließen sich nicht gegenseitig aus, sondern sind im Sinne der Curriculum-Instruktion-Assessment-Triade (Pellegrino, 2010) aufeinander abzustimmen. Wenngleich viele Kodierungen Passagen betreffen, die auf die Regelwerke und Ordnungsmittel nur unspezifisch eingehen oder ausschließlich nennen, gibt es auch Passagen, die speziell auf Diagnostik ausgerichtet sind – wie etwa in einem Mastermodul, in welchem u. a. „spezielle rechtliche und ethische Aspekte der Diagnostik im betrieblichen Kontext“ (Georg-August-Universität Göttingen, 2019, S. 150) thematisch behandelt werden.

In insgesamt 28 Modulen werden, laut der Dokumentenanalyse, (unspezifische) Formulierungen zu den *Voraussetzungen der Lehrenden* (Kategorie 1.2.1) gemacht, explizit wird die *diagnostische Kompetenz* (Kategorie 1.2.2) in sieben Module genannt. Bei der Interpretation dieses

Ergebnisses ist zu beachten, dass in den unspezifischen Formulierungen der verschiedenen Standorte die Anforderung der diagnostischen Kompetenz dennoch enthalten sein kann. Des Weiteren ist zu berücksichtigen, dass die diagnostische Kompetenz nicht immer so benannt sein muss und es daher möglich ist, dass Module diese zwar enthalten, die Inhalte jedoch aufgrund der unspezifischen Bezeichnung in andere (Sub-)Kategorien kodiert wurden. Wird Diagnostik beispielsweise in der Pädagogik behandelt (Kategorie 5), ist beim Ziel des Moduls davon auszugehen, dass auch die entsprechende Kompetenz der Lernenden (potentielle Lehrkräfte) entwickelt werden soll.

Da das Thema der Kompetenzorientierung relevant in der beruflichen Bildung ist (vgl. Kapitel 4.2.2), wird im Rahmen dieser Arbeit davon ausgegangen, dass das Thema der Erstellung von (Prüfungs-)Aufgaben zur Erfassung von (Problemlöse-)Kompetenz im Zuge der aktuellen Fragestellungen (Kategorie 7) behandelt wird (vgl. Kapitel 3.2). Aktuelle Fragestellungen werden in 45 Modulen der analysierten Dokumente (an 16 Standorten) behandelt, wobei an acht Standorten diese sowohl im Bachelor- als auch im Masterstudium thematisiert werden.

Wird die Häufigkeit der Überschneidungen der Kategorien betrachtet, also wie häufig Themenstellungen in einem Modul vorkommen¹⁰⁶, fällt auf, dass sich die Kodierungen einer Kategorie mindestens einmal mit einer anderen Kategorie überschneiden. Das heißt, dass meist in einem Modul nicht nur eines der untersuchten Themenfelder separat betrachtet wird, sondern mehrere gleichzeitig. Dies scheint nachvollziehbar, wenn es um bestimmte Themen geht. Wenn beispielsweise in einem Modul das Thema Lernerfolgskontrolle (Kategorie 5.5.4) behandelt wird, ist dafür auch die Entwicklung von Aufgaben (Kategorie 5.2) und die Bewertung dieser (Kategorie 5.3) sowie die Berücksichtigung rechtlicher Regelungen (Kategorie 1.1) relevant. Die Anzahl der Überschneidungen variiert zwischen den Kategorien, wobei nur mindestens einmal kodierte Kategorien berücksichtigt wurden. Während der geringste Überschneidungswert bei neun in der Kategorie 4.5 *Handlungsorientierung in der Outputdiagnostik* liegt, liegt der höchste Wert bei 146 in der Kategorie 5.5.3 *Leistungsmessung und -diagnostik*. Diese Werte zeigen, dass am häufigsten Themen zur *Leistungsmessung und -diagnostik*, mit anderen, zur Beantwortung der Forschungsfrage als relevant erachteten Themenfeldern gemeinsam thematisiert werden. Dieses Ergebnis ist nicht unabhängig davon, dass zum Thema der Leistungsmessung und -diagnostik häufiger kodiert wurde (13 Kodierungen), im Vergleich zur Handlungsorientierung in der Outputdiagnostik (eine Kodierung). Die Werte aus dem Code-Relations-

¹⁰⁶ Hierfür steht die Code-Relation-Browser-Funktion in MAXQDA zu Verfügung.

Browser werden nicht als ein Indiz fehlender Trennschärfe zwischen den Kategorien interpretiert, da die Kappa-Werte (κ) mindestens im guten Bereich liegen (vgl. Kapitel 4.2.2), die Kodierregeln es erlauben oder gar bedingen, Mehrfachkodierungen vorzunehmen (vgl. Kapitel 4.2.2) und die Interpretation der Ergebnisse und die Beispiele zeigen, dass das gemeinsame Vorkommen von Kodierungen plausibel ist.

Den höchsten Überschneidungswert stellt der Wert 16 zwischen den beiden Kategorien *3.5 Handlungsorientierung im Prozess* und *3.5 Problemorientierung im Prozess* dar. Ein Beispiel hierfür ist das Mastermodul *Unterrichtspraktische Studien II* der Universität Mainz. Zu den Inhalten zählen u. a. die „Anwendung didaktischer Prinzipien und Umsetzung zeitgemäßer didaktischer Konzepte (Kompetenzorientierung, Konstruktivismus, Problemorientierung, Lernfeldkonzept, Komplexe Lehr-Lernarrangements, Handlungsorientierung, Situatives Lernen)“ (Johannes Gutenberg-Universität Mainz [JGU], 2019, S. 4). Dies beiden Kategorien weisen Parallelen auf (vgl. Kapitel 2.5) und werden daher nachvollziehbar oft gemeinsam thematisiert.

Mit 14 Überschneidungen findet sich der zweitgrößte Wert zwischen den Kategorien *2.1 Diagnostik der Lern- und Leistungsvoraussetzungen* und *3.1. Diagnostik im Lernprozess der Lernenden*. Jedoch ist in beiden Fällen nicht bestimmt, was genau unter *Diagnostik* gefasst wird. Als Beispiel dient das Mastermodul *Schwerpunktmodul Praxissemester Berufskolleg II*, in welchem als Lernziel formuliert ist, dass Studierende „Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern ... beschreiben und in Ansätzen ... diagnostizieren“ (Universität zu Köln, 2019c, S. 48) können sollen.

Die Kategorien *1.1 Vorgaben Richtlinien und Regelungen* und *5.5.3 Leistungsmessung und –diagnostik* überschneiden mit einem Wert von zwölf. Mit Blick auf das Material, findet sich hierzu beispielhaft das Mastermodul *Lernen, Entwickeln und Lehren im berufs- und wirtschaftspädagogischen Kontext 2* der Christian-Albrechts-Universität zu Kiel [CAU]. In diesem wird darauf gezielt, dass Studierende Curricula analysieren und dabei „Leistungsmessungen und -bewertungen unter besonderer Berücksichtigung der Förderfunktion in mikro- und makrodidaktische Planungen“ (CAU, 2018, S. 13) integrieren können sollen.

Ein gemeinsames Bild für alle Standorte ist kaum möglich, da die Streuung je nach Kategorie unterschiedlich ist. In fast allen analysierten Dokumenten zum Bachelor- und Masterstudium Wirtschaftspädagogik wird mindestens einmal ein Thema behandelt, das für die hier gestellte Forschungsfrage als relevant erachtet wird (vgl. Tabelle 12). Während an einigen Standorten nahezu alle (Sub-)Kategorien mindestens einmal angesprochen wurden, wurde lediglich in drei

analysierten Dokumenten (zwei im Bachelor- und eines zum Masterstudium Wirtschaftspädagogik) keine Passage kodiert. Im Bachelorstudium ergibt sich ein Mittelwert von $M = 5.14$ ($SD = 4.85$), im Masterstudium ein Mittelwert von $M = 9.19$ ($SD = 10.00$). Die Werte sind im Masterstudium deutlich höher als im Bachelorstudium. Das bedeutet, dass im Vergleich zum Bachelorstudium im Masterstudium mehr Module kodiert wurden, die mindestens ein für die Forschungsfrage relevantes Thema behandeln. Dies ist nicht unabhängig davon, dass in die Analyse mehr Mastermodule eingeflossen sind als Bachelormodule (vgl. Tabelle 6). Auch die Ergebnisse der Fragebogenstudie bestätigen dies weitestgehend, trotzdem hierin gleich viele Module analysiert wurden (51 Master- und 51 Bachelormodule).

5.1.3 Ergebnisse zum Referendariat für das Lehramt an kaufmännisch-beruflichen Schulen

Bei der Ergebnisdarstellung wird auch hier auf eine Aufschlüsselung nach Bundesländern verzichtet. Vielmehr soll ein Gesamtbild der Ergebnisse zu den analysierten Dokumenten über alle Bundesländer gezeichnet werden, wenngleich Passagen aus Dokumenten einzelner Bundesländer bzw. aus den Transkripten der Interviews zitiert werden. Wird aus Transkripten zitiert, wird aus o. g. Grund und aus Gründen des Datenschutzes nicht angegeben, aus welchem Bundesland der*die Teilnehmende ist.

Zunächst werden in Tabelle 15 die deskriptiven Werte aus der Dokumentenanalyse zu den jeweiligen Kategorien dargestellt, bevor diese im Anschluss näher betrachtet und um die Ergebnisse der Interviewstudie ergänzt werden. Die Werte in der Tabelle dienen dazu, einen ersten Überblick über die Kodierungen in den jeweiligen Kategorien zu schaffen.

Tabelle 15:

Übersicht der deskriptiven Werte der Kodierungen in der Dokumentenanalyse zum Referendariat

		Module	Standorte gesamt
1. Input	N _{Codes;Input}	48	6
	Min _{Codes;Input}	0	
	Max _{Codes;Input}	12	6
	M _{Codes;Input}	1.60	
	SD _{Codes;Input}	2.70	
2. Eingangs- diagnostik	N _{Codes;StatusD}	24	6
	Min _{Codes;StatusD}	0	
	Max _{Codes;StatusD}	9	6
	M _{Codes;StatusD}	0.57	
	SD _{Codes;StatusD}	1.64	
3. Prozess- diagnostik	N _{Codes;ProzessD}	70	6
	Min _{Codes;ProzessD}	0	
	Max _{Codes;ProzessD}	7	6
	M _{Codes;ProzessD}	1.67	
	SD _{Codes;ProzessD}	1.98	
4. Ouput- diagnostik	N _{Codes;OutputD}	13	4
	Min _{Codes;OutputD}	0	
	Max _{Codes;OutputD}	3	6
	M _{Codes;OutputD}	0.31	
	SD _{Codes;OutputD}	0.72	
5. Diagnostik in der Pädagogik	N _{Codes;Diagnostik}	52	6
	Min _{Codes;Diagnostik}	0	
	Max _{Codes;Diagnostik}	8	6
	M _{Codes;Diagnostik}	0.87	
	SD _{Codes;Diagnostik}	1.36	
6. Handlungs- und Problemorientierung	N _{Codes;HPO}	0	0
	Min _{Codes;HPO}	0	
	Max _{Codes;HPO}	0	5
	M _{Codes;HPO}	--	
	SD _{Codes;HPO}	--	
7. Aktuelle Fragestellungen	N _{Codes;Aktuell}	0	0
	Min _{Codes;Aktuell}	0	
	Max _{Codes;Aktuell}	0	6
	M _{Codes;Aktuell}	--	
	SD _{Codes;Aktuell}	--	
Gesamt	N _{Codes;GESAMT}	207	6
	Min _{Codes;GESAMT}	14	
	Max _{Codes;GESAMT}	58	6
	M _{Codes;GESAMT}	34.50	
	SD _{Codes;GESAMT}	17.96	

(Eigene Darstellung)

Analog zur Analyse der Dokumente des Studiums Wirtschaftspädagogik wurde auch zum Referendariat in der ersten Kategorie nach dem *Input* analysiert. Hierbei wurden für alle Bundesländer, dessen Dokumente analysiert wurden, mindestens drei- und maximal zwölfmal Passagen kodiert, in welchen Themen zu *Vorgaben, Richtlinien und/oder Regelungen* (in Form von Curricula) behandelt wurden. Auch hier wurde in den Kategorien nicht danach unterschieden, welche Vorgaben, Richtlinien und/oder Regelungen behandelt werden, da dies nicht zwangsweise zu einem trennscharfen Ergebnis führen würde. Vielmehr wurde der Blick auf die Kodierungen und die einzelnen kodierten Segmente (Passagen) gerichtet und eingeschätzt, ob diese sich konkret auf die pädagogische Diagnostik im Sinne der Prüfungen beziehen. Auch, wenn sich die meisten Passagen eher auf den Lehr-Lern-Prozess beziehen (vgl. z. B. Amt für Lehrerbildung Hessen, 2012, S. 67), gibt es auch Passagen in Bezug auf Assessments. Als Beispiel dienen folgende zwei Passagen: „Die Lehrkraft erfasst und beurteilt die Leistungen von Schülerinnen und Schülern unter Anwendung der geltenden rechtlichen Vorgaben“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 15) und „Sie entwickeln ... die notwendigen Instrumente der Leistungserfassung gemeinsam in schulischen Gremien auf der Grundlage rechtlicher Vorgaben“ (Niedersächsisches Kultusministerium, 2013b, S. 19)¹⁰⁷. Insgesamt wurden zu dieser Kategorie insgesamt 34 Module aus sechs Bundesländern kodiert. Auch in der Interviewstudie zeigte sich, dass curriculare Vorgaben Beachtung finden, die u. a. Vorgaben zum (kompetenzorientierten) Lernen und Prüfen beinhalten (z. B. Lars, Lara und Lamia, Pos. 21, 44; Lorenz, Pos. 10). Interessant ist, dass auch im Interview die wenig komplexe Gestaltung der Prüfungen angesprochen wurde, welche im Kontrast zu den curricularen Anforderungen steht (vgl. Kapitel 3.2). So beschreibt Laurentia dies als Herausforderung:

„Und dann wird da auch Thema das Spannungsfeld von einerseits, den Prüfungsaufgaben, ... die ja in einigen Berufen recht stark einen heimlichen Lehrplan vorgeben und die Bedarfe der SchülerInnen gehen dahin, dass sie natürlich im Unterricht, in den Klassenarbeiten, genau diesen Prüfungstyp ... wiederfinden wollen, den sie nachher in der Abschlussprüfung auch sehen. Und dagegen steht dann der Anspruch der Leitlinie schulisches Curriculum, die eine sehr viel komplexere Aufgabenstellung verlangt Und da sind dann die Ansprüche aus dem schulischen Curriculum komplexer als die Ansprüche aus Prüfungsaufgaben und das dann zusammenzubringen, ist auch Gegenstand der fachdidaktischen Seminare, aber

¹⁰⁷ Das Jahr der Veröffentlichung des Dokuments ist als solches nicht im Dokument zu finden. Dieses wurde in einer persönlichen Nachricht der Mitarbeiterin bekanntgegeben (Leonard, 2021).

nicht in einer Ausführlichkeit über drei, vier Sitzungen, sondern eher ein, zwei Sitzungen.“
(Laurentia, Pos. 16)

Die Probandin geht auch auf die zeitlichen Herausforderungen im Referendariat ein. Dies lässt die Interpretation zu, dass Bedarf einer tiefergehenden Behandlung mit dem Thema besteht, dies jedoch an organisatorische Grenzen stößt. Ihre Aussage bestätigt dies.

„Im Verhältnis zu den anderen Dingen, die sie [die LiV] hier für ihr Referendariat und für ihre eigenen Prüfungsleistungen ... [leisten] mussten, fällt das nicht in das Gewicht, wie es eigentlich fallen müsste, wenn man sich die berufliche Wirklichkeit im Anschluss bei einer vollen Stelle anschaut.“ (Laurentia, Pos. 16)

Zwar findet sich in den Dokumenten der meisten Bundesländer (fünf von sechs) mindestens eine Passage, in der von Voraussetzung der Lehrenden gesprochen wird, diese werden jedoch in den meisten Fällen nicht konkret benannt. In den Dokumenten zu vier Modulen von lediglich zwei Bundesländern wird die diagnostische Kompetenz konkret angesprochen und als solche kodiert. In einer Passage steht „Kompetenzen: Diagnostik und Beratung“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 6). Ähnlich wie in den Analysen zum Studium, muss bei der Interpretation hier auch berücksichtigt werden, dass in unkonkreten Formulierungen der Voraussetzungen der Lehrkräfte die diagnostische Kompetenz inbegriffen sein kann. Gleichzeitig ist zu beachten, dass die Kompetenz nicht immer als solche ausgeschrieben sein muss. Vielmehr zeigt die Analyse des Materials, dass Passagen zu den einzelnen Tätigkeiten in der pädagogischen Diagnostik bzw. den einzelnen Phasen enthalten sind. Die Kodierungen lassen darauf schließen, dass Module zum Thema des untersuchten Bereichs durchaus die Kompetenz voraussetzen oder anstreben können, selbst wenn diese nicht explizit genannt wird. Diese Annahme bestätigen auch die Ergebnisse der Interviewstudie. In Interviewtranskripten von vier Teilnehmenden wurden Aussagen gefunden, die die diagnostische Kompetenz als Anforderung benennen und diese als wichtig für die Konzeption problemhaltiger Aufgaben erachten (vgl. Kapitel 0). Luisa sagte zur Aufgabenkonzeption,

„dass unterschiedliche Fähigkeiten eine Rolle spielen, dass Dinge erklärt beziehungsweise bewertet werden müssen. Auf der einen Seite aber, dass vielleicht auch Zusammenhänge erfasst werden müssen. Und das sind verschiedene, sagen wir mal Fähigkeiten, die dann auch in der Konzeption vorhanden sein sollten.“ (Luisa, Pos. 22)

Bei der Betrachtung der Diagnostik und den diagnostischen Teilaspekten in den verschiedenen Phasen (Eingangs-, Prozess- und Outputdiagnostik) zeigen sich Unterschiede, die in Tabelle 16 dargestellt werden:

Tabelle 16:

Anzahl der kodierten Module in der Dokumentenanalyse zum Referendariat zu den systematisch angelegten Kategorien der Eingangs-, Prozess- und Outputdiagnostik

Phase Kategorie	Anzahl	Eingangs- diagnostik <i>(Kategorie 2)</i>	Prozess- diagnostik <i>(Kategorie 3)</i>	Output- diagnostik <i>(Kategorie 4)</i>
Diagnostik <i>(Subkategorie 1)</i>	Module	22	24	4
	Standorte	6	6	2
Aufgaben- erstellung <i>(Subkategorie 2)</i>	Module	0	4	3
	Standorte	--	2	3
Bewertung und Beurteilung <i>(Subkategorie 3)</i>	Module	2	7	5
	Standorte	2	4	3
Gütekriterien <i>(Subkategorie 4)</i>	Module	0	2	1
	Standorte	--	1	1
Handlungs- orientierung <i>(Subkategorie 5)</i>	Module	0	19	0
	Standorte	--	6	--
Problem- orientierung <i>(Subkategorie 6)</i>	Module	0	14	0
	Standorte	--	4	--

(Eigene Darstellung)

In jedem analysierten Dokument der betrachteten Bundesländer findet sich mindestens ein Modul zum Thema Diagnostik. Während insgesamt 22 Module der Kategorie 2.1 *Diagnostik der Lern- und Leistungsvoraussetzungen* zugeordnet werden konnten, wurden zu den anderen Subkategorien der Kategorie 2 nur vereinzelte Passagen kodiert. So finden sich in den Dokumenten keine Aussagen zu *Aufgabenerstellung zu eingangsdagnostischen Zwecken* (Kategorie 2.1). Ebenso wurden die Kategorien *Gütekriterien* sowie zu *Handlungs- und Problemorientierung* (Kategorien 2.4 bis 2.6) nicht angesprochen. In den Dokumenten zweier Bundesländer wurde die *Bewertung und Beurteilung* zu dieser Phase der Diagnostik (Kategorie 2.3) jeweils einmal kodiert. Es kann festgehalten werden, dass das Thema der *Eingangsdagnostik* durchaus in jedem der analysierten Bundesländer behandelt wird. So wird das Thema etwa im *Fachcurriculum Wirtschaftswissenschaften* als „Lehr- und Lernausgangslagen wahrnehmen, Potenziale erkennen, diagnostisch erfassen“ (Zentrum für schulpraktische Lehrerbildung Leverkusen,

o.A.b, S. 1) bezeichnet. Unklar ist jedoch, was genau hierbei thematisiert wird und welche Arbeits- bzw. Durchführungsschritte bei der Diagnostik erforderlich sind. Aus den Transkripten der Interviews wurde keine Passage zu Kategorie 2 kodiert.

Wie auch bei der Analyse der Dokumente des wirtschaftspädagogischen Studiums, wurden auch in den Dokumenten zum Referendariat am häufigsten Passagen zu Prozessdiagnostik kodiert – im Vergleich zu den anderen beiden Phasen (Eingangs- und Outputdiagnostik). Die Kategorie *3.1 Diagnostik im Lernprozess der Lernenden* wird in 24 Modulen (an sechs Standorten) angesprochen. Zur Aufgabenerstellung in dieser Phase finden sich in zwei Bundesländern und insgesamt vier Modulen relevante Passagen (vgl. z. B. Ministerium der Justiz Rheinland-Pfalz, 2012, S. 9). Passagen aus sieben Modulen wurden der Subkategorie *3.3 Bewertung und Beurteilung der Lernenden im Lehr-Lern-Prozess* zugeordnet. In vier der sechs Bundesländer wurde mindestens ein Modul hierzu gefunden und dieses kodiert. In einem dieser Bundesländer wurde das Thema der *Gütekriterien* in der Phase der *Diagnostik* (Kategorie 3.4) in zwei Modulen berührt, in den anderen ließ sich eine solche Passage nicht finden. *Handlungsorientierung im Prozess* (Kategorie 3.5) wurde in allen Dokumenten der analysierten Bundesländer mindestens ein- bis höchstens siebenmal kodiert (19 Module), wobei sich diese i. d. R. auf die Unterrichtsgestaltung beziehen (vgl. z. B. Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 12). Passagen zu *Problemorientierung im Prozess* (Kategorie 3.6) wurden in Dokumenten von vier Bundesländern (14 Module) kodiert. Auch diese beziehen sich i.d.R. auf den Unterricht – z. B. „Problemorientiertes Lernen“ (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, o.A., Modul *Unterrichtskonzeptionen an beruflichen Schulen*). Insgesamt kann zur Prozessdiagnostik festgehalten werden, dass sie in allen Bundesländern behandelt wird, wobei die Prozessdiagnostik in einigen Bundesländern expliziter beschrieben wird.

Besondere Beachtung scheint die *Bewertung und Beurteilung im Lehr-Lern-Prozess* zu finden. Auch die didaktische Gestaltung des Prozesses wird beschrieben, wobei es sich auf die unterrichtliche Gestaltung fokussiert. Die Anzahl der Kodierungen in der Dokumentenanalyse zeigen, dass das Thema *Outputdiagnostik* insgesamt weniger stark behandelt wird als die Themen *Eingangs-* und *Prozessdiagnostik* (vgl. Tabelle 16). Es finden sich zu dieser Kategorie insgesamt lediglich in den Dokumenten der Hälfte der Bundesländer Passagen und somit Kodierungen. In vier Modulen an zwei Standorten wird das Thema der *Outputdiagnostik* unspezifisch thematisiert (Kategorie 4.1). In drei Modulen von drei Bundesländern wird *Aufgabenerstellung zur Outputdiagnostik* (Kategorie 4.2) explizit genannt, wie in der Beschreibung des Ministeri-

ums für Kultus, Jugend und Sport Baden-Württemberg (2018) zum Kapitel *Leistungen feststellen*: „[LiV] beachten bei den Aufgabenstellungen sachliche Zusammenhänge mit anderen, bereits behandelten Themenbereichen“ (S. 123).

Die Aussage von Luisa bestätigt das Ergebnis der Dokumentenanalyse:

„Es ist so, dass wir hier ja eine Unterscheidung nicht direkt, aber implizit vornehmen. ... einmal ... [die] Gestaltung von Lernaufgaben für den Unterricht ... Auf der anderen Seite ... diese Prüfungsgestaltung bzw. Leistungsbewertungsperspektive.“ (Luisa, Pos. 16)

Auch ging aus dem Interview hervor, dass der Ausbildungsschwerpunkt auf der Gestaltung des Lehr-Lern-Prozesses liegt. Dies schließt allerdings Aufgaben zu summativen Zwecken nicht aus. So zeigen Passagen, dass durchaus Synergien bestehen, wie durch folgende Aussage:

„Also ist da eine enge Koppelung ... zwischen der Leistungsfeststellung am Ende ... einer problemorientierten Lernaufgabe ... eine problemorientierte Aufgabe, die im Rahmen der Überprüfung stattfindet Insofern ist es das, ... dass man das nicht völlig losgekoppelt voneinander betrachten kann ja, also die die Vorbereitung darauf findet ... dann entsprechend mit der Ausarbeitung von solchen Lernaufgaben statt.“ (Luisa, Pos. 16)

Aus den Interviews geht auch hervor, dass die Aufgabenerstellung zu unterschiedlichen Phasen „*analog*“ (Lukas, Pos. 10) läuft, womit durch die Behandlung der Aufgabenerstellung für die eine Phase auch das Gelernte auf die Aufgabenerstellung für eine andere Phase übertragen wird (vgl. hierzu auch Klotz, 2015, S. 75). Dies entspricht gleichzeitig dem Gedanken der Curriculum-Instruktion-Assessment-Triade (Pellegrino, 2010). Den Interviews konnte zudem entnommen werden, dass die Anforderungen an Lehrkräfte, Aufgaben zu formativen und summativen Zwecken entwickeln zu können, bereits in der Ausbildungsphase nicht auf Klausuraufgaben beschränkt wird. So sagte etwa Lukas, dass die Ausbildung im Referendariat auch vor dem Hintergrund geschieht, dass „*sie, ... als Lehrer eine Prüfungsaufgabe später [er]stellen müssen*“ (Pos. 24). Auch für die *Aufgabenerstellung zur Outputdiagnostik* (Kategorie 4.2) zeigt sich, dass diese explizit in den Studienseminaren thematisiert wird. Dies steht im Kontrast zum Ergebnis der Dokumentenanalyse. Der Übereinstimmungswert im Code-Relation-Browser in MAXQDA zeigt, dass die Aufgaben zu summativen Zwecken in der Regel kompetenzorientiert, handlungsorientiert (Überschneidungen in fünf Dokumenten) und/oder problemorientiert (Überschneidungen in fünf Dokumenten) behandelt werden. Anders als aus dem Ergebnis der

Dokumentenanalyse interpretiert, werden Handlungs- und Problemorientierung im Referendariat im Zusammenhang mit der Outputdiagnostik thematisiert. So wird auch explizit auf Problemlösekompetenz im summativen Assessment bzw. in der Outputdiagnostik eingegangen:

„... in den Prüfungsaufgaben oder in einer Prüfungssituation ... fangen Sie von dem Level her ja schon eher ein bisschen weiter unten an und daher macht sowas durchaus Sinn, immer vom Fokus aus zu betrachten, dass in jede Leistungssituation so ein bisschen von allem rein gehört. Aber einfach aufgrund dieser höheren Wertigkeit. Und das verknüpfen wir eben mit der Erkenntnis aus der Wissenspsychologie auf einem höheren taxonomischen Level Transfer und Problemlösen Also nicht nur in Abschlussprüfungssituationen, sondern tatsächlich immer, also absolut in jeder Klausur, die es gibt, sollte von allem etwas drin sein.“
(Lukas, Pos. 26)

Die bei der Erstellung von problemhaltigen Prüfungs-/Klausuraufgaben zu berücksichtigenden Merkmale wurden teilweise konkret benannt, wie die folgende Aussage zeigt:

„Und wenn wir uns den anschauen, dann können wir aus der empirischen Forschung eben ableiten, dass es darum geht, authentische Aufgaben, problemhaltige Aufgaben und Aufgaben aus dem späteren Beruf zu integrieren in die Prüfungsaufgaben und die Formate der Prüfungsaufgaben. Also das wäre so der Anteil der allgemeinen Pädagogik zu sagen, es soll möglichst authentisch, lebensnah, berufsnah sein und da möglichst situiert ... mit authentischen Belegen.“ (Lara, Pos. 10)

Die *Bewertung und Beurteilung* in der Phase des Outputs (Kategorie 4.3) wird im Rahmen der Dokumentenanalyse in insgesamt fünf Modulen (verteilt auf drei Bundesländer) thematisiert, Aussagen zu *Gütekriterien* finden sich nur in einem Modul. Hierzu steht, dass die angehende Lehrkraft „eine Klassenarbeit unter Beachtung der Gütekriterien auf[stellt] und korrigiert“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26). Auch dazu liefert die Interviewstudie ein besseres Ergebnis, indem von Teilnehmenden aus vier Standorten berichtet wird, dass Gütekriterien in Zusammenhang mit der Outputdiagnostik behandelt werden. Das Thema der *Bewertung und Beurteilung* in dieser Phase wird an fünf Standorten behandelt, indem die Teilnehmer über „[die] Korrektur einer Klassenarbeit über Notentransparenz“ (Lorenz, Pos. 10), *Leistungsbewertung* (Levin, Pos. 56), *Aufwand zur Auswertungs- bzw. Beurteilung* (Luisa, Pos. 22) und „*Erwartungshorizonte vom Hintergrund der Gütekriterien*“ (Lukas, Pos. 14) sprechen.

Während in der Dokumentenanalyse zum Referendariat relativ viele Kodierungen zur Kategorie 5 (*Diagnostik in der Pädagogik*) vorgenommen werden konnten, sind es in der Interviewstudie im Vergleich weniger. In den Dokumenten wurden Passagen in fünf Modulen an vier Standorten gefunden, die das Thema *Diagnostik* unspezifisch thematisieren und diese damit keiner Phase zuordenbar ist (Kategorie 5.1; vgl. z. B. Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, o. A.). In der Interviewstudie gab es hierzu zwei Passagen. Zur *Aufgabenerstellung zu diagnostischen Zwecken* (Kategorie 5.2) wurden Passagen in fünf Modulen dreier Standorte gefunden. Auch hier ist unklar, zu welcher Phase diese thematisiert werden. Als Beispiel dient die Frage: „Wie konzipiere ich Aufgabenstellungen des Faches kriterien- und adressatengerecht und unter Nutzung moderner Medien?“ (Zentrum für schulpraktische Lehrerbildung Leverkusen, o.A.b, S. 3). Auch in den Interviewtranskripten finden sich Passagen, in denen über die Aufgabenerstellung zu diagnostischen Zwecken im Allgemeinen gesprochen wird (fünf Passagen aus drei Transkripten). Dies betrifft z. B. „*Schwierigkeiten bei Formulierungen*“ (Luisa, Pos. 20). An fünf Standorten und in elf Modulen wurde das Thema der *Bewertung und Beurteilung* aufgeführt. Beispielfhaft kann folgende Passage aus dem Dokument des Ministeriums Rheinland-Pfalz (2012) aufgeführt werden, in dem es heißt, dass die „Anwärterinnen und Anwärter lernen ... unterschiedliche Formen der Leistungsbeurteilung zu verstehen, verantwortungsbewusst anzuwenden und zu reflektieren“ (S. 5). Auch in den Transkripten aller Interviewteilnehmenden wurden Passagen hierzu kodiert. Die Ergebnisse der Dokumentenanalyse lassen damit den Schluss zu, dass das Thema *Bewertung und Beurteilung von Leistungen* ausreichend behandelt wird. Aus Aussagen der Interviewteilnehmenden geht jedoch hervor, dass diese „in der Regel relativ kurz“ behandelt werden „und dann vertiefend, wenn es als Bedarf der Teilnehmenden wiederauftaucht“ (Laurentia, Pos. 16).

In den Dokumenten wurden *Gütekriterien* in drei Modulen an zwei von sechs Standorten thematisiert, wobei auch hier keine Zuordnung zu einer Phase erfolgt (Kategorie 5.4). Auch dieses Bild bestätigte sich in der Interviewstudie, indem vier Passagen an drei Standorten in diese Kategorie kodiert wurden. Die Ergebnisse der Interviewstudie zeigen, dass Gütekriterien häufiger im Rahmen diagnostischer Verfahren im Zusammenhang mit der Outputdiagnostik thematisiert werden. Die Anzahl der kodierten Passagen zu Gütekriterien in den jeweiligen Phasen gestaltet sich wie folgt: 2.4 *Eingangsdagnostik der Eignungsdiagnostik*: keine Kodierung; 3.4 *Gütekriterien in der Prozessdiagnostik*: eine Passage aus einem Standort und 4.4 *Gütekriterien in der Outputdiagnostik*: sieben Passagen, verteilt auf drei Standorte. Durch die Anwendung kann das Lernen positiv beeinflusst werden (vgl. hierzu z. B. Dehnbostel, 2000, S. 107; Krille, 2019, S. 21; siehe auch Kapitel 6.3), was didaktisch sinnvoll ist. Des Weiteren wurden in den

Dokumenten Diagnoseinstrumente bzw. -verfahren genannt, die keiner der Phasen zugeordnet werden konnten (Kategorie 5.5 bzw. Subkategorien 5.5.1 bis 5.5.4). So werden in vier Modulen an zwei Standorten Diagnoseinstrumente unspezifisch genannt (*5.5.1 Diagnoseinstrumente und -verfahren allgemein*). In drei Modulen eines Standorts werden die Instrumente zur Kompetenzmessung bzw. -diagnostik (Kategorie 5.5.2), in 16 Modulen an fünf Standorten Instrumente zur Leistungsmessung bzw. -diagnostik (Kategorie 5.5.3) und in fünf Modulen an drei Standorten Instrumente zur Lernerfolgskontrolle bzw. -diagnostik (Kategorie 5.5.4) beschrieben. Ankerbeispiele hierfür finden sich im Kodierleitfaden (vgl. Anhang A). Dieses Ergebnis verdeutlicht, dass Diagnostik und diagnostische Verfahren in der Ausbildung von Lehrer*innen thematisiert werden, jedoch kann nicht immer klar definiert werden, zu welcher Phase bzw. zu welchem Zweck die diagnostische Tätigkeit behandelt wird. Ebenso ist aufgrund der unkonkret in Stichpunkten dargestellten Information nicht immer klar, welche Inhalte zum jeweiligen Verfahren gelehrt und gelernt werden. Die Ergebnisse der Interviews zeigen ein anderes Bild, denn es gibt wesentlich mehr Kodierungen, die einer bestimmten Phase zugeordnet werden können. Anders als in der Dokumentenanalyse, sind lediglich in einem Transkript Passagen zu finden, die zwar Diagnoseinstrumente nennen, diese jedoch keiner bestimmten Phase der Diagnostik zuordnen. In dieser Passage werden Formen von Verfahren und Instrumenten zur Leistungsfeststellung genannt (vgl. Luisa und Leon, Pos. 16, 17).

Zu den Subkategorien *6.1 Handlungsorientierung* und *6.2 Problemorientierung* fanden sich in den analysierten Dokumenten keine kodierbaren Passagen. Das bedeutet, dass Handlungs- und Problemorientierung, wenn diese in den Dokumenten thematisiert wurden, den analysierten Phasen (Eingang, Prozess und/oder Output) zugeordnet werden konnten. Dass in den analysierten Inhalten zu Handlungs- und Problemorientierung zu finden sind, kann aus den vorherigen Kategorien *3.5 Handlungsorientierung im Prozess* und *3.6 Problemorientierung im Prozess* entnommen werden. Zu den anderen Phasen gibt es in den Dokumenten keine Kodierung, wohl aber in den Transkripten (vgl. Tabelle 16). Im Umkehrschluss bedeutet das, dass basierend auf der Dokumentenanalyse Handlungs- und Problemorientierung zwar thematisch aufgeführt werden, diese aber stets die Gestaltung des Lehr-Lern-Prozesses betreffen. Die Ergebnisse der Interviewstudie zeigen bereits, dass die Themen auch in Bezug auf die Outputdiagnostik behandelt werden. Zusätzlich konnten in den Transkripten einige wenige Passagen gefunden werden, die der Kategorie *6.1 Handlungsorientierung* (zwei Kodierungen) bzw. *6.2 Problemorientierung* (vier Kodierungen) zugeordnet werden konnten. Dabei wurde die *Problemorientierung als fachdidaktisches Prinzip* (Laurentia, Pos. 18) vorgestellt und das Verständnis der Hand-

lungs- und/oder Problemorientierung wiedergegeben. Letzteres trug zur Validierung der vorgestellten Studie bei, indem es zeigte, dass das Verständnis der Interviewteilnehmenden zu den Gemeinsamkeiten und Unterschieden der Problemlöse- und Handlungskompetenz der vorliegenden Arbeit entspricht (vgl. Kapitel 2.5). Folgendes Beispiel soll dies verdeutlichen: „*Problemorientierung ja nicht losgekoppelt ... von dem Thema oder der übergeordneten Kompetenz der Handlungsorientierung. Also, das muss man ja in einer Verbindung miteinander betrachten*“ (Luisa und Leon, Pos. 35; ähnlich auch in z. B. Laurentia, Pos. 18).

Dies bestätigte sich auch darin, dass Teilnehmende häufig von *Kompetenzorientierung* sprachen, weniger von *Handlungs- bzw. Problemlösekompetenz* (z. B. Lorenz, Pos. 10; Lukas, Pos. 6). Die entsprechenden Passagen wurden daher in die jeweiligen Kategorien beider Kompetenzen kodiert (auch in den entsprechenden Subkategorien der jeweiligen Phase), um eine Unterschätzung zu vermeiden. Dennoch gab es auch Teilnehmende, die sich konkret auf die Problemlösekompetenz bezogen.

Zu 7. *Aktuellen Fragestellungen* konnten sowohl anhand der Dokumente als auch in den Interviewtranskripten keine Passagen kodiert werden. Das muss nicht bedeuten, dass im Referendariat keine Auseinandersetzung mit aktuellen Fragen der Berufs- bzw. Wirtschaftspädagogik geschieht.

Im nächsten Schritt wurde die Übereinstimmung im Code-Relations-Browser in MAXQDA betrachtet. Ziel war es herauszufinden, ob und wie sich die Kodierungen in den Dokumenten überschneiden. Da das jeweilige Modul als Kodiereinheit festgelegt wurde, konnte die gemeinsame Behandlung eines die für die Forschungsfrage relevanten Themas in einem Modul betrachtet werden. Wie auch in der Dokumentenanalyse zum Studium zeigt sich, dass kaum ein betrachtetes Thema überschneidungsfrei ist, denn die Module wurden mehreren Kategorien zugeordnet.

In den Dokumenten zum Referendariat finden sich die häufigsten Überschneidungen mit der Kategorie *1.1 Vorgaben, Richtlinien und Regelungen*. Diese überschneidet sich mit je sieben Kodierungen bzw. kodierten Modulen mit den Kategorien *2.1 Diagnostik der Lern- und Leistungsvoraussetzungen* und *3.5 Handlungsorientierung im Prozess*, jedoch am häufigsten (acht Kodierungen) mit der Kategorie *3.1 Diagnostik im Lernprozess der Lernenden*. Letztgenannte (Kategorie 3.1) ist die sich am zweithäufigsten mit anderen Kategorien überschneidende. Neben der Kategorie 1.1 überschneidet sich Kategorie 3.1 am häufigsten (acht Kodierungen) mit der Kategorie 2.1. Demnach werden häufig in einem Modul Themen zur Diagnostik im Lernpro-

zess der Lernenden, zu curricularen Vorgaben sowie zur Diagnostik der Lern- und Leistungsvoraussetzungen von Lernenden behandelt. Bei den Kategorien zur Diagnostik (Kategorien 2.1 und 3.1) wird allerdings nur die Diagnostik der Phase zugeordnet, nicht aber konkret beschrieben, was genau dazu gelehrt wird. Beispielhaft heißt es in einem Modul, dass angehende Lehrkräfte „Entwicklungsstände, Lernpotentiale, Lernfortschritte und Lernhindernisse [ermitteln]“ (Niedersächsisches Kultusministerium, 2013b, S. 22). Im *Handlungsfeld L*, das vom Zentrum für schulpraktische Lehrerbildung Leverkusen (o. A.a, S. 5) beschrieben wird, werden sogar alle drei Kategorien (Kategorien 1.1, 2.1 und 3.1) genannt. Hierin finden sich u. a. folgende, stichpunktartig aufgeführte Inhalte:

- „Vorgaben des Schulgesetzes und der Ausbildungs- und Prüfungsordnungen in den unterschiedlichen Bildungsgängen
- Richtlinien und Lehrpläne/ Bildungspläne, schuleigene Vereinbarungen zur Leistungsbewertung
- Erstellung einer Lehr-Lernkontrolle in Bezug auf einen Wiwi-Inhalt unter Berücksichtigung der anzustrebenden Kompetenzen
- ...
- Lernprozessanalyse und Leistungsfeststellung
- Zeitpunkte der Lernerfolgskontrolle im Rahmen einer Lernsituation: Lernprozess/ Lernergebnis“ (Zentrum für schulpraktische Lehrerbildung Leverkusen, o. A.a, S. 5)

Diese Überschneidung ist nachvollziehbar, gerade mit Blick darauf, dass in curricularen Richtlinien Kriterien formuliert werden, anhand welcher Bezugsnormen gebildet und für diagnostische Zwecke genutzt werden (vgl. Kapitel 2.4.3). Geht es in einem Modul, wie etwa im Modul *Handlungsfeld L* (Zentrum für schulpraktische Lehrerbildung Leverkusen, o. A.a), insgesamt um das Thema der Leistungserfassung und -bewertung bzw. -beurteilung, scheint es auch hier verständlich, dass diese im Kontext mehrerer Phasen behandelt werden, da sich hierdurch inhaltlich Parallelen ergeben.

Abgesehen von Kategorien, zu denen es keine Kodierungen gab (vgl. Tabelle 15), liegt die kleinste Summe der Überschneidungen mit den Kategorien *3.4 Gütekriterien in der Prozessdiagnostik* und *4.1 Gütekriterien in der Outputdiagnostik*. Dieses Ergebnis ist nicht unabhängig davon, dass zu Kategorien der Gütekriterien insgesamt kaum Kodierungen vorgenommen werden konnten (vgl. Tabelle 16; zu Kategorie 5.4 finden sich drei Kodierungen).

Abschließend ist festzuhalten, dass sich die Standorte untereinander unterscheiden. Fachleitungen teilten mit, dass „sehr bedarfsorientiert gearbeitet wird“ (Laurentia, Pos. 14) und auch

Möglichkeiten geboten werden, etwa „zu Beginn des Seminars ... aktuelle Probleme“ (Lukas, Pos. 16) zu besprechen. Des Weiteren sind die Unterschiede im Angebot der Wahlmodule begründet. Folgende Passage verdeutlicht, dass über den Standortvergleich hinaus auch kaum ein Vergleich der Jahrgänge eines Seminars möglich ist:

„zusätzlich zu diesen beiden Curricula haben wir noch Wahlmodule, also ... haben wir dann auch noch einmal etwas zum Thema Aufgabenerstellung für diejenigen, die das dann eben anwählen. Das ist allerdings ein Angebot, was auch immer abhängig davon ist, wie wir die Referenten bekommen. Weil wir da dann auch Lehrkräfte aus dem Schuldienst mit ins Seminar holen Und darüber hinaus haben wir dann manchmal noch so Projektstage Das heißt, da versuchen wir, das Thema dann punktuell auch wiederaufzunehmen, je nachdem, wie die Referentenlage einerseits, die Bedarfe andererseits dann sind.“ (Laurentia, Pos. 8)

5.1.4 Limitationen der Studien zur Angebotsperspektive

Bei der Interpretation und weiteren Nutzung der Ergebnisse sind Limitationen zu beachten.

Limitationen zur Dokumentenanalyse

Die Suche und Auswahl der in der Dokumentenanalyse verwendeten Dokumente wurde in Kapitel 4.2.1 beschrieben. Module – und teilweise sogar ganze Studiengänge, die zum Lehramt an kaufmännisch-beruflichen Schulen berechtigen können – zu denen keine Informationen verfügbar waren, konnten nicht in die Analyse aufgenommen werden. Des Weiteren wurden nicht für alle Standorte sowohl die Dokumente der entsprechenden Bachelor- als auch der Masterstudiengänge analysiert. Grund hierfür ist die teilweise Nicht-Verfügbarkeit der Dokumente (z. B. Technische Universität München) oder aber ein anderer Studiengang im Masterstudium als Wirtschaftspädagogik (z. B. Leuphana Universität Lüneburg). Zudem ist festzuhalten, dass nicht immer alle Module, die dem wirtschaftspädagogischen Masterstudium zugeordnet werden, in den analysierten Dokumenten zu finden waren (z. B. Universität Rostock, 2014; Beschreibung nur zu den Wahlpflichtmodulen im Bereich Bildungswissenschaft). Zur ersten Phase der Lehrer*innenbildung wurden beide Studienrichtungen betrachtet. Um Doppelkodierungen der Module, die für beide Studienrichtungen angeboten werden, wurden diese an den Standorten jeweils einmal in die Analyse einbezogen (vgl. Tabelle 5). Des Weiteren ist zu berücksichtigen, dass manche Standorte zwar verschiedene Wahlmodule anbieten, die Studierenden jedoch nicht alle belegen müssen. Bei den Analysen zur zweiten Phase der Lehrer*innenbildung wurden nicht Dokumente aller Bundesländer bzw. Standorte berücksichtigt.

Zwar wäre es möglich gewesen, bei der Ergebnisdarstellung zur zweiten Phase der Lehrer*innenbildung die behandelten Inhalte der Module in Allgemeinpädagogik und Fachdidaktik zu unterscheiden. Auf diese Darstellung wurde jedoch verzichtet, da nicht zu jedem Bundesland die Modulbeschreibungen zur fachdidaktischen Ausbildung öffentlich verfügbar waren. Eine Unterscheidung bei der Darstellung der Ergebnisse zum Studium wäre auch nicht möglich, da in den Modulen zum Studium nicht stets danach unterschieden wird.

Wie die Dokumente zum Studium, gestalteten sich auch die Dokumente zur zweiten Phase der Lehrer*innenbildung unterschiedlich. Dies wiederum hatte einen Einfluss darauf, wie viele Lerneinheiten (hier: Module) kodiert wurden. So war etwa in der Beschreibung der Ausbildung angehender Lehrkräfte in Hessen das *Unterrichten im Fach Politik und Wirtschaft* (Amt für Lehrerbildung Hessen, 2012, S. 79) im entsprechenden Dokument zu finden, während in Bayern die *Hauptseminarmodule* auf der Homepage aufgeführt wurden (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, o. A.). Die Module zur *Fachdidaktik Wirtschaftswissenschaften/Wirtschaft und Verwaltung* hingegen wurden nicht veröffentlicht (vgl. M. Amend, persönl. Mitteilung, 29.10.2021). Dies führte zu einer Verzerrung im Hinblick auf die Anzahl der analysierten Module und demzufolge auch auf die Anzahl der Kodierungen. Aus diesem Grund wurde darauf geachtet, die Ergebnisse nicht nur in ihrer Anzahl der Kodierungen vorzustellen, sondern auch in den inhaltlichen Aussagen der Passagen. Hier muss einschränkend festgehalten werden, dass in den Dokumenten keine komplexen Ausführungen zum Thema bzw. Inhalt des Moduls folgten. Vielmehr wurden kurze Aussagen und Stichworte verwendet. Dies erschwerte einen tieferen Einblick darin, wie die angehenden Lehrkräfte auf die Erstellung von problemhaltigen (Prüfungs-)Aufgaben vorbereitet bzw. dafür ausgebildet werden.

Unklare und unkonkrete Aussagen fanden sich auch in den Dokumenten zum Studium, welche einen Interpretationsspielraum bieten. Ebenso wurde nicht immer exakt angegeben, welche Inhalte in den Modulen behandelt werden. So fanden sich in den Dokumenten Bezeichnungen wie „Mögliche Inhalte“ (z. B. Goethe-Universität Frankfurt, 2015b). Zudem unterschieden sich auch hier die Dokumente in ihrem Aufbau und in ihren Bezeichnungen. So waren beispielsweise nicht nach jeder Inhaltswiedergabe die Ziele der Veranstaltung zu finden. Auch die Zielbezeichnungen unterschieden sich. Diese hießen „Lernziele“ bzw. „Lernziele und Kompetenzen“ (FAU, 2019a), „Lernergebnisse“ (Universität Mannheim, 2018) und „Qualifikationsziele“

bzw. „Qualifikationsziele/Kompetenzen“ (JGU, 2018)¹⁰⁸. An der Leuphana Universität Lüneburg (2019) wurden Inhalte und z. T. Ziele in einer Spalte angegeben, auch wenn die Spalte lediglich als „Inhalt“ bezeichnet wurde. Ähnliche Aussagen ließen sich über die Beschreibung der Mastermodule an der Universität Konstanz (o. A.) treffen.

Die analysierten Dokumente können jederzeit aktualisiert werden, insofern sollte die Aktualität der Ergebnisse zum Zeitpunkt der Weiternutzung kritisch überprüft werden. Die Dokumentenanalyse zur zweiten Phase der Lehrer*innenbildung für kaufmännisch-berufliche Schulen erfolgte in der ersten Jahreshälfte 2020. Bereits in der zweiten Jahreshälfte wurde eine neue Fassung der Beschreibung der Ausbildung in Bayern veröffentlicht. Stichprobenartig wurden Vergleiche der Module beider Dokumente vorgenommen, mit dem Ergebnis, dass zwischen der analysierten und der später herausgegebenen Fassung keine für diese Arbeit relevanten Unterschiede bestehen (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, 2019, 2020a, 2020b). Insbesondere wurde dabei auf die Module des ersten Ausbildungsjahres *Pädagogische Diagnostik I* und *Pädagogische Diagnostik II* geachtet, da hierzu viele Kodierungen vorliegen.

Um herauszufinden, welche Taxonomiestufen zu den jeweiligen Inhalten bestrebt und Lerngelegenheiten hierzu gegeben werden, wurden die Lern- bzw. Qualifikationsziele der Module – gemäß in die einer *skalierenden Strukturierung* nach Gläser-Zikuda (2013, S. 145-146) – in diese Stufen zugeordnet. Jedoch erwies sich dies als schwierig, da sich oft keine Inhalte in den Zielformulierungen wiederfanden. Ein Beispiel stellt das Bachelormodul *Grundlagen der beruflichen Fachdidaktiken* am Standort Magdeburg dar. Unter den Aufzählungen der Lehrinhalte finden sich „Prüfungen in der beruflichen Bildung“ (Otto von Guericke Universität Magdeburg, 2019, S. 10) wieder, nicht aber in den Qualifikationszielen. Daher konnte die Erweiterung der Studie nicht stattfinden.

Es wäre möglich, die Inhalte der Module anhand der Credit-Angaben zu gewichten. Allerdings wurden in den hier betrachteten Modulen meist nicht ausschließlich für die Forschungsfrage relevanten Themen behandelt, sondern auch andere. Die in den Modulen behandelten Inhalte wurden in den Dokumenten nicht gewichtet, weshalb es nicht möglich war, den Anteil der für diese Arbeit relevanten Themen zu bestimmen.

¹⁰⁸ Wobei sich auch innerhalb eines Dokuments die Bezeichnungen ändern können. So werden Überschriften zu den Zielen einer Veranstaltung an der Johannes Gutenberg-Universität Mainz [JGU] (2018) beispielsweise manchmal „Qualifikationsziele/Kompetenzen“ (z. B. S. 89) und manchmal „Qualifikationsziele/Lernergebnisse/Kompetenzen“ (z. B. S. 11) genannt.

Um einen tieferen Einblick in die Ausbildungsinhalte beider Phasen zu gewinnen, könnte zudem eine weiterführende Studie durchgeführt werden, in welcher die in der Veranstaltung eingesetzten Materialien (z. B. Veranstaltungsfolien) analysiert werden. Dies würde einen Rückschluss darauf ermöglichen, was tatsächlich in dem jeweiligen Modul gelehrt und den Lernenden zur Verfügung gestellt wird. In der vorliegenden Arbeit wurde diese Idee nicht weiterverfolgt. Stattdessen wurde für einen tieferen Einblick in die Thematik die Dokumentenanalyse um die Fragebogenstudie zur ersten Phase (vgl. Kapitel 4.2.3 und 5.1.2) bzw. eine Interviewstudie zur zweiten Phase der Lehrer*innenbildung (vgl. Kapitel 4.2.5 und 5.1.3) erweitert.

Limitationen der Fragebogenstudie

Zur Teilnahme an der Fragebogenstudie wurden alle (vertretenden) Studiengangsleiter*innen der Wirtschaftspädagogik in Deutschland eingeladen, teilgenommen haben 56 % der Eingeladenen. Mit einer früheren Erinnerungsmail oder durch mehrmalige Erinnerungen (vgl. Rost, 2007, S. 93) hätte ggf. eine höhere Rücklaufquote erreicht werden können. Von einigen Studiengangsleiter*innen kamen Rückmeldungen zur Nicht-Teilnahme. Ein*e Studiengangsleiter*in verwies auf das Modulhandbuch, dieses sei zur Beantwortung der Forschungsfrage ausreichend. Ein*e andere*r Studiengangsleiter*in beantwortete die Fragen aus dem Fragebogen in einem Online-Gespräch. Das Besprochene konnte nicht ausgewertet werden, da die Form der Erhebung nicht dem Studiendesign entspricht. Zudem wurde das Gespräch nicht aufgezeichnet, dies hätte insbesondere datenschutzrechtliche Vorbereitungen vorausgesetzt. Notizen wären zwar hilfreich gewesen, diese bergen jedoch ein hohes Fehler- bzw. Verlustpotential (vgl. Vogel & Funck, 2018) – während der aktiven Teilnahme an einem Gespräch konnte nicht jedes Wort notiert werden. Eine weitere Rückmeldung besagte, dass die Stelle des*r Studiengangsleiters*in aktuell nicht besetzt ist und keine Person des Instituts bzw. Lehrstuhls die Fragen beantworten kann.

Begründet werden kann die Nicht-Teilnahme teilweise auch damit, dass kurz vor dem Erhebungszeitraum der erste Lockdown der Corona-Pandemie verhängt wurde. Die Maßnahmen in der Corona-Pandemie waren für Universitäten nicht unbedeutend, da diese beispielsweise recht kurzfristig von Präsenz- auf virtuelle Lehrformate umsteigen mussten. Es liegt nahe, dass eine freiwillige Teilnahme an einer Studie zu diesem Zeitraum keine Priorität hatte.

Eine weitere deutliche Limitation liegt in der Beantwortung der Items in der Fragebogenstudie. Teilnehmende haben zum einen mit „s.o.“ auf ihre Antworten zu vorherigen Fragen verwiesen. Zum anderen wurde der Hinweis gegeben, dass eine andere Person (Mitarbeiter*in des Lehr-

stuhls oder des Instituts bzw. ein*e Lehrbeauftragte*r) für das Modul verantwortlich ist – insgesamt betrifft dies acht Module. Eine Kontaktaufnahme aller Mitwirkenden bzw. Zuständigen konnte nicht erreicht werden, sodass bewusst lediglich die Studiengangsleiter*innen ausgewählt wurden, da davon ausgegangen wurde, dass diese einen Überblick über alle angebotenen Lehrveranstaltungen haben.

Die mit dem Fragebogen ermittelten Informationen sind hilfreich für die Beantwortung der Forschungsfrage, zumindest geben diese einen ersten Einblick in die relevanten Studiengänge einiger Bundesländer. Bei der Auswertung der Daten wurde jedoch festgestellt, dass einige Teilnehmende die Fragen nach der (*Aufgaben-*)*Bewertung* missverstanden haben. Es liegt die Vermutung nahe, dass sie verstanden haben, dass es um die Bewertung der Aufgabe geht und nicht um die Bewertung der Schüler*innenleistungen zu einer Aufgabe. Eine Anpassung der Formulierung ist hier erforderlich. Eine weitere – jedoch als unwesentlich für das Ergebnis erachtete – potentielle Schwierigkeit besteht bei der Benennung der Inhalte in den Fragen. Zwar wurde im Fragebogen *Prüfung* als Wortteil in Klammern gesetzt (z. B. Item 14: „Bitte geben Sie an, ob ... Inhalte zur (Prüfungs-) Aufgabenerstellung ... behandelt werden.“), doch es wäre auch möglich gewesen, dass Teilnehmende den Fokus bei der Beantwortung der Fragen ausschließlich auf das Prüfen legen. Die Ergebnisse und insbesondere die Kodierungen der offenen Fragen bestätigen dies jedoch nicht. Die Teilnehmenden bezogen sich auch auf die anderen Phasen der Diagnostik (vgl. Kapitel 5.1.2).

*Limitationen der Interviewstudie mit Studiengangsleiter*innen*

Mit der Interviewstudie zur zweiten Phase der Lehrer*innenbildung wurde das Ziel verfolgt, tiefere Einblicke in die Ausbildungsinhalte aus der Angebotsperspektive zu gewinnen. Für die Kontaktaufnahme der zuständigen Fachleiter*innen der Studienseminare für kaufmännisch-berufliche Schulen war in einigen Bundesländern eine Genehmigung seitens der zuständigen Ministerien notwendig. Da nicht alle Anträge genehmigt wurden und weil in einem Bundesland trotz Genehmigung des Antrags Fachleiter*innen der Einladung nicht folgten, konnten nicht in allen betrachteten Bundesländern Fachleiter*innen interviewt werden (vgl. Kapitel 4.2.5). Dies stellt eine wesentliche Limitation der Studie in Bezug auf die Stichprobe dar.

Bei der Auswertung der Studie fiel zudem auf, dass sich die teilnehmenden Fachleiter*innen ausschließlich auf die Prozess- und Outputdiagnostik bezogen, nicht aber auf die Eingangsdagnostik. Dies kann daran liegen, dass nicht explizit nach der Eingangsdagnostik gefragt wurde, sondern nur zur Prozess- und Outputdiagnostik (vgl. Kapitel 5.1.3). Dass Teilnehmende

des Interviews nicht von sich aus das Thema ausführen, kann daran liegen, dass die Eingangsdiagnostik bereits bei der Zuteilung in die Schulform an der Berufsschule geschieht und somit weniger häufig von den Lehrkräften ausgeführt wird.¹⁰⁹ Gleichzeitig ist davon auszugehen, dass die Kompetenz im Hinblick auf Prozess- und Outputdiagnostik auch in der Eingangsdiagnostik Anwendung findet. Ebenso können auch Synergien zwischen Prozess- und Outputdiagnostik bestehen.

Zusammenfassend zeigt sich, dass eine standortübergreifende Aussage weder zur ersten noch zur zweiten Phase möglich ist. Teilweise können sich sogar die Jahrgänge eines Standorts in ihren angebotenen Ausbildungsinhalten unterscheiden, etwa durch die inhaltliche Gestaltung der Wahlmodule (vgl. Kapitel 5.1.2 und 5.1.3).

Die Module, im Sinne der Lehre der ausbildenden Institute, stellen lediglich ein Angebot dar. Dieses bedingt nicht, dass die Studierenden und LiV die Inhalte tatsächlich erlernen bzw. das Lern-/Kompetenzziel erreichen. Vielmehr sind diese als Lerngelegenheiten zu verstehen (vgl. hierzu Baumert & Kunter, 2006, S. 477; Kunter et al., 2011, S. 62-63), die von den Lernenden unterschiedlich in Anspruch genommen und genutzt werden können. Insofern ist die Betrachtung der Nachfrageperspektive von großer Relevanz für die vorliegende Arbeit (vgl. Kapitel 4.3). Die Ergebnisse hierzu werden in Kapitel 5.2.1 vorgestellt.

5.2 Ergebnisse der Studie zur Nachfrageperspektive

5.2.1 Ergebnisse der Interviewstudie

Die Interviewstudie mit angehenden Lehrkräften betätigt größtenteils die Ergebnisse der Vorstudien. So gaben alle sechs Teilnehmenden (Masterstudierende und LiV, die mindestens eine Klausur erstellt haben) an, dass sie das Vorgehen bei der Klausurerstellung nicht bzw. kaum während der Ausbildung gelernt haben. Sie bezeichnen ihre Vorbereitung auf die Aufgaben pädagogischer Diagnostik als unzureichend: „*also das kalte Wasser, wo man reinfliegt, ist tatsächlich sehr kalt*“ (Raphael, Pos. 55).

Diese Aussage trifft auf beide Ausbildungsphasen zu. Anhand folgender zwei Zitate als Antwort auf die Frage, wo die Teilnehmenden das bisherige Vorgehen zur Erstellung einer Klausur gelernt haben, soll beispielhaft ein näherer Einblick gegeben werden:

¹⁰⁹ Begründet wird das damit, dass für eine betriebliche Ausbildung rechtlich kein bestimmter Schulabschluss als Voraussetzung definiert wird (vgl. Bundesagentur für Arbeit, 2023; Bundesinstitut für Berufsbildung [BIBB], o. A.). Demnach ist von einer heterogenen Schülerschaft an Berufsschulen – auch in Klassen bestimmter Ausbildungsberufe – auszugehen.

„Gar nicht. ... Also mein [anonymisiert] ist auch ... Lehrer ... da habe ich das natürlich auch schon so ein bisschen mitbekommen.“ (Rebecca, Pos. 23)

„Wie ich Klassenarbeiten erstelle? Gar nicht. Also, das habe ich, ich habe gegoogelt nach Klassenarbeiten. Und habe vorher auch meine Kollegin gefragt ... wir haben uns dann beide überlegt, wie man so eine Klassenarbeit erstellt und wie wir so ein Bewertungsraster erstellen und haben eben, sind da in die Richtung vorgegangen.“ (Svana, Pos. 28)

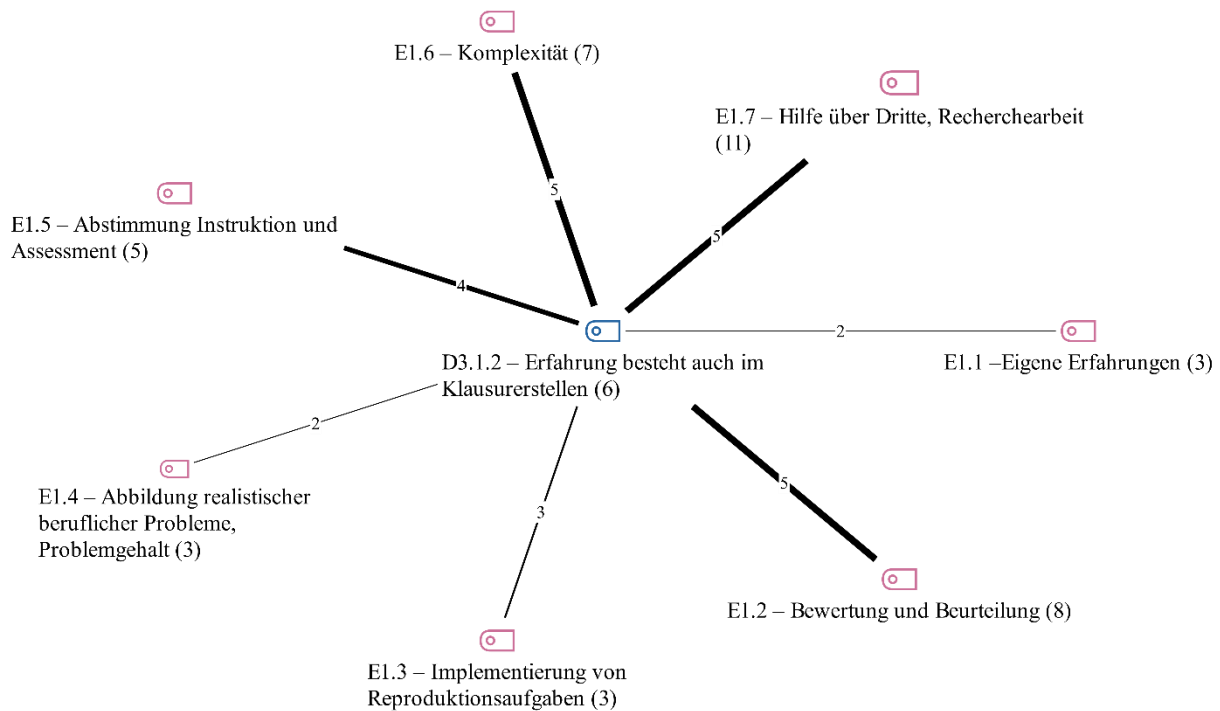
In Bezug auf das Studium sagte Ricardo (Pos. 35) „Nein, in der Tat habe ich dazu ... nicht einen Kurs besucht. Gar nichts“ und berichtete weiter zum Referendariat kritisch:

„ich glaube, da habe ich jetzt auch noch nicht wirklich viel erfahren. ... Wir erstellen ja auch Arbeitsblätter Und das kann ich natürlich auch für Prüfungsaufgaben adaptieren. Aber auch da bin ich gefühlt erst am Anfang und fehlt mir noch so ein bisschen der Überblick und (unv.) Informationen, die ich aber noch nicht konkret in Aufgabenstellungen einfließen lassen kann.“ (Ricardo, Pos. 37)

Abbildung 13 stellt die Übereinstimmung der angesprochenen Kategorien zum Vorgehen bei der Klausurerstellung der erfahrenen Teilnehmer*innen dar. Anhand der Übereinstimmung der Kategorien *D3.1.2 Erfahrung besteht auch im Klausurerstellen* und *E1.6 Komplexität* soll die Grafik beispielhaft erläutert werden. Die Häufigkeit des Vorkommens (hier: 6 und 7) findet sich hinter den Kategorien-Codes (hier: D3.1.2 und E1.6) wieder. Die Zahl auf den Linien (hier: 5) steht für die Anzahl der Dokumente, bei denen mindestens einmal zu den beiden verbundenen Kategorien kodiert wurde. Bei der Interpretation der Zahlen zu der Anzahl der Kodierungen (in der Grafik in Klammern neben der jeweiligen Kategorie gesetzt) ist zu beachten, dass in einem Transkript mehrere Kodierungen zu einem der genannten Codes zu finden sein können.

Abbildung 13:

Code-Relations-Modell (MAXQDA) zum bisherigen Vorgehen beim Klausurerstellen



(Eigene Darstellung)

Zwei Interviewte berichteten, dass sie bei der Erstellung der Aufgaben auf bisher gesammelte *Eigene Erfahrungen* zurückgreifen und aus ihrer gewonnenen Expertise profitieren. Fünf Interviewte berichteten, dass sie beim Erstellen der Aufgaben gleich an die *Bewertung und Beurteilung* (mit-)denken und sich deshalb einen Erwartungshorizont anlegen. Dabei wurde davon berichtet, „*einfach und schnell ... aber auch objektiv korrigieren zu können*“ (Sarah, Pos. 30) und die Flexibilität in den Lösungen bewusst einzuschränken, um „*möglichst zügig*“ (Ricardo, Pos. 13) auswerten zu können. Zum letztgenannten Zitatausschnitt ergänzte der Interviewte kritisch, dass die erstellten Aufgaben „*von daher ... eher eine Wissensabfrage wie problemorientierte Lösungsaufgaben*“ (Ricardo, Pos. 13) sind. Auch andere Interviewte berichteten, dass sie bewusst *Reproduktions- bzw. Wissensaufgaben* in die Arbeit integrieren. Zwei Interviewte achten bei der Aufgabenerstellung darauf, die *berufliche Realität der Schüler*innen* abzubilden bzw. versuchen, die Aufgabe *problemhaltig* zu gestalten. Rebecca erzählte, dass sie sich hierzu „*erst mal fragen [musste]: Was ist ... wichtig, was für Probleme können die [Schüler*innen] auch in der Realität haben, mit denen sie umgehen müssen*“ (Pos. 11). Nur eine Person gab an, bei der Erstellung von Klausuren IHK-Prüfungsaufgaben als Hilfe bzw. Orientierung zu nutzen: „*Ja, ich richte mich doch stark nach dem, was später in der IHK gefragt wird als Prüfungsfragen und versuche die Schülerinnen und Schüler darauf schon so ein bisschen vorzubereiten und*

ein bisschen, ja, eigentlich eine Wissensabfrage“ (Ricardo, Pos. 13). Auf der anderen Seite gab es auch Teilnehmende, die sich bewusst nicht an den Aufgaben der IHK orientieren. So sagte Saika kritisch:

„Nö da [in die IHK-Prüfungsaufgaben] hab‘ ich gar nicht reingeschaut, weil ich die nicht also in Erinnerung hatte ich sie nicht so problemorientiert, sondern eher so abfragemäßig. Das ist ja fast immer nur nennen und aufzählen und ganz selten mal begründen. Oder man soll eine Reihenfolge herstellen, oder so was.“ (Saika, Pos. 111)

Vier Teilnehmende berichteten, dass sie die Aufgaben der Klausur auf den Unterricht und z. T. die Berufspraxis beziehen, es gab aber keine Person an, das Curriculum (nochmals) betrachtet zu haben. Im Sinne der *Curriculum-Instruktion-Assessment-Triade* (Pellegrino, 2010) wäre dies jedoch sinnvoll. Dass Proband*innen dies nicht nannten, kann damit zusammenhängen, dass das bei der Unterrichtskonzeption bereits betrachtet wird und der Rückbezug auf die Instruktion die Beachtung curricularer Vorgaben einschließt.

Unter der Kategorie der *Komplexität*, zu welcher sich Passagen in fünf Dokumenten (fünf Interviewte) finden, wurden Schritte zum Vorgehen festgehalten, die die Bearbeitungsdauer und Schwierigkeit einer Aufgabe sowie die zugelassenen Hilfsmittel zur Bearbeitung der Aufgabe umfassten. Auch damit setzen sich die Interviewten bei der Erstellung von Aufgaben bzw. Konzeption von Klausuren auseinander. Die meisten Kodierungen dazu finden sich in der Kategorie E1.7. Die Interviewten berichteten, dass sie ihre Mentor*innen beim Erstellen der Aufgaben um *Hilfe* gebeten haben und sich beispielsweise „zwei bis drei [Alt-]Klausuren zukommen lassen“ (Seila, Pos. 16) haben, mit denen sie gearbeitet haben – diese wurden modifiziert, ergänzt und/oder abgeändert. Ebenfalls holen sie sich Unterstützung von Kolleg*innen, indem beispielsweise „diese [selbst erstellte] Klassenarbeit einer Kollegin gezeigt [wird], die auch [Fach anonymisiert] unterrichtet“ (Svana, Pos. 20). Es wurde auch davon berichtet, (Fach-)Bücher zu nutzen. Dass fünf von sechs erfahrenen Teilnehmer*innen bei der Erstellung von Aufgaben Unterstützung einholen und Recherchen betreiben, kann zum einen auf die fehlende Expertise zurückgeführt werden. Zum anderen kann dies aber auch damit erklärt werden, dass die Vorbereitung auf diese Tätigkeit einer Lehrkraft durch die bisherige Ausbildung der Interviewten nicht ausreichend ist. Zu Letzterem wird im nächsten Absatz konkreter eingegangen.

Auch wenn im Studium bzw. Referendariat Inhalte hierzu behandelt wurden, gab die große Mehrheit der Teilnehmenden (13 Personen, davon zwei LiV) an, sich kaum bzw. nicht (ausreichend) auf die Aufgaben pädagogischer Diagnostik vorbereitet zu fühlen. Das wahrgenommene Problem berichtet ein LiV wie folgt:

„Ja, pädagogische Diagnostik hatten wir in der Tat auch schon Sitzungen zu ... Das Problem ist vielleicht auch die zeitliche Komponente wieder, das findet ja alles nur in kurzen Abständen statt, wird alles angerissen und ... im Alltag werde ich dann mit diesen Dingen oftmals alleine gelassen und muss schauen, wie ich damit zurecht komme. Also ich kriege einen Input und bin dann angehalten, mich selbst damit zu befassen. Und da würde ich tatsächlich erwarten, dass sich dafür ein bisschen mehr Zeit genommen würde, dass das sich auch anzugucken und konkreter zu werden.“ (Ricardo, Pos. 69)

Konkret bezieht sich die Kritik auf das Erstellen von (problemhaltigen) Aufgaben. Obwohl im Studium das Thema der Erstellung von (problemhaltigen) Aufgaben – auch außerhalb der Interventionsstudie (vgl. Kapitel 7) – behandelt wird (dies wurde von neun Studierenden und einer*m LiV bestätigt), diente diese Unterrichtszwecken. Zudem geht aus den Transkripten hervor, dass das Thema im Studium kaum Anwendung findet. Ein Student sagte hierzu: *„Wir sollten uns Gedanken machen über Aufgaben, selbst erstellt haben wir ja keine, das war nur was für uns gute ... Aufgaben ausmacht“ (Samuel, Pos. 60).*

Nicht unabhängig davon berichteten dieselben neun Studierenden und zwei LiV davon, das Gefühl zu haben, dass das Erstellen (problemhaltiger) Aufgaben im Studium nicht ausreichend behandelt wurde, insbesondere zu summativen Zwecken. Das Ergebnis wird durch die Einschätzung von Sina gestützt, die sagte, dass das Thema der *„Aufgabenerstellung ... noch mehr behandelt werden [muss]“* und ergänzte, dass es zwar *„in Diagnostik ein kleines bisschen dazu [gab], aber ... zu kurz [kam]“ (Pos. 76).* Ein ähnliches Bild zeigt sich zum Referendariat. Zwar gaben die interviewten drei LiV an, dass das Thema der (problemhaltigen) Aufgabenerstellung Bestandteil des Referendariats ist (wenn auch zu unterrichtlichen Zwecken). Gleichzeitig wurde dieses von allen drei LiV kritisiert bzw. als nicht ausreichend empfunden. So sagte etwa Raphael folgendes:

„Ja, ... wir wissen das alles auf dem Papier, Lebensweltbezug und lebenslanges Lernen und das ist alles schön und gut. Aber wie man das dann eben doch umsetzt bzw. wie komme ich denn an die Informationen, die ich brauche, um so eine Aufgabe realitätsnah und quasi irgendwie die Lebenswelt der Schüler betreffend abbilden kann. Ich brauche Informationen dazu ... das sagt einem erstmal keiner.“ (Raphael, Pos. 59)

Positivere Rückmeldungen kamen dazu, dass das Thema der *Bewertung und Beurteilung* (acht Studierende) sowie das Thema der *Gütekriterien* (neun Studierende), wenn auch z. T. theoretisch bzw. als *„die ersten Ansätze“ (Sibel, Pos. 48),* im Studium behandelt wurde. In Bezug auf das Referendariat gaben LiV an, das Thema der *Bewertung und Beurteilung* (zwei LiV) und

der *Gütekriterien* (ein LiV) behandelt zu haben. Für beide Ausbildungsphasen wurden jedoch Einschränkungen dahingehend berichtet, dass das Wissen hierzu nicht nachhaltig sei. Die Gründe hierfür sind das Prüfungsformat¹¹⁰, der geringe Umfang und/oder die fehlende bzw. zu kurz kommende, tiefergehende Behandlung des Themas (z. B. Ricardo, Pos. 65; Samira, Pos. 82; Sarah, Pos. 57).

Auch wenn ein Bezug auf die Praxis durch Praktika im Studium von sieben Teilnehmenden (davon zwei LiV) erwähnt wurde, ist aus den Interviews zu entnehmen, dass dieser nicht als ausreichend empfunden wird. Interviewte kritisierten den hohen Theorie- und den geringen Praxisanteil zur pädagogischen Diagnostik im Studium. Besonders wurde hier Bezug auf die Erstellung von (problemhaltigen) Aufgaben zu Zwecken des summativen Assessments genommen. Folgende Passagen verdeutlichen dies:

„... also ich habe ganz viel Theorie gelernt in meinem Studium, aber mir fehlt die praktische Umsetzung. Und in den Modulen, wo es primär um die Praxis ging, ging es halt eher um die Unterrichtsgestaltung ... es ging eigentlich nur darum, wie man Lehr-Lern-Prozesse gestalten kann.“ (Saika, Pos. 48)

„Auf Prüfungsaufgaben außerhalb von diesem Seminar [indem die Intervention stattfand] eigentlich gar nichts. ... im Studium lag der Schwerpunkt halt immer sehr auf den Methoden ... gerade in der Praktikumsvorbereitung.“ (Sven, Pos. 60)

Die Kritik der fehlenden Praxis bezieht sich aber auch auf die (Fach-)Didaktik – die Aussage von Sabrina fasst dies zusammen:

„Also insgesamt fühle ich mich vom Didaktischen her. Ja, ich kenne die Konzepte, aber ich glaube, die Umsetzung wird im Referendariat dann noch mal schwer, weil man .. im Studium einfach viel zu wenig diese praktischen Erfahrungen hat. Man lernt die ganzen theoretischen Konzepte und man erstellt mal eine Aufgabe, aber bis ... auf die beiden Praktika hat man keinerlei Praxiserfahrung. Und ich glaube, das ist schwierig.“ (Sabrina, Pos. 52)

Interessant ist, dass zwei der drei LiV auch in Bezug auf das Referendariat die fehlende/geringe Praxis bemängelten. Rebecca sagte hierzu eindeutig, dass die *„Verknüpfung von Theorie und Praxis fehlt“* (Pos. 39). Zwar teilten Teilnehmende mit, dass sie das Thema der Kompetenzor-

¹¹⁰ Z. B. sagte Sabrina (Pos. 62) dazu: *„Und ja nach der Klausur ist es ja so, dass man nur noch die Hälfte weiß, wenn überhaupt“*.

ientierung im Studium (acht Teilnehmende) bzw. im Referendariat (zwei Teilnehmende) behandelt haben, dies aber passend zu den oben dargestellten Zitaten eher theoretisch geschah, kaum Anwendung fand und der Fokus auf das Unterrichten lag:

„Aber das waren eher so Lernen im Modell, Handlungskompetenzen, Handlungsorientierung ... und so was. Es ging eher darum kam, wie gestalte ich meinen Unterricht ... und nicht wie konzipiere ich auf Aufgaben.“ (Sami, Pos. 56)

Zwar wurde im Studium über dieses Thema „gesprochen“ (Raphael, Pos. 47), und auch im Referendariat liegt der „Fokus immer auf den Kompetenzen“ (Rebecca, Pos. 39), doch die Umsetzung bzw. Anwendung dessen kommt zu kurz¹¹¹. Diese und ähnliche Aussagen, die neben der pädagogischen Diagnostik die (fach-)didaktische Ausbildung ansprechen, sind deshalb besonders relevant, da sie die Konzeption von Aufgaben, auch zu summativen Zwecken, betreffen (vgl. Kapitel 2.5). Samis Aussage dazu, wo er das hypothetische Vorgehen zur Erstellung einer Klausur¹¹² gelernt hat, bilden die Meinung vieler Teilnehmenden ab (elf Teilnehmende zum Studium, davon zwei LiV; alle drei Teilnehmenden zum Referendariat):

„das habe ich auch durch tatsächlich jetzt durch Fachdidaktik¹¹³ gelernt. Also deswegen finde ich das Modul auch sehr interessant, weil es mal wieder ein praxisnahes Modul ist und wir auch wirklich Aufgaben erstellen müssen und wir da auch mal diese Herangehensweise lernen. ... und wo wir das unter diesen diagnostischen Gesichtspunkten und auch vor allem unter der Problemlösekompetenz unter diesen Kriterien sozusagen Aufgaben prüfen, ob das denn überhaupt einsetzbar ist.“ (Sami, Pos. 52)

Zur Verbesserung des Studiums schlug er vor

„[zu] versuchen ... mehr Beispiele aus dem Feld zu nehmen, also auch aus der Praxis, ... vielleicht auch das anonymisierte Daten aus der Praxis ... Dass man auch weiß, so okay, das und das wird dann wirklich geantwortet von den Schülerinnen und Schüler jetzt gerade auf die Sache mit den Aufgabenkonzipierungen oder auch in der Didaktik“ und ergänzte *„Theoretisch sind wir dann wahrscheinlich gut ausgebildet, aber praktisch würde ich mir wünschen. Ich weiß auch, dass es so in meinem Umkreis so ist, dass sich das viele wünschen würden, dass es sehr viel praktischer angelegt wird.“* (Sami, Pos. 92)

¹¹¹ Rebecca sagte dazu: *„Also diese Verknüpfung von Theorie und Praxis fehlt mir dann oft, damit ich wirklich weiß, was gemeint ist“* (Pos. 39). Ricardo (Pos. 53, 57) berichtete, dass er sich bei der Umsetzung schwertut.

¹¹² Diese Frage wurde Teilnehmenden gestellt, die bisher keine Aufgaben zu summativen Zwecken erstellt haben.

¹¹³ Im Rahmen der genannten Veranstaltung fand die Intervention (vgl. Kapitel 7) statt.

Insgesamt wurde von Teilnehmenden der Interviewstudie der Wunsch geäußert, dass Veranstaltungen angeboten werden, die einen höheren Praxisanteil haben und das Thema der Erstellung problemhaltiger Aufgaben (zu summativen Zwecken) stärker und früher im Studium einzubeziehen:

„Also ... ich fand dieses Modul [indem die Intervention stattfand] richtig cool mit den problemhaltigen Aufgaben und ich würde mir das schon wünschen, dass man das vielleicht im Bachelor auch schonmal macht, dass man sich noch mehr mit der ... Aufgabenerstellung beschäftigt und auch mit der Diagnostik. Also, dass es ... nicht ganz so forschungsorientiert und theoretisch ist. Ich verstehe, dass ... man das braucht als Grundlage. Aber vielleicht, dass man noch mehr in die Praxis geben kann, was einen wirklich dann Hilfen an die Hand gibt für den späteren Job, weil das das fehlt mir so im Studium, so ein bisschen. Man fühlt sich trotzdem noch komplett unvorbereitet auf so einen Lehrerberuf.“ (Saika, Pos. 68)

Es muss berücksichtigt werden, dass acht Teilnehmende (fünf Studierende, drei LiV) angaben, noch nicht viele (wirtschafts-)pädagogische Module belegt zu haben, da sie sich noch zu Beginn des Masterstudiums/des Referendariats befinden und/oder an ihrem Standort im Studium nicht viele pädagogische Module angeboten werden. Weiterhin zeigte sich eine Einschränkung in Bezug auf die Wahlmodule. So teilte Raphael mit, dass die (problemhaltige) Aufgabenerstellung in seinem Studium *„kein verpflichtender Schwerpunkt“* ist und ergänzte dies wie folgt:

„Das wird es mit Sicherheit irgendwo in einer Universität gegeben haben. Aber die Modulfelder nannten sich in der in der Pädagogik immer [Modulnamen anonymisiert]. Und als Beispiel [Modulname anonymisiert] hatte dann die Wahl aus acht bis zehn verschiedenen Seminaren, die man ... belegen konnte. Da musste man dann vielleicht eins bis zwei machen, um dieses Modul abzuschließen. Jetzt war das in meinem Fall beispielsweise ... niemals irgendwas mit Diagnostik.“ (Raphael, Pos. 45)

5.2.2 Limitationen der Interviewstudie mit angehenden Lehrkräften

Das Ergebnis der Studie zeigt homogene Erfahrungen der angehenden Lehrkräfte, sodass vorsichtig auch auf die Erfahrungen anderer nicht an der Studie teilgenommenen Personen der gleichen Gruppe (Studierende der Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen) geschlossen werden kann (vgl. Flick, 2010, S. 282). Eine allgemeingültige Aussage für das wirtschaftspädagogische Studium und das Referendariat für das Lehramt an kaufmännisch-beruflichen Schulen in der Bundesrepublik Deutschland ist aufgrund der geringen Stichprobe (13 Studierende, drei LiV) jedoch nicht möglich. Vielmehr bietet sie einen Einblick in

die Ausbildung beider Phasen an ausgewählten Standorten. Die befragten LiV konnten Äußerungen zu den Kategorien in Bezug auf die Studieninhalte treffen (Kategorie S des Kodierleitfadens im Anhang D), Masterstudierende allerdings nicht zum Referendariat (Kategorie REF des Kodierleitfadens im Anhang D). Das heißt, dass lediglich Aussagen der LiV in die Kategorien zu den Inhalten des Referendariats kodiert wurden. Eine Sättigung konnte dahingehend nicht erreicht werden. Um allgemeingültige Aussagen für beide Phasen der Ausbildung der Zielgruppe treffen zu können, müsste die Stichprobe erweitert werden. Anhand der erweiterten Stichprobe kann zudem überprüft werden, ob der Kodierleitfaden auch für Interviews mit angehenden Lehrkräften anderer Standorte funktioniert oder etwa einer Erweiterung bedarf (z. B. aufgrund weiterer Verbesserungswünsche zum Studium).

Die Rückmeldungen zu den Ausbildungsinhalten der Teilnehmenden sind als Selbsteinschätzung zu verstehen. Selbiges gilt auch für ihre Wahrnehmung der unzureichenden Vorbereitung in Bezug auf Aufgaben der pädagogischen Diagnostik allgemein und auf die Erstellung von problemhaltigen Prüfungsaufgaben im Speziellen (F_{1-3}). Da es sich um subjektive Auskünfte handelt, besteht die Gefahr, dass die erlangten Informationen nicht sicher sind (vgl. Terhart, 2002, S. 37). Eine weitere Einschränkung zeigt sich dahingehend, dass sich sieben Teilnehmende¹¹⁴ im ersten Masterjahr befanden und fünf davon angaben, noch nicht viele pädagogische Module belegt zu haben. LiV betonen, die Ausbildung noch nicht abgeschlossen zu haben: „*Ich bin im Lernprozess*“ (Ricardo, Pos. 31). Insofern sind die Aussagen der Proband*innen zu den Ausbildungsinhalten mit Vorsicht zu interpretieren, wie auch die folgende Aussage zeigt:

„*Gut, da muss man jetzt natürlich sagen Wir haben im Mai angefangen, wir sind ja noch lange keine Profis und noch weit ab von einer Profession. Wir sind ja auch noch am Lernen und wir machen ja auch Fehler.*“ (Raphael, Pos. 53)

Samuel (Pos. 37) ergänzte zu seiner Antwort, dass er sich durch das Studium „*noch nicht besonders gut*“ auf Aufgaben der pädagogischen Diagnostik vorbereitet fühlt, da er noch nicht alle vorgesehenen Module des Masterstudiums belegt hat. Bei allen Interviewten handelt es sich also um Personen, die sich noch in der Ausbildungsphase befinden – ob im Studium oder Referendariat. Das heißt auch, dass sie noch nicht die Möglichkeit hatten, an allen vorgesehenen Lernangeboten der Universität und/oder des Studienseminars teilzunehmen. Somit ist es den Teilnehmenden noch nicht möglich, abschließende Aussagen zu ihrer Ausbildung zu tref-

¹¹⁴ Davon waren eine Person im ersten und sechs im zweiten Mastersemester.

fen. Vielmehr bildet das Ergebnis den aktuellen Stand der Wahrnehmung zur bisherigen Ausbildung ab, auch wenn einige Teilnehmende in ihrer Ausbildungsphase fortgeschritten sind bzw. eine der Phasen bereits abgeschlossen haben (LiV). Die Ergebnisse zeigen jedoch einen hohen Grad an Übereinstimmung mit den Vorstudien, gerade in Bezug auf das Studium. Unter Berücksichtigung dieser und der hier ausgeführten Limitationen in Bezug auf die Stichprobe kann dennoch vorsichtig davon ausgegangen werden, dass die Ergebnisse zuverlässig sind und sich nicht (wesentlich) verändern würden, wenn Personen befragt werden, die ihre Ausbildung bereits abgeschlossen haben.

Die insgesamt relativ häufig geäußerte Kritik in Bezug auf die fehlende Praxisnähe kann auch mit dem Zeitraum der Ausbildung zusammenhängen. Die Interviews wurden in den Jahren 2021 und 2022 durchgeführt. Die Teilnehmenden waren in ihrer Ausbildungsphase in dieser Zeit – und in den Semestern davor (etwa März 2020) – von der Corona-Pandemie und den damit verbundenen Maßnahmen betroffen. Aufgrund dessen, dass die Lehre an den Universitäten, Studienseminaren und Schulen überwiegend online/digital stattfand, könnte der Praxisanteil, insbesondere im Studium, (noch) geringer ausgefallen sind als davor.

Bosse (2012, S. 13-16) zeigte bereits vor der Pandemie für die erste Phase der Lehrer*innenbildung, dass sich die Bundesländer in den praktischen Anteilen unterscheiden. Zwar beziehen sich die Ergebnisse auf das Gymnasium, dennoch scheinen sie auch für Lehrkräften an beruflichen Schulen zutreffend zu sein. So sind beispielsweise im Masterstudium Wirtschaftspädagogik an der Universität Duisburg-Essen (2019b, S. 39) ein Praxissemester mit einem Arbeitsaufwand von insgesamt 390 Stunden vorgesehen, während im Masterstudium Wirtschaftspädagogik an der Goethe-Universität Frankfurt (2015b, S. 41) das Praktikum über zwei Semester geht (acht Stunden Hospitation, ein dreiwöchiges Blockpraktikum sowie weitere acht Stunden Hospitation).

Des Weiteren ist zu berücksichtigen, dass nicht alle Aspekte, die im Zusammenhang mit der Beantwortung der Forschungsfrage(n) zusammenhängen, berücksichtigt werden konnten. In Bezug auf die Erstellung von (Prüfungs-)Aufgaben äußerte beispielsweise eine Studentin Kritik dahingehend, dass der Umgang mit curricularen Vorgaben kaum gelernt wird:

„Das, wenn man dann wirklich in der Klasse ist und selbständig Aufgaben erarbeiten muss und man zwar den Rahmenlehrplan zur Verfügung hat, aber so im gesamten Studium immer nur ein Bereich vielleicht mal kurz angeschaut hat, aber nicht wirklich gelernt hat, wie man zum Beispiel im Rahmenlehrpläne oder Prüfungskataloge liest und wie man daraus auch

aufbauen, dann auch wirklich ein ganz[es] Konzept für ein ganzes Schuljahr erstellt. Ich glaube, das wird herausfordernd.“ (Sabrina, Pos. 52)

Da dieser Aspekt nicht häufig genannt wurde, wurde hierzu keine Subkategorie gebildet, sondern die Aussage in die Kategorie *S2.3 – Nicht gut/ unvorbereitet auf pädagogische Diagnostik* eingeordnet. Bei einer größeren Stichprobe – und damit mehr Teilnehmenden – wäre dies möglicherweise wiederholt genannt worden. Entsprechend birgt die geringe Anzahl an Studienteilnehmer*innen die Gefahr, dass bestimmte Aspekte unberücksichtigt oder zumindest unterschätzt bleiben. Dass einige Themen kaum Gegenstand der Interviews waren, kann auch damit zusammenhängen, dass sie von den Teilnehmenden als kaum oder nicht relevant eingeschätzt wurden, die (Nach-)Fragen im Interview nicht präzise genug waren oder die Teilnehmenden bei der Beantwortung der ihnen zunächst unbekannteren ad-hoc (Nach-)Fragen nicht daran gedacht haben.

Dass die interviewten Personen vorab an der Interventionsstudie teilnahmen (vgl. Kapitel 7), bringt Vor- und Nachteile mit sich. Die Teilnehmenden wurden so schon für das Thema der Erstellung problemhaltiger Prüfungsaufgaben sensibilisiert. So sind sie ggf. kritischer mit den bisherigen Ausbildungsinhalten zum Thema umgegangen als Personen sein würden, die vorab nicht darüber reflektiert haben. Zudem haben die Teilnehmenden im Rahmen der Intervention entsprechende Aufgaben formuliert. Dadurch haben sie die relevanten Aspekte bei der Erstellung solcher Aufgaben vermutlich entdeckt. Entsprechend könnten sie dadurch konkretere Anforderungen formuliert haben.

Zudem ist zu berücksichtigen, dass die Teilnahme an der Interventionsstudie in Form eines Trainings für die LiV freiwillig war. Es ist daher ein Selbstselektionseffekt anzunehmen und davon auszugehen, dass nur diejenigen teilgenommen haben, die ohnehin Interesse am Thema haben. So besteht die Vermutung, dass Personen, die ihre Vorbereitung durch das Studium und/oder das Referendariat auf die Erstellung von problemhaltigen (Prüfungs-)Aufgaben als ausreichend gut wahrnehmen, eher nicht an dieser Maßnahme teilgenommen haben. Es ist auch zu bemerken, dass die Veranstaltung, im Rahmen dessen die Intervention stattfand, eine (Wahl-)Pflichtveranstaltung für die Studierenden war, sodass es auch hier zu Selbstselektionseffekten gekommen sein kann (vgl. Kapitel 8.1.4).

In der vorliegenden Arbeit werden die Ergebnisse der Vorstudien (vgl. Kapitel 5.1) der konkreten Standorte nicht mit den Ergebnissen aus den Interviews mit Teilnehmenden der jeweiligen Standorte verglichen. Begründet wird dies damit, dass zum einen die Ergebnisse nicht auf Standortebene vorgestellt werden, sondern versucht wird, ein Gesamtbild abzubilden (vgl. z. B.

Kapitel 4.2.1). Auf der anderen Seite ist es nicht Ziel der Studie, die Ausbildungsstandorte zu beiden Phasen der Ausbildung zu evaluieren. Vielmehr ging es um die Gewinnung eines Einblicks in die Ausbildung angehender Lehrkräfte kaufmännisch-beruflicher Schulen. Wie bereits dargestellt, ist es sinnvoll, die Studie um weitere Studierende und LiV auch anderer Standorte zu erweitern, um der Nachfrageperspektive stärker Rechnung zu tragen. Dies würde zur Annäherung an ein allgemeingültiges Ergebnis beitragen.

5.3 Zusammenfassende Implikationen aus den Studien zu den Ausbildungsinhalten

Je nach Standort beider Phasen unterscheiden sich die Ausbildungsinhalte, sodass eine allgemeingültige Aussage nicht möglich ist. Zusammenfassend kann jedoch festgehalten werden, dass die erste Phase der Lehrer*innenausbildung primär ein theoretisches Fundament legt, die Vertiefung und Erweiterung der Inhalte (z. B. durch Anwendung) geschieht eher in der zweiten Phase¹¹⁵. Themen zur Outputdiagnostik werden zwar in der zweiten Phase der Lehrer*innenbildung stärker behandelt als in der ersten, jedoch scheint auch dies nicht hinreichend mit Blick auf die Erstellung problemhaltiger Aufgaben zu Zwecken der Outputdiagnostik aus Sicht der angehenden Lehrkräfte zu sein (vgl. Kapitel 5.2). Vielmehr dienen Themen zur Erstellung kompetenzorientierter (handlungs- und/oder problemorientiert) Aufgaben in beiden Phasen häufig zu Zwecken der Gestaltung von Lehr-Lern-Prozessen, etwa in Form von Lernsituationen und/oder Arbeitsblättern. Auch wenn diese Inhalte nicht irrelevant für die Erstellung von (problemhaltigen) Aufgaben zu Zwecken der *Outputdiagnostik* sind, ist festzustellen, dass diese kaum konkret thematisiert werden.

Die Ergebnisse der Interviewstudie mit den Seminarfachleitungen zeigen bessere Ergebnisse als die der Dokumentenanalyse und der Interviewstudie mit LiV (vgl. Kapitel 5.1.3 und 5.2.1). Themen der *Bewertung und Beurteilung*, der *Gütekriterien* und der *Berücksichtigung curriculärer Anforderungen* sind Bestandteil beider Ausbildungsphasen, so die Ergebnisse der hier vorgestellten Analysen. Diese Themen beziehen sich nicht immer konkret auf die Gestaltung von Aufgaben zu diagnostischen Zwecken zur Ermittlung des Outputs, dennoch sind sie als bedeutende Grundlage dafür zu betrachten (vgl. Kapitel 2.4).

Die Kritik in Bezug auf die Theorie-Praxis-Verzahnung in der Lehrer*innenbildung existiert bereits länger (z. B. Schaefers, 2002), doch es wird auch noch aktuell dazu geforscht, beispielsweise im Rahmen der *Qualitätsoffensive Lehrerbildung* (Bresges, Harring, Kauertz, Nordmeier & Parchmann, 2019, S. 5). Die Forderungen der angehenden Lehrkräfte in der Interviewstudie

¹¹⁵ Dieses Ergebnis überschneidet sich mit den curricularen Anforderungen aus Kapitel 2.2.

bezüglich einer Erhöhung der Praxisanteile, insbesondere im Studium (vgl. Kapitel 5.2), können nicht ohne Weiteres umgesetzt werden. Studien zur Praxis in der ersten Phase der Lehrkräftebildung zeigen nämlich, dass die Erhöhung des Praxisanteils nicht unmittelbar mit einer Verbesserung der Kompetenzen angehender Lehrkräfte einhergeht. Vielmehr ist die Qualität dieses entscheidend (vgl. z. B. Gröschner, 2012; Gröschner & Hascher, 2019; Müller & Dieck, 2011; siehe auch Terhart, 2002, S. 33). Insofern ist zu überlegen, wie die Qualität der Lehrer*innenbildung, insbesondere in der ersten Phase, für angehende Lehrkräfte verbessert werden kann, um einen guten Einblick und ggf. (weitere)¹¹⁶ Erprobungen in der Praxis effektiv zu ermöglichen.

Durch die Praxis wiederum ist zu erwarten, dass auch die Handlungskompetenz angehender Lehrkräfte verbessert wird, da das Handlungswissen in kompetentes Handeln mündet (vgl. Dubs, 2008, S. 13-14). Terhart (2002, S. 2) hält fest, dass der Praxisanteil sowie der Anteil pädagogischer Inhalte in der Lehrer*innenbildung für untere Jahrgangsstufen im Vergleich zu oberen höher ist. Entsprechend könnten, neben dem Trainingskonzept (Kapitel 7), auch Ideen zur Umsetzung einer praxisnäheren Ausbildung entwickelt werden, indem (allgemein-)pädagogische Inhalte (z. B. der pädagogischen Diagnostik) aus der Ausbildung von beispielsweise Grundschullehrer*innen adaptiert werden. Dabei sollte der Fokus insbesondere auf die Erstellung von problemhaltigen Aufgaben zu Zwecken der Outputdiagnostik, im Sinne der validen Prüfung von Kompetenzen in der kaufmännisch-beruflichen Bildung (vgl. Kapitel 3.2), geschärft werden.

Gleichzeitig findet sich in der Literatur auch eine Forderung nach der Differenzierung in den Ausbildungsphasen. So unterstreicht Neuweg (2010), dass „eine zentrale Aufgabe der Grundausbildung [aller Lehrämter] in der Grundlegung der fachlichen Kompetenz liegt“ (S. 41). Insofern ist die Frage berechtigt, ob und wie den Anforderungen entsprochen und auf die Forderungen einer praxisnäheren Ausbildung eingegangen werden kann, ohne dafür im fachwissenschaftlichen Bereich Einbußen hinzunehmen.

Die Vorschläge bzw. Kritikpunkte der Teilnehmenden, das theoretisch Gelernte anzuwenden und die anschließende Reflexion dessen – die auch Müller und Dieck (2011, S. 49) als eine Herausforderung folgern – können auch auf die Erstellung von problemhaltigen Aufgaben zu Zwecken des summativen Assessments übertragen werden. Dabei kann an Ertl-Schmuck

¹¹⁶ Es ist zu berücksichtigen, dass durch das Praktikum im Studium bereits Praxiseinblicke und erste Erfahrung ermöglicht sein sollten.

(2019) anlehnend „die Schulpraxis über authentische Fälle“¹¹⁷ (S. 50) kritisch reflektiert werden, im Falle der Erstellung von Aufgaben zu summativen Zwecken durch etwa Klausuraufgaben von Lehrkräften. Dass die Reflexion der Inhalte relevant ist, geht auch aus der Forschung zur Lehrkräftefortbildung hervor (vgl. Kapitel 6.3) sowie auch in Bezug auf die Ausbildung von Lehrkräften in der ersten Phase (Dubs, 2008, S. 18-21)¹¹⁸. Weitere Ansätze zur Gestaltung der universitären Lehrpraxis zur Ausbildung von angehenden Lehrkräften finden sich in Dubs (2008, S. 21-24). Da bei der Aufgabenerstellung fachwissenschaftliche Aspekte eine zentrale Rolle spielen, wird davon ausgegangen, dass die fachwissenschaftliche Perspektive beim Behandeln dieses Themas nicht zu kurz kommt. Gerade vor dem Hintergrund, dass im wirtschaftspädagogischen Studium wirtschaftswissenschaftliche Module einen wesentlichen Bestandteil bilden (als *Fachwissenschaft* in Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, 2014) und diese in der Regel pädagogischen Modulen vorgeschaltet sind¹¹⁹, kann davon ausgegangen werden, dass im Rahmen der Aufgabenerstellung – auch im Studium – dieses fachwissenschaftliche Wissen im pädagogischen Kontext umgesetzt werden kann.

¹¹⁷ Ähnlich auch in Artelt und Kunter (2019, S. 407).

¹¹⁸ Dubs (2008, S. 18-21) bezieht sich in diesem Zusammenhang auf die Ausbildung für das Unterrichtspraxis. Dennoch ist es naheliegend, dass seine geschilderten Gedanken auch auf das Entwickeln von Aufgaben zu summativen Zwecken übertragen werden kann, was als Bestandteil der *Beurteilungsaufgabe* einer Lehrkraft (Dubs, 2008, S. 17) zugeordnet werden würde (vgl. Kapitel 2.4 zur diagnostischen Kompetenz von Lehrkräften).

¹¹⁹ Hierzu fand keine systematische Prüfung statt. Vielmehr war es eine Beobachtung aus ausgewählten Studien(verlaufs)plänen, z. B. Friedrich-Alexander-Universität Erlangen-Nürnberg [FAU] (2019b, S. 34-35); Georg-August-Universität Göttingen (o.A.); Mercator School of Management (o. A., S. 2); Technische Universität Dresden (2016).

6 Stand der Trainingsforschung

6.1 Hinführung zu den Einflussfaktoren

Im Rahmen dieser Arbeit wird das Verständnis zur Wirksamkeit von Bildungsmaßnahmen nach Lipowsky und Rzejak (2021, S. 12) verwendet. In diesem wird eine Fort-/Weiterbildungsmaßnahme¹²⁰ dann als wirksam erachtet, wenn „die Fortbildungen wissenschaftlich belegt zur Weiterentwicklung wichtiger Kompetenzen von Lehrpersonen beitragen und/oder sich positiv auf die Weiterentwicklung des Unterrichts und auf das Lernen von Schülerinnen und Schülern auswirken“ (Lipowsky & Rzejak, 2021, S. 12). In Bezug auf die Wirksamkeitsfaktoren einer Maßnahme unterscheidet Lipowsky (2009, S. 347) zwischen unterschiedlichen *Merkmalen der Weiterbildungsangebote*, die *Art und Intensität der Nutzung der Weiterbildungsangebote*, die *personenbezogenen Voraussetzungen* und *schulkontextbezogenen Einflussfaktoren*. Diese werden in der Arbeit zusammengefasst als *teilnehmendenbedingte*, *trainingsbedingte* und *kontextbezogene Faktoren*. Es ist anzunehmen, dass diese nicht unabhängig voneinander sind. Die Faktoren beziehen sich in der Literatur meist auf Fort- und Weiterbildungen von Lehrkräften. Dieses Verständnis wird in der vorliegenden Arbeit jedoch auch auf die Gruppe der Masterstudierenden Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen als angehenden Lehrkräfte angewendet. Begründet wird dies damit, dass die Merkmale nicht unabhängig der Merkmale aus den Lerntheorien sind, wie an entsprechender Stelle ausgeführt wird. Zudem wird davon ausgegangen, dass insbesondere LiV durch den Lernort Schule die Möglichkeit haben, das Gelernte in die Praxis umzusetzen. Somit herrschen für diese Gruppe ähnliche Bedingungen wie für ausgebildete Lehrkräfte. Die nachfolgend aufgeführten Faktoren erheben aufgrund der Vielfältigkeit dieser keinen Anspruch auf Vollständigkeit.

6.2 Teilnehmendenbedingte Faktoren

Zu den personen- bzw. teilnehmendenbedingten Faktoren zählt die *Motivation* (Lipowsky, 2009, S. 346-349; Richter & Vigerske, 2011, S. 13). Sie ist bereits bei der Entscheidung für oder gegen eine Trainingsteilnahme relevant. Die Motivation hängt dabei vom *erwarteten Nutzen* und der *subjektiven Wahrnehmung der Relevanz* ab (Lipowsky & Rzejak, 2021, S. 60). Dies überschneidet sich mit den von Krainer und Posch (2010) unter dem Aspekt der *individuellen Voraussetzungen* zusammengefassten Punkten „intrinsic Werthaltungen und Motive“, „Bewusstsein, durch Fortbildung einer Rechenschaftspflicht zu entsprechen“ sowie „Interesse,

¹²⁰ Die Autor*innen beschränken sich in dem Bericht und ihrer Definition auf Fortbildungen (Lipowsky & Rzejak, 2021, S. 12), wobei sie den Grund hierfür auch erläutern (vgl. Lipowsky & Rzejak, 2021, S. 16). Dieses Verständnis wird für die vorliegende Arbeit erweitert genutzt und auch auf andere Situationen bzw. Bildungsphasen übertragen.

durch Fortbildung Vorteile zu erhalten oder Nachteilen zu entgehen“ (S. 483). Neben der intrinsischen Motivation, oder auch gleichzeitig, können extrinsische Motivationsfaktoren vorliegen, indem beispielsweise das Training als *Mittel zum Zweck*¹²¹ gesehen wird (vgl. z. B. Brady et al., 2009, S. 445; Harvey, 2005) bzw. *Normen* (z. B. rechtliche Vorschriften) die Teilnahme bedingen (Krainer & Posch, 2010, S. 482). Lipowsky und Rzejak (2021) kommentieren in Bezug auf die Bedingungen in der Bundesrepublik Deutschland, „dass die Teilnahme an Fortbildungen weder mit karrierebezogenen Vorteilen und Anreizen für die Lehrkräfte verbunden ist noch besondere Anerkennung erfährt“ (S. 72). Sie führen auf, dass „In Zeiten des Lehrermangels ... Lehrpersonen für die Teilnahme an längerfristig angelegten Fortbildungen mitunter nicht freigestellt [werden] und ... teilweise nicht mehr in der Unterrichtszeit stattfinden [dürfen]“ (Lipowsky & Rzejak, 2021, S. 72). Auch die Motivation zum Transfer – d.h. Anwendung des Gelernten in die Praxis – gilt als relevanter Faktor (Lipowsky, 2010, S. 63; Lipowsky & Rzejak, 2021, S. 35). Ebenso ist *Freiwilligkeit* ein wichtiger Faktor für die Motivation der Teilnehmenden (Vigerske, 2016, S. 61), denn die Wirksamkeitserwartung an einer Maßnahme ist höher, wenn Personen freiwillig daran teilnehmen (Vigerske, 2016, S. 116, 137, 140-141; Wagnitschek, Katschnig, Auferbauer & Prorok, 2020, S. 748).

Der (persönliche) Trainingserfolg wird auch von *persönlichen Faktoren* beeinflusst, wie etwa die Lehrerfahrung eines Teilnehmenden (Richter, Engelbert, Weirich & Pant, 2013, S. 195-196; vgl. z. B. Richter, Kunter, Klusmann, Lüdtke & Baumert, 2011; für einen Überblick siehe Krille, 2019, S. 66-69; insgesamt zur Weiterbildungsteilnahme nach Alter siehe z. B. Bundesministerium für Bildung und Forschung, 2021, S. 61, 2022; Kuwan, Bilger, Gnahn & Seidel, 2006, S. 91). Hierzu zählt beispielsweise das Alter. Krille (2019, S. 66) hält fest, dass eine inverse U-Form vorliegt, d. h. dass junge, unerfahrene¹²² sowie ältere, erfahrenere Lehrkräfte am wenigsten an (Fort-/Weiter-)Bildungsmaßnahmen¹²³ teilnehmen. Weitere Untersuchungen zeigen, dass jüngere Teilnehmende im Vergleich mit älteren eine höhere Motivation für die Teilnahme an Maßnahmen aufweisen (Brady et al., 2009, S. 441), wobei auch die funktionale Rolle

¹²¹ Hierunter könnte beispielsweise eine notwendige Schulung zur Einführung in eine Schulsoftware fallen.

¹²² Hier wird die Wortwahl der Autorin „inexperienced“ in übersetzter Form verwendet (vgl. Krille, 2019, S. 66). Dabei kann dieser Wortwahl nicht ganz zugestimmt werden, da auch Lehrkräfte, die jung und/oder neu in dem Beruf sind, nicht als *unerfahren* angenommen werden, sondern als *wenig* oder *weniger* erfahren. So berichten Richter, Kunter, Klusmann, Lüdtke und Baumert (2011, S. 119), aus dessen Beitrag auch die Autorin an genannter Stelle zitiert, von einem Mindestwert von einem Jahr Lehrerfahrung.

¹²³ Im Beitrag wird von *Professional Development* gesprochen (Krille, 2019, S. 41). Der Begriff wird von Richter et al. (2011, S. 116), in Anlehnung an Baumert und Kunter (2006) sowie Kunter et al. (2007), definiert als die Teilnahme an formellen und informellen Lernangeboten zur Vertiefung und Erweiterung der beruflichen Kompetenz von Lehrkräften, einschließlich des Wissens, der Überzeugung und Motivation sowie der Fähigkeit zur Selbstregulation (aus dem Englischen übersetzt). Auf eine wörtliche Übersetzung des Begriffs wird hier verzichtet und (Fort- bzw. Weiter-)Bildungsmaßnahmen verwendet.

der Maßnahme die Motivation extrinsisch beeinflusst (vgl. Brady et al., 2009, S. 445; siehe auch im vorherigen Absatz). Seifried und Wuttke (2017, S. 120) zeigen, dass sowohl die Trainingsmotivation als auch der Trainingserfolg u. a. von den Trainingserwartungen und deren Erfüllung beeinflusst werden. Zudem gilt das Anknüpfen an Voraussetzungen als ein wichtiger Faktor (Lipowsky, 2004, S. 474; Lipowsky & Rzejak, 2021, S. 56).

Der Erfolg einer Fort- und Weiterbildung hängt auch von den *kognitiven Voraussetzungen*, z. B. dem Vorwissen der Teilnehmenden, ab (Lipowsky, 2004, S. 474, 2009, S. 347). Unter *Kohärenz* führt die Autorgruppe Garet, Porter, Desimone, Birman und Yoon (2001, S. 927) u. a.¹²⁴ das Vorwissen der an den Fortbildung teilnehmenden Lehrkräften auf und stellt in einer Studie fest, dass dieses den Lern- und Kompetenzzuwachs – und weiter noch die Veränderung der Lehrpraxis – positiv beeinflusst (Garet et al., 2001, S. 931-934). Zudem können *volitionale Faktoren* (z. B. Ausdauer) auf den Weiter-/Fortbildungserfolg einwirken (Lipowsky, 2009, S. 347), so wie auch die *Selbstwirksamkeitserwartung*, „also ihre Überzeugung, die Entwicklung ihrer Schülerinnen und Schüler auch dann positiv beeinflussen zu können, wenn die Bedingungen ungünstig sind – und in Verbindung damit ein Mindestmass an Verantwortlichkeit für das Lernen der Schülerinnen und Schüler betrachtet werden“ (Lipowsky, 2009, S. 347-348). Die (bisherige) Erfahrung einer *Person wirksam zu sein*, begünstigt nach Krainer und Posch (2010, S. 483) den Zugang zu Fortbildungsangeboten. Weiterhin werden *Überzeugungen* der Teilnehmenden aufgeführt (Lipowsky, 2004, S. 474, 2009, S. 347). Um nach einer Trainingsmaßnahme Änderungen im Verhalten (Behavior)¹²⁵ zu zeigen, ist nach Kirkpatrick und Kirkpatrick (2006, S. 23) ein Änderungswunsch erforderlich sowie die Überzeugung, zu wissen, was wie hierfür getan werden muss.

6.3 Trainingsbedingte Faktoren

Für die Wirksamkeit von Weiterbildungsmaßnahmen sollte die *Strukturierung und Planung* eines Trainings berücksichtigt werden (Wanitschek et al., 2020, S. 755). Hierzu gehören verschiedene Faktoren, die im Folgenden vorgestellt werden.

Für eine wirksame Weiter- bzw. Fortbildung von Lehrkräften sollten die *Bedarfe, Ziele und Erwartungen* aufeinander abgestimmt sein (Göb, 2017, S. 20; Wanitschek et al., 2020). Lipowsky (2004, S. 462) berichtet, dass Bundesländer bzw. Schulen immer mehr *dezentralisierte*,

¹²⁴ Daneben führen die Autor*innen zur Kohärenz den Einklang der Inhalte und der Pädagogik mit den nationalen, staatlichen und lokalen Standards, Rahmenwerken und Assessments sowie die Unterstützung der Lehrkräfte zur nachhaltigen und fortlaufenden Kommunikation mit anderen Lehrkräften, die ihre unterrichtliche Praxis auf ähnlicher Weise verändern möchten, auf (Garet, Porter, Desimone, Birman & Yoon, 2001, S. 927).

¹²⁵ Das Verhalten bzw. Behavior ist die dritte Ebene der Evaluation nach Kirkpatrick und Kirkpatrick (2006) und wird im Kapitel 6.5 erläutert.

interne Schulungen anbieten und dies mit der Erwartung geschieht, dass ein Vorort-Training erfolgreicher ist. Auf die Bedürfnisse der Lehrkräfte sowie der Schule kann besser eingegangen und das Training ist speziell auf die Zielgruppe zugeschnitten werden. Die *Orientierung an den Bedarfen der Teilnehmenden* stellt damit einen weiteren Faktor für die Wirksamkeit von Fortbildungen dar (Kirkpatrick & Kirkpatrick, 2005, S. 58; Krille, 2019, S. 26; Vigerske, 2016, S. 142; Wanitschek et al., 2020, S. 744).

Auf inhaltlicher Ebene ist „die *Fokussierung auf curriculare und fachbezogene Inhalte* [Herv. d. Verf.], die die Lehrpersonen auch tatsächlich unterrichten“ (Lipowsky, 2009, S. 352) ein wichtiges Element (vgl. hierzu auch Lipowsky, 2010, S. 63; Lipowsky & Rzejak, 2012, S. 240). Zudem zielen solche Maßnahmen auf „die Erweiterung fachdidaktischen und diagnostischen Lehrerwissens“ (Lipowsky & Rzejak, 2012, S. 240) ab. Auch nach Garet et al. (2001, S. 930-933) beeinflusst der Fachbezug einer Maßnahme den Lern- und Kompetenzzuwachs sowie die Veränderung der Lehrpraxis positiv. Die Ergebnisse der Studie von Johannmeyer und Cramer (2021, S. 1197) zeigen, dass Fortbildungen zu den Themen Abschlussprüfungen und Beurteilung von Leistungen eine stärkere Nachfrage erfahren – dies ist für die vorliegende Arbeit von großer Relevanz.

Auf Basis von bisherigen Studien wird festgehalten, dass Teilnehmende den Nutzen einer Fortbildung daran bemessen, „inwiefern sie neue Impulse und Anregungen für ihren alltäglichen Unterricht erhalten“ (Lipowsky & Rzejak, 2012, S. 237). Krille (2019) bestätigt dies und führt „Neue Ideen und Inspirationen für Unterricht“ (S. 18) als einen Grund für die Fortbildungsteilnahme auf. Sowohl die Transferierbarkeit des Gelernten (vgl. Lipowsky & Rzejak, 2021, S. 61 sowie Kapitel 6.5) als auch die wahrgenommene Relevanz der Thematik eines Trainings sind entscheidend für eine Fortbildung (Krille, 2019, S. 18; Lipowsky & Rzejak, 2021, S. 59-61).

Auch die *Dauer bzw. Zeit* des Fort- und Weiterbildungsangebots beeinflussen die Effektivität dessen (Gräsel et al., 2004, S. 133; Lipowsky, 2004, S. 473, 2009, S. 350-351; Rzejak, Lipowsky & Bleck, 2020, S. 19; Seifried, Wuttke, Türling, Krille & Paul, 2015, S. 181). Es wird davon ausgegangen, dass kurze Maßnahmen oder auch „One-Shot-Fortbildungen“ (Lipowsky & Rzejak, 2012, S. 239) in der Regel wenig erfolgreich sind. Studien zeigen, dass kurze Maßnahmen keinen signifikanten Effekt haben bzw. ungünstig für die Wirksamkeit von Trainings sind (vgl. z. B. Göb, 2017, S. 19-20; Lipowsky & Rzejak, 2021, S. 56; Morningstar & Benitez, 2013; Müller, Kemethofer, Andreitz, Nachbaur & Soukup-Altrichter, 2019, S. 113, 129; Yoon, Duncan, Wen-Yu Lee, Scarloss & Shapley, 2007, S. 3). Dem Ergebnis der Studie von Johannmeyer und Cramer (2021, S. 1196) zufolge werden Fortbildungen mit bis zu einer

vier-Stunden-Dauer weniger nachgefragt und besucht als längere. Es ist jedoch nicht generell davon auszugehen, dass längere Maßnahmen im Vergleich zu kurzen¹²⁶ Maßnahmen erfolgreicher sind (Lipowsky & Rzejak, 2012, S. 239, 2021, S. 56). Die Wahl einer Maßnahmendauer ist situationsbedingt und hängt etwa von „den Voraussetzungen der teilnehmenden Lehrpersonen ..., vom Fortbildungsthema, den Zielen der Fortbildung, der Qualität der Fortbildungsangebote sowie von deren effektiver Nutzung durch die Teilnehmerinnen und Teilnehmer“ (Lipowsky & Rzejak, 2021, S. 56) oder aber auch der Komplexität der Umsetzung ab (Lipowsky & Rzejak, 2021, S. 56).

Doch nicht der Zeitfaktor allein ist für die Wirksamkeit von Maßnahmen entscheidend, sondern auch die entstandenen *Lerngelegenheiten* – also die Möglichkeiten für Teilnehmende, sich mit den Inhalten zu beschäftigen und diese zu erproben bzw. anzuwenden (Göb, 2017, S. 20; Ingvarson, Meiers, Beavis & Adrian, 2005, S. 15; Lipowsky, 2004, S. 473, 2010, S. 64; Schröder, Thienen, Müller & Homann-Vorderbrück, 2020, S. 246¹²⁷). Dieser Aspekt ist allerdings nicht unabhängig der Dauer einer Maßnahme. Garet et al. (2001, S. 930-933) zeigen, dass die Lernaktivitäten bzw. Lerngelegenheiten die Dauer einer Veranstaltung beeinflussen. Lerngelegenheiten, insbesondere aktive, tragen zur Erweiterung des Wissens und der Kompetenzen der Lehrkräfte bei. Diese wiederum beeinflussen wesentlich die Umsetzung des Gelernten und damit die Lehrpraxis (Garet et al., 2001, S. 931-934). Diese Verbindung wird auch in anderen Bildungsphasen thematisiert. Beispielsweise beeinflussen in Bezug auf die Unterrichtsgestaltung die Dauer und die Möglichkeiten der Lernenden zum (aktiven) Lernen die Ergebnisse dieser (z. B. Ditton, 2000, S. 79-83; Helmke, 2002, S. 261-262; Helmke & Lenske, 2013, S. 216-217). Übertragen auf die Lehrkräftebildung, kann daher angenommen werden, dass diese Faktoren nicht nur für dritte Phase der Lehrer*innenbildung bedeutend sind, sondern ebenso bereits in den vorherigen Phasen.

Zudem ist aus der Forschung bekannt, dass nicht nur Anwendungsphasen erstrebt werden sollten, sondern vielmehr ein *Wechsel zwischen Input- und Anwendungsphasen* (Cademartori, Seifried, Wuttke, Krille & Salge, 2017, S. 32, 43, angelehnt an die Studie von Piwowar, Thiel & Ophardt, 2013; Krille, 2019, S. 21; Lipowsky & Rzejak, 2021, S. 50-52). Ein weiteres Merkmal stellt das in Fortbildungsmaßnahmen gegebene *Feedback* dar (Cademartori et al., 2017, S. 42; Lipowsky, 2009, S. 354; Lipowsky & Rzejak, 2021, S. 53-55). Dies scheint gerade in Bezug

¹²⁶ Lipowsky und Rzejak (2021) schreiben hier von „wenige[n] Stunden“ (S. 56).

¹²⁷ Ingvarson, Meiers, Beavis und Adrian (2005) berichten aus einer Studie in Australien. Da die Gegebenheiten in Australien nicht gleich dem in Deutschland sind, sind die Ergebnisse aus der Studie mit Vorsicht zu übertragen.

auf die aktiven Lerngelegenheiten nachvollziehbar, in denen die Teilnehmenden das Gelernte anwenden, ausprobieren und ein förderliches Feedback dazu erhalten. Auch die Möglichkeit zur Reflexion der Teilnehmenden ist für Weiterbildungsmaßnahmen relevant (Lipowsky, 2009, S. 355; Lipowsky & Rzejak, 2021, S. 50-52). Nicht zuletzt ist die *Lernumgebung* ein wichtiger Aspekt für den Maßnahmenerfolg. Die Lernumgebungsqualität beschreibt Schlicht (2012, S. 166) in Anlehnung an Ditton (2009, S. 615) als „die Eignung der Lernumgebung zur Förderung von Lernerfolg und Akzeptanz“ (Schlicht, 2012, S. 166). Auf die Wichtigkeit dieser Faktoren für die Trainingsevaluation wird in Kapitel 6.5 näher eingegangen.

6.4 Kontextbezogene Faktoren

In der vorliegenden Arbeit wird die Zielgruppe der angehenden Lehrkräfte betrachtet, sodass der Fokus der Wirksamkeit eines Trainings auf den Schulkontext gerichtet wird. Kirkpatrick und Kirkpatrick (2006, S. 23-24) führen hierzu neben den persönlichen Bedingungen der Teilnehmenden weitere Kontexte auf, die eine Verhaltensänderung¹²⁸ begünstigen bzw. hemmen können (u. a. Arbeitsklima, Belohnung bzw. Förderung zur Veränderung).

Kirkpatrick und Kirkpatrick (2006, S. 23-24) zufolge kann das *Arbeitsklima* verhindernd, entmutigend, neutral, ermutigend oder fordernd sein. Als verhindernd gilt ein Klima, wenn das Gelernte in der Praxis nicht angewendet werden darf. In einem entmutigenden Klima wird nicht offen ausgesprochen, dass das Gelernte nicht angewendet werden soll, sondern den Teilnehmenden implizit vermittelt – etwa dadurch, dass der*die Vorgesetzte*r selbst auf die Änderung des Verhaltens verzichtet. Den Autoren zufolge ist die Möglichkeit, ein Training (besser: die Inhalte des Trainings) auf der Arbeit anzuwenden, wenn ein verhinderndes oder entmutigendes Arbeitsklima herrscht, gering oder nicht gegeben. Werden die Inhalte der Maßnahmen von Vorgesetzten ignoriert, entsteht ein neutrales Klima. Dieses kann zu einem verhindernden oder entmutigenden Arbeitsklima führen, wenn durch die Veränderung des Verhaltens der Teilnehmenden etwas Negatives aus Sicht des*r Vorgesetzten geschieht. Unter einem neutralen Arbeitsklima hängt die Veränderung des Verhaltens laut den Autoren davon ab, ob die teilnehmende Person eine Veränderung möchte, weiß was und wie etwas zu tun ist und für die Veränderung (intrinsisch oder extrinsisch oder sowohl als auch) belohnt/gefördert wird. In einem ermutigenden Arbeitsklima hingegen wird das Lernen und die Anwendung des Gelernten auf der Arbeit gefördert. Dies geschieht beispielsweise, indem der*die Vorgesetzte Interesse am Gelernten und die Bereitschaft zeigt, behilflich bei der Umsetzung zu sein. In einem fordernden Arbeitsklima

¹²⁸ Das Verhalten bzw. „Behavior“ ist die dritte Ebene der Evaluation nach Kirkpatrick und Kirkpatrick (2006). Auf diesen wird in Kapitel 6.5 eingegangen.

weiß der*die Vorgesetzte*r also, was die Mitarbeitenden im Training gelernt haben und versucht, den Transfer des Gelernten sicherzustellen. In einem ermutigenden oder fördernden Arbeitsklima hängt die Veränderung des Verhaltens vom Wunsch der teilnehmenden Person zur Veränderung und dem Wissen darüber, was (und wie) zu tun ist, ab. Um eine Veränderung des Verhaltens erreichen zu können, empfehlen Kirkpatrick und Kirkpatrick (2006, S. 24-25), dass Trainer*innen zu Beginn Informationen über das Arbeitsklima der Teilnehmer*innen einholen und entsprechend handeln. Als ein Vorschlag zur Förderung eines positiven Arbeitsklimas wird empfohlen, auch Vorgesetzte zu den Trainingsmaßnahmen einzuladen und/oder diese nach den Bedürfnissen der Mitarbeiter*innen zu befragen. Eine weitere Möglichkeit bietet die explizite Information der Vorgesetzten über den Inhalt der Trainings.

Die Nachfrage nach Weiterbildungsmaßnahmen wird anlehnend an Kirkpatrick und Kirkpatrick (2006, S. 24) durch *Belohnungen* intrinsisch und/oder extrinsisch beeinflusst – intrinsisch durch z. B. die eigene Zufriedenheit, extrinsisch z. B. durch Lob von Vorgesetzten.¹²⁹ Doch nicht nur bei der Umsetzung des im Training Gelernten können diese Faktoren einen Einfluss haben. Ebenso beeinflussen nach Krainer und Posch (2010, S. 482-483) *strukturelle Voraussetzungen* in Form von Normen (z. B. rechtliche Vorgaben, vgl. Kapitel 6.2) und Ressourcen die Nachfrage der Lehrkräfte nach einer Fortbildung. Die Ressourcen unterteilen Krainer und Posch (2010, S. 483) in materielle (z. B. finanziell für die Teilnahme an Fortbildungen) und immaterielle Ressourcen (z. B. Regelungen und organisatorische Maßnahmen, bei denen Zeitressourcen freigeschaltet werden). Gerade in Bezug auf die zeitlichen Ressourcen gibt es, wie weiter oben aufgeführt, in der Bundesrepublik Deutschland Knappheit (vgl. Krille, 2019, S. 18; Lipowsky & Rzejak, 2021, S. 72; vgl. Kapitel 6.2).

Auch das *Schulumfeld* beeinflusst den Erfolg einer Maßnahme. Hierzu zählt neben dem Arbeitsklima „die Innovationskraft und der Reformeifer im Kollegium, das Interesse der Schulleitung an einer unterrichtsnahen Weiterbildung der Kolleginnen und Kollegen, die Schaffung von Gelegenheiten zur Anwendung des Gelernten im Schulalltag, die Begleitung und Unterstützung der Lehrpersonen in der Phase nach Abschluss der Fortbildung sowie das Feedback durch Schulleitung und Kollegium“¹³⁰ (Lipowsky, 2009, S. 349). Die Ergebnisse der Studie von Garet et al. (2001, S. 931-934) zeigen, dass die Kohärenz der Maßnahme den Lern- und Kompetenzzuwachs sowie die Veränderung der Lehrpraxis positiv beeinflussen. Hierzu gehört

¹²⁹ Da diese intrinsische und extrinsische Belohnung im Zusammenhang mit dem Schulkontext stehen können, werden diese hierunter erläutert.

¹³⁰ Vgl. hierzu auch Lipowsky und Rzejak (2021, S. 19-20) sowie Kirkpatrick und Kirkpatrick (2006, S. 23).

u. a. die Unterstützung der Lehrkräfte zur nachhaltigen und fortlaufenden Kommunikation mit anderen Lehrkräften, die die unterrichtliche Praxis in ähnlicher Weise verändern möchten.

6.5 Trainingsevaluation

Evaluation wird definiert als „die systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung des Konzepts, des Designs, der Umsetzung und des Nutzens sozialer Interventionsprogramme“ (Wottawa, 2006, S. 662). Diese wird genutzt, um eine Beurteilung und Verbesserung von Interventionsmaßnahmen – auch in der Bildung – vorzunehmen (Wottawa, 2006, S. 662). Während beispielsweise der monetäre Nutzen und die Wirtschaftlichkeit einer Maßnahme präzise gemessen werden können, gestaltet sich dies bei der Bestimmung des Nutzens im Sinne einer *pädagogischen Wirksamkeit* komplizierter (Schlicht, 2012, S. 15). Bezugnehmend auf Nistor, Schnurer und Mandl (2005) sowie Reinmann-Rothmeier, Mandl und Prenzel (1997) hält Schlicht (2012, S. 13) folgende Kriterien hierfür fest:

- Lernerfolg: zielbezogener Zuwachs an Wissen-, Können- und Wollenszuwachs
- Akzeptanz des Lernangebots: der von Teilnehmenden subjektiv wahrgenommener Nutzen des Lernangebots
- Lernumgebungsqualität: Qualität des Lernens, Lehrens und die Bedingungen

Typischerweise erfolgt die Evaluation von Trainingsmaßnahmen¹³¹ basierend auf den vier Stufen nach Kirkpatrick und Kirkpatrick (2006), welche sich mit den von Schlicht (2012, S. 13) genannten überschneiden. Diese unterteilen sich in Reaktion (Reaction), Lernen (Learning), Verhalten (Behavior) und Ergebnisse (Results) (Kirkpatrick, 2007; Kirkpatrick & Kirkpatrick, 2006). Von Lipowsky (2010) werden diese auf den Kontext der Fortbildung von Lehrer*innen übertragen und im Folgenden eingehend beschrieben.

In der ersten Ebene (Reaktion) geht es insbesondere um die Zufriedenheit der Teilnehmenden, d. h. sie verfügen über eine positive Reaktion (Kirkpatrick & Kirkpatrick, 2006, S. 21-22). Lipowsky (2010, S. 52) ergänzt neben der Zufriedenheit die Akzeptanz der Fortbildungsmaßnahme von den Teilnehmenden sowie deren eingeschätzte Relevanz bezüglich der Maßnahme. Zwar bedingt eine positive Reaktion nicht zwingend Lernen, so Kirkpatrick und Kirkpatrick (2006, S. 22)¹³², doch wenn Teilnehmende keine positive Reaktion zeigen, sind sie möglicherweise auch nicht motiviert zu lernen. Auch Lipowsky (2010, S. 54) kommt zu dem Schluss,

¹³¹ Kirkpatrick und Kirkpatrick (2006, z. B. S. 19) schreiben selbst von „training programs“.

¹³² Es wird beispielsweise aus der Studie von Goldschmidt und Phelps (2007) berichtet, dass die wahrgenommene Qualität – welches Ebene 1 der Evaluation zuzuordnen wäre – mit dem Wissenszuwachs nicht zusammenhängt.

dass ein Mindestmaß an Akzeptanz erforderlich ist, auch wenn dies das Lernen und Transferieren nicht bedingt. Eine Evaluation auf dieser Ebene kann sowohl (formativ) während des Trainings (vgl. z. B. Wagner, Lüftenegger, Finsterwald, Schober & Spiel, 2012) als auch danach (summativ) durchgeführt werden. Um etwaige Anpassungen an die Bedürfnisse der Teilnehmenden vornehmen zu können, wird eine kontinuierliche Evaluation während des Trainings empfohlen (Kirkpatrick & Kirkpatrick, 2005, S. 59-60).

Kirkpatrick und Kirkpatrick (2006, S. 22) beschreiben die Ebene des Lernens wie folgt: „Learning can be defined as the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending the program“. *Lernen*, das in dieser Ebene angesprochen wird, geht über den reinen Wissenserwerb hinaus. Es schließt die Einstellungen und Fähigkeiten der Teilnehmenden ein. Diese Beschreibung kann auch als Vermeidung von trägem Wissen verstanden werden. Träges Wissen ist dabei als eine Art von Wissen zu verstehen, welches nicht über den Lehr-Lern-Kontext bzw. der darin befindlichen Situation hinausgeht und keine Anwendung findet (Gruber, Mandl & Renkl, 2000, S. 139; Renkl, 1994; Vonken, 2005, S. 38). Aus dieser Erklärung ist zu schließen, dass das Gelernte nicht im (Leben- und Berufs-) Alltag angewendet werden kann und nicht zur Handlungskompetenz beiträgt. Bei der Weiter- und Fortbildung von Lehrkräften wird demnach nicht auf den reinen Wissenszuwachs bzw. auf die kognitive Dimension¹³³ beschränkt, vielmehr werden auch die affektiv-motivationalen Dimensionen berücksichtigt (Lipowsky, 2009, S. 346, 2010, S. 54; Lipowsky & Rzejak, 2012, S. 238; vgl. auch Achtenhagen & Baethge, 2007, S. 58¹³⁴) – wie etwa die Selbstwirksamkeit (Lipowsky & Rzejak, 2012, S. 238). In der Evaluationsforschung von Bildungsmaßnahmen für Lehrkräfte wird demnach die Kognition von Lehrkräften im weiten Sinne verstanden, das „Überzeugungen und subjektive Theorien, ... fachliche, fachdidaktische, pädagogisch-psychologische und diagnostische Wissen von Lehrerinnen und Lehrern“ (Lipowsky, 2010, S. 54) einschließt. Zur Evaluation auf dieser Ebene werden u. a. Tests als geeignet beschrieben (Kirkpatrick & Kirkpatrick, 2005, S. 61).

Die Ebene des *Verhaltens* (Behavior) bezeichnet Lipowsky (2010, S. 57) als *unterrichtspraktisches Handeln*. Um eine Veränderung auf dieser Ebene zu erreichen, muss mindestens eine Änderung der Einstellung, ein Wissenszuwachs oder eine Verbesserung der Fähigkeiten stattfinden (Kirkpatrick & Kirkpatrick, 2006, S. 22; vgl. hierzu auch Kirkpatrick & Kirkpatrick,

¹³³ Zu diesen zählen Lipowsky und Rzejak (2012, S. 238) Wissen und Überzeugungen.

¹³⁴ Zwar beziehen sich Achtenhagen und Baethge (2007, S. 58) mit dieser Aussage auf die berufliche Bildung, dennoch ist davon auszugehen, dass dies nicht nur auf den Bereich der beruflichen Erstausbildung zutreffend ist, sondern auch darüber hinaus gilt.

2005, S. 58 sowie die in Kapitel 6.3 beschriebene Studie von Garet et al., 2001, S. 931-934). Diese Veränderung ist jedoch auch abhängig von den Arbeitsbedingungen bzw. dem Arbeitsklima (Kirkpatrick & Kirkpatrick, 2006, S. 23-24, vgl. Kapitel 6.4). In Bezug auf die Evaluation von Weiter- und Fortbildungsmaßnahmen liegen methodische Herausforderungen vor, weshalb Evaluationen auf dieser Ebene selten durchgeführt werden (Lipowsky & Rzejak, 2012, S. 238).

Die *Ergebnisse* (Results) der Trainingsmaßnahme definieren Kirkpatrick und Kirkpatrick (2006) als „final results that occurred because the participants attended the program“ (S. 25). Im Unternehmenskontext sind dies beispielsweise eine gestiegene Produktion, bessere Qualität und steigende Verkaufszahlen (Kirkpatrick & Kirkpatrick, 2006, S. 25). Die Autoren weisen darauf hin, dass nicht alle Ergebnisse einfach gemessen werden können. Dazu zählen beispielsweise Themen zur Führung, Kommunikation, Motivation und Zeitmanagement (Kirkpatrick & Kirkpatrick, 2006, S. 26). Auf den Kontext der Lehrer*innen(fort-)bildung übertragen, bezeichnet Lipowsky (2010, S. 59) die vierte Ebene als *Effekte auf Schüler*innen*. Hierzu gehört der schulische Erfolg der Schüler*innen in Bezug auf z. B. ihre Leistungen und Motivationen (Lipowsky, 2010, S. 59, 63).

7 Trainingsstudie mit Studierenden der Wirtschaftspädagogik und LiV an kaufmännisch-berufsbildenden Schulen

7.1 Relevanz des Trainingsinhalts und Vorstellung des Trainingskonzepts

Fachdidaktische und diagnostische Kompetenzen von Lehrkräften sind für die Erstellung problemhaltiger Aufgaben zu Zwecken summativen Assessments entscheidend (vgl. Kapitel 0). Die Vorstudien zeigen, dass relevante Themen hierzu insgesamt in der ersten und zweiten Phase der Ausbildung der angehenden Lehrkräfte kaufmännisch-beruflicher Schulen thematisiert werden – auch wenn sich die Umsetzung zwischen den Standorten unterscheidet und sich die Inhalte zumeist auf Lehr-Lern-Prozesse richten (Kapitel 5). Assessmentaufgaben zu summativen Zwecken werden hingegen kaum fokussiert, problemhaltige Assessmentaufgaben zu summativen Zwecken noch weniger bis gar nicht. Angehende Lehrkräfte berichten in der Interviewstudie darüber, sich kaum auf das Erstellen von problemhaltigen Aufgaben zu summativen Zwecken vorbereitet zu fühlen. Ein aktueller Blick in die Forschungsergebnisse zeigt, dass Prüfungsaufgaben als summative Assessments nicht hinreichend Problemlösekompetenz erfassen (Wuttke et al., 2022, vgl. Kapitel 3.2). Die Ergebnisse zeigen Verbesserungspotenzial auf: Zum einen mit Blick auf die Problemhaltigkeit der Aufgaben in Prüfungssituationen und deren Validität (Wuttke et al., 2022, S. 44; in Bezug auf die Validität siehe auch z. B. Deutscher & Winther, 2019; Klotz & Winther, 2012 und Kapitel 3.2) und zum anderen mit Blick auf die Diskrepanz zwischen den an Lehrkräfte gestellten Kompetenzanforderungen (vgl. Kapitel 2) und den aktuellen Gegebenheiten der Ausbildung (vgl. Kapitel 5). Diese Erkenntnisse ermöglichen, das Training bedarfsorientiert zu gestalten bzw. auszuwählen, was wiederum dessen Wirksamkeit zum Positiven beeinflusst (z. B. Krille, 2019, S. 26; vgl. Kapitel 6.3).

Im Projekt TeKoP¹³⁵ wurde ein Training zum technologiebasierten und problemhaltigen Prüfen konzipiert, durchgeführt und evaluiert (Wuttke et al., 2023). Das Training richtete sich an Personen, die an der Erstellung von Prüfungsaufgaben beteiligt sind oder sein können. Das Trainingskonzept wurde für die vorliegende Trainingsstudie zugrunde gelegt. Während sich im TeKoP-Training Teilnehmende in ihrer Profession und weiteren Merkmalen unterschieden (Wuttke et al., 2023), wurde in der Trainingsstudie der vorliegenden Arbeit eine im Vergleich homogenere Gruppe¹³⁶ angesprochen – wirtschaftspädagogische Studierende sowie LiV an

¹³⁵ Im Projekt TeKoP wurde die Bildung/Qualifizierung von Lehrkräften als möglicher Erklärungsansatz zu den vorliegenden Befunden zur Prüfungsaufgabenanalyse gesehen (Wuttke et al., 2022, S. 47; Wuttke et al., 2023). Hierfür wurde allerdings keine systematische Analyse der Ausbildungsinhalte der ersten und zweiten Phase der Lehrer*innenbildung durchgeführt, wie es in der vorliegenden Arbeit der Fall ist. Stattdessen wurden Angebote zur Schulung des Personals durchsucht und herausgefunden, dass ein entsprechendes Angebot nicht vorliegt (Wuttke et al., 2022, S. 46).

¹³⁶ Die Beschreibung der Stichprobe erfolgt in Kapitel 7.3.3.

kaufmännisch-beruflichen Schulen als angehende Lehrkräfte kaufmännisch-beruflicher Schulen. Das Trainingskonzept wurde im Hinblick auf Inhalt, Aufbau und Evaluationsinstrumente nicht verändert, da die ersten Projektergebnisse zum Training¹³⁷ darauf schließen ließen, dass Teilnehmende (N = 74) mit dem Training zufrieden sind und den Trainingsnutzen sowie den Lernerfolg als hoch bis sehr hoch einschätzen¹³⁸. Zudem konnte ein signifikanter Lernzuwachs im Pre-Posttest-Vergleich zum Wissen über Merkmale problemhaltiger Aufgaben nachgewiesen werden ($t(119) = -3.473$; $p < .001$) (Wuttke et al., 2021, Folie 7-8; ähnliche Ergebnisse konnten auch in einer vorherigen Präsentation mit Daten von N = 66 Masterstudierenden der Wirtschaftspädagogik vorgestellt werden, vgl. Geiser, Turhan, Wuttke, Seeber & Meiners, 2021, Folie 11-17). Das Ergebnis mit insgesamt 152 Teilnehmenden bestätigte dies zum späteren Zeitpunkt. So zeigten sich die Teilnehmenden der Trainings im Rahmen des TeKoP-Projekts weiterhin mit dem Training (M = 3.00; SD = 0.65) und der Instruktion (M = 3.40; SD = 0.51) zufrieden bis sehr zufrieden¹³⁹. Sie schätzten den Nutzen des Trainings als hoch bis sehr hoch ein (M = 3.00; SD = 0.77) sowie auch ihren Lernzuwachs (M = 3.20; SD = 0.606). Letzteres bestätigte auch der Pre-Posttest-Vergleich des Wissenstest zu den Merkmalen problemhaltiger Aufgaben, der zu einem signifikanten Ergebnis führte ($t(119) = -3.473$; $p < .05$). Zum Wissen über diagnostische Gütekriterien konnte kein signifikanter Lernzuwachs verzeichnet werden ($t(142) = -0.290$; $p > .05$) (Goethe-Universität Frankfurt & Georg-August-Universität Göttingen, 2023, S. 15-16)¹⁴⁰. Die Überlegungen des multiprofessionellen Teams¹⁴¹ zur inhaltlichen und methodisch-didaktischen Gestaltung des Konzeptes, unter Berücksichtigung der Merkmale effektiver Trainings (vgl. Wuttke et al., 2023)¹⁴², werden auch für angehende Lehrkräfte als zielführend eingestuft. Auf Letzteres wird im Folgenden näher eingegangen sowie trainingsbedingte Aspekte (vgl. Kapitel 6.3) berücksichtigt.

¹³⁷ Zu der Stichprobe zählten auch Masterstudierende der Wirtschaftspädagogik, was für das Anwenden des Konzepts sprach.

¹³⁸ Für die genannten Skalen wurde eine Likert-Skala von 1 = „trifft gar nicht zu“ bis 4 = „trifft zu“ eingesetzt.

¹³⁹ Für die Skalen zur Trainings- und Instruktionzufriedenheit sowie zum eingeschätzten Trainingsnutzen und Lernzuwachs wurde eine Likert-Skala von 1 = „trifft gar nicht zu“ bis 4 = „trifft zu“ eingesetzt.

¹⁴⁰ Im Rahmen des TeKoP-Projekts wurden auch andere Aspekte (z. B. Erwartungen der Teilnehmenden) evaluiert (Goethe-Universität Frankfurt & Georg-August-Universität Göttingen, 2023). Diese Ergebnisse werden in der vorliegenden Arbeit nicht aufgeführt, sondern der Fokus auf die Forschungsfragen der Trainingsstudie und der darin formulierten Schwerpunkte gerichtet.

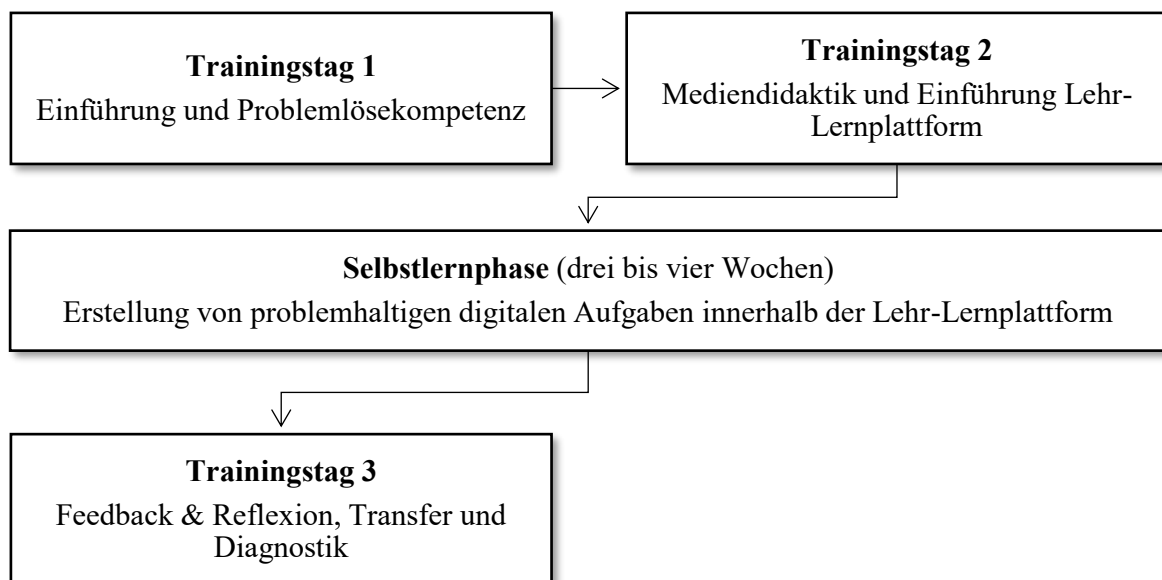
¹⁴¹ Das Team bestand aus Wirtschaftspädagoginnen, Wirtschaftsinformatikern und einem Experten für Mediendidaktik und pädagogische Psychologie.

¹⁴² Da das Trainingskonzept übernommen wird, ist es an dieser Stelle nicht vermeidbar, dass Aspekte zur inhaltlichen und didaktisch-methodischen Gestaltung aus den genannten Quellen wiederholt werden. Die Inhalte des Kapitels gehen allerdings über eine reine Wiederholung hinaus, in dem diese mit Fokus auf die genannte Zielgruppe beleuchtet werden, wodurch auch eine Legitimation der Anwendung des Trainingskonzepts erfolgen soll.

Das Training besteht aus mehreren Bausteinen: *Erstellen problemhaltiger Aufgaben, Mediendidaktik, Einführung in die Lehr-Lernplattform, Feedback und Reflexion, Transfer und Diagnostik*. Das Trainingskonzept wird in Abbildung 14 grafisch dargestellt. Bei der Beschreibung wird auf die Bausteine zur *Erstellung problemhaltiger Aufgaben* (Trainingstag 1) sowie *Feedback und Reflexion, Transfer* und insbesondere *Diagnostik* (Trainingstag 3) fokussiert, da sich der Blick der Evaluation hierauf richtet (vgl. Kapitel 8). Ebenso wird auf die dazwischen stattfindende Selbstlernphase eingegangen. Die Bausteine zu *Mediendidaktik* und *Einführung in die Lehr-Lernplattform* (Trainingstag 2) werden aus vorher genanntem Grund in den Hintergrund gerückt.

Abbildung 14:

Trainingsstudie des Forschungsprojektes „TeKoP“



(Übernommen und angepasste Darstellung aus Wuttke et al., 2021, Folie 4¹⁴³)

Das Training startet mit dem Aufzeigen der Relevanz von Problemlösekompetenz. Dies soll die Teilnehmenden für die Thematik sensibilisieren. Hierfür wird insbesondere auf Anforderungen einer dynamischen Arbeitswelt eingegangen (z. B. Funke et al., 2018, S. 41; Vuori et al., 2019, S. 238; § 1 BBiG; vgl. Kapitel 1). Das Aufzeigen der Relevanz soll die Wirksamkeit des Trainings positiv beeinflussen (z. B. Krille, 2019, S. 18; vgl. Kapitel 6.3). Die Problemlösekompetenz wird als curriculare Anforderung am Beispiel des Ausbildungsberufs Industriekauf-

¹⁴³ Neben dem, dass die Formatierung leicht geändert wurde, wurde das Element der Grafik zum „Feedback- und Follow-Up-Veranstaltung“ aus Wuttke et al. (2021, Folie 4) gelöscht, da dies in der vorzustellenden Trainingsevaluation keine Beachtung fand. Dadurch sollen Irritationen vermieden werden.

mann*frau erläutert (Bundesinstitut für Berufsbildung, 2002; KMK, 2002). Um eine theoretische Grundlage für alle Teilnehmenden zu schaffen, wird das Konstrukt der Problemlösekompetenz vorgestellt (vgl. Kapitel 2.5.2) sowie Handlungs- und Problemlösekompetenz gegenübergestellt (vgl. Kapitel 2.5.5). Anschließend wird auf Merkmale problemhaltiger Aufgaben eingegangen (vgl. Kapitel 2.5.3). Als Werkzeug (Tool) wird den Teilnehmenden ein Raster zu Merkmalen problemhaltiger Aufgaben (Anhang E) bereitgestellt. Sie machen sich während des Trainings damit vertraut, können aber auch darüber hinaus damit arbeiten. Dieser Trainingsbaustein schließt mit einer kritischen Reflexion bestehender Prüfungsaufgaben für die Ausbildungsberufe Industriekaufmann*frau und Kaufmann*frau für Büromanagement ab. Hierfür werden die Ergebnisse der Prüfungsaufgabenanalyse von Wuttke et al. (2022) vorgestellt. Dies dient dem Aufzeigen der Notwendigkeit eines solchen Trainings für die Praxis und soll einen positiven Effekt auf die Trainingswirksamkeit haben (vgl. Kapitel 6.3). Außerdem werden beispielhaft Einflussfaktoren auf das Problemlösen vorgestellt (vgl. Kapitel 2.5.4).

Der zweite Trainingstag beginnt mit der Vorstellung relevanter Aspekte der Mediendidaktik, die für die technologiebasierte Aufgabenerstellung in Lernmanagementsystemen zu berücksichtigen sind. Anschließend findet eine Einführung in das Lernmanagementsystem ILIAS statt. Das technologiebasierte Arbeiten ermöglicht es, der Kritik der fehlenden Authentizität bestehender Prüfungsaufgaben (z. B. Ressel & Kiesecker, 2019, S. 22-23; Winther et al., 2015, S. 6; vgl. Kapitel 3.2) zu begegnen, indem beispielsweise mit Videos und Darstellungen von E-Mail-Postfächern realitätsnahe Szenarien dargestellt werden (Wuttke et al., 2023).

Während des gesamten Trainings findet ein Wechsel zwischen Input- und Anwendungsphasen statt. Die Teilnehmende können das theoretisch Gelernte anwenden, ausprobieren und reflektieren. Auch dies soll sich positiv auf die Wirksamkeit des Trainings auswirken (z. B. Lipowsky & Rzejak, 2021, S. 50-52; vgl. Kapitel 6.3). Eine positive Auswirkung wird zudem durch die Arbeit mit realen Unterlagen (Curricula der Ausbildungsberufe, einsatzfähiges Lernmanagementsystem) angestrebt, denn der Erfolg und die Zufriedenheit der Studierenden hängt u. a. von der Anwendbarkeit und Erkennbarkeit der Relevanz für die Schulpraxis ab (Pasternack, Baumgarth, Burkhardt, Paschke & Thielemann, 2017, S. 90).

Ein Vorteil des Konzepts ist die eingebettete Selbstlernphase. Diese wird in vielerlei Hinsicht den Anforderungen von Pasternack et al. (2017, S. 158-159)¹⁴⁴, gerecht. Die Selbstlernphase bietet den Teilnehmenden einen individuellen Gestaltungsspielraum. Dadurch können sie selbst

¹⁴⁴ In Anlehnung an Berendt, Voss und Wildt (2002), Landfried und Senger (2004) und Markowitsch, Messerer und Prokopp (2004).

entscheiden, wann sie wie viel arbeiten. Zudem bietet diese die Möglichkeit, das Gelernte aktiv anzuwenden und so trägem Wissen (Renkl, 1994; Whitehead, 1929) vorzubeugen. Durch potentiell¹⁴⁵ authentische Probleme aus der Schul- bzw. Prüfungspraxis, nämlich dem Erstellen einer problemhaltigen (und technologiebasierten) Prüfungsaufgabe, wird eine authentische Problemsituation für Studierende und LiV kreiert. Dozent*innen begleiten die Teilnehmenden tutoriell, sodass angehende Lehrkräfte die Möglichkeit haben, fall- und erfahrungsbasiert zu lernen (vgl. Pasternack et al., 2017, S. 158-159). Die Teilnehmenden können jederzeit Fragen stellen und bei Bedarf Unterstützung einholen. Zudem erhalten Sie unterstützende Materialien: die Foliensätze, ein Handbuch zur Erstellung problemhaltiger Aufgaben (2022) und ein Handbuch zum Umgang mit ILIAS (o. A.)¹⁴⁶. Über ILIAS haben die Studierenden und LiV Zugang zu *Learning Nuggets*, dabei handelt es sich um kurze Erklärvideos zum Umgang mit ILIAS und darin speziell zu Inhalten für das Erstellen von Testfragen. Auch werden ihnen die Ausbildungsrahmenlehrpläne, Ausbildungsordnungen und Prüfungskataloge der Ausbildungsberufe zur Verfügung gestellt. Dadurch wird an die berufliche Realität angeknüpft und die Authentizität und Relevanz der Inhalte des Trainings aufgezeigt.

Am dritten Trainingstag werden zunächst die Produkte aus der Selbstlernphase, in Form von problemhaltigen technologiebasierten Prüfungsaufgaben, reflektiert und dazu Feedback gegeben. Auch dies soll die Wirksamkeit des Trainings positiv beeinflussen (z. B. Cademartori et al., 2017, S. 42; vgl. Kapitel 6.3). Das Feedback (seitens der Dozierenden sowie der Teilnehmenden) ist individualisiert. Es soll gleichzeitig zur Reflexion anregen und die Wirksamkeit positiv beeinflussen (vgl. Lipowsky, 2009, S. 355 sowie Kapitel 6.3). Im Anschluss daran findet der Transfer statt, indem die Teilnehmenden die Aufgaben, die sie zunächst für die Zielgruppe der Industriekaufleute erstellen, für die Zielgruppe der Kaufleute für Büromanagement adaptieren. Die Themen der Aufgaben sind daher so ausgewählt, dass zu beiden Ausbildungsberufen curricular formulierte Anforderungen bestehen. Dies soll ebenfalls trägem Wissen entgegenwirken (Renkl, 1994; Whitehead, 1929). Diese Situation wird als authentisch eingeschätzt und ist im Sinne der Wirksamkeit des Trainings (vgl. Kapitel 6.3), da Lehrkräfte an beruflichen Schulen potentiell ebenfalls mehrere Berufsgruppen unterrichten und prüfen.

Der dritte Trainingstag endet mit dem Thema der diagnostischen Gütekriterien. Hierfür werden diese zunächst vorgestellt und ihre Relevanz erläutert (vgl. Kapitel 2.4.5). Dabei wird der Fokus

¹⁴⁵ Zwar hat die aktuelle Prüfungspraxis in dieser Hinsicht Optimierungspotenzial (vgl. Kapitel 3.2), dennoch wird die Aufgabenstellung als *potentiell authentisch* bezeichnet, da gerade mit Blick auf die Curriculum-Instruktion-Assessment Triade von Pellegrino (2010) Aufgaben der Anforderung einer problemhaltigen (und technologiebasierten) Gestaltung gerecht werden sollten (Wuttke et al., 2022).

¹⁴⁶ Zu den Handbüchern sind die jeweils finalen Versionen als Quellen angegeben.

auf den Hauptgütekriterien (Objektivität, Reliabilität, Validität) gerichtet. Teilnehmende erhalten den Auftrag, die erstellte Aufgabe auf ihre Auswertungsobjektivität sowie inhaltliche Validität kritisch zu überprüfen. Während für die Auswertungsobjektivität ein Scoringverfahren angewendet wird (Klotz, 2015, S. 72, 86; vgl. Kapitel 2.6), werden für die Inhaltsvalidität curriculare Anforderungen berücksichtigt und die Beachtung dieser in der Aufgabe kritisch geprüft (vgl. Klotz, 2015, S. 85-86; Maier, 2015, S. 19 sowie Kapitel 2.6). Auch hier findet ein Wechsel zwischen Input- und Anwendungsphase statt und Teilnehmende werden zur Reflexion angehalten (vgl. Kapitel 6.3).

Zwar ist in der Trainingsforschung nicht eindeutig geklärt, welche Dauer ein Training haben sollte (Cademartori et al., 2017, S. 31-32; Lipowsky, S. 350-351), dennoch wird davon ausgegangen, dass sehr kurze Maßnahmen keine langfristige Wirkung haben (Lipowsky, 2009, S. 350-351; Lipowsky & Rzejak, 2021, S. 56-57; Seifried & Wuttke, 2017, S. 117-118; vgl. Kapitel 6.3). Das vorgestellte Trainingskonzept ist über drei Trainingstage angelegt, wobei zwischen Trainingstag 2 und 3 eine ca. vierwöchige Selbstlernphase eingebettet ist. Die Dauer des Trainings wird basierend auf der Empfehlung von Lipowsky und Rzejak (2021, S. 58), längere Fortbildungen bei Teilnehmenden mit wenig Vorwissen und Erfahrung einzusetzen, als angemessen eingeschätzt. Ausgehend von den Vorstudien zu den Ausbildungsinhalten aus der Angebots- und Nachfrageperspektive (vgl. Kapitel 5) und den Ergebnissen der Prüfungsaufgabenanalyse (Wuttke et al., 2022; vgl. Kapitel 3.2) wird davon ausgegangen, dass die Teilnehmenden diese Kriterien erfüllen, also über wenig Vorwissen sowie Erfahrungen verfügen.

7.2 Ziel und Beschreibung des methodischen Vorgehens der Trainingsevaluation

Ziel der Trainingsstudie ist es herauszufinden, ob und wie das im Projekt TeKoP entwickelte Training (vgl. Kapitel 7.1) auch für angehende kaufmännisch-berufliche Lehrkräfte in der ersten und zweiten Phase der Lehrer*innenbildung geeignet ist. Dabei liegt der Fokus auf der Zufriedenheit und dem Lern-/Wissenszuwachs¹⁴⁷. Dies ist den ersten beiden Evaluationsebenen (Reaktion und Lernen) nach Kirkpatrick und Kirkpatrick (2006) zuzuordnen. Die Forschungsfragen hierzu lauten:

*F₂: Wie erfolgreich ist das Training für angehende Lehrkräfte an kaufmännisch-beruflichen Schulen/Wirtschaftspädagog*innen?*

*F₂₋₁: Wie zufrieden sind angehende Lehrkräfte an kaufmännischen-beruflichen Schulen/Wirtschaftspädagog*innen mit dem Training?*

¹⁴⁷ Zur besseren Lesbarkeit wird in Folgenden nur die Bezeichnung *Lernzuwachs* genutzt.

*F₂₋₂: Wie schätzen angehende Lehrkräfte an kaufmännischen-beruflichen Schulen/Wirtschaftspädagog*innen ihren Lernzuwachs durch das Training ein?*

F₂₋₃: Welchen Lernzuwachs haben angehende Lehrkräfte durch das Training?

F₃: Wie kann die Zufriedenheit mit dem Training erklärt (verbessert) werden?

F₄: Wie kann der (wahrgenommene) Lernzuwachs durch Training erklärt (verbessert) werden?

Für die Studie wurde ein quasi-experimentelles Design gewählt, indem neben einer Treatment-/Interventionsgruppe eine Kontrollgruppe untersucht wurde (vgl. Döring & Bortz, 2016, S. 199; McMillan & Schumacher, 2010, S. 22). Es wurden Pre- und Posttests eingesetzt. Während im Pretest persönliche Einflussfaktoren erhoben wurden (vgl. hierzu Kapitel 6.2)¹⁴⁸, wurde nach Ende des Trainings ein Posttest durchgeführt. Hierbei wurden die Instrumente aus dem Pretest (z. T. adaptiert) eingesetzt und um Fragebogenteile ergänzt, die der Evaluation auf der ersten Ebene nach Kirkpatrick und Kirkpatrick (2006) dienen. Auf die einzelnen Instrumente wird in Kapitel 7.3.1 näher eingegangen. Dadurch wurde das Trainings quantitativ evaluiert.

Für eine qualitative Evaluation des Trainings wurden Trainingsteilnehmende interviewt. Dies ermöglichte es, tiefergehende Einblicke in Bezug auf die Forschungsfragen (*F₂₋₁*, *F₂₋₂*, *F₃* und *F₄*) zu erhalten. In Bezug auf die Wahrnehmungen der Teilnehmenden und somit tiefergehende Einblicke sind qualitative Daten aussagekräftig - insbesondere mit Blick auf die Validität (vgl. Mayring, 2012, S. 297). Mit der Begründung, dass „Evaluation ... vor allem dann aufschlussreich sein [wird], wenn es ihr gelingt, die unterschiedlichen – subjektiven – Bewertungen verschiedener Beteiligter zu erfassen und über deren Vergleich und Kontrastierung zu einer Bewertung zu gelangen“, stuft Flick (2009, S. 13), in Anlehnung an Flick (2006), Interviews als passendes Instrument ein. Durch die Erweiterung der quantitativen Studie um Interviews als qualitativer Zugang, wurde ein Mixed-Methods-Ansatz zur Beantwortung der Forschungsfragen verfolgt. Vorteilhaft ist dabei, dass quantitativ erhobene Daten und erzielte Ergebnisse um qualitative ergänzt werden und somit bestimmte Ergebnisse gestützt und erklärt werden können (vgl. z. B. Gläser-Zikuda, Seidel, Rohlf, Gröschner & Ziegelbauer, 2012, S. 8; McMillan & Schumacher, 2010, S. 25). Das Vorgehen bei der Auswertung der Daten wird in den Kapiteln 7.3.3 und 7.4.2 eingehend beschrieben.

Mit Blick auf den objektiven Lernzuwachs (*F₂₋₃*) wurden die Interventions- und Kontrollgruppe anhand der erzielten Werte in den Wissenstests im Pre- und Posttest verglichen – entsprechend der Empfehlung von Kirkpatrick (2007, o. A.). Da sich in den vorherigen Studien die Tendenz

¹⁴⁸ In den Test wurden nicht alle möglichen Einflussfaktoren erhoben, da dies zu einem umfangreicheren Test geführt hätte, was zulasten der Datenqualität führen könnte.

zeigte, dass Studierende und LiV nicht hinreichend auf die problemhaltige Aufgabenerstellung zu summativ-diagnostischen Zwecken vorbereitet werden bzw. sich nicht ausreichend gut vorbereitet wahrnehmen (vgl. Kapitel 5.1 und 5.2), war davon auszugehen, dass der Lernzuwachs zwischen den beiden Messzeitpunkten in der Kontrollgruppe weniger stark ist als in der Interventionsgruppe. Zudem zeigte sich in den Vorstudien, dass das Thema der diagnostischen Gütekriterien eher theoretisch behandelt wird und kaum bzw. keine Anwendung stattfindet (vgl. Kapitel 5). In der Intervention wurde das Thema sowohl theoretisch als auch praktisch durch Anwendung behandelt (vgl. Kapitel 7.1), sodass zu vermuten war, dass der Lernzuwachs in der Interventionsgruppe stärker ausgeprägt ist als in der Kontrollgruppe. Folgende Hypothesen wurden, basierend auf diesen Erwartungen, abgeleitet:

- H1: Teilnehmende der Interventionsgruppe haben zum zweiten Messzeitpunkt (Posttest) einen stärkeren Lernzuwachs in Bezug auf das Wissen zu den Merkmalen problemhaltiger Aufgaben als Teilnehmende der Kontrollgruppe.
- H2: Teilnehmende der Interventionsgruppe haben zum zweiten Messzeitpunkt (Posttest) einen stärkeren Lernzuwachs in Bezug auf das Wissen zu diagnostischen Gütekriterien als Teilnehmende der Kontrollgruppe.

Da auch teilnehmendenbedingte Faktoren den Erfolg einer Intervention beeinflussen (vgl. Kapitel 6.2), wurde der Ausbildungsstand der Teilnehmenden für den Lernzuwachs berücksichtigt. Die Vorstudien zeigten, dass insbesondere im Studium das Thema der Erstellung von problemhaltigen Aufgaben zu summativen Zwecken nicht hinreichend behandelt wird (vgl. Kapitel 5). Daher wurde angenommen, dass kein signifikanter Unterschied im Lernzuwachs zwischen den Teilnehmenden höherer und niedrigerer Fachsemester besteht. Das Fachsemester hat demnach keinen Effekt auf die Leistung im Posttest. Hierzu wurde die Hypothese H3a formuliert:

- H3a: Teilnehmende, die sich in einer fortgeschritteneren Ausbildungsphase befinden, haben einen stärkeren Lernzuwachs in Bezug auf das Wissen zu Merkmalen problemhaltiger Aufgaben als Teilnehmende, die in ihrer Ausbildungsphase noch nicht fortgeschritten sind.

Themen zu diagnostischen Gütekriterien werden laut Vorstudien (vgl. Kapitel 5) stärker behandelt – wenn auch nicht immer im Zusammenhang mit Assessments und mit wenig Anwendungsmöglichkeiten. Daher wurde davon ausgegangen, dass der Lernzuwachs bei Studierenden

höherer Fachsemester niedriger ist als bei Studierenden, die in den Fachsemestern weniger fortgeschritten sind. Begründet mit dem *Deckeneffekt* (vgl. Döring & Bortz, 2016, S. 738) ergab sich folgende Hypothese H3b.

H3b: Teilnehmende, die sich in einer fortgeschritteneren Ausbildungsphase befinden, haben einen niedrigeren Lernzuwachs in Bezug auf das Wissen zu diagnostischen Gütekriterien als Teilnehmende, die in ihrer Ausbildungsphase noch nicht fortgeschritten sind.

Mit dem *Deckeneffekt* sind auch die Hypothesen H4a und H4b zu begründen. So wurde davon ausgegangen, dass der erreichte Pretest-Wert im jeweiligen Wissenstest den erreichten Wert im Posttest beeinflusst.

H4a: Teilnehmende, die ein besseres Vorwissen zu den Merkmalen problemhaltiger Aufgaben (gemessen an der erreichten Punktezahl im Pretest) haben, haben einen niedrigeren Lernzuwachs in Bezug auf das Wissen zu den Merkmalen problemhaltiger Aufgaben.

H4b: Teilnehmende, die ein besseres Vorwissen zu den diagnostischen Gütekriterien (gemessen an der erreichten Punktezahl im Pretest) haben, haben einen niedrigeren Lernzuwachs in Bezug auf das Wissen zu den diagnostischen Gütekriterien.

Zur Beantwortung der Forschungsfragen F_3 und F_4 wurde basierend auf bestehender Literatur (vgl. Kapitel 6.5) die Annahme getroffen, dass die Trainingszufriedenheit und der subjektive Lernzuwachs, im Sinne eines wahrgenommenen Lernzuwachses, korrelieren. Hierzu wurde die Hypothese H5 formuliert.

H5: Die Trainingszufriedenheit der Teilnehmenden korreliert positiv mit ihrem eingeschätzten Lernzuwachs.

7.3 Methodisches Vorgehen in der quasi-experimentellen Studie

7.3.1 Eingesetzte Instrumente

Im Rahmen der quasi-experimentellen Studie bearbeiteten beide Gruppen (Interventions- und Kontrollgruppe) einen Pre- und Posttest (vgl. Reichardt, 2019, S. 115). Analog zum Training im Projekt TeKoP wurden Instrumente zu Evaluationszwecken genutzt, jedoch lag der Fokus auf den vorgestellten Forschungsfragen.¹⁴⁹ Es fand keine Änderung an den Instrumenten statt,

¹⁴⁹ Im bzw. für das Projekt TeKoP wurden zusätzlich zu den beschriebenen Instrumenten weitere eingesetzt. Zu diesen gehören u. a. Fragebögen zu Erwartungen der Teilnehmenden (adaptiert von Krille (2017), die die Items

da diese bereits pilotiert und geprüft wurden (vgl. Kapitel 7.1) und sich das Training inhaltlich nicht veränderte. Lediglich ein Teil des biografischen Fragebogens wurde verändert und an die Zielgruppe angepasst. So wurden die Teilnehmenden z. B. nicht nach ihrer Tätigkeit in der Schule befragt, sondern welche Tätigkeit sie nach ihrem Studium bzw. Referendariat anstreben. Items aus dem Fragebogen zu den Biografien der Teilnehmenden wurden zur Stichprobenbeschreibung und für statistische Berechnungen, z. T. als Prädiktoren, genutzt (vgl. Kapitel 7.3.3).

Zur Erfassung der Teilnehmerreaktion und damit einer Evaluation auf der ersten Ebene nach Kirkpatrick und Kirkpatrick (2006, S. 21-22), wurden die folgenden vierstufigen Skalen eingesetzt (vgl. Tabelle 17).

Tabelle 17:

Skalen und Beispielim(s) zur Evaluation auf der ersten Ebene

Skala	Beispielim(s)	Quelle
Instruktionsqualität	„Der erste Trainingsteil hat mir geholfen zu verstehen, wie komplexe problemhaltige Lern- und Prüfungsaufgaben entwickelt werden können.“ „Ich fühlte mich während des Trainings willkommen, um Hilfe/Rat zu fragen.“	In Anlehnung an Ahmed (2010) entwickelt und an die Trainingsinhalte angepasst
Zufriedenheit	„Ich werde das Training in guter Erinnerung behalten.“	Grohmann und Kauffeld (2013, S. 142)
Nutzen	„Die Teilnahme an dem Training ist nützlich für meine (zukünftige) Tätigkeit, z. B. als (ehrenamtliche/-r) Prüfer/-in oder bei der Erstellung von Tests und Klassenarbeiten.“	In Anlehnung an Grohmann und Kauffeld (2013, S. 142) entwickelt

Anmerkung zur Skala: 1 = „trifft gar nicht zu“, 2 = „trifft eher nicht zu“, 3 = „trifft eher zu“, 4 = „trifft voll zu“ (Eigene Darstellung)

Die Evaluation auf der ersten Ebene wurde nur zum Ende der Maßnahme durchgeführt, folglich war der Fragebogenteil Bestandteil des Posttests. Es wäre möglich gewesen, bereits während der Maßnahme zu evaluieren (vgl. Kirkpatrick & Kirkpatrick, 2005, S. 59-60, vgl. Kapitel 6.5). Darauf wurde jedoch verzichtet, da das Training aufgrund der Vorstudien möglichst bedarfsorientiert gestaltet wurde (vgl. Kapitel 7.1). Ein weiterer Grund ist, dass eine Anpassung während der Trainings die Objektivität der Studie gefährdet hätte, denn die Studie wurde in mehreren Gruppen zu unterschiedlichen Zeiträumen durchgeführt.

auf Basis des VIE-Ansatzes nach Vroom (1964) und anderen, bestehenden Instrumenten von z. B. Noe und Schmitt (1986), Tharenou (2001) und Zaniboni, Fraccaroli, Truxillo, Bertolino und Bauer (2011) entwickelte), zur Selbsteinschätzung der Medienkompetenz (Niegemann & Korbach, 2020) sowie zu Akzeptanz neuer Medien (Davis, Bagozzi & Warschaw, 1989).

Zur Evaluation auf der zweiten Ebene (Lernen bzw. Lernzuwachs) wurden Items eingesetzt, die den eingeschätzten Lernzuwachs in Form einer Selbstauskunft abfragen - entsprechend der Empfehlung von Kirkpatrick und Kirkpatrick (2005, S. 61). In Tabelle 18 wird die vierstufige Skala zur Selbstauskunft mit Beispielimens vorgestellt.

Tabelle 18:

Skala und Beispielimens zur Evaluation auf der zweiten Ebene

Skala	Beispielimens(s)	Quelle
Eingeschätzter Lernzuwachs	„In dem Training habe ich viel über das Erstellen problemhaltiger Lern- und Prüfungsaufgaben und ihre mediendidaktische Umsetzung gelernt.“ „Ich fühle mich in der Lage, problemhaltige Lern- und Prüfungsaufgaben für andere als die exemplarisch betrachteten Ausbildungsberufe zu entwickeln.“	In Anlehnung an Grohmann und Kauffeld (2013, S. 142) entwickelt und an die Trainingsinhalte angepasst

Anmerkung zur Skala: 1 = „trifft gar nicht zu“, 2 = „trifft eher nicht zu“, 3 = „trifft eher zu“, 4 = „trifft voll zu“ (Eigene Darstellung)

Zum anderen wurden mittels Wissenstests der objektive Lernzuwachs zu den Merkmalen problemhaltiger Aufgaben und diagnostischen Gütekriterien im Pre- und Posttest gemessen. Die Wissenstests wurden im Projekt TeKoP eigens entwickelt (Wuttke et al., 2023). Zur Wiederverwendbarkeit der in der Studie eingesetzten Fragebogen und Tests sind diese nicht angehängt und publiziert.

7.3.2 Durchführung und Sampling

Zur Beantwortung der Forschungsfragen wurde das Training (1) mit Masterstudierenden der Wirtschaftspädagogik an den Universitäten Frankfurt und Göttingen sowie mit (2) LiV an kaufmännisch-beruflichen Schulen durchgeführt. Die Erhebungen fanden im Sommersemester 2021, Wintersemester 2021/2022 und Sommersemester 2022 statt. Das Training wurde im Rahmen der wirtschaftspädagogischen Masterveranstaltung *Didaktik der beruflichen Aus- und Weiterbildung, DBWA* (Goethe-Universität Frankfurt) bzw. *Vertiefende Fachdidaktik und Unterrichtsforschung Wirtschaftswissenschaften, VFD* (Georg-August-Universität Göttingen) angeboten. Somit handelt es sich um eine Feldstudie, für welche quasi-experimentelle Designs empfehlenswert sind (vgl. Döring & Bortz, 2016, S. 199-200). Anzumerken ist, dass an beiden Universitäten zu den genannten Modulen eine Parallelveranstaltung stattfand (mit Ausnahme der DBWA im Sommersemester 2022). Das jeweilige Modul war für die Studierenden keine

Pflicht- sondern eine (Quasi-)Wahlpflichtveranstaltung^{150,151,152}. Die Veranstaltungen an beiden Universitäten wurden als Blockveranstaltung (inkl. einer Selbstlernphase) angeboten. Am Pre- und Posttest nahmen insgesamt 58¹⁵³ Personen der Interventionsgruppe teil.

LiV an kaufmännisch-beruflichen Schulen wurden zur freiwilligen Teilnahme an der Trainingsveranstaltung eingeladen. Die Einladung fand über Facebook-Gruppen, Flyer an kaufmännisch-berufliche Studienseminare bundesweit, Gespräche und Telefonate mit Mitarbeitenden der Studienseminare, eine virtuelle Vorstellung des Programms an einem Studienseminar sowie die Veröffentlichung auf der Seite der *Goethe Lehrkräfteakademie* statt. Die Veranstaltung wurde von der *Hessischen Lehrkräfteakademie* zertifiziert und fand über drei synchrone Trainingstage statt, wobei zwischen Trainingstag zwei und drei die Selbstlernphase stattfand (vgl. Kapitel 7.1). Somit wurden sowohl die Veranstaltungen für die Studierenden als auch für LiV im gleichen Format und unter möglichst ähnlichen Rahmenbedingungen angeboten (vgl. Döring & Bortz, 2016, S. 201). Die Anzahl der teilnehmenden LiV am Pre- und Posttest liegt bei fünf Personen.

Zur Bildung der Kontrollgruppe wurden verschiedene Maßnahmen ergriffen. Die Studierenden des Masterstudiengangs Wirtschaftspädagogik erhielten über einen Verteiler bestimmter Veranstaltungen der Universitäten Informationen über die Teilnahmemöglichkeit. Zudem wurde im Rahmen verschiedener Veranstaltungen vor Ort für die Teilnahme geworben und Flyer mit QR-Codes verteilt. Dabei wurden Studierende angesprochen, die nicht das Modul absolvierten, in dem das Training durchgeführt wurde. Des Weiteren wurde über *StudyDrive* ein Beitrag über die Studie gepostet. An der Erhebung (Pre- und Posttest) nahmen insgesamt 30 Personen (Kontrollgruppe) teil.

Die LiV wurden über die Studienseminare kontaktiert sowie über die Verteiler dieser angeschrieben und zur Teilnahme an der Fragebogenstudie als Kontrollgruppe eingeladen. Zudem

¹⁵⁰ An der Goethe-Universität Frankfurt (2015b, S. 39) wird das Modul als eine Pflichtveranstaltung ausgewiesen. Es werden aber zwei parallele Veranstaltungen angeboten, mit jeweils einem schulischen und einem betrieblichen Schwerpunkt. Daher wurde hier der Begriff der *Quasi-Wahlpflichtveranstaltung* genutzt. Zu der Veranstaltung, in der das Training stattfand, wurde angegeben, dass diese einen schulischen Schwerpunkt hat und welche Inhalte in der Veranstaltung behandelt werden.

¹⁵¹ An der Georg-August-Universität Göttingen (2019, S. 154-155) ist das Modul als Wahlpflichtveranstaltung ausgewiesen (Georg-August-Universität Göttingen, 2019, S. 12) Zur Veranstaltung, in der das Training stattfand, wurden die Inhalte dessen angegeben.

¹⁵² Die Studierenden beider Standorte konnten also zwischen den Veranstaltungen wählen.

¹⁵³ Diese und die im Nachfolgenden aufgeführten Zahlen zu den Teilnehmenden der Pre- und Posttests beider Gruppen umfassen nur die Anzahl der Teilnehmenden, die sowohl am Pre- als auch am Posttest teilnahmen und der Verarbeitung der Daten zustimmten. Daten von Personen, die an keinem oder nur einem der Tests teilnahmen und/oder der Datenschutzerklärung nicht zustimmten, wurden nicht berücksichtigt.

wurden private Kontakte genutzt und Beiträge in Facebook-Gruppen gepostet. Der Einladung folgten fünf Personen.

Teilnehmende der Kontrollgruppen wurden über die rechtlichen, organisatorischen und inhaltlichen Gegebenheiten informiert. Zudem wurde ihnen mitgeteilt, dass die Teilnahme am Pre- und Posttest vergütet wird. Die Erhebungen fanden im Zeitraum vom April 2021 bis Juli 2022 mit mehreren Gruppen statt. Einen Überblick hierzu geben die nachfolgenden Tabellen 19 und 20.

Tabelle 19:*Durchführungs- und Erhebungszeiträume der Trainings, inkl. Pre- und Posttestbefragung (Interventionsgruppe)*

	LiV		Masterstudierende			
	Gruppe 1	Gruppe 2	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
N _{vollständige Teilnahme}	2	4	19	9	23	10
N _{Pre-Posttest}	1	4	18	8	22	10
Trainingstag 1 (inkl. Pretest)	03.05.2021	29.09.2021	16.04.2021 (Pretest) 30.04.2021 ¹⁵⁴ (Trainingstag 1)	21.04.2021	29.10.2021	13.04.2022
Trainingstag 2	04.05.2021	02.10.2021	07.05. + 14.05.2021	05.05.2021	05.11.2021	01.06.2022
Selbstlernphase	05.05. - 06.06.2021	03.10. - 31.10.2021	17.05. - 14.06.2021	06.05. - 30.05.2021	06.11. - 26.11.2021	28.04. - 18.05.2022
Trainingstag 3	24.06.2021	13.11.2021	25.06.2021	16.06.2021	10.12.2021	01.06.2022
Posttest (ggf. Zeitraum)	30.07. - 13.08.2021	17.12. - 31.12.2021	16.07.2021	14.07.2021	14.01.2021	29.06.2022

(Eigene Darstellung)

Tabelle 20:*Zeiträume der Pre- und Posttestbefragung (Kontrollgruppe)*

	LiV		Masterstudierende			
	Gruppe 1		Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
N _{Pre-Posttest}	5		4	2	9	15
Pretest	17.09. - 03.10.2021		30.04. - 09.05.2021	11.05. - 23.05.2021	21.10. - 01.11.2021	11.04. - 27.04.2022 ¹⁵⁵
Posttest	03.12.2021 - 19.12.2021		19.07. - 01.08.2021		10.01. - 23.01.2022	25.06. - 10.07.2022

(Eigene Darstellung)

¹⁵⁴ Da die Veranstaltungstermine des Standorts in diesem Semester halbtags geplant waren, musste der Pretest eine Woche vor Trainingstag 1 stattfinden, und Trainingstag 2 musste auf zwei Veranstaltungstage aufgeteilt werden.

¹⁵⁵ Da das Semester in Göttingen später startet als in Frankfurt, wurde den Studierenden in Göttingen ein längerer Zeitraum zur Bearbeitung gegeben.

Die Zuteilung zu den Interventions- und Kontrollgruppen verlief nicht zufällig bzw. randomisiert (vgl. Döring & Bortz, 2016, S. 199). Die Proband*innen hatten die Möglichkeit, sich für das jeweilige Modul (Studierende) bzw. die Fortbildung (LiV) anzumelden und an der Intervention teilzunehmen oder aber an der vergüteten Befragung. Aufgrund der Nicht-Randomisierung bestand die Wahrscheinlichkeit, dass sich die Teilnehmenden beider Gruppen schon vor der Intervention unterscheiden (Reichardt, 2019, S. 113), etwa in ihrem Interesse zum Erstellen problemhaltiger Aufgaben. Da das Training nicht verpflichtend war, wurde bereits bei der Rekrutierung bzw. Anmeldung eine Selbstselektion seitens der potenziellen Teilnehmer*innen erwartet. Dies kann etwa durch die Motivation zur Trainingsteilnahme (allgemein und speziell zu dem Thema) bedingt sein (z. B. Krille, 2019; vgl. Kapitel 6.2). Das muss aber nicht bedeuten, dass die Interventionsgruppe (intrinsisch) motivierter ist als die Nicht-Teilnehmenden (Kontrollgruppe). So hatte das Training, zumindest für Masterstudierende, auch eine funktionale Bedeutung bzw. einen extrinsischen Motivationsfaktor (vgl. z. B. Brady et al., 2009, S. 445; Harvey, 2005), da dieses im Rahmen einer Lehrveranstaltung angeboten wurde und die Studierenden dafür Credit Points erhielten. Die LiV haben keine auf ihre Ausbildung anrechenbare Credit Points o. Ä. für die Teilnahme an der Veranstaltung erhalten. Ihnen wurde für die Teilnahme am Pre- und Posttest mit einem Gutschein im Wert von 30 Euro als Incentive gedankt.

Selektionsunterschiede können die interne Validität gefährden (Döring & Bortz, 2016, S. 200; Reichardt, 2019, S. 113; Rost, 2007, S. 111-112). Um personenbezogene und untersuchungsgebundene Störvariablen zu berücksichtigen und somit die Validität der Studie zu unterstützen, wurden die von Döring und Bortz (2016, S. 200-201) genannten Techniken (möglichst) beachtet. So wurde versucht, in den Interventions- und Kontrollgruppen beider Zielgruppen Personen zu erreichen, die im selben Ausbildungsgang und -stadium sind und sich hinsichtlich dieser Merkmale ähneln (vgl. hierzu die deskriptive Beschreibung der Gruppen in Kapitel 7.3.3). Weiter wurden die Bedingungen während der Erhebungen möglichst gleich gehalten (Konstanthalten; vgl. Kapitel 7.3.1). Ausfälle konnten vermieden werden (vgl. Döring & Bortz, 2016, S. 201). Zudem wurden unabhängige Variablen (etwa das Geschlecht, Geburtsjahr) abgefragt, die als potentielle Störvariablen gelten könnten (Döring & Bortz, 2016, S. 200; Reichardt, 2019, S. 115; Rost, 2007, S. 71). Durch die Pre- und Posttests erfolgte eine Messwiederholung (Döring & Bortz, 2016, S. 200-201).

7.3.3 Beschreibung der Stichprobe der Interventions- und Kontrollgruppe

An der Interventionsstudie im Rahmen der universitären Lehrveranstaltungen haben 58 Masterstudierende der Wirtschaftspädagogik vollständig teilgenommen. In der Kontrollgruppe waren es 30 Masterstudierende. In Tabelle 21 wird die Stichprobe genauer beschrieben.

Tabelle 21:

Beschreibung der Stichprobe der Masterstudierenden

	Standort	N	Geschlecht	Alter M (SD)	Studienfach Wirtschaftspädagogik ^{a)}		Fachsemester M (SD)	Angestrebte Tätigkeit ^{a)}	
					Studienrichtung I	Studienrichtung II		in Schule	andere
IG	Insgesamt	58	m = 22 w = 36	27.72 (4.30)	11	29	^{b)} 2.67 (0.92)	44	11
	Frankfurt	18	m = 6 w = 12	30.94 (6.10)	4	14	2.50 (1.15)	14	2
	Göttingen	40	m = 16 w = 24	26.28 (1.92)	7	15	^{c)} 2.79 (0.74)	30	9
KG	Insgesamt	30	m = 15 w = 15	26.63 (2.72)	15	14	2.17 (0.99)	17	13
	Frankfurt	14	m = 4 w = 10	26.00 (2.63)	6	8	2.64 (1.08)	8	6
	Göttingen	16	m = 11 w = 5	27.19 (2.76)	9	6	1.75 (0.68)	9	7

a) Abweichungen der Häufigkeiten vom angegebenen N-Werte sind auf Missings zurückzuführen.

b) N = 51; 13 Missings.

c) N = 28; 12 Missings.

(Eigene Darstellung)

Während in der Interventionsgruppe mehr Frauen vertreten waren, war die Verteilung der Geschlechter in der Kontrollgruppe gleich. Das Durchschnittsalter beider Gruppen zeigte keine große Abweichung. Bei allen Studierenden handelte es sich um Masterstudierende der Wirtschaftspädagogik, die im zweiten und dritten Fachsemester waren. Der Großteil der Studierenden beider Gruppen gab an, eine Tätigkeit an der Schule anzustreben, wobei relativ betrachtet mehr Teilnehmende der Interventionsgruppe als Teilnehmende der Kontrollgruppe diese Intention hatten. Dies kann mit dem Schwerpunkt der Veranstaltungsgruppe, in der das Training angeboten wurde, erklärt werden (vgl. Kapitel 7.3.2). Insgesamt gaben 15 Teilnehmende der Interventionsgruppe an, keine Berufsausbildung absolviert zu haben ($N_{IG|Frankfurt} = 9$; $N_{IG|Göttingen} = 6$). 24 haben eine kaufmännische Ausbildung abgeschlossen ($N_{IG|Frankfurt} = 8$; $N_{IG|Göttingen} = 16$), eine Person eine Ausbildung in einem anderen Bereich ($N_{IG|Frankfurt} = 1$). Zu dieser Frage lagen 18 Missings vor. Bei den Proband*innen der Kontrollgruppe hatten 13 Teilnehmende keine Berufsausbildung ($N_{KG|Frankfurt} = 9$; $N_{KG|Göttingen} = 4$) und 16 Proband*innen eine kaufmännische Berufsbildung absolviert ($N_{KG|Frankfurt} = 5$; $N_{KG|Göttingen} = 11$). Ein*e Proband*in hat eine andere Berufsausbildung abgeschlossen ($N_{KG|Göttingen} = 1$).

Die Stichprobe der LiV wird in Tabelle 22 dargestellt.

Tabelle 22:

Beschreibung der Stichprobe der LiV

	N	Geschlecht	Alter M (SD)	Studienfach Wirtschaftspädagogik ^{a)}		Ausbildungsjahr M (SD)	Angestrebte Tätigkeit	
				Studienrichtung I	Studienrichtung II		in Schule	andere
IG	5	m = 3 w = 2	33.00 (4.85)	1	3	1.20 (0.45)	5	0
KG	5	m = 2 w = 3	33.60 (5.77)	2	3	1.80 (0.45)	4	1

a) Abweichungen der Häufigkeiten vom angegebenen N-Werte sind auf Missings zurückzuführen. (Eigene Darstellung)

Bei beiden Stichproben sind die Geschlechterverteilung und das Durchschnittsalter ähnlich. Alle Teilnehmende der Interventions- und Kontrollgruppe gaben an, studiert zu haben. Vier der fünf Teilnehmenden der Interventionsgruppe gaben den Master als den höchsten Bildungsabschluss an, wie auch alle fünf Teilnehmenden der Kontrollgruppe. Alle Teilnehmenden befanden sich noch in der Ausbildungszeit (im ersten und zweiten Ausbildungsjahr). Bis auf eine, gaben alle Personen aus der Kontrollgruppe an, eine Tätigkeit an der Schule anzustreben. Vier LiV der Interventionsgruppe berichteten davon, eine kaufmännische oder eine gewerblich-technische Berufsausbildung absolviert zu haben. In der Kontrollgruppe gab es zwei Proband*innen ohne berufliche Ausbildung, die drei anderen haben eine kaufmännische Berufsausbildung absolviert. Die LiV der Interventionsgruppe gaben an, Industriekaufleute mindestens einmal unterrichtet zu haben, ein*e LiV hat bereits mindestens einmal Kaufleute für Büromanagement unterrichtet. Eine*r der fünf LiV hat keine Erfahrung mit dem Unterrichten der beiden Ausbildungsberufsgruppen. Ein*e LiV der Kontrollgruppe gab ebenfalls an, mindestens einmal Industriekaufleute unterrichtet zu haben, zwei hatten Erfahrung mit dem Unterrichten von Kaufleuten für Büromanagement. In der Kontrollgruppe hatte ein*e LiV beide Berufe mindestens einmal unterrichtet, ein*e LiV hatte keine Erfahrung im Unterrichten der Ausbildungsgruppen.

7.3.4 Auswertung der quantitativen Daten zur Trainingsevaluation

Zur Evaluation auf der ersten und zweiten Ebene nach Kirkpatrick und Kirkpatrick (2006) und zur Beantwortung der Forschungsfragen F_{2-1} und F_{2-2} wurden die Reliabilitätswerte der Skalen mittels SPSS berechnet. Hierbei wurden die Daten der Masterstudierenden und LiV berücksichtigt. Es wurden (nahezu) gute bis sehr gute Reliabilitätswerte erreicht (vgl. Bortz & Döring, 2006, S. 725). Dies ist u. a. darauf zurückzuführen, dass die Skalen bzw. Items aus bestehenden

Fragebögen adaptiert wurden. Die Skala zur wahrgenommenen Nützlichkeit mit lediglich zwei Items blieb unberücksichtigt, da hier ein Cronbachs Alpha-Wert von $\alpha = .597$ erreicht wurde und dieser im unzureichenden Bereich liegt (vgl. Blanz, 2021, S. 250 nach Cronbach, 1951). Tabelle 23 bildet die Reliabilitätswerte der verbleibenden Skalen und die Anzahl der zugehörigen Items ab. Der Bericht zu den Ergebnissen erfolgt deskriptiv.

Tabelle 23:

Reliabilitätswerte der eingesetzten Skalen

Skala	Anzahl der Items	α
Instruktionsqualität	5	.825
Zufriedenheit mit dem Training	2	.846
Wahrgenommener Lernzuwachs	4	.731

(Eigene Darstellung)

Für die Wissenstests beider Zielgruppen wurden zunächst die erreichten Pre- und Posttestwerte deskriptiv betrachtet. Zur Auswertung der vorhandenen Daten der studentischen Kontroll- und Interventionsgruppe wurde weiterhin eine Kovarianzanalyse (ANCOVA) durchgeführt. In dieser wurde als abhängige Variable die erreichte Punktzahl im Posttest des jeweiligen Wissenstests (zu den Merkmalen problemhaltiger Aufgaben und zu den diagnostischen Gütekriterien) definiert und die erreichte Punktzahl im Pretest zum jeweiligen Wissenstest als Kovariate genutzt. Dadurch wurde bestrebt, den erreichten Punktwert im Posttest um den Wert im Pretest zu bereinigen, um so den Effekt bzw. die Unterschiede der Intervention- und Kontrollgruppe „eindeutig auf das Treatment zurückführen zu können“ (Döring & Bortz, 2016, S. 722; vgl. hierzu auch McMillan & Schumacher, 2010, S. 317; Reichardt, 2019, S. 121; Rost, 2007, S. 194). Mit diesem Vorgehen sollten die Hypothesen 1 und 2 überprüft werden. Zur Erfüllung der Voraussetzungen wurden die Gruppen auf Normalverteilung mit dem Shapiro-Wilk-Test geprüft. Auch wurde die Homogenität der Kovariaten und der Regression sowie die Varianzhomogenität mittels eines Levene-Tests überprüft (Bortz & Schuster, 2010, S. 311-314; Bühner & Ziegler, 2017, S. 380-381; Cohen, 1988, S. 493-496; Field, 2009, 399, 413-415). Tabelle 24 bildet die Ergebnisse ab.

Tabelle 24:*Voraussetzungsprüfung für ANCOVA*

Voraussetzungsprüfung für ANCOVA	Kovariate	Homogenität der Regressionssteigung	Varianz
Wissenstest Merkmalen problemhaltiger Aufgaben ¹⁵⁶	$F_{AM}(1, 96) = .696;$ $p_{AM} > .05$	$F_{AM}(1, 84) = 2.531;$ $p > .05$	$p_{AM} > .05$
Wissenstest Gütekriterien	$F_{GK}(1, 96) = 3.545;$ $p_{GK} > .05$	$F_{GK}(1, 85) = .009;$ $p_{GK} > .05$	$p_{GK} > .05$

(Eigene Darstellung)

Es konnte keine Normalverteilung der betrachteten Werte im Pre- und Posttest der Wissenstests ermittelt werden (in allen Fällen $p < .01$). Da jedoch die Kovarianzanalyse als robust gilt (Bortz & Schuster, 2010, S. 311; Rost, 2007, S. 188; zur Robustheit vgl. Wilcox, 2005), wurde trotz Verletzung der Voraussetzung mit dem Verfahren gearbeitet. Nach der Prüfung auf signifikante Unterschiede zwischen der Interventions- und Kontrollgruppe und unter Berücksichtigung des Pretests, wurden die Effektstärken anhand der Formel aus Field (2009, S. 416) berechnet.

Da die Stichprobengröße der LiV zu klein war, wurde keine ANCOVA durchgeführt. Auf eine gemeinsame ANCOVA mit der studentischen Gruppe wurde verzichtet, da ein größerer Gruppenunterschied zwischen Studierenden und LiV vorlag. Die beiden Gruppen waren unterschiedlich groß und befanden sich in unterschiedlichen Ausbildungsphasen und Erfahrungen im Unterrichten und Klausurerstellen (vgl. Kapitel 7.3.3). Dies könnte zu einer Verfälschung der Ergebnisse führen. Stattdessen wurde mittels Betrachtung der Mittelwerte der LiV-Gruppen eine Tendenz abgeleitet.

Um zu erklären, welchen Beitrag einzelne Prädiktoren – hier: die unabhängigen Variablen in den Hypothesen 3a bis 4b – für den erreichten Wert im Posttest leisten, wurde eine multiple Regressionsanalyse durchgeführt (vgl. Bauer, 2020, S. 8-9; Bühner & Ziegler, 2017, S. 677-678; Döring & Bortz, 2016, S. 626). Das Analyseverfahren konnte aufgrund der geringen Stichprobe der LiV nur für die Daten der Studierenden angewendet werden. Aus oben genannten Gründen wurde auf eine Zusammenführung der Daten beider Gruppen verzichtet (vgl. auch

¹⁵⁶ Da die Verfahren für die Variablen zum Wissenstest zu den Merkmalen problemhaltiger Aufgaben und zum Wissenstest zu diagnostischen Gütekriterien beschrieben werden, werden zu den einzelnen Werten die Abkürzungen „AM“ für den Wissenstest zu den Merkmalen problemhaltiger Aufgaben und „GK“ für den Wissenstest zu diagnostischen Gütekriterien eingefügt. Dadurch soll eine genauere Zuordnung der Werte erleichtert werden.

Kapitel 8.1.4). Im Folgenden wird daher die Regressionsanalyse unter Berücksichtigung der studentischen Daten beschrieben.

Zunächst wurden die notwendigen Voraussetzungen geprüft (vgl. z. B. Baltes-Götz, 2022, S. 90-92; Bortz & Schuster, 2010, S. 348; Bühner & Ziegler, 2017, S. 720-739; Field, 2009, S. 220-222), die im Nachgang vorgestellt und angewandt werden. Als Kriterium (abhängige Variable) wurde jeweils der erzielte Wert im Wissenstest zu den Merkmalen problemhaltiger Aufgaben und den diagnostischen Gütekriterien betrachtet. Der Empfehlung von Bühner und Ziegler (2017, S. 739) folgend, wurden zunächst alle interessierten Prädiktoren (hier: das Fachsemester, der erreichte Wert im Pretest und die Gruppenzugehörigkeit) sowie die Störvariablen (Alter und Geschlecht) bestimmt. Es zeigte sich, dass die Störvariablen als Prädiktoren in beiden Fällen keinen signifikanten Einfluss haben (jeweils $p > .05$) – weder für das Kriterium zum erreichten Posttestwert im Wissenstest zu den Merkmalen problemhaltiger Aufgaben noch zu den diagnostischen Gütekriterien. Verglichen mit den Modellen zu den Merkmalen problemhaltiger Aufgaben ohne Störvariablen (vgl. Tabelle 29 in Kapitel 8.1.3: korrigiertes $R^2_{AM} = .375$ in Modell 1 bzw. korrigiertes $R^2_{AM} = .380$ in Modell 2), nahm im Modell mit Störvariablen gar der Wert des korrigierten R^2 (korrigiertes $R^2_{AM} = .363$) ab. Demnach hat sich das Modell durch die Hinzunahme der Störvariablen als Prädiktoren nicht verbessert. Für das Modell mit Kriterium des Posttestwerts im Wissenstest zu den diagnostischen Gütekriterien lag zwar der Wert mit einem korrigierten $R^2_{GK} = .207$ im Modell mit Störvariablen höher als im Modell ohne Störvariablen (vgl. Tabelle 30 in Kapitel 8.1.3: korrigierte $R^2_{GK} = .189$ in Modell 1); allerdings konnte im Modell ohne Störvariablen und ohne den Prädiktor *Fachsemester* ein besserer Wert errechnet werden (vgl. Tabelle 30 in Kapitel 8.1.3: korrigierte $R^2_{GK} = .209$ in Modell 2). Daher wurde auf die Berücksichtigung der Störvariablen verzichtet.

Zur Prüfung der Ausreißer wurden die mittels SPSS berechneten studentisierten ausgeschlossenen Residuen betrachtet. Diese lagen für die Modelle beider Kriterien¹⁵⁷ im akzeptablen Wertebereich, d. h. es lagen keine Ausreißer vor (Eid, Gollwitzer & Schmitt, 2016, S. 194; Field, 2009, S. 216-217). Anschließend wurden die Hebelwerte (englisch: Leverage) geprüft. Auch hierfür wurde der mit SPSS berechnete Wert genutzt und so die Wertebereiche überprüft. Keiner der Werte war $> .2$, somit liegen keine Ausreißer vor (Huber, 1981, S. 192-195; Stahel, 2017, S. 94-95). Der in SPSS berechnete Wert zur Cook-Distanz, welche die Hebelwirkung und studentisierte Residuen berücksichtigt (Bühner & Ziegler, 2017, S. 723-724 nach Cook, 1977),

¹⁵⁷ Im Modell zum Kriterium des Posttestwerts des Wissenstest zu den diagnostischen Gütekriterien lag ein Wert bei -3.348 . Dieser Datensatz wurde aber nicht eliminiert, da die Abweichung keine große war.

bestätigte die vorherigen Ergebnisse. Es konnten keine Ausreißer bestimmt werden, da die Werte bei < 1 lagen (Backhaus, Erichson, Gensler, Weiber & Weiber, 2021, S. 133; Field, 2009, S. 217). Schließlich wurde mittels eines Durbin-Watson-Tests eine mögliche Autokorrelation der Residuen ausgeschlossen. Für das Kriterium des Posttests zu den Merkmalen problemhaltiger Aufgaben wurde ein Wert von 1.952, für das Kriterium der erreichten Punkte im Posttest zu diagnostischen Gütekriterien von 1.902 berechnet. Da beide Werte nah am Wert 2 waren, wurde auf eine Nichtkorrelation geschlossen (Backhaus et al., 2021, S. 118; Field, 2009, S. 220-221). Dass die Anzahl der Prädiktoren geringer ist als der Umfang der Stichprobe, deutet zudem auf eine Nicht-Multikollinearität hin (Bortz & Schuster, 2010, S. 354). Die Multikollinearität, also die Korrelation der Werte, wurde zum einen mittels des Korrelationskoeffizienten nach Pearson berechnet und zum anderen über die VIF-Werte geprüft. Die Ergebnisse finden sich in Tabelle 25.

Tabelle 25:*Ergebnisse zur Multikollinearitätsprüfung*

Korrelation der Prädiktoren nach Pearson		Punktzahl im Posttest zu Aufgabenmerkmale	Punktzahl im Pretest zu Aufgabenmerkmale	Fachsemester	Intervention (IG =1; KG =0)	VIF_{AM}
Kriterium: Punktzahl im Posttest zu Aufgabenmerkmale ¹⁵⁸	Punktzahl im Posttest zu Aufgabenmerkmale		.430	.255	.516	
	Punktzahl im Pretest zu Aufgabenmerkmale	.430		.171	.147	1.042
	Fachsemester	.255	.171		.257	1.092
	Intervention (IG =1; KG =0)	.516	.147	0.257		1.083
Korrelation der Prädiktoren nach Pearson		Punktzahl im Posttest zu Gütekriterien	Punktzahl im Pretest zu Gütekriterien	Fachsemester	Intervention (IG =1; KG =0)	VIF_{GK}
Kriterium: Punktzahl im Posttest zu Gütekriterien	Punktzahl im Posttest zu Gütekriterien		.452	-.037	.189	
	Punktzahl im Pretest zu Gütekriterien	.452		.072	.221	1.052
	Fachsemester	-.037	.072		.257	1.071
	Intervention (IG =1; KG =0)	.189	.221	.257		1.120

(Eigene Darstellung)

Das Ergebnis zeigt über alle Variablen einen Korrelationskoeffizienten einen Wert von $< .6$. Auch die ermittelten VIF-Werte (alle > 1.0) weisen auf keine Multikollinearität hin (Bühner & Ziegler, 2017, S. 735-736). Zur Prüfung der Homoskedastizität erfolgte eine grafische Betrachtung mittels Scatterplots (vgl. Bühner & Ziegler, 2017, S. 725-729). Während für das Kriterium der erreichten Posttestpunkte im Wissenstest zu den diagnostischen Gütekriterien eine annähernd gute Verteilung über die x-Achse zu beobachten ist, entspricht die Streuung für die erreichten Punkte im Posttest des Wissenstests zu den Merkmalen problemhaltiger Aufgaben nicht dem Ideal. Bühner und Ziegler (2017, S. 725) merken an, dass die Homoskedastizität eine selten erfüllte Voraussetzung ist. Das Ergebnis zur Homoskedastizität wurde daher toleriert, auch weil die anderen Voraussetzungen erfüllt waren. Zuletzt wurden die Residuen auf Normalverteilung geprüft. Dies geschah grafisch mittels Histogramme und P-P-Plots. Es konnte eine relative Normalverteilung der Residuen für das Modell zum Wissenstest der Merkmale

¹⁵⁸ Abgekürzte Bezeichnung für den Wissenstest zu den *Merkmale* problemhaltiger *Aufgaben*.

problemhaltiger Aufgaben festgestellt werden. Die Grafiken zum Wissenstest zu den diagnostischen Gütekriterien zeigt keine Normalverteilung. Zudem wurde der Wert unter Anwendung eines Shapiro-Wilk Tests überprüft. Entgegen dem grafischen Ergebnis zeigte der Wert, dass die Daten mit $p_{AM} > .05$ und $p_{GK} > .05$ normalverteilt sind (vgl. Field, 2009, S. 144). Die multiplen Determinationskoeffizienten, die Aufschluss über die Varianzaufklärungen geben (Döring & Bortz, 2016, S. 835-836; Field, 2009, S. 237; Stahel, 2017, S. 129), lagen ohne Störvariablen bei $R^2_{AM} = .400$ (korrigiertes $R^2_{AM} = .375$) bzw. $R^2_{GK} = .222$ (korrigiertes $R^2_{GK} = .189$). Die ermittelte Werte sind nach Cohen (1988, S. 412-414) als hoch bzw. moderat anzusehen. Anhand der Werte aus der ANOVA-Tabelle ($F_{AM}(3, 72) = 16.025$, $p_{AM} < .001$; $F_{GK}(3, 72) = 6.841$, $p_{GK} < .001$) konnte bestimmt werden, dass die Prädiktoren (Gruppenzugehörigkeit¹⁵⁹, erreichter Wert im Pretest, Fachsemester) die erreichte Punktzahl im jeweiligen Post-Wissenstest statistisch signifikant voraussagen (vgl. Baltes-Götz, 2022, S. 96; Field, 2009, S. 236-237). Zur Berechnung des Korrelationskoeffizienten zur Prüfung der Hypothese H5 wurden die Koeffizienten nach Pearson in SPSS berechnet. Die Interpretation erfolgte nach Cohen (1988, S. 80).

¹⁵⁹ Damit ist die Dummy-Variable gemeint, die angibt, ob die Teilnehmenden der Interventions- oder Kontrollgruppe angehörten.

7.4 Methodisches Vorgehen in der Interviewstudie zur Trainingsevaluation

7.4.1 Durchführung und Sampling

Die Proband*innen des Interviews waren angehende Lehrkräfte, die im Rahmen des Interviews auch zu den Ausbildungsinhalten aus der Nachfrageperspektive befragt wurden (vgl. Kapitel 4.3). Das Vorgehen entspricht auch hier dem episodischen Interview. Es wurde ein teilstandardisierter Leitfaden eingesetzt (vgl. Kapitel 4.3.1) – zu Evaluationszwecken ist dies nicht unüblich (vgl. Flick, 2011, S. 278). Auf den Durchführungszeitraum der Intervention wurde bereits in Kapitel 7.3.2 eingegangen. An den Interviews nahmen Personen beider Teilnehmer*innengruppen der Trainingsstudie teil: Masterstudierende der Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen.

Die Teilnehmenden der Interviews waren also gleichzeitig auch Teilnehmende der Trainingsstudie (vgl. Kapitel 7.3.2). Da (a) die Teilnehmendengruppen nicht gleich groß waren (vgl. Kapitel 7.3.3) und (b) die Teilnahme am Interview auf Freiwilligkeit beruhte, kam es zu Selektionseffekten (Interesse und Bereitschaft für die Teilnahme). Die bestrebte Anzahl von jeweils zehn Personen pro Zielgruppe konnte nicht erreicht werden, was nicht unabhängig der dargestellten Gründe (a) und (b) ist. Schließlich wurden drei LiV an kaufmännisch-beruflichen Schulen und 13 Masterstudierende der Wirtschaftspädagogik interviewt. Um Redundanzen zu vermeiden, wird zur Stichprobenbeschreibung auf Kapitel 4.3.3 verwiesen.

Nach Ende des Trainings bzw. am letzten Trainingstag wurden die Teilnehmenden dazu eingeladen, an einem ca. einstündigen Interview teilzunehmen. Für die Teilnahme erhielten sie 30 Euro, was angekündigt wurde. Die Einladung zum Interview fand zum Ende der Veranstaltung statt. So sollte vermieden werden, dass die Teilnehmenden während der Intervention einen Rollenwechsel von Teilnehmer*in zum*r Gutachter*in erleben und sich auf die Evaluation durch das Interview vorbereiten. Die Interviews wurden (überwiegend) innerhalb von zwei Wochen nach Abschluss des Trainings durchgeführt. Auch wenn der Leitfaden für die beiden Zielgruppen gleich war, gab es Fragen, die sich entsprechend des aktuellen Status (Student*in oder LiV) des Interviewten änderten. Um diese Änderungen nicht im Rahmen der Interviews ad-hoc berücksichtigen zu müssen, wurden zwei Leitfäden angefertigt (vgl. Anhang F). Im Interview wurde eine Tischvorlage mit dem Trainingsaufbau eingeblendet, um die Teilnehmenden darin zu unterstützen, zu einzelnen Trainingsbausteinen Rückmeldung zu geben (vgl. Anhang F). Auf die Beschreibung der Strukturierung des Leitfadens wird an dieser Stelle verzichtet, da dies bereits in Kapitel 4.3.1 erfolgte. Die Interviews wurden von der Dozentin der Veranstaltung

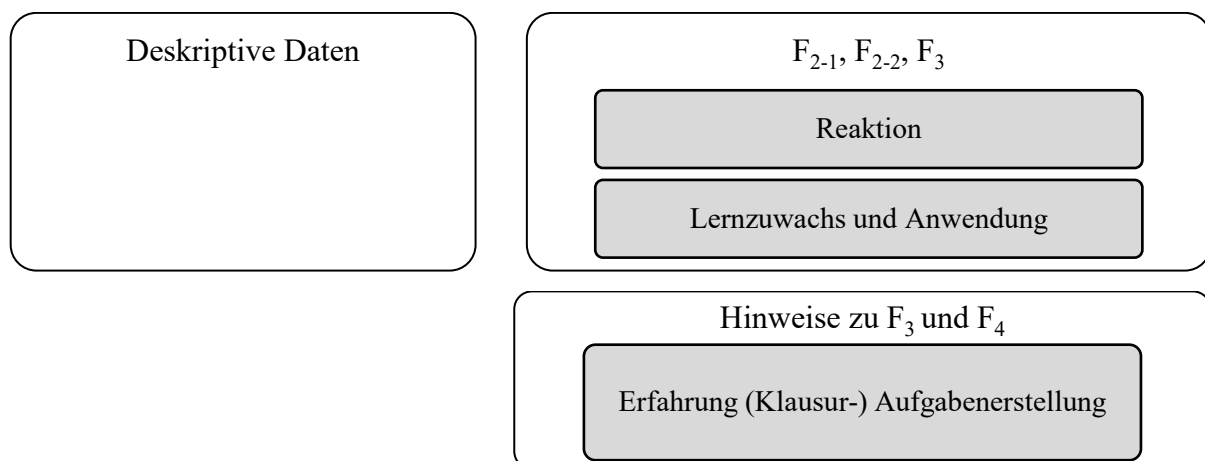
durchgeführt. Dies könnte zu (vermeintlich) sozial erwünschten¹⁶⁰ Aussagen geführt haben. Um dem entgegenzuwirken, wurde versucht, eine offene Stimmung herzustellen (Misoch, 2019, S. 68) und stets betont, dass auch kritische Rückmeldungen zu Evaluationszwecken wertvoll sind.

7.4.2 Auswertung der qualitativen Daten zur Trainingsevaluation

Die Daten der Interviewstudie wurden mittels der zusammenfassenden qualitativen Inhaltsanalyse nach Mayring (2016, S. 96) in MAXQDA ausgewertet.¹⁶¹ Diese Methode ist sinnvoll, da sie „um Praxisverbesserung bemüht [ist]“ (Gläser-Zikuda, 2013, S. 137). So konnte die (fehlende) Wirksamkeit des Trainings erklärt und Implikationen zur Optimierung abgeleitet werden. Als deduktives Element, das *Selektionskriterium* für die Bildung der Kategorien (Mayring, 2016, S. 115-116), wurde insbesondere auf die Stufen der Evaluation nach Kirkpatrick und Kirkpatrick (2006) zurückgegriffen. Der Fokus der Evaluation wurde auf die erste und zweite Stufe gelegt, zu denen u. a. Subkategorien zu den Trainingsbausteinen gebildet wurden. Abbildung 15 dient der Zuordnung der Kategorien nach Forschungsfragen, die im Rahmen der Studie u. a.¹⁶² beantwortet werden sollten.

Abbildung 15:

Hauptkategorien zur Trainingsevaluation nach Forschungsfragen



(Eigene Darstellung)

¹⁶⁰ Um den Fokus der Arbeit zu wahren, wird an dieser Stelle auf das Begriffsverständnis von sozialer Erwünschtheit, deren Gründe und Auswirkungen und weitere Aspekte nicht eingegangen. An dieser Stelle sei auf das Werk von Wolter (2012), insb. S. 37-45, verwiesen.

¹⁶¹ Da auch bei der Interviewstudie zur Nachfrageperspektive zu den Ausbildungsinhalten die Auswertung nach der zusammenfassenden qualitativen Inhaltsanalyse nach Mayring (2016) geschah, sind Wiederholungen zum Verfahren an dieser Stelle kaum vermeidbar. Auf einen bloßen Verweis auf Kapitel 4.3.2 wird verzichtet, um die Transparenz zur Studie zu gewährleisten.

¹⁶² Die Ergänzung „u. a.“ deshalb, weil in der Interviewstudie auch nach den Ausbildungsinhalten aus der Nachfrageperspektive (vgl. Kapitel 4.3) gefragt wurde.

Zusätzlich wurden – zugunsten der Offenheit – induktiv Kategorien gebildet (vgl. Gläser-Zikuda, 2013, S. 137; Mayring, 2016, S. 115-117). Dafür wurde das Material kleinschrittig durchgegangen, Kategorien aus dem Material abgeleitet und diese stellenweise zusammengefasst. Zu den Kodierregeln wird auf die bereits aufgeführten Regeln 1, 2, 3 und 5 in Kapitel 4.3.2 verwiesen.

Sechs Transkripte (fünf von Interviews mit Studierenden und eines vom Interview mit einem*r LiV) wurden zunächst doppelt kodiert. Dies entspricht 37,5 % des gesamten Materials und liegt im gesetzten Rahmen von Mayring (2016, S. 117). An notwendigen Stellen wurden die Kodierregeln genauer formuliert und Kategorien besprochen, was zur Validität dieser beitrug. Anschließend wurden die Kategorien revidiert und das Vorgehen wie folgt begründet:

- Bei den Feedbacks zu den einzelnen Interventionsbausteinen wurden die Bausteine zur Mediendidaktik und technischen Gestaltung zusammengefasst betrachtet. Dies liegt darin begründet, dass der Fokus der vorliegenden Arbeit auf die Gestaltung problemhaltiger Aufgaben zu diagnostischen Zwecken liegt, und diese beiden Bausteine als Mittel dafür genutzt werden.
- Positives Feedback zu einzelnen Bausteinen sowie zum Aufbau der Intervention wurde zusammenfassend betrachtet und in einer Kategorie abgebildet. Negative Punkte und Verbesserungsvorschläge wurden hingegen differenziert aufgeführt, um den Fokus auf das Optimierungs-/Verbesserungspotenzial zu legen. Ein weiterer Grund dafür ist, dass bei einer differenzierten Betrachtung positiver Aspekte häufig Überschneidungen zu finden waren, was die Trennschärfe der Kategorien gefährdete.
- Im Kodierleitfaden wurde zunächst nach *Lernzuwachs* und *Nutzen* differenziert. Diese schließen sich jedoch nicht gegenseitig aus, da Nutzen auch den Lernerfolg einschließt (vgl. Schlicht, 2012, S. 13). Die Schwierigkeit der Differenzierung war auch im Kodierprozess festzustellen, weshalb die Unterscheidung aufgehoben wurde. Stattdessen wurde der Nutzen zur Kategorie des Lernzuwachses zugeordnet. Es wurde differenziert zwischen *Nutzen/Lernzuwachs durch einzelne Bausteine* und *Kein Nutzen/Lernzuwachs durch einzelne Bausteine* sowie *Anwendbarkeit/Transfer in Praxis* und *Schwierigkeiten und Hilfe in Anwendbarkeit/Transfer in Praxis*. Die Anwendung/der Transfer sind zwar nicht unabhängig von Nutzen und Lernzuwachs, wurden aber in den Interviews explizit zum Ausdruck gebracht, weshalb eine eigene Subkategorie hierzu sinnvoll war. Die ge-

meinsame übergeordnete Kategorie wurde *Lernzuwachs durch Intervention und Anwendung* genannt. Auch wurden hier die Bausteine zur Mediendidaktik und zur technischen Umsetzung zusammengefasst.

- Studentische Proband*innen, die keine Erfahrung in der Erstellung von Klausuren (als summatives Assessment) hatten, wurden gefragt, wie sie eine Klausur erstellen würden. Da diese Frage nach der Interventionsstudie gestellt wurde, wurden die Rückmeldungen zu den Subkategorien *Anwendbarkeit/Transfer in Praxis* bzw. *Schwierigkeiten und Hilfe in Anwendbarkeit Transfer in Praxis* kodiert. Dieses Vorgehen wurde gewählt, da die Teilnehmenden ihr Vorgehen mit der Interventionsteilnahme erklärten.

Der finale Kodierleitfaden (Anhang G) besteht aus nachfolgenden Hauptkategorien:

- *D – Deskriptive Daten*
- *E – Erfahrung (Klausur-)Aufgabenerstellung*
- *R – Reaktion zur Intervention*
- *LA – Lernzuwachs durch Intervention und Anwendung*

In Anhang F sind die Zuordnung der Leitfragen des Interviewleitfadens zu den einzelnen Forschungsfragen und Kategorien zu finden.

Der entstandene Kodierleitfaden wurde anschließend dafür genutzt, die bereits kodierten Dokumente erneut zu sichten und alle relevanten Passagen zu berücksichtigen, die vorher möglicherweise untergegangen sind. Im Anschluss an die Doppelkodierung wurde die Reliabilität der (Sub-)Kategorien ermittelt. In Tabelle 26 werden die mittels MAXQDA errechneten Kappa-Werte (κ) der Kategorien dargestellt. Die Werte zu allen Kategorien der unteren Ebenen sind im Leitfaden (vgl. Anhang G) zu finden.

Tabelle 26:*Kappa-Werte (κ) zu den Kategorien des Interviews zur Trainingsevaluation*

Kür- zel	(Sub-)Kategorien	κ
D	Siehe Kapitel 4.3.2	.83
E	Siehe Kapitel 4.3.2	.75
R	R1 – Reaktion zur Gestaltung einzelner Bausteine	.80
	R2 – Reaktion zum Aufbau der Intervention	.74
	R3 – Reaktion zum Material der Intervention	.64
	R4 – Reaktion zur Instruktion/Hilfestellung	.73
LA	L1 - Nutzen/Lernzuwachs durch einzelne Bausteine	.83
	L2- Kein Nutzen/Lernzuwachs durch einzelne Bausteine	.88
	A1 – Anwendbarkeit/Transfer in Praxis	.71
	A2 – Schwierigkeiten und Hilfe in Anwendbarkeit/Transfer in Praxis	.84

(Eigene Darstellung)

Alle Werte lagen im mindestens *substanziellen* bis *nahezu perfekten* Bereich¹⁶³ (vgl. Landis & Koch, 1977, S. 165). Das verbleibende Material wurde dann unter Verwendung des finalen Kodierleitfadens kodiert.

¹⁶³ Die Bezeichnungen der Bereiche sind auch hier aus Landis und Koch (1977, S. 165) und wurden aus dem Englischen ins Deutsche übersetzt.

8 Ergebnisse, Limitationen und Implikationen der Trainingsstudie mit Studierenden der Wirtschaftspädagogik und LiV an kaufmännisch-beruflichen Schulen

8.1 Ergebnisse der quasi-experimentellen Studie und der Interviewstudie

8.1.1 Vorwort zur Ergebnisdarstellung

Im Nachfolgenden werden die Ergebnisse der Studien zur Trainingsevaluation vorgestellt. Dabei werden die Ergebnisse der quantitativen Daten mit den qualitativen verknüpft, um einen tieferen Einblick in die Ergebnisse zu erlangen. Die Darstellung erfolgt nach der ersten und zweiten Evaluationsstufe nach Kirkpatrick und Kirkpatrick (2006) – entsprechend der Auswertung der quantitativen und qualitativen Daten (vgl. Kapitel 7.3.3 und 7.4.2). Mittels der Evaluationsergebnissen sollen primär Optimierungspotenziale des Trainings aufgezeigt werden (vgl. Kirkpatrick & Kirkpatrick, 2006, S. 19).

Zur ersten Evaluationsebene werden zunächst die quantitativen Ergebnisse deskriptiv vorgestellt. Anschließend werden die Ergebnisse der Interviewstudie mit Teilnehmenden konkret in Bezug auf die einzelnen Bausteine präsentiert. Auch zur zweiten Evaluationsebene werden die quantitativen Auswertungen um die Ergebnisse der qualitativen Studie ergänzt. So wird neben den objektiven Ergebnissen über den Lernzuwachs auch ein Einblick in die subjektive Wahrnehmung der Teilnehmenden gewährt.

8.1.2 Ergebnisse der Trainingsevaluation auf Ebene 1

Alle Trainingsteilnehmenden sind sowohl mit der Qualität der Instruktion als auch mit dem Training insgesamt zufrieden bis sehr zufrieden. Über die genannten Skalen hinweg zeigen Studierende eine höhere Zufriedenheit als LiV. Tabelle 27 stellt die Ergebnisse bezogen auf die Skalen dar, die Limitationen finden sich in Kapitel 8.1.4.

Tabelle 27:*Deskriptive Ergebnisdarstellung zu Ebene 1*

	Skala	N	M	SD
Master- studie- rende	Instruktionsqualität	58	3.45	0.50
	Zufriedenheit mit dem Training	58	3.41	0.54
LiV	Instruktionsqualität	5	3.31	0.50
	Zufriedenheit mit dem Training	5	3.30	0.57

Anmerkung zur Skala: 1 = „trifft gar nicht zu“, 2 = „trifft eher nicht zu“, 3 = „trifft eher zu“, 4 = „trifft voll zu“
(Eigene Darstellung)

Die Ergebnisse bestätigen die aus den qualitativen Daten gewonnenen Ergebnisse. Mit 22 kodierten Passagen von elf Teilnehmenden finden sich viele positive Rückmeldungen, die jedoch keinem speziellen Baustein zugeordnet werden konnten. Die Teilnehmenden sagten beispielsweise Folgendes:

„Ich fand einfach das Modul toll ... Also ich hoffe, dass Sie auch weiterhin so machen werden ... es war sehr bereichernd ... und ich hoffe ... für nachkommenden Studenten, dass die auch so ein Modul belegen dürfen.“ (Sarah¹⁶⁴, Pos. 143)

„Ich fand, das war bisher ... die interessanteste Master-Veranstaltung, die ich in WiPäd hatte.“ (Saika, Pos. 147)

„Aber wie gesagt, die gesamte Lehrveranstaltung hat mir auf jeden Fall gut gefallen und hat mir persönlich viel für die Praxis gebracht und interessante Einblicke gegeben.“ (Steven, Pos. 182)

Um diese Rückmeldungen einzubeziehen, wurde die Subkategorie *R1.7.1 – Positives* gebildet. Eine Subkategorie für kritische Aussagen wurde nicht gebildet, da keine passende Passage im Material gefunden wurde. Dies wird als Indiz dafür betrachtet, dass die Teilnehmenden insgesamt (mindestens) zufrieden mit der Veranstaltung waren. Wenn sie Kritik äußerten, konnten sie diese auf einen konkreten Baustein beziehen. Im Folgenden werden die Ergebnisse für die Bausteine aufgezeigt, wobei die Bausteine zur Mediendidaktik- und ILIAS-Schulung in einem betrachtet werden.

¹⁶⁴ Wie in Kapitel 4.3.1 beschrieben, wurde das Interviewmaterial nach Vornamen beginnend mit „R“ für LiV und „S“ für Studierende des Masterstudiengangs Wirtschaftspädagogik benannt. Diese sind Pseudonyme.

Anzahlsmäßig finden sich über die Transkripte hinweg mehr Kodierungen zu positiven Aspekten im Baustein zum Erstellen problemhaltiger Aufgaben (16 Kodierungen, 11 Proband*innen, davon zwei LiV) als zu dem, was kritisiert bzw. als Verbesserungsvorschlag formuliert wurde (drei Kodierungen, drei Proband*innen). Der Baustein wurde von elf Teilnehmenden, darunter zwei LiV, als „gut“ (z. B. Raphael, Pos. 81; Sina, Pos. 124) bis „sehr gut“ (z. B. Saika, Pos. 127; Steven, Pos. 128) – oder mit ähnlichen Worten – bewertet. Es wurde weiter aufgeführt, dass etwa das zum Baustein eingesetzte „*Raster ... anschaulich ... und hilfreich*“ (Sina, Pos. 124)¹⁶⁵ und praxisorientiert war (Sabrina, Pos. 98). Dies geschah etwa dadurch, dass Teilnehmende eigenständig „*Aufgaben ... erstellt[en]*“ und „*sich erstmal langsam [an das Erstellen problemhaltiger Aufgaben] rantasten konnte[n]*“ (Sina, Pos. 140). Des Weiteren wurden die stattfindenden „*Gruppenarbeitsphase[n]*“ (z. B. Saika, Pos. 139; Samuel, Pos. 123) positiv hervorgehoben.

Kritikpunkt zum Baustein zu problemhaltigen Aufgaben gab es im Hinblick auf den Anteil der Anwendungsphasen (im Team) und des Austausches, beides wurde mehr gewünscht (fünf Teilnehmende, davon ein LiV). Hierzu sagte Samuel (Pos. 157), dass er es besser fände „*noch mehr selbst ... diese Beispielaufgaben zu machen und mehr Gruppenarbeiten und mehr, ja, das Umformulieren und wie man es problemhaltiger machen kann*“. Ricardo betonte die Bedeutung des Arbeitens in der Gruppe und begründete dies damit, dass das „*ein neues Feld ... war, weil es ... noch nicht in der Uni behandelt wurde, im Studienseminar behandelt wurde, in der Schule auch nicht wirklich behandelt wurde*“ (Pos. 105).

Auch zum Baustein zur Mediendidaktik und zum technologiebasierten Arbeiten finden sich mehr positive Rückmeldungen als Verbesserungsvorschläge, auch wenn quantitativ betrachtet die Relation im Vergleich zum vorherigen Baustein höher liegt (13 Teilnehmende mit insgesamt 15 Kodierungen vs. acht Teilnehmende mit insgesamt 12 Kodierungen). Konkret wurde hierzu der „*spannende Inhalt*“ (Raphael, Pos. 81) positiv hervorgehoben, der als „*richtig gut*“ bezeichnet wurde (Ricardo, Pos. 85). Durch diesen entstand ein Bewusstsein dafür, „*wie man auf was ... medial achten sollte*“ (Sarah, Pos. 117). Des Weiteren wurde positiv bewertet, dass bei der Schulung zur Technik „*der Bildschirm geteilt wurde, man Schritt für Schritt gesehen hat, wie das funktioniert*“ (Sabrina, Pos. 118). Kritik zu den Bausteinen zur Mediendidaktik und Schulung in ILIAS fand sich u. a. im Hinblick auf die fehlende Kompatibilität der Inhalte. Folgende Passage stellt dies beispielhaft dar:

¹⁶⁵ Zum Raster problemhaltiger Aufgaben kam vielfach Rückmeldung, auch dazu, wie es künftig Einsatz finden soll. Dazu wird im Verlauf der Arbeit eingegangen.

„Ich [war] so ein bisschen überlastet von diesen Informationen, aber auch, weil es dann um so Kleinigkeiten ja viel ging Und manchmal kann man das im ILIAS ja nicht so umsetzen ... das fand ich ein bisschen schade. das habe ich dann quasi in ILIAS auch ein bisschen unter den Tisch fallen lassen einfach.“ (Saika, Pos. 131)

Des Weiteren wurde zum mediendidaktischen Teil der Veranstaltung rückgemeldet, dass der Vortrag kürzer hätte gefasst werden können (Sabrina, Pos. 118). Sowohl zur Mediendidaktik als auch zur Technischulung wurde der Wunsch geäußert, mehr und ggf. längere Arbeits- und Übungsphasen zu haben (z. B. Saika, Pos. 127; Seila, Pos. 84; Raphael, Pos. 81). Die Technischulung wurde auch als *„ein bisschen viel auf einmal“* (Samuel, Pos. 157) bezeichnet.

Zur Selbstlernphase der Veranstaltung kamen überwiegend positive Rückmeldungen (13 Teilnehmende, 15 Kodierungen). Konkretisiert wurde dies allerdings kaum, die Aussagen hierzu waren *„in Ordnung“* (Sina, Pos. 124), *„gut“* (z. B. Samuel, Pos. 157; Steven, Pos. 128) bzw. *„sehr gut“* (z. B. Seila, Pos. 115; Sabrina, Pos. 98). Konkrete Rückmeldungen bezogen sich auf die Möglichkeit, das theoretisch Gelernte eigenständig anzuwenden, wodurch *„man ... sich einfach noch mal ganz anders Gedanken über Themen“* machen konnte (Shirin, Pos. 107). Eine Teilnehmerin beschrieb ihre Wahrnehmung wie folgt:

„Die Selbstlernphase war natürlich auch sehr spannend, weil man sich eben diese Erstellung der Aufgabe komplett selbst erarbeitet hat und man dann vielleicht auch noch mal vor anderen Problemen steht Und allgemein dieser Prozess, ... von dem Rohzustand der Aufgabe bis zur finalen Aufgabe, ... war sowieso sehr spannend und auch nochmal ein bisschen anders, als wenn ich jetzt Prüfungsaufgaben für eine der Klasse erstelle, weil ich halt wusste, das ist zu Übungszwecken, also da habe ich dann selber, ich persönlich, nicht so Leistungsdruck; das muss jetzt auf jeden Fall funktionieren, sondern man konnte einfach sich mal ausleben sozusagen.“ (Rebecca, Pos. 91)

Vier Teilnehmende (sechs Kodierungen) haben Verbesserungsvorschläge bzw. Kritikpunkte genannt. Diese beziehen sich auf die Umsetzungsschwierigkeiten in ILIAS bzw. die technischen Probleme (Raphael, Pos. 83; Ricardo, Pos 73, 83). Des Weiteren kam von einem Teilnehmer die Rückmeldung, dass er die Selbstlernphase in Gruppen besser finden würde als in Einzelarbeit (Ricardo, Pos. 89). So könnten die Teilnehmenden *„in Austausch kommen mit verschiedenen Personen, die verschiedene Meinungen haben und verschiedene Ansätze, verschiedene Hintergründe“* (Ricardo, Pos. 89). Dies würde allerdings zu Lasten der Möglichkeit gehen, individuell zu arbeiten. Andere Teilnehmende äußerten nicht den Wunsch, die Selbstlernphase in Gruppen stattfinden zu lassen.

Aus studentischer Perspektive kam von Steven (Pos. 132) der Verbesserungsvorschlag, „*mehr Beispiele*“ in Form von Aufgaben für den speziellen Beruf zur Verfügung zu stellen, woran „*man sich ... orientieren kann, wie man sowas grundsätzlich machen könnte*“. Dies sei gerade dann wichtig, wenn der Beruf von den Lehrkräften selbst nicht erlernt wurde und diese über keine Unterlagen dafür verfügt. Bei der Gruppe der LiV ist davon auszugehen, dass diese selbst Unterlagen erstellt bzw. Erfahrung mit verschiedenen Berufen haben, da sie in diesen bereits unterrichten. Auch haben sie (eher) Zugang zu Unterlagen aus der Praxis über die Schule. Allen Teilnehmenden wurden zwar Prüfungskataloge, Rahmenlehrpläne und Ausbildungsordnungen zur Verfügung gestellt. Sina (Pos. 124) schlug allerdings vor, weitere Informationsmaterialien in Form von Lehrbüchern für die Themen der Berufe zur Verfügung zu stellen, revidierte dies aber mit den Worten, „*dass das eigentlich auch nicht geht*“. Sie ergänzte: „*das übersteigt meiner Meinung nach auch den Rahmen. Dann hätten Sie da 30 verschiedene Kapitel im Buch, müssten Sie hochladen*“. Das Verfügbarmachen von Lehrbüchern wäre mit einem Aufwand für die kritische Prüfung der Inhalte seitens der Teilnehmenden verbunden. Die Bereitstellung seitens der Dozierenden könnte implizieren, dass die Materialien geprüft sind, auch wenn dies anders (ausdrücklich) kommuniziert worden wäre. Zudem gibt es im Internet Austauschplattformen und andere Homepages, auf die zurückgegriffen werden kann, um an Informationen für das jeweilige Thema des jeweiligen Berufs zu gelangen.¹⁶⁶

Dass den Teilnehmenden Zeit zum Feedbackgeben und zur Reflexion der erstellten Aufgabe eingeräumt wurde, wurde insgesamt als positiv (acht Kodierungen, sechs Teilnehmende) und als (sehr) hilfreich und/oder gelungen bewertet (z. B. Rebecca, Pos. 91; Samuel, Pos. 157; Sarah, Pos. 121). Es kam die Rückmeldung, dass es „*bisher in keinem Modul so erlebt [wurde], dass so intensiv [auf das Feedback] eingegangen worden ist*“ (Sarah, Pos. 121). Negative Passagen wurden zu diesem Baustein nicht gefunden.

Auch im Baustein zum Transfer konnten ausschließlich positiv bewertete Passagen gefunden werden (sieben Kodierungen, sechs Teilnehmende). Rebecca (Pos. 91) bezeichnete den Baustein als „*sehr spannend, weil das auch für ... Berufsalltag ... gezeigt hat, wie viel man aus einer Aufgabe auch für andere Bereiche machen kann*“. Auch aus studentischer Perspektive kam die Rückmeldung, der Baustein sei praxisrelevant,

¹⁶⁶ Dies wurde im Rahmen des Projekts nicht ausdrücklich systematisch geprüft, sondern stellt eine Vermutung dar. Da die Internetwelt sich sehr dynamisch gestaltet, wäre hierzu auch keine finale Aussage möglich und müsste regelmäßig auf Aktualität geprüft werden. Gerade vor dem Hintergrund, dass das Training in mehreren Runden zu unterschiedlichen Zeiten stattfand, würde der Aufwand den Rahmen der Arbeit sprengen.

„... weil das ist ja einfach ... in der Realität so [ist], dass man unterschiedlichste Klassen ... unterrichten wird. ... Und ... wenn man weiß, dass man Dinge transferieren kann und dem war ich tatsächlich auch nicht immer so bewusst gewesen, dass man da auch schnell und gut agieren kann, ja, wenn man da immer wieder diese Querverbindungen auch selbst herstellt, also das fande ich tatsächlich sehr gut.“ (Shirin, Pos. 107)

Der Optimierungsvorschlag, in den Übungsphasen die zu transferierende Aufgabe in Gruppen zu besprechen, um den Transfer zu erleichtern, wurde von Sabrina (Pos. 98, 110) gegeben.

Im Baustein Diagnostik wurden hauptsächlich Gütekriterien besprochen und angewendet. Insgesamt wurden hierzu neun positive Passagen von insgesamt acht Teilnehmenden kodiert. Insbesondere wurde die Anwendung des theoretisch Gelernten hervorgehoben (Sarah, Pos. 121; Samuel, Pos. 139; Shirin, 107). Auch hier wurde ein Vergleich zu anderen Veranstaltungen gezogen:

„Diagnostik und insbesondere Gütekriterien, die waren beide tatsächlich für mich auch ganz wichtig, fande ich ... auch sehr wichtig und sehr sehr gut, dass Sie wirklich noch einmal zurückgegangen sind, auch wenn Sie vielleicht sagen können, es müsste eigentlich jeder gehört haben oder in anderen Veranstaltungen vielleicht auch im Bachelor schonmal damit vertraut gewesen zu sein. Aber für mich war auch da die Vorstellung so wie Sie es gemacht haben, viel praxisnah als ich das in der Vergangenheit erlebt habe. Weil wir haben ja wirklich viel mit Beispielen gearbeitet und auch mal das auf unsere Themen noch mal übertragen, ... das hat mir wirklich auch viel weitergeholfen. Ich hatte ... in der Vergangenheit das Gefühl, dass ich diese Kriterien weniger verstanden habe, als dass ich sie jetzt mit einer wirklichen praktischen Anwendung dann auch verstanden habe.“ (Shirin, Pos. 107)

Die fehlende Praxis scheint sich aber nicht nur auf (ausgewählte) Module im Studium zu beziehen. So berichtete Rebecca (Pos. 93), dass sie zwar den „Teil Diagnostik und Gütekriterien ... auch noch mal interessant [fand]“ und ergänzte ihren Eindruck, dass die Inhalte „natürlich sehr wichtig für die Erstellung von Prüfungsaufgaben [sind]. Aber in der Praxis beschäftigt man sich ... zu selten damit“¹⁶⁷. Gleichzeitig finden sich im Interviewmaterial sieben Passagen von sieben Teilnehmenden, die sich mehr Inhalt bzw. Übungen hierzu gewünscht hätten. So kam der Vorschlag, das Gelernte an konkreten Praxisbeispielen anzuwenden (Sami, Pos. 132; Svana, Pos. 91). In weiteren sechs Passagen von insgesamt fünf Teilnehmenden fand sich Kritik

¹⁶⁷ Auch die Befunde zu den Ausbildungsinhalten aus der Nachfrageperspektive (vgl. Kapitel 5.2) zeigen, dass angehenden Lehrkräften der Praxisanteil zu kurz kam.

dahingehend, dass der Baustein als komplex und/oder anspruchsvoll (z. B. Samira, Pos. 153) und/oder trocken (Raphael, Pos. 83) empfunden wurde.

Auf den Baustein zur Diagnostik bezogen sich auch die Umpositionierungsvorschläge von drei Teilnehmenden, die dies damit begründeten, dass dieser „*spät am Tag gewesen*“ sei (Raphael, Pos. 83). Einen weiteren Aspekt für die frühere Behandlung des Themas betraf die Priorisierung dessen.

„Also vielleicht hätte ich dann jetzt auch irgendwie gedacht, ... das ist ein wichtiger Aspekt, den kann ich gar nicht so unterfallen lassen. Aber dadurch, dass es so am Ende war, war es halt für mich so, okay, das muss jetzt auch noch irgendwie machen so ... meine Empfehlung [wäre], das vielleicht vorzuziehen.“ (Seila, Pos. 139)

Ergänzend wurde der Wunsch geäußert, Zeit und Raum für die Anwendung der erstellten Aufgaben (Sarah, Pos. 130), der Besprechung von Alltagsbeispielen (Ricardo, Pos. 95, 99) und die Betrachtung weiterer Ausbildungsberufe (Sven, Pos. 168) zu bekommen. Auch wurde vorgeschlagen, den Schwerpunkt der Veranstaltung zu erweitern, sodass auch Aufgaben für den Unterricht entwickelt werden (Saika, Pos. 143). Die häufigste Kritik (acht Teilnehmende, 15 Kodierungen) in Bezug auf den Aufbau der Veranstaltung bezog sich auf das Blockformat. Stattdessen wurden „*wöchentlich[e] und dafür etwas kürzer[e]*“ (Samira, Pos. 168; ähnlich auch von z. B. Seila, Pos. 112; Sven, Pos. 171) Sitzungen bzw. die Reduzierung des Umfangs einzelner Bausteine (Samuel, Pos. 155) vorgeschlagen. Trotz der Kritik und der Verbesserungsvorschläge, waren die meisten Teilnehmenden zufrieden mit dem Interventionsaufbau; so gaben zwölf Teilnehmende (29 Kodierungen) hierzu positive Rückmeldung. Hieraus ergibt sich insbesondere, dass der Wechsel zwischen Input- und Anwendungsphasen gelungen (z. B. Rebecca, Pos. 138; Samira, Pos. 157), „*der rote Faden*“ zu erkennen (Shirin, Pos. 111; ähnliche Aussage auch von z. B. Sina, Pos. 136; Sven, Pos. 152) und der „*Aufbau ... genau richtig [war]*“ (Samira; Pos. 157; ähnliche Aussage auch von z. B. Sina, Pos. 152). Auch wurden die Positionierung und zeitliche Planung zur Selbstlernphase als angemessen empfunden, was die Aussage von Sami (Pos. 128), sich auf die Selbstlernphase „*sehr gut vorbereitet*“ gefühlt zu haben, bestätigt.

Über die in der Intervention verwendeten Materialien finden sich von zwölf Interviewteilnehmenden Rückmeldungen. Positive Aussagen finden sich meist in Bezug auf das Raster zu den Merkmalen problemhaltiger Aufgaben (z. B. Raphael, Pos. 85; Samira, Pos. 71; zum Raster vgl. Anhang E). Es wurden Aussagen getroffen, dass Teilnehmende diese auch für die aktuelle berufliche Praxis genutzt haben bzw. vermutlich künftig nutzen werden (z. B. Rebecca, Pos.

99; Seila, Pos. 35). Auf Letzteres wird bei der Ergebnisdarstellung zur zweiten Evaluations-ebene eingegangen (vgl. Kapitel 8.1.3). Weiterhin wurde der individuelle Feedbackbogen, der den Teilnehmenden online zur Verfügung gestellt wurde, positiv hervorgehoben (z. B. Saika, Pos. 127). Eine Studentin kommentierte wie folgt:

„Und Feedbackbogen in ILIAS ... ich ... fand diesen Feedbackbogen mega ... natürlich ... habe ich mir Notizen gemacht, als nach der Präsentation Feedback gegeben worden ist. Aber es noch mal so in grob zu sehen, ... dieses ... Raster, ... finde ich eine sehr gute Idee und auch so fürs Nachhinein. ... ich werde es mir auf jeden Fall abspeichern, um auch ... in einem halben Jahr oder Jahr und in zwei Jahren noch mal darüber nachzudenken, was hatte ich in meiner Problemlöseaufgabe denn für Punkte, die ich nicht, ja, so nicht intensiv genug oder anders bedacht habe. Und was könnte ich noch mal anders machen? Also so als Reflektion nochmal finde ich sehr gut.“ (Sarah, Pos. 121)

Die Teilnehmenden berichteten, dass sie die zur Verfügung gestellten Materialien (u. a. Präsentationsfolien und Handbücher) für die Vor- bzw. Nachbereitung genutzt haben (z. B. Samuel, Pos. 133; Sibel, Pos. 120; Shirin, Pos. 107). Die Teilnehmenden gaben an, dass die dauerhafte Bereitstellung dieser es ihnen ermöglichte, die Materialien dann zu nutzen, *„wenn man es ... wirklich braucht“* (Svana, Pos. 107).

Gelobt wurde von zwölf Teilnehmenden (23 Kodierungen) die Instruktion und Hilfestellung seitens der Dozierenden. Die Instruktion sowie die Hilfestellung, etwa während der Begleitung in der Selbstlernphase, wurden als *gelingen* (z. B. Raphael, Pos. 83; Saika, Pos. 127; Sibel, Pos. 147) bzw. *hilfreich* (z. B. Sarah, Pos. 101, Samira, Pos. 126; Seila, Pos. 80) bewertet. Besonders wurde von Teilnehmenden das Feedback zur erstellten Aufgabe in der Selbstlernphase positiv hervorgehoben. So sagte etwa Sabrina:

„Also ich finde es immer schön, wenn man wirklich Aufgaben hat, die man selber bearbeitet und wo man sich dann ausprobieren kann und ohne dass man dann im Prinzip sich erst in einer Klausur ausprobiert oder einer Hausarbeit, sondern dass man vorher schon mal irgendwo ein Feedback bekommt und darüber spricht, wie man das am besten umsetzt. Das fand ich immer schön.“ (Sabrina, Pos. 118)

Dass die Veranstaltung von mehreren Dozierenden begleitet wurde, wurde als *„sehr sehr gut“* (Shirin, Pos. 107) empfunden und hervorgehoben, dass der*die jeweilige Dozent*in *„[thematisch] extrem tief drin steckt ... eingebunden [wurde]“* (Shirin, Pos. 107).

8.1.3 Ergebnisse der Trainingsevaluation auf Ebene 2

Die Interviews zeigen, dass die Teilnehmenden durch die Veranstaltung für die Thematik der Erstellung von problemhaltigen Aufgaben zu summativen Diagnostikzwecken sensibilisiert wurden. Sibel (Pos. 128) sagte, sie könne das Gelernte aus der Veranstaltung „auf jeden Fall anwenden“ und habe nun ein „Bewusstsein dafür, was auf [sie] ... zukommen wird“. Die Anwendbarkeit des Gelernten¹⁶⁸ aus der Veranstaltung wurde mit anderen Modulen verglichen und es folgten Aussagen wie beispielsweise:

„Entschuldigung, wenn ich das so platt sage, aber es gibt Module, die belegt man, nimmt die CP mit und macht einen Haken dahinter ... Aber hier weiß ich, dass ich da auch immer wieder reinschauen werde und dass ich das weiter brauchen werde.“ (Sibel, Pos. 128)

„abgesehen von ... unserem Modul jetzt, hatte ich kein Modul, das mich auf die Praxis wirklich vorbereitet hat.“ (Sarah, Pos. 51)

Wie sich konkret die Anwendbarkeit gestaltet und den Teilnehmenden zufolge gestalten kann, wird im Folgenden dargestellt.

Der subjektiv wahrgenommene Lernzuwachs spiegelt sich auch in den Ergebnissen der quantitativen Evaluation wider. So erreichte die Skala zum wahrgenommenen Lernzuwachs bei den Studierenden einen Mittelwert von $M = 3.41$ ($SD = 0.38$) und bei den LiV einen Mittelwert von $M = 3.08$ ($SD = 0.23$). Bei einer vierstufigen Likert-Skala (1 = „trifft gar nicht zu“ bis 4 = „trifft voll zu“) ist der wahrgenommene Lernzuwachs beider Gruppen als hoch bis sehr hoch zu bewerten, wobei die Studierenden den eigenen Lernzuwachs höher wahrnahmen als die LiV. Deskriptiv erreichten die Studierenden in den Wissenstests folgende Werte (vgl. Tabelle 28):

¹⁶⁸ Zwar berichteten Teilnehmende hier von der Anwendbarkeit des Gelernten, da allerdings die Anwendung in der Praxis nicht beobachtet wurde, wurden solche Aussagen der Evaluationsebene 2 zugeordnet.

Tabelle 28:*Punkte im Pre- und Posttest zum Wissen der Studierenden*

Wissenstest zu		Merkmale problemhaltiger Aufgaben				Diagnostischen Gütekriterien			
Studierende	N	Pretest		Posttest		Pretest		Posttest	
		M	SD	M	SD	M	SD	M	SD
IG	58	3.29	1.21	4.29	0.88	5.64	1.81	5.33	1.91
KG	30	2.97	1.27	2.97	1.27	4.73	1.93	4.67	2.16

Maximal erreichbare Punktzahl im Wissenstest zu den Merkmalen problemhaltiger Aufgaben = 5 Punkte

Maximal erreichbare Punktzahl im Wissenstest zu diagnostischen Gütekriterien = 8 Punkte

(Eigene Darstellung)

Die Studierenden der Interventions- und Kontrollgruppe unterschieden sich im Pretest (vor dem Training) nicht signifikant im Wissen zu den Merkmalen problemhaltiger Aufgaben ($M_{AMPre|Stud|IG} = 3.29$, $SD_{AMPre|Stud|IG} = 1.21$; $M_{AMPre|Stud|KG} = 2.97$, $SD_{AMPre|Stud|KG} = 1.27$; $t_{AM|Stud}(86) = 1.176$, $p_{AM|Stud} > .05$). Zwischen den Gruppen gab es einen (wenn auch nicht hohen) signifikanten Unterschied zugunsten der Interventionsgruppe in Bezug auf das Wissen zu den Gütekriterien ($M_{GKPre|Stud|IG} = 5.64$, $SD_{GKPre|Stud|IG} = 1.81$; $M_{GKPre|Stud|KG} = 4.73$, $SD_{GKPre|Stud|KG} = 1.93$; $t_{GK|Stud}(86) = 2.171$, $p_{GK|Stud} = .033$)¹⁶⁹.

Ausgehend davon, und um mittels der quantitativen Daten eine objektive Aussage zum Lernzuwachs zu treffen, wurde zunächst anhand einer Kovarianzanalyse überprüft, ob sich der erreichte Wert in den Posttests beider studentischen Gruppen unterscheidet, wenn um die erreichten Werte in den Pretests als Kovariate kontrolliert wird. Mit Kontrolle der erreichten Punkte, konnte für den Wissenstest zu den Merkmalen problemhaltiger Aufgaben ein signifikanter Unterschied zwischen den Gruppen festgestellt werden: $F(1, 85) = 32.135$, $p < .001$, partial $\eta^2 = .274$. Das partielle η^2 „gibt den Anteil der aufgeklärten Variabilität der Messwerte relativ zur Summe der Quadratsummen des systematischen Effekts und des Residuums auf der Ebene der Stichprobe an“ (Rasch, Friese, Hofmann & Naumann, 2014, S. 50). D. h., dass in diesem Fall ein Anteil von 27,4 % der Variation der erreichten Posttest-Punkte im Wissenstest zu den Merkmalen problemhaltiger Aufgaben, unter Berücksichtigung der erreichten Punkte im Pretest, durch die Interventionsmaßnahme erklärt werden kann. Die Effektstärke der erreichten

¹⁶⁹ Zum Vergleich der Pretest-Werte in den Wissenstests wurde ein t-Test für unabhängige Stichproben durchgeführt. Mittels der G*Power-Analyse wurde eine *notwendige* (von den Autor*innen „Stichprobenumfangsplanung“ genannt; vgl. Rasch, Friese, Hofmann und Naumann, 2021, S. 66) Stichprobengröße von 84 ($d = 0.8$, $\alpha = .05$, Power = 0.95) berechnet. Zwar liegt die Stichprobengröße für die Stichprobe der Studierenden über diesem Wert ($N_{Stud} = 88$), allerdings sind die Gruppen nicht gleich groß ($N_{Stud|IG} = 58$; $N_{Stud|KG} = 30$). Dennoch wurde der t-Test durchgeführt und dies damit begründet, dass in beiden Gruppen die Mindeststichprobengröße von 30 Proband*innen erreicht wurde (vgl. hierzu das zentrale Grenzwerttheorem; Bortz & Schuster, 2010, S. 86-87; Döring & Bortz, 2016, S. 641).

Punkte im Pretest zu den Merkmalen problemhaltiger Aufgaben beträgt $r_{\text{PreAM}|\text{PostAM}} = .403$. Dieser Wert ist mit $p_{\text{PreAM}|\text{PostAM}} < .01$ signifikant. Für das Training zeigt sich ein signifikanter Wert von $r_{\text{Training}|\text{PostAM}} = .524$ ($p_{\text{Training}|\text{PostAM}} < .01$). Die Effektstärken sind nach Cohen (1988, S. 80) als mittel ($r_{\text{PreAM}|\text{PostAM}}$) bzw. stark ($r_{\text{Training}|\text{PostAM}}$) zu interpretieren.

Auch wurde zum Wissenstest zu den diagnostischen Gütekriterien anhand der studentischen Daten eine ANCOVA durchgeführt. Diese zeigte, dass sich die erreichten Punkte im Posttest beider studentischen Gruppen, unter Kontrolle der erreichten Punkte im Pretest zum Wissen über diagnostische Gütekriterien, nicht signifikant unterscheiden: $F(1, 86) = .079$, $p > .05$, partial $\eta^2 = .001$. Das partial η^2 zeigt, dass weniger als 1 % der Variation der erreichten Punkte im Wissenstest zu den Gütekriterien durch das Training erklärt werden kann, wenn für die erreichten Punkte im Pretest des Wissenstests kontrolliert wird. Die signifikante Effektstärke des Pretests zu den Gütekriterien liegt bei $r_{\text{PreGüte}|\text{PostGüte}} = .475$ ($p_{\text{PreGüte}|\text{PostGüte}} < .01$) und ist nach Cohen (1988, S. 80) als mittel zu werten. Die Effektstärke des Trainings liegt bei $r_{\text{Training}|\text{PostGüte}} = .030$ und ist damit unter der Grenze eines schwachen Effekts (vgl. Cohen, 1988, S. 79). Aufgrund des p-Wertes ($p_{\text{Training}|\text{PostGüte}} > .05$) ist auch von keiner Signifikanz auszugehen.

Mittels der ANCOVA konnte für die Gruppe der Studierenden gezeigt werden, dass zum Wissen über Merkmale problemhaltiger Aufgaben, unter Kontrolle des Vorwissens mittels des Pretests, zwischen der Interventions- und Kontrollgruppe zum zweiten Messzeitpunkt (nach der Intervention, im Posttest) ein signifikanter Unterschied vorliegt. Somit wird die Hypothese H1 für die studentischen Teilnehmenden angenommen. In Bezug auf das Wissen zu den diagnostischen Gütekriterien konnte kein solcher Unterschied errechnet werden: Demnach hat das Training in Bezug auf das Wissen zu den diagnostischen Gütekriterien zu keinem signifikanten Lernzuwachs im Vergleich der Gruppen geführt. Die Hypothese H2 ist demnach abzulehnen.

Mithilfe der multiplen linearen Regressionsanalyse wurde untersucht, welchen Beitrag einzelne Prädiktoren auf die Leistung im Posttest haben (vgl. Bühner & Ziegler, 2017, S. 677-678; Döring & Bortz, 2016, S. 626). Die Regressionsanalyse der erreichten Punkte der Studierenden im Wissenstest zu den Merkmalen problemhaltiger Aufgaben lieferte das nachfolgende Ergebnis (vgl. Tabelle 29).

Tabelle 29:

Ergebnis der Regressionsanalyse zur Erklärung der erreichten Punkte im Wissenstest zu den Merkmalen problemhaltiger Aufgaben

	Modell 1			Modell 2			Modell 3		
	B	SE	β	B	SE	β	B	SE	β
(Konstante)	1.735	.390		1.970	.300		2.967	.188	
Intervention (IG = 1; KG = 0)	1.112	.238	.443***	1.217	.215	.482***	1.326	.231	.526***
Punktzahl im Pretest zu Merkmalen problemhaltiger Aufgaben	.339	.090	.351***	.336	.083	.345***			
Fachsemester	.104	.121	.082						
R ²		.400			.394			.277	
korrigiertes R ²		.375			.380			.268	

Anmerkung: *** $p < .001$; Modell 1 mit $N = 76$ (nur Studierende); Modell 2 und Modell 3: $N = 88$ (nur Studierende); das abweichende N in Modell 1 ist auf die Missings in der Variable *Fachsemester* zurückzuführen (Eigene Darstellung)

In Modell 1 wurden die Gruppenzugehörigkeit (Interventions- oder Kontrollgruppe), die erreichten Punkte im Pretest zu den Merkmalen problemhaltiger Aufgaben und das aktuelle Fachsemester als Prädiktoren berücksichtigt. Das Modell erklärt 37,5 % der Varianz (korrigiertes $R^2 = .375$). Das Fachsemester, in welchem sich die Studierenden befinden, hat keinen signifikanten Einfluss auf die erreichten Punkte im Posttest zum genannten Wissenstest. Dieses Ergebnis deckt sich mit dem der Vorstudien (vgl. Kapitel 5), sodass die Hypothese H3a abgelehnt wird. Hingegen haben die Teilnahme an der Intervention und die im Pretest erreichten Punkte einen signifikanten Einfluss auf die Ergebnisse des Wissenstests. Auch in Modell 2, in welchem der Prädiktor Fachsemester unberücksichtigt bleibt, haben die Teilnahme an der Intervention und die im Pretest erreichten Punkte einen signifikanten Einfluss auf die Ergebnisse des Wissenstests. Modell 2 erklärt 38 % der Varianz. In Modell 3 wurde lediglich die Teilnahme an der Intervention als Prädiktor berücksichtigt. Der Einfluss der Teilnahme ist signifikant, gleichzeitig sinkt jedoch die erklärte Varianz auf 26,8 %. Dass die Teilnahme an der Intervention in allen Modellen einen signifikanten Einfluss zeigt, bestätigt das Ergebnis der ANCOVA. Dies bestätigt die Annahme der Hypothese H1. Hingegen zeigen Modell 1 und 2, dass mit höherer Punktzahl im Pretest eine signifikant höhere Punktzahl im Posttest einhergeht. Dieses Ergebnis führt zu dazu, dass die Hypothese H4a abgelehnt wird. Es zeigt sich, dass anhand aller Modelle eine starke Varianzaufklärung möglich ist (vgl. Cohen, 1988, S. 414).

Unter Berücksichtigen der Prädiktoren Gruppenzugehörigkeit, erreichte Punkten im Pretest zum Wissen über diagnostische Gütekriterien und Fachsemester, wurde eine Regressionsanalyse für die im Posttest erreichten Punkte im Wissenstest zu den diagnostischen Gütekriterien als Kriterium durchgeführt. Tabelle 30 bildet die Ergebnisse der drei berechneten Modelle ab.

Tabelle 30:

Ergebnis der Regressionsanalyse zur Erklärung der erreichten Punkte im Wissenstest zu den diagnostischen Gütekriterien

	Modell 1			Modell 2			Modell 3		
	B	SE	β	B	SE	β	B	SE	β
(Konstante)	3.002	.758		2.339	.591		4.667	.365	
Intervention (IG = 1; KG = 0)	.469	.434	.119	.216	.413	.051	.661	.449	.157
Punktzahl im Pretest zu diagnosti- schen Gütekrite- rien	.442	.109	.433***	.492	.104	.463***			
Fachsemester	-.197	.214	-.099						
R ²		.222			.227			.025	
korrigiertes R ²		.189			.209			.013	

Anmerkung: *** $p < .001$; Modell 1 mit $N = 76$ (nur Studierende); Modell 2 und Modell 3: $N = 88$ (nur Studierende); das abweichende N in Modell 1 ist auf die Missings zur Variable *Fachsemester* zurückzuführen (Eigene Darstellung)

Modell 1 erklärt 18,9 % der Varianz und zeigt, dass das Fachsemester keinen signifikanten Einfluss auf die erreichten Punkte im Posttest zum hier betrachteten Wissenstest hat. Somit kann die Hypothese H3b abgelehnt werden. Dass über alle drei Modelle die Teilnahme an der Intervention keinen signifikanten Einfluss hat, bestätigt die Ergebnisse der ANCOVA und führt damit zur Ablehnung der Hypothese H2. In den Modellen 1 und 2 zeigt sich zudem, dass lediglich der Prädiktor zu den erreichten Punkten im Pretest einen signifikanten Einfluss auf die erreichten Punkte im Wissenstest (Posttest) hat. Dieses Ergebnis führt zur Ablehnung der Hypothese H4b. Die größte Varianzaufklärung findet sich in Modell 2 (korrigiertes $R^2 = .209$). Während Modell 3 eine schwache Varianzaufklärung zeigt, ist diese in den Modellen 1 und 2 als moderat einzuschätzen (vgl. Cohen, 1988, S. 413).

Da aufgrund der geringen Stichprobengröße der LiV keine komplexeren Berechnungen durchgeführt werden konnten, werden in Tabelle 31 die Ergebnisse der LiV-Gruppen deskriptiv vorgestellt. Anhand dieser Ergebnisse sollen Tendenzen abgeleitet werden.

Tabelle 31:*Punkte im Pre- und Posttest zum Wissen der LiV*

Wissenstest zu		Merkmale problemhaltiger Aufgaben				Diagnostischen Gütekriterien			
LiV	N	Pretest		Posttest		Pretest		Posttest	
		M	SD	M	SD	M	SD	M	SD
IG	5	3.00	2.35	4.20	0.84	3.60	3.05	2.80	1.30
KG	5	3.00	1.41	3.60	1.67	3.80	2.86	3.80	3.19

Maximal erreichbare Punktzahl im Wissenstest zu den Merkmalen problemhaltiger Aufgaben = 5 Punkte

Maximal erreichbare Punktzahl im Wissenstest zu diagnostischen Gütekriterien = 8 Punkte

(Eigene Darstellung)

Die Betrachtung der Mittelwerte beider Gruppen (Kontroll- und Interventionsgruppe) zu beiden Wissenstests zeigen für LiV kaum einen Unterschied ($M_{AMP_{Pre|LiV|IG}} = 3.00$, $SD_{AMP_{Pre|LiV|IG}} = 2.35$; $M_{AMP_{Pre|LiV|KG}} = 3.00$, $SD_{AMP_{Pre|LiV|KG}} = 1.41$; $M_{GK_{Pre|LiV|IG}} = 3.60$, $SD_{GK_{Pre|LiV|IG}} = 3.05$; $M_{GK_{Pre|LiV|KG}} = 3.80$, $SD_{GK_{Pre|LiV|KG}} = 2.86$)¹⁷⁰. Die Vergleiche der jeweiligen Pre- und Posttestwerte deuten auf einen Lernzuwachs der Teilnehmenden des Trainings in Bezug auf das Wissen zu den Merkmalen problemhaltiger Aufgaben. Da auch die Kontrollgruppe im Posttest einen höheren Wert als im Pretest erzielte, kann Lernzuwachs der Interventionsgruppe nicht allein auf das Training zurückgeführt werden. Dennoch ist zu vermuten, dass das Training, wenn auch nicht vollständig, zum Lernzuwachs in diesem Bereich beigetragen hat. Dies wird zum einen durch den höheren Mittelwert der Interventionsgruppe im Posttest (trotz gleichem Mittelwert im Pretest der Interventions- und Kontrollgruppe) erklärt. Zum anderen damit, dass die zu Beginn vorhandene (relativ große) Streuung ($SD_{IG|PreAM} = 2.35$) im Posttest stark abnimmt ($SD_{IG|PostAM} = 0.84$). Im Wissenstest zu den diagnostischen Gütekriterien zeigt sich eine Wissensabnahme in der Interventionsgruppe. Zwar nimmt die Streuung zwischen Pre- und Posttests ebenfalls ab ($SD_{IG|PreGK} = 3.05$ vs. $SD_{IG|PostGK} = 1.30$), doch deuten die Mittelwerte auf ein Optimierungspotenzial der Intervention hin. Hingegen zeigt die Kontrollgruppe im Wissenstest zu den diagnostischen Gütekriterien einen konstanten Mittelwert mit steigender Streuung ($SD_{KG|PreGK} = 2.86$ vs. $SD_{KG|PostGK} = 3.19$). Dass der Vergleich des Pre- und Posttests zum Wissenstest zu den diagnostischen Gütekriterien ein schlechteres Ergebnis zeigt als der zu Merkmalen problemhaltiger Aufgaben, deckt sich mit den Ergebnissen der studentischen Gruppen. Zusammen mit den Rückmeldungen der Teilnehmenden (siehe Kapitel 8.1.2), wird die Vermutung bestärkt, dass das Training in diesem Bereich methodisch optimierungsbedürftig ist. Diese Tendenzen entsprechen den vorgestellten Ergebnissen zu den studentischen Teilnehmenden,

¹⁷⁰ Die Stichprobengröße der LiV lag deutlich unter der berechneten Stichprobengröße mittels der G*Power-Analyse, weshalb kein t-Test berechnet wurde.

auch wenn dadurch keine eindeutigen Aussagen möglich sind. In Bezug auf die Hypothesen ergeben sich demnach hieraus keine Änderungen.

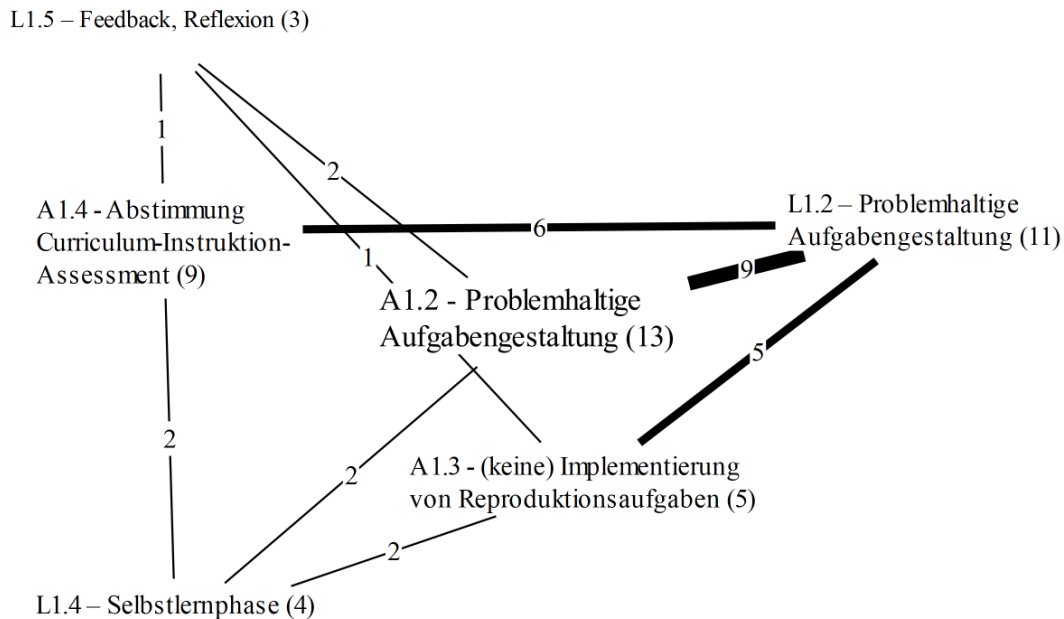
Die Ergebnisse der Interviewstudie zum Training zeigen, dass die Teilnehmenden etwas lernen bzw. daraus für ihre (potentielle) Tätigkeit als Lehrkraft „mitnehmen“ (z. B. Raphael, Pos. 11) konnten. Hierzu kamen von 13 Teilnehmenden (davon zwei LiV) unkonkrete Aussagen, die zunächst keinem Baustein des Trainings zugeordnet werden konnten. Diese Personen teilten mit, dass sie einen Lernzuwachs wahrgenommen haben. Zu welchem Thema konkretisierten sie (erst einmal) nicht, sondern i. d. R. erst durch Rückfragen oder das Einblenden der Tischvorlage (vgl. Anhang F). In der Interviewstudie wurden die Teilnehmenden auch befragt, welches Wissen bzw. welche Kompetenzen aus welchen Bausteinen sie erlangten und in die Praxis umsetzen könnten. Es ist anzumerken, dass die hypothetische Vorgehensbeschreibungen der Teilnehmenden, die bisher keine Klausur erstellt haben, auch in die Kategorie zum Lernzuwachs kodiert wurden. Begründet ist dies damit, dass sie ggf. von der Intervention beeinflusst wurden und/oder sozial erwünscht antworteten. Die kodierten Passagen können daher als weiterführende Hinweise für die Evaluation der Ebene des Lernens verstanden werden. Der Gedanke hierbei ist, möglichst über die Ebene des Wissens hinaus Ansätze zur Anwendbarkeit des Wissens in die Praxis zu gewinnen, um so etwa träges Wissen zu vermeiden (vgl. Renkl, 1994; Whitehead, 1929). Die Teilnehmenden konnten – anders als bei der quantitativen Erhebung mit geschlossenen Fragen – zur potentiellen Anwendbarkeit oder aktuellen Anwendung des Gelernten in der Praxis konkrete Beispiele geben. Dies erleichterte die Zuordnung¹⁷¹. Abbildung 16 zeigt die Relation der Kategorien zum selbst eingeschätzten Lernzuwachs und Anwendung der Inhalte nach Vorkommen in einem Dokument. Die Verbindungen bzw. Liniestärken zeigen, in wie vielen Dokumenten die verbundenen Kategorien gleichzeitig vorkamen. Dies wurde als ein Hinweis dafür verstanden, dass Teilnehmende die gelernten Inhalte aus einem Baustein anwenden wollen bzw. erwarten, diese anwenden zu können. Hierfür wurden die Kategorien zum eingeschätzten Lernzuwachs aus den Bausteinen zur Erstellung von problemhaltigen Aufgaben (L1.2) sowie der Selbstlern- (L1.4) und Reflexionsphase (L1.5) betrachtet und diese in Relation zu den entsprechenden Kategorien in Bezug auf die Anwendung gesetzt (A1.2, A1.3, A1.4). Die in Klammern hinter der Kategorienbezeichnung abgebildete Zahl zeigt, in wie

¹⁷¹ Beispiele zur konkreten Anwendung könnten der Evaluationsstufe 3 zum *Verhalten* (vgl. Kirkpatrick, 2007; Kirkpatrick & Kirkpatrick, 2006; Kapitel 6.5) zugeordnet werden. Hierauf wurde verzichtet, da kein direkter Einblick in das Praxisgeschehen möglich war und auch nicht allen Teilnehmenden in der Zeit während bzw. unmittelbar nach der Intervention die Anwendung in der realen Praxis möglich war, sondern sie von ihrem geplanten Vorgehen berichteten. Unter diesen Bedingungen ist eine solide Evaluation auf der dritten Ebene nicht möglich.

vielen Dokumenten Kodierungen zu der entsprechenden Kategorie enthalten sind. Dies entspricht auch der Anzahl der Proband*innen, die sich zu den jeweiligen Inhalten der Kategorien äußerten.

Abbildung 16:

Code-Relations-Modell (MAXQDA) zum eingeschätzten Lernzuwachs (L) und zur Anwendung (A)



(Eigene Darstellung)

Neun Teilnehmende (davon zwei LiV), die von einem selbsteingeschätzten Lernzuwachs durch den Baustein zum *Erstellen problemhaltiger Aufgaben* berichteten, schätzen auch, dass sie das Gelernte anwenden können. In diesem Zusammenhang wurde insbesondere das zur Verfügung gestellte Raster zu den Merkmalen problemhaltiger Aufgaben positiv erwähnt, analog zur ersten Ebene der Evaluation (vgl. Kapitel 8.1.2). Teilnehmende teilten mit, dass sie das Raster bereits nutzen, es ihnen in anderen Situationen bereits nutzen würde und/oder sie sich vorstellen könnten, dieses künftig zu verwenden: „*Dieses Schema für die problemlösenden Aufgaben fand ich da sehr hilfreich, dass man da einmal so einen Überblick hatte, das hätte mir zum Beispiel persönlich auch extrem in meinem Praktikum geholfen*“ (Steven, Pos. 104), wozu er im Gesprächsverlauf ergänzte, dass er dieses „*auch später im Referendariat mit Sicherheit noch irgendwo in meinen Unterlagen nutzen können*“ (Steven, Pos. 128). Eine Referendarin kommentierte:

„Gerade dieses Raster zur Problemhaltigkeit, was ihr im Training uns auch zur Hand gegeben habt, finde ich sehr hilfreich. Und das habe ich tatsächlich jetzt auch schon zur Überprüfung von meinen eigenen Lernsituationen eingesetzt, um zu gucken ... an welchen Stell-schrauben könnte ich da vielleicht noch drehen. Und das ist, glaube ich so eine Grunderkenntnis, die ich auch aus dem Training mitgenommen habe.“ (Rebecca, Pos. 99)

Sechs Teilnehmende (davon ein LiV), die von einem selbsteingeschätzten Lernzuwachs durch den Baustein zum *Erstellen problemhaltiger Aufgaben* berichteten, schätzen auch, dass sie die Lerninhalte zur Abstimmung von Curriculum, Instruktion und Assessment anwenden konnten/werden. Dies ist im Sinne der Triade von Pellegrino (2010). Fünf Teilnehmende (davon ein LiV) gaben an, dass sie sich durch diesen Baustein nun bewusst für oder gegen die Implementierung von Reproduktionsaufgaben entscheiden würden. Dieses Thema wurde im genannten Baustein behandelt, da im Kontext von Prüfungsaufgaben die kritische Reflexion über das Implementieren wenig problemhaltiger Aufgaben bzw. Wissensabfragen nicht irrelevant ist. Prüfungsaufgaben sollen ein breites Spektrum an Leistungen abbilden und in ihrer Schwierigkeit variieren (AkA, 2023b; ZPA Nord-West, 2021; vgl. Kapitel 3.2). Der durch die Teilnehmenden selbst eingeschätzte Lernzuwachs aus den Bausteinen der *Selbstlernphase* (vier Studierende) und dem *Feedback* bzw. der *Reflexion* (drei Studierende) zeigt ebenfalls Relationen zu den genannten Anwendungsgebieten, sie sind allerdings weniger stark ausgeprägt – es finden sich jeweils eine oder zwei Überschneidungen.

Während für den Baustein zum *Erstellen problemhaltiger Aufgaben* alle Teilnehmenden über einen Lernzuwachs berichteten, gaben fünf Interviewte – darunter ein LiV – an, einen weniger starken Lernzuwachs durch den Baustein zur *Mediendidaktik und zum technologiebasierten Arbeiten* wahrzunehmen (Kategorie L2.1). Zwei der fünf Personen teilten mit, dass sie zwar einen Lernzuwachs wahrgenommen haben, dieser aber im Vergleich weniger stark ausgeprägt ist. Unter anderem schlugen sie eine Optimierung vor. Sabrina sagte, dass zwar

„dieser ... Vortrag ... bezüglich, wie man das am besten farblich gestaltet und welche Schriftarten und Ähnliches ... wichtig“ sei, sie aber „persönlich ... da jetzt nicht so viel drauf gemacht [habe] bei der Aufgabenstellung. [Sie] habe natürlich geguckt, dass es leserlich ist und dass vielleicht auch Sachen hervorgehoben werden. Aber das habe [sie] gemacht, weil [sie] es auch so gemacht hätte, auch ohne den Vortrag.“ (Sabrina, Pos. 110)

Sie schlug vor, statt den Inhalten zur Textgestaltung, rechtliche Aspekte zur Nutzung von Bildern tiefergehend zu behandeln (Sabrina, Pos. 110). Trotz der konstruktiv-kritischen Rückmeldungen, empfanden fünf Studierende und zwei LiV einen Lernzuwachs durch diesen Baustein.

Die Aussagen bezogen sich sowohl auf die *Mediendidaktik* als auch auf den *Umgang mit ILIAS*. Weiter ist den Transkripten zu entnehmen, dass Teilnehmende das Gelernte anwenden bzw. anwenden werden (elf Teilnehmende, davon ein LiV). Seila (Pos. 35) beispielsweise sagte, „*dass [sie] auch gerade durch dieses Seminar ein[en] Lernzuwachs*“ wahrgenommen hat und sie „*weiß, worauf [sie] achten muss*“. Sie ergänzte folgendes:

„Ich weiß nur nicht, ob ich ILIAS als Software so unbedingt verwenden würde oder ob es da auch noch andere Möglichkeiten gibt oder auch bessere Software. Aber diese Möglichkeit gehabt zu haben ... zu sehen, wie das aussehen kann, finde ich gut ... also, dass ich schon gucke, dass es halt schon irgendwie bisschen ansprechender auch ist, wie wir halt auch in ILIAS versucht haben mit diesen ganzen Logos oder interaktiven Bildern und so weiter. Das ist halt nicht irgendwie so etwas ist, ja gut ich mache jetzt was technologiebasiertes, aber nur, weil ich irgendwie es im PC haben will und nicht auf dem Blatt, aber ist eigentlich genau das Gleiche.“ (Seila, Pos. 35)

Es ist davon auszugehen, dass Teilnehmende die Möglichkeiten technologiebasierten Arbeitens für Prüfungszwecke sowie die Bedeutung der mediendidaktischen Umsetzung erkannt haben. Die Selbstlernphase war für den Umgang mit ILIAS besonders hilfreich, wie zwei der vier Teilnehmenden (Samuel, Pos. 143; Saika, Pos. 131) angaben. Beide nahmen einen Lernzuwachs aus der Phase wahr.

Zum Baustein des *Feedbacks* bzw. der *Reflexion* wurden kaum Aussagen über einen wahrgenommenen Lernzuwachs getroffen – in diese Kategorie wurden lediglich aus Transkripten von drei Studierenden kodiert. Dies ist insofern nachvollziehbar, als dass in diesen Bausteinen keine neuen Inhalte vermittelt wurden, sondern diese dem Festigen der Inhalte aus den Bausteinen zum *Erstellen problemhaltiger Aufgaben*, zur *Mediendidaktik* und zum *Umgang mit ILIAS* dienten. Sowohl die *Selbstlernphase* als auch das anschließende *Feedback* und die *Reflexion* zur erstellten Aufgabe können als Mediatoren verstanden werden, die das Lernen und Anwenden der Inhalte aus den vorherigen Bausteinen begünstigen. Dieses Verständnis bestärken auch die Erkenntnisse zur ersten Evaluationsebene, in welcher die Möglichkeiten zum Anwenden des Gelernten, insbesondere in der Selbstlernphase, sowie das Feedback und die Reflexion positiv bewertet wurden (vgl. Kapitel 8.1.2).

Zum Baustein zum *Transfer* berichteten lediglich zwei Personen explizit von einem Lernzuwachs (Saika, Pos. 131; Sami, Pos. 136). Verglichen mit den anderen Bausteinen, kann diese daran liegen, dass sich die Inhalte auf einen Veranstaltungsteil beschränkten. Die Studierende haben sich während ihrer Hausarbeit tiefergehend mit dem *Transfer* beschäftigt. Dass lediglich

zwei Teilnehmende einen Lernzuwachs wahrnahmen, dies von den anderen jedoch nicht berichtet wurde, wird als ein Zeichen dafür verstanden, dass der Baustein zumindest keinen negativen Effekt ausübte. Eine Studentin gab an, das Gelernte aus dem Baustein anwenden zu können (Seila, Pos. 128, 135). Es ist hervorzuheben, dass diese Studentin sich nicht zum Lernzuwachs aus dem Baustein äußerte.

Der Baustein zur *Diagnostik* wurde im Vergleich mit den anderen Bausteinen häufig kritisiert (vgl. Kapitel 8.1.2). Auch bei der Betrachtung der quantitativen Ergebnisse besteht ein Verbesserungspotential. Dennoch gab es einige wenige Aussagen dazu, dass daraus gelernt wurde (Seila, Pos. 128; Shirin, Pos.30; Sami, Pos. 136). Zwei Teilnehmende bewerteten den Lernzuwachs als eingeschränkt bzw. nicht langfristig. Dazu sagte Seila:

„bei Diagnostik der Gütekriterien war es halt so, okay, ich setze mich jetzt noch einmal mit dem Inhalt auseinander, aber das ist so, hört sich jetzt auch ein bisschen Lump an, aber ich habe es halt nur gemacht, weil ich es weiß, ich muss es jetzt in der Hausarbeit machen. Und ich versuche mich da jetzt kurz reinzufuchsen und irgendwie zu verstehen, was dahintersteckt. Aber das ist so eine Sache, die werde ich wahrscheinlich in drei Wochen wieder vergessen haben.“ (Seila, Pos. 128)

Dennoch finden sich in den Transkripten Passagen dazu, dass die aus dem Baustein gewonnene Kompetenz in der Tätigkeit als Lehrkraft Einsatz finden wird. Anhand der Konkretisierung der Teilnehmenden konnte die Rückmeldungen zur Anwendbarkeit in zwei Kategorien unterteilt werden: Komplexität (A1.5) und Bewerten und Beurteilen (A1.6). Beide Elemente sind bedeutende Bestandteile pädagogischer Diagnostik (vgl. Kapitel 2). Bei der Komplexität bezogen sich die Teilnehmenden auf den Lernstand der Lernenden, die Aufgabenschwierigkeit, die geschätzte Bearbeitungsdauer sowie die Hilfsmittel zur Aufgabenbearbeitung. Insgesamt fanden sich dazu Passagen von sieben Teilnehmenden, davon ein LiV. Bei der Bewertung und Beurteilung bezogen sich neun Studierende und ein LiV darauf, wie sie die Schüler*innenlösungen bewerten und welchen Erwartungshorizont sie anlegen werden. Dies steht unmittelbar mit der Beachtung der Gütekriterien, insb. der Objektivität, in Verbindung. Sami (Pos. 40) sagte beispielsweise, dass er sich bei der Konzeption einer Klausur auch Gedanken um den Erwartungshorizont machen und sich um Objektivität bemühen würde. Auch in Bezug auf die Validität kamen Aussagen der Teilnehmenden, wenn auch implizite. So sagte beispielsweise Sabrina (Pos. 32), dass es *„ganz wichtig [ist], dass man sich vorher natürlich auch überlegt, was man erwartet, bevor man die Aufgaben erstellt, während man die Aufgaben erstellt, damit man na-*

türlich auch weiß, worauf man hinaus will bei der Aufgabe“. Gleichzeitig meldete ein Studierender zum Baustein *Diagnostik* zurück, dass Inhalte zu den Gütekriterien zwar in einer vorherigen Veranstaltung behandelt wurden, diese aber „*doch immer noch schwer [sind]*“ (Sami, Pos. 128). Als Lösung hierfür schlug er mehr Praxisbeispiele vor, wodurch das Thema „*anschaulicher wird und man das besser greifen kann*“ (Sami, Pos. 132). Diese Forderung überschneidet sich mit der Kritik der Studierenden über den geringen Praxisanteil im Studium (vgl. Kapitel 5.2.1).

LiV teilten einschränkend zum Lernzuwachs durch die Intervention mit, dass das Potential nicht vollständig ausgeschöpft wurde. Sie begründeten dies mit der Doppelbelastung aufgrund des parallelen Aufwandes in den Studienseminaren und Schulen. Beispielsweise sagte Raphael folgendes:

„Deswegen war das [Erstellen problemhaltiger Prüfungsaufgaben], ich denke für uns alle, ja Neuland und konnten so ein bisschen was mitnehmen. Dazu muss man allerdings sagen, wir hätten gerne alle noch mehr mitgenommen. Das war aber jetzt nicht irgendwie dir oder der Veranstaltung geschuldet, sondern das Drumherum bei uns aktuell ..., dass das zeitlich einfach irgendwie ein bisschen ... schiefgelaufen ist.“ (Raphael, Pos. 11)

Rebecca schlug vor, ein solches Programm in die Ausbildung angehender Lehrkräfte zu integrieren, um das Lernpotential besser auszuschöpfen:

„Und ich könnte mir gut vorstellen, dass man wirklich als Referendarin oder Referendar besser auf die Berufspraxis vorbereitet werden würde, wenn so ein Training, wie ihr das anbietet, auch mal im Vorbereitungsdienst integriert wäre, also verpflichtend integriert wäre, zum Beispiel.“ (Rebecca, Pos. 142)

Weiter wurde zur Anwendbarkeit des Gelernten von den Teilnehmenden die Rückmeldung gegeben, dass sie Hilfe von Dritten benötigen (würden) (A2.1; sieben Teilnehmende, davon zwei LiV) und nur eingeschränkte Möglichkeiten für die Prüfungspraxis sehen (A2.2; zwei Teilnehmende, davon ein LiV). In Bezug auf die Hilfe von Dritten wurden etwa Mentor*innen und erfahrene Kolleg*innen genannt. Die Aufgaben würden ihnen zwecks Einholung eines Feedbacks (Raphael, Pos. 37) und/oder von Ratschlägen vorgelegt (Sabrina, Pos. 20; Samuel, Pos. 25; Sina, Pos. 28; Svana, Pos. 20). Weiter wurden Rechercharbeiten genannt, etwa über das Internet bzw. Altklausuren (Sabrina, Pos. 20; Sina, Pos. 28, Steven Pos. 44). Die eingeschränkte Möglichkeit zur Anwendung des Gelernten wurden auch mit der Pflicht bzw. Verantwortung

zur Vorbereitung auf die Prüfung begründet. Damit wurde das *teaching to the test*-Phänomen angesprochen, auch wenn dieses gleichzeitig kritisiert wurde:

„Ich, also ich denke besonders, wenn das eine Klasse ist, die im letzten Ausbildungsjahr ist, die wirklich vor den Abschlussprüfungen sind, ich glaube, selbst wenn man ... die Aufgaben nicht ganz so toll findet, ist es glaube ich, ist man irgendwie auch in der Pflicht, die darauf vorzubereiten, bei manchen ist es vielleicht nicht sicher, ob Sie übernommen werden und es kommt auf die Note drauf an.“ (Samuel, Pos. 29) und

„... ich glaube, im Unterricht kann man das sehr gut anwenden, zum Teil. Aber letztendlich muss man ja auch auf die Prüfungen vorbereiten, es sollte nicht so sein. Aber letztendlich ist das ja die Realität. ... Und da muss man halt auch ein Stück weit dann pragmatisch sein und gucken, was dann am Ende abgefragt wird ... es ist vielleicht bisschen demotivierend, aber ich glaube, bei zumindest bei einigen Schülern ist wichtiger, dass sie letztendlich eine gute Note am Ende dann haben, als wenn der Unterricht noch toll ist und noch problemhaltiger und so weiter. Es ist deprimierend das so zusagen.“ (Samuel, Pos. 147; ähnliche Aussagen kamen auch von Ricardo, Pos. 41).

Der Aussagen von Kirkpatrick und Kirkpatrick (2006, S. 21-22) und Lipowsky (2010, S. 54) dazu, dass positive Reaktionen zum Training das Lernen begünstigen (wenn auch nicht bedingen), kann aufgrund der qualitativen Daten zugestimmt werden. Dies zeigt der Code-Relations-Browser (Tabelle 32) zur positiven Reaktion auf die Trainingsbausteine und zum selbsteingeschätzten Lernzuwachs der Teilnehmenden durch die einzelnen Bausteine. Der Browser wurde über MAXQDA so eingestellt, dass das Vorkommen der Codes im gleichen Dokument, d. h. von der gleichen Person, angezeigt wird. In den Zellen ist die Anzahl der Dokumente angegeben, in denen sich die jeweiligen Kategorien überschneiden. Kategorien, die denselben Trainingsbaustein betreffen, sind grau hervorgehoben.

Tabelle 32:

Code-Relation-Browser (MAXQDA) zum Lernzuwachs (L) durch und zur Reaktion (R) auf einzelne Bausteine

Kategorie	L1.2	L1.3	L1.4	L1.5	L1.6	L1.7	L1.1
R1.1.1	8	5	4	2	2	2	9
R1.2.1	8	4	4	3	3	2	11
R1.3.1	9	5	4	3	3	2	12
R1.4.1	4	3	1	1	1	0	6
R1.5.1	5	2	2	2	2	1	8
R1.6.1	5	2	2	3	3	1	6
R1.7.1	8	4	4	3	3	2	9

(Eigene Darstellung)

Der Browser spiegelt die oben beschriebenen Ergebnisse wider, die häufigsten Überschneidungen betreffen den Baustein zum *Erstellen problemhaltigen Aufgaben* (L1.2 und R1.1.1). Die wenigsten Kodierungen und Überschneidungen finden sich zu den Bausteinen zum *Feedback, Reflexion* (L1.5 und R1.4.1) und *Transfer* (L1.7 und R1.6.1), gefolgt vom Baustein der *Diagnostik* (L1.6 und R1.5.1). Wie auch in vorherigen Abschnitten dargestellt, wurde zu diesem Baustein konstruktive Kritik geäußert, da etwa die Inhalte als komplex wahrgenommen wurden. Entsprechend wurde auch weniger häufig von einem Lernzuwachs gesprochen, was sich mit den Ergebnissen der quantitativen Daten deckt. Der Korrelationskoeffizient der Skalen zur wahrgenommenen Zufriedenheit der Teilnehmenden mit dem Training und dem wahrgenommenen Lernzuwachs ist mit $r = .565$ ($p < .001$) nach Cohen (1988, S. 80) als stark zu interpretieren. Beide Ergebnisse führen zur Entscheidung, die Hypothese H5 anzunehmen, d. h. die Zufriedenheit der Trainingsteilnehmenden mit dem Training korreliert mit dem von ihnen subjektiv empfundenen Lernzuwachs. Die Daten lassen allerdings keine Aussage über einen kausalen Zusammenhang zu.

8.1.4 Limitationen der Trainingsevaluation

Die Trainingsevaluation erfolgte nach dem Mixed-Methods-Ansatz, indem quantitative und qualitative Daten erhoben und ausgewertet wurden. Dies ermöglicht einen umfassenden Einblick, die Studien und Ergebnisse unterliegen jedoch auch Limitationen.

Die im quasi-experimentellen Design gestaltete Studie bringt die Einschränkung mit sich, dass diese „im Vergleich zu einer experimentellen Untersuchung ... eine geringere interne Validität [im Original hervorgehoben]“ aufweist (Döring & Bortz, 2016, S. 199-200). Das ist u. a. damit zu begründen, dass sich die Teilnehmenden an der Interventions- und Kontrollgruppe nicht nur darin unterscheiden haben, ob sie an der Intervention teilgenommen haben oder nicht, sondern

die Wahrscheinlichkeit der Unterscheidung der Personen in weiteren Merkmalen bestand. Bei der Zuordnung in Gruppen konnte dies nicht umfassend berücksichtigt werden, da die Zuordnung nicht durch die Forscherin geschah, sondern durch die Proband*innen selbst. Das heißt, es kam zu Selektionseffekten. Mittels der ANCOVA und der Regressionsanalyse wurden zwar einige Variablen als Einflussfaktoren bzw. Störvariablen berücksichtigt, dennoch kann nicht ausgeschlossen werden, dass weitere Merkmale relevant wären, wie etwa die Teilnahmemotivation. Aus welcher Motivation die Teilnehmenden an der Veranstaltung teilgenommen haben, wurde weder in der quasi-experimentellen noch in der Interviewstudie berücksichtigt, da (1) der Schwerpunkt der Forschung nicht darauf lag, (2) es zumindest für einen Teil der Stichprobe eine (Wahl-)Pflichtveranstaltung darstellte und somit die Frage eher pragmatisch zu beantworten ist und (3) nicht alle Teilnehmenden ihre Teilnahmemotivation mitteilten, was zu einer Verzerrung führen könnte. Eine weitere Einschränkung zeigte sich in der vorliegenden Studie, (4) dass zumindest die LiV einen Großteil oder gar die kompletten Veranstaltung in ihrer Freizeit absolvieren mussten, was wiederum zu zeitlichen Einschränkungen und zu Terminüberschneidungen führen könnte. Um die Frage nach der (fehlenden) Teilnahmemotivation, zumindest für die Zielgruppe der LiV, zu klären, würde es sich anbieten, neben den Personen, die an der Intervention teilgenommen haben, auch diejenigen zu befragen, die sich gegen die Teilnahme entschieden haben.

Eine wesentliche Limitation der Ergebnisse durch Selbstselektionseffekte zeigt sich insbesondere in Bezug auf die teilnehmenden LiV, für die die Teilnahme am Training freiwillig war und das Training keine Funktion in Form von etwa Credits für die Ausbildung darstellte. In diesem Zusammenhang wird der Anreiz durch Incentives für die Teilnahme am Training bzw. am Pre- und Posttest als marginal angenommen oder gar ausgeblendet. Begründet wird dies damit, dass die Anzahl der Anmeldungen zum Training ($N_{\text{AnmeldungenLiV|Runde1}} = 5$; $N_{\text{AnmeldungenLiV|Runde2}} = 6$) im Vergleich zum Rekrutierungsaufwand (vgl. Kapitel 7.3.2) nicht hoch war. Die Anzahl derjenigen, die das Training aufgenommen ($N_{\text{AufgenommenLiV|Runde1}} = 3$; $N_{\text{AufgenommenLiV|Runde2}} = 6$) und abgeschlossen haben ($N_{\text{AbgeschlossenLiV|Runde1}} = 1$; $N_{\text{AbgeschlossenLiV|Runde2}} = 4$) war noch niedriger. Daten von Personen, die das Training nicht abgeschlossen haben, blieben unberücksichtigt. Zwar meldeten sich die Personen in der Regel mit einer Begründung ab, dennoch wäre es interessant, systematisch den Grund zu erfahren, indem beispielsweise von einer dritten Person (nicht Dozierende des Trainings) Gespräche geführt werden. Es kann angenommen werden, dass bedingt durch den Alltagsstress und die Belastung im Referendariat (vgl. z. B. Darius, Bunzel, Ehms-Ciechanowicz & Böckelmann, 2020; Drüge, Schleider & Rosati, 2014; Pereira Kastens, Döring-Seipel & Nolle, 2020) das Training, das kein Bestandteil der Ausbildung in

der zweiten Phase und freiwillig war, uninteressant schien oder abgebrochen wurde, um somit (auch zeitlichen) Mehraufwand zu vermeiden. Dass die Teilnahme auf Freiwilligkeit beruhte, geht auf der anderen Seite mit der Erwartung einher, dass das Training für Teilnehmende, die das Training vollständig besuchten, wirksam sein wird und sich auch auf die Transferaktivitäten des Gelernten positiv auswirkt (vgl. Kapitel 6, insb. Vigerske, 2016, S. 116, 137, 140–141). Anders war es bei Masterstudierenden der Wirtschaftspädagogik. Da das Modul, in dem das Training stattfand, Bestandteil des Masterstudiums und akkreditiert war, hatte die Veranstaltung für Masterstudierende auch einen funktionalen Zweck. Teilnehmende Masterstudierende beider Standorte hatten allerdings die Wahl zwischen zwei parallellaufenden Veranstaltungen. Anzumerken ist, dass die Veranstaltung, im Rahmen dessen das Training angeboten wurde, sich (primär) an diejenigen richtete, die die schulische Karriere anstreben und/oder Wirtschaftspädagogik in Studienrichtung II studieren (vgl. Kapitel 7.3.2).

Die Gruppengrößen der LiV war sehr klein (mit jeweils fünf Teilnehmenden in beiden Gruppen), weshalb die Auswertung und Ergebnisinterpretation mit deskriptiven Werten stattfand (vgl. Kapitel 7.3.4, 8.1.2 und 8.1.3). Aufgrund der geringen Anzahl der Teilnehmenden wurden die Einschätzungen der einzelnen LiV verglichen mit den Studierenden höher gewichtet. Die Gruppengröße der Studierenden ermöglichte hingegen weitergehende Verfahren (vgl. Kapitel 7.3.4, 8.1.2 und 8.1.3). Zwar waren die studentische Interventions- und Kontrollgruppe nicht gleich groß, dennoch konnte ein Vergleich der Gruppen aufgrund der Varianzhomogenität (Kapitel 7.3.2) gezogen werden. Erstrebenswert wäre es in einer weiterführenden Studie, die Stichprobengrößen sowohl der Gruppe der LiV als auch der Studierenden zu erhöhen.

Zur Prüfung der Hypothesen 3a, 3b, 4a und 4b wurden die Daten der Studierenden und der LiV separat betrachtet. Zur Prüfung der Hypothesen 3a und 3b (Ausbildungsphase) wäre es möglich, die Anzahl der Module zu betrachten, die den Fortschritt der Ausbildungsphase kennzeichnen. Es wurde jedoch lediglich die Gruppe der Studierenden nach den belegten Modulen befragt. Da es sich bei den Studierenden um Studierende zweier Standorte handelt und diese aufgrund der unterschiedlichen Studienverlaufspläne und angebotenen Module nicht auf dieselben Lernangebote zurückgreifen können, wurde auf eine Betrachtung der belegten Module verzichtet. Zwar wäre es denkbar, die Analyse hierfür nach Standorten vorzunehmen, allerdings würde das zu kleineren Stichproben führen. Dies wiederum würde zu erschwerten Analysen oder gar zu Fehlermeldungen führen und die Aussagekraft der Ergebnisse wäre geringer. Da davon ausgegangen wurde, dass Studierende höherer Semester in der Regel mehr (wirtschafts-)pädagogische Module belegt haben und im Studium fortgeschrittener sind als Studierende niedrigerer

Semester, wurde davon ausgegangen, dass die Anzahl der belegten Module implizit in der Variable zum Fachsemester enthalten ist. Dies konnte mittels des Korrelationswerts zwischen der Anzahl der belegten Module und dem Fachsemester bestätigt werden ($r = .731, p < .001$). Nach Cohen (1988, S. 80) liegt hierbei eine starke Korrelation vor. Einschränkend ist hier zu beachten, dass weder die Angabe zu den belegten Modulen noch zum Fachsemester eine sichere Auskunft über (nachhaltiges) Wissen/Kompetenz bzw. das Gelernte geben (vgl. Baumert & Kunter, 2006, S. 477, 490).

Eine weitere Einschränkung stellt der Vergleich der Studierenden mit LiV, die das Studium abgeschlossen haben, dar. Die Anzahl der teilnehmenden LiV ist zu gering, als dass die Gruppe in die Analyse eingebunden werden konnte. Aufgrund der Intervallskalierung der Variable Fachsemester, war es nicht möglich, die Daten der LiV zu ergänzen, indem in Semesterabständen weitergezählt wird. Dies ist damit zu begründen, dass die Organisation des Referendariats nicht immer nach Semestern verläuft, die LiV ihre aktuelle Ausbildungsphase in Jahren angeben und es unklar war, ob LiV unmittelbar nach dem Masterstudium mit dem Referendariat starteten. Ein weiterer Grund gegen das Zusammenführen der studentischen Daten und der Daten der LiV war, dass wesentliche Unterschiede zwischen den Zielgruppen angenommen wurden, etwa durch die Praxiserfahrung, was bei einer Zusammenführung der Daten zur Verfälschung führen könnte.

Bezüglich der Regressionsanalyse wird in der Literatur darauf hingewiesen, diese an derselben Stichprobe, inkl. der Anzahl der Fälle, durchzuführen (vgl. Bauer, 2020, S. 39). Wie zu den Tabellen 29 und 30 angemerkt, waren die Stichproben im jeweils ersten Modell kleiner als die der anderen Modelle, was mit den Missings der Variable *Fachsemester* zu begründen ist. Dies stellt eine Limitation der Studie dar. Gleichwohl war es theoretisch wichtig, auch die Variable zum *Fachsemester* zu berücksichtigen. In den Modellen 2 und 3 nur die Fälle zu berücksichtigen, die auch in Modell 1 integriert waren, würde zu einer geringeren Stichprobengröße in den Modellen 2 und 3 führen. Dies würde die Qualität einschränken, indem nicht die gesamte Stichprobe, die an der Studie und den Pre- und Posttest teilnahmen, berücksichtigt werden. Daher fiel die Entscheidung dafür aus, die Modelle mit unterschiedlich großen N zu berechnen.

Durch die Studie war es möglich, Aussagen zu den ersten beiden Ebenen der Evaluation nach Kirkpatrick und Kirkpatrick (2006) zu treffen. Zu den Ergebnissen zum Wissenstest diagnostischer Gütekriterien muss einschränkend ergänzt werden, dass die Interventionsgruppe mit einem besseren Mittelwert im Pretest startete. Zwar war der Effekt der Intervention in den Berechnungen ohnehin nicht signifikant. Aber wenn es dies wäre, müssten weitere Limitationen

damit verbunden werden. Weiter ist einschränkend zum wahrgenommenen und objektiven Lernzuwachs zu berichten, dass nicht klar ist, ob das Gelernte ausschließlich auf das Training zurückzuführen ist. Dies ist eher zu verneinen, da alle Teilnehmenden, sowohl Studierende als auch LiV, sich in der Ausbildungsphase befinden und eine pädagogische Ausbildung absolvieren. Daher ist davon auszugehen, dass auch bestehende bzw. in einem anderen Kontext erworbene Kompetenzen¹⁷² genutzt wurden. Dennoch ist – zumindest in Bezug auf das Wissen über Merkmale problemhaltiger Aufgaben – mittels quantitativer Daten zu bestätigen, dass die Intervention einen signifikanten Beitrag dazu geleistet hat (vgl. Kapitel 8.1.3). Dass Teilnehmende auch (quantitativ und qualitativ) von einem wahrgenommenen Lernzuwachs berichten, bestätigt die Ergebnisse (vgl. Kapitel 8.1.3), weshalb *Testungseffekte* (Rost, 2007, S. 109) durch die Pre- und Posttest-Messung als marginal angenommen werden. Dass in Bezug auf das Wissen zu diagnostischen Gütekriterien kein signifikanter Lernzuwachs ermittelt werden konnte, deutet zum einen auf das dahingehende Optimierungspotential der Inhalte hin (vgl. Kapitel 8.2). Auf der anderen Seite kann es sein, dass es bei der Erhebung (auch) zu Ermüdungseffekten und/oder fehlender Motivation an (weiterer) Teilnahme an der Erhebung gekommen sein kann (vgl. Döring & Bortz, 2016, S. 729; Klöckner & Friedrichs, 2014, S. 677), gerade vor dem Hintergrund, dass der Test zum Wissen über diagnostische Gütekriterien relativ zum Schluss des gesamten Fragebogens kam. Auf der anderen Seite überschneidet sich das genannte Ergebnis mit dem, das im Projekt TeKoP errechnet wurde (vgl. Goethe-Universität Frankfurt & Georg-August-Universität Göttingen, 2023, S. 15-16; Kapitel 7.1), was den Optimierungsbedarf des Bausteins verdeutlicht.

Wie Baumert und Kunter (2006) es bereits feststellten, ist es im Vergleich zur Erfassung des Wissens nicht trivial die „Kombination von Wissen und praktischen Können“ (S. 486) zu erfassen. Im Projekt wurden zwar Wissenstests zu relevanten Inhalten eingesetzt, jedoch kann nicht begutachtet werden, wie die tatsächliche Umsetzung dessen in der Praxis geschieht. Hierfür können weiterführende Studien angesetzt werden, um das Praxishandeln zu erfassen und somit auf der dritten Ebene zu evaluieren. Im Kontext der Lehrer*innen(fort-)bildung wird die vierte Ebene, die Ebene der Ergebnisse (Results), als „Effekte auf Schüler/innen“ (Lipowsky, 2010, S. 59) bezeichnet, wobei hierzu der schulische Erfolg der Schüler*innen in Bezug auf etwa deren Leistungen und Motivation zählen (Lipowsky, 2010, S. 59, 63). Anhand der Daten aus den Teilstudien war es nicht möglich, auf der vierten Ebene zu evaluieren. Dieser würde eine weiteren Studie bedarfen, in der, in Anlehnung an die Beschreibung der Ebene von

¹⁷² Vgl. hierzu auch die Beschreibung von „Reifung und Entwicklung“ von Rost (2007, S. 108-109), die Bezeichnung ist im Original auf S. 108 hervorgehoben.

Lipowsky (2010, S. 59, 63) folgend, beispielsweise Schüler*innen einer Lehrkraft befragt werden, wobei die Lehrkraft vorher an der Intervention teilgenommen haben muss. Hierbei könnte etwa erfasst werden, wie sich die Motivation der Schüler*innen zum Problemlösen geändert hat oder gar ob und wie sich die Problemlösekompetenz der Schüler*innen ändert. Weiterführende Studien könnten eine Evaluation auf diesen höheren Ebenen anstreben und u. a. kontextbezogene Faktoren auf die Wirksamkeit des Trainings berücksichtigen (vgl. Kapitel 6.4), wengleich die Evaluation auf diesen Ebenen nicht trivial und mit einem höheren Aufwand verbunden ist (vgl. Lipowsky, 2010, S. 57-62).

Die qualitative Evaluation in Form von Interviews erfolgte von der Autorin des Werkes und gleichzeitig der (Mit-)Dozierenden und Betreuenden des Trainings. Auch wenn den Teilnehmenden im Vorfeld und zum Teil im Gespräch betont wurde, dass das Gespräch keinerlei Nachteile für sie haben wird und Kritik willkommen ist (vgl. Kapitel 7.4.1), kann nicht ausgeschlossen werden, dass Teilnehmende sozial erwünscht sprachen. Gleichzeitig muss allerdings gesagt werden, dass Teilnehmende des Interviews auch kritisch waren und den Anschein machten, offen Feedback zu geben (vgl. Kritikpunkte an den Trainingsbausteinen in den Kapiteln 8.1.2 und 8.1.3). Zudem ergänzten zum Teil Teilnehmende zu ihrem Lob an der Veranstaltung, dass sie dies aufrichtig meinen. So wurde z. B. kommentiert: *„Und will das jetzt auch nicht so, dass ich jetzt weil du vor mir bist. Aber ich bin wirklich ein Fan davon muss ich sagen. Ich fand das Modul wirklich sehr, sehr gut“* (Sami, Pos. 148).

8.2 Implikationen aus der Trainingsstudie, insb. -evaluation

Zusammenfassend wurde das Training überwiegend als positiv wahrgenommen und führte – zumindest in Bezug auf das Wissen zu problemhaltigen Aufgaben – zu einem Lernzuwachs (vgl. Kapitel 8.1.2 und 8.1.3). Aufgrund der Datenanalyse ergeben sich im Hinblick auf die Hypothesen folgende Ergebnisse (vgl. Tabelle 33):

Tabelle 33:

Angenommene und abgelehnte Hypothesen

Hypothese	Ergebnis
H1: Teilnehmende der Interventionsgruppe haben zum zweiten Messzeitpunkt (Posttest) einen stärkeren Lernzuwachs in Bezug auf das Wissen zu den Merkmalen problemhaltiger Aufgaben als Teilnehmende der Kontrollgruppe.	✓
H2: Teilnehmende der Interventionsgruppe haben zum zweiten Messzeitpunkt (Posttest) einen stärkeren Lernzuwachs in Bezug auf das Wissen zu diagnostische Gütekriterien als Teilnehmende der Kontrollgruppe.	✗
H3a: Teilnehmende, die sich in einer fortgeschritteneren Ausbildungsphase befinden, haben einen stärkeren Lernzuwachs in Bezug auf das Wissen zu Merkmalen problemhaltiger Aufgaben als Teilnehmende, die in ihrer Ausbildungsphase noch nicht fortgeschritten sind.	✗
H3b: Teilnehmende, die sich in einer fortgeschritteneren Ausbildungsphase befinden, haben einen niedrigeren Lernzuwachs in Bezug auf das Wissen zu diagnostischen Gütekriterien als Teilnehmende, die in ihrer Ausbildungsphase noch nicht fortgeschritten sind.	✗
H4a: Teilnehmende, die ein besseres Vorwissen zu den Merkmalen problemhaltiger Aufgaben (gemessen an der erreichten Punktezahl im Pretest) haben, haben einen niedrigeren Lernzuwachs in Bezug auf das Wissen zu den Merkmalen problemhaltiger Aufgaben.	✗
H4b: Teilnehmende, die ein besseres Vorwissen zu den diagnostischen Gütekriterien (gemessen an der erreichten Punktezahl im Pretest) haben, haben einen niedrigeren Lernzuwachs in Bezug auf das Wissen zu den diagnostischen Gütekriterien.	✗
H5: Die Trainingszufriedenheit der Teilnehmenden korreliert positiv mit ihrem eingeschätzten Lernzuwachs.	✓

(Eigene Darstellung)

Zwar werden in beiden Phasen der Lehrer*innenbildung wichtige Inhalte behandelt, die für pädagogisch-diagnostische und fachdidaktische Zwecke relevant sind – basierend auf den Ergebnissen der vorgelegten Studien besteht jedoch Optimierungspotenzial. Dies gilt insbesondere für das Erstellen von problemhaltigen Aufgaben (zu summativen Zwecken). Diese Annahme lässt sich bereits aus Vorstudien (vgl. Kapitel 5) ableiten, konnte jedoch durch die durchgeführten Studien gestützt werden. Dass das Fachsemester der Teilnehmenden zum Messzeitpunkt keinen signifikanten Effekt auf das relevante Wissen hat (vgl. H3a und H3b und Kapitel 8.1.3), ist mit dem Optimierungspotential der Ausbildung sowie der Effektivität und Bedarfsorientierung des Konzepts zu erklären.

Insgesamt zeigten die Rückmeldungen eine positive Wahrnehmung der Teilnehmenden (sowohl Studierende als auch LiV) bezüglich der Intervention. Es wurde auch das Potential zum Lernzuwachs und zur Anwendung in der Praxis aufgezeigt. Es scheint sinnvoll, die Intervention weiterhin (zumindest) in die Ausbildung angehender Lehrkräfte in die erste und zweite Phase der Lehrer*innenbildung zu integrieren. Inwieweit den Rückmeldungen der Teilnehmenden, das Thema zum Erstellen von problemhaltigen (Prüfungs-)Aufgaben früher in der Ausbildung (z. B. bereits im Bachelorstudium) zu behandeln gefolgt werden sollte, sollte näher betrachtet werden. Hierzu kann kein Ergebnis berichtet werden, da das Training mit Masterstudierenden und LiV stattfand. Die Vorschläge der Teilnehmenden, die Inhalte auf die Unterrichtsgestaltung (Instruktion) zu übertragen, könnte mit dem Vorschlag kombiniert werden, auch andere Ausbildungsberufe zu betrachten und für sie Aufgaben zu erstellen (vgl. Kapitel 8.1.2). Dadurch würde das Training erweitert werden.

Der Vorschlag zur Gruppenarbeit bei der Aufgabenerstellung (Ricardo, Pos. 89) ist mit der Prüfungspraxis vereinbar, denn Prüfungsaufgaben werden in der Regel in Fachausschüssen mit mehreren Personen gebildet (vgl. Kapitel 3.2). In Bezug auf die Praxis für die Erstellung von Klausuren kann im Rahmen dieser Arbeit keine abschließende Aussage getroffen werden, es bedarf einer näheren Betrachtung der Gegebenheiten. Sollten Klausuren als dezentrale Aufgaben (überwiegend) in Einzelarbeit erstellt werden, so scheint es sinnvoll, das vorliegende Konzept beizubehalten und die individuelle Arbeit der Teilnehmenden zu unterstützen.

Im Rahmen der vorgestellten Studie erfolgte die Evaluation nicht auf allen vier Ebenen nach Kirkpatrick und Kirkpatrick (2006) – auf diese Limitation wurde bereits in Kapitel 8.1.4 eingegangen. Dennoch wird erwartet, dass sich das Training auch positiv auf die Praxis auswirken kann (Ebene 3: Behavior; Ebene 4: Results), da insbesondere fachdidaktische und diagnostische Inhalte vermittelt und solche Kompetenzen gefördert werden (vgl. Lipowsky & Rzejak, 2012, S. 240). Unter Berücksichtigung der Erkenntnisse aus der Trainingsforschung (vgl. Kapitel 6), wurde die Maßnahme praxisnah und (schul-)alltagstauglich gestaltet (vgl. Krille, 2019, S. 18; Lipowsky & Rzejak, 2012, S. 237; Kapitel 6.3). Gerade dem Aspekt der Praxisnähe, welcher insbesondere in Bezug auf das Studium seitens der angehenden Lehrkräfte in der Vorstudie bemängelt wurde (vgl. Kapitel 5.2), wird eine hohe Bedeutung zugesprochen (vgl. Kapitel 8.1.2 und 8.1.3). Rückmeldungen der Teilnehmenden, die parallel zum Training in der Praxis/an Schulen arbeiteten, die Inhalte des Trainings bereits zu nutzen, bestätigen die Effektivität des Trainings. Die anderen Teilnehmenden verkündeten ihre Intention, das Gelernte in der künftigen Praxis anzuwenden zu wollen (vgl. Kapitel 8.1.3).

Bei der Integration des Konzepts in die Ausbildung angehender Lehrkräfte sollte das kritische Feedback in Bezug auf die Inhalte zu diagnostischen Gütekriterien berücksichtigt werden (vgl. Kapitel 8.1.2 und 8.1.3). Durch die Intervention konnte kein signifikanter Lernzuwachs diesbezüglich errechnet werden (vgl. H2 und Kapitel 8.1.3) und es besteht ein Verbesserungspotenzial. Dies könnte durch das Einbeziehen praxisnaher Beispiele und einer weniger komplexen Gestaltung der Beispiele und Übungen erreicht werden (vgl. insbesondere die Rückmeldungen der Teilnehmenden zum Trainingsbaustein in Kapitel 8.1.2).

Insgesamt ist festzuhalten, dass das Konzept trotz der Kritik zum Großteil wirksam ist. Es kann davon ausgegangen werden, dass dessen Inhalte auch in der Praxis Umsetzung finden werden. Dem beschriebenen Verständnis von Gräsel (2008) entsprechend, kann das Konzept als *Implementation*, also „als Verbreitung einer Neuerung in einer ersten Phase verstanden“ (S. 201) werden. Als Reaktion auf die bestehende Kritik an der kaufmännischen Prüfungspraxis (vgl. Kapitel 3.2) und an die nicht-hinreichend erfüllten Anforderungen an (angehende) Lehrkräfte zur Erstellung problemhaltiger Aufgaben zu summativen Assessmentzwecken (vgl. Kapitel 0 und Kapitel 5) wird erwartet, dass mit der Implementierung des Konzepts folgende Ziele erreicht werden:

- (1) (Angehende) Lehrkräfte an kaufmännisch-beruflichen Schulen werden bereits in der ersten und zweiten Phase der Ausbildung praxisnah darin geschult, schriftliche problemhaltige Prüfungsaufgaben für (ausgewählte) kaufmännische Berufe zu entwickeln,
- (2) die diagnostischen Gütekriterien genügen.
- (3) Dadurch werden, im Sinne der Curriculum-Instruktion-Assessment-Triade (Pellegrino, 2010), Aufgaben, die zu Assessmentzwecken erstellt werden, den curricularen Anforderungen¹⁷³ gerecht, wodurch insbesondere die Validität von Prüfungsaufgaben unterstützt wird.
- (4) Langfristig wird erwartet, dass durch den Transfer¹⁷⁴ die Prüfungspraxis im Sinne der validen Erfassung von Problemlösekompetenz verbessert wird. Lehrkräfte erstellen nämlich

¹⁷³ Die Ebene der Instruktion wurde im Rahmen der Arbeit nicht fokussiert. Dennoch zeigen Ansätze, dass auf instruktionaler Ebene die (Problemlöse-)Kompetenzorientierung stattfindet (etwa in Lernfabriken, vgl. z. B. Fasshauer, Wilbers & Windelband, 2021; sowie als Planspiele, vgl. z. B. Beiträge in Blötz, 2008). Die Annahme ist naheliegend, dass (angehende) Lehrkräfte die Kompetenzen zum Erstellen problemhaltiger Aufgaben nicht nur für Prüfungszwecke (im Sinne eines summativen Assessments) nutzen werden, sondern u. a. auch für formative Zwecke. Diese Annahme wird bestärkt durch die Aussagen der Teilnehmenden, die die Erweiterung der Inhalte auf Unterrichtsebene wünschten (vgl. Kapitel 8.1.2). Auch dies ist im Sinne der Curriculum-Instruktion-Assessment-Triade von Pellegrino (2010). Zudem ist zu berücksichtigen, dass bereits in der Ausbildung von angehenden Lehrkräften Themen pädagogischer Diagnostik – insbesondere im Zusammenhang mit der Prozessdiagnostik bzw. der Instruktion – behandelt werden (vgl. Kapitel 5).

¹⁷⁴ Zur Beschreibung des Transferverständnisses sei an dieser Stelle auf Gräsel (2008, S. 201-204) sowie Gräsel, Jäger und Willke (2005) verwiesen. Das Ziel des Transfers wurde auch im TeKoP-Projekt, aus dem das Trainingskonzept stammt (vgl. Kapitel 7.1), gesetzt (Wuttke, Seeber & Schumann, o. A.).

nicht nur für die eigenen Klassen und Schulen Assessmentaufgaben, sondern sind auch Mitglieder paritätisch besetzter Prüfungsausschüsse und beteiligen sich an der Erstellung von schriftlich-kaufmännischen Prüfungsaufgaben (Kapitel 2.3).

- (5) Dass diese Aufgaben nicht nur problemhaltig sind und Anforderungen diagnostischer Gütekriterien entsprechen, sondern auch technologiebasiert umgesetzt werden, ist (mindestens) im Sinne der authentischen Gestaltung und der Nutzung technologischer Potentiale, die u. a. das Auswerten von Leistungen der Auszubildenden genutzt werden können.¹⁷⁵

¹⁷⁵ Da der Fokus der Arbeit nicht auf dem technologiebasierten Arbeiten lag, wird hierzu nicht in die Tiefe eingegangen. Hierzu sei auf den Beitrag von Wuttke et al. (2023) verwiesen.

9 Diskussion und Forschungsdesiderat

In einer sich nicht nur durch den technologischen Fortschritt wandelnden Arbeitswelt werden Routinetätigkeiten von Fachkräften substituierbar. Probleme, die nicht automatisiert bearbeitet werden können, sind von kompetenten Fachkräften zu lösen. Um die komplexen Anforderungen der Arbeitswelt bewältigen zu können, ist berufliche Problemlösekompetenz erforderlich (vgl. Kapitel 1). Um das in den Curricula formulierte Ziel der beruflichen Problemlösekompetenz durch die berufliche Ausbildung zu erreichen, bedarf es, neben entsprechenden Lerngelegenheiten, Instrumente zur Prüfung dieser Kompetenz (vgl. Kapitel 1). Darin liegt der Fokus der vorliegenden Arbeit begründet, welche sich mit dem schriftlichen Prüfen als Outputdiagnostik, einer besonderen Form der Statusdiagnostik (vgl. Kapitel 2.4.3), beschäftigt.

An bisherigen schriftlich-kaufmännischen Prüfungsaufgaben wird u. a. bemängelt, dass diese überwiegend Wissen abfragen und kaum Probleme abbilden. Letzteres ist jedoch zur Erfassung der beruflichen Problemlösekompetenz erforderlich (vgl. Kapitel 3.2). Wird nach möglichen Erklärungsansätzen gesucht, kann u. a. die notwendige Kompetenz der an der Erstellung von schriftlichen Prüfungsaufgaben (potentiell) beteiligten Lehrkräften (vgl. Kapitel 2.3) kritisch betrachtet werden. Theoretisch, curricular und praktisch begründet, sollen Lehrkräfte die Problemlösekompetenz der Lernenden nicht nur fördern, sondern auch die notwendigen Kompetenzen aufzeigen können, problemhaltige Aufgaben zu summativen Zwecken zu erstellen (vgl. Kapitel 0).

In der vorliegenden Arbeit wurde die Zielgruppe der angehenden Lehrkräfte kaufmännisch-beruflicher Schulen in der ersten (Studium) und zweiten Phase (Referendariat) der Lehrer*innenbildung fokussiert. Die Forschungsfragen der Arbeit können in zwei Forschungsbereiche unterteilt werden: *Ausbildungsinhalte* und *Trainingsevaluation*. In allen Teilstudien wurde auf Grundlage theoretischer Erkenntnisse gearbeitet und der aktuelle Stand der Forschung berücksichtigt. Im Bereich der Ausbildungsinhalte wurde ein mehrmethodischer Forschungsansatz gewählt (vgl. Kapitel 4) und dadurch die Ergebnisse (vgl. Kapitel 5) gesichert. Durch die Mehrperspektivität wurde zudem der Blick nicht auf die Angebotsperspektive (Perspektive der Lehrenden und der bildenden Institutionen) beschränkt (vgl. Kapitel 4.2), sondern um die der Nachfrage (Perspektive angehender Lehrkräfte als Lernende) erweitert (vgl. Kapitel 4.3). Zur ersten Ausbildungsphase (Studium) wurden Dokumente aller Universitätsstandorte in Deutschland analysiert, die den Studiengang Wirtschaftspädagogik anbieten. Des Weiteren wurden alle (vertretenden) Studiengangleiter*innen der Wirtschaftspädagogik zur Teilnahme an der Fragebogenstudie eingeladen. Zur zweiten Ausbildungsphase (Referendariat) wurden Dokumente von

sechs Bundesländern zum Referendariat an kaufmännisch-beruflichen Schulen analysiert. Die Beschränkung auf die Bundesländer ist damit begründet, dass an der Trainingsstudie (vgl. Kapitel 7) teilnehmende LiV aus diesen Bundesländern erwartet waren. In der Interviewstudie mit Fachleitungen konnten Interviews mit Vertreter*innen aus vier Bundesländern geführt werden, da der Zugang zu den Teilnehmenden der anderen Bundesländer nicht möglich war (vgl. Kapitel 4.2.5 und 4.2.6). Die Ergebnisse zeigen, dass eine standortübergreifende Aussage zu den jeweiligen Phasen der Ausbildung kaum möglich ist, da sich die Standorte untereinander unterscheiden. Daher wird ein Gesamtbild für die jeweilige Phase vorgestellt.

Zur ersten Phase der Ausbildung kann festgehalten werden, dass insbesondere der Fokus auf der Gestaltung von Unterricht liegt und Themen zur Outputdiagnostik kaum behandelt werden (vgl. Kapitel 5.1.2). Im Vergleich dazu wird das Thema der Outputdiagnostik in der zweiten Phase der Ausbildung stärker beleuchtet (vgl. Kapitel 5.1.3). Handlungs- und/oder Problemorientierung werden in beiden Phasen thematisiert, auch wenn weniger im Zusammenhang mit der Outputdiagnostik bzw. mit der Erstellung von Aufgaben zu summativen Zwecken (vgl. Kapitel 5.1). Während die Fragebogenstudie mit (vertretenden) Studiengangsleiter*innen die Erkenntnisse aus der Dokumentenanalyse zur ersten Phase der Lehrer*innenbildung weitestgehend bestätigt (vgl. Kapitel 5.1.2), liefert die Interviewstudie mit Fachleiter*innen der Studienseminare differenziertere Ergebnisse im Vergleich zur Dokumentenanalyse zur zweiten Phase der Lehrer*innenbildung und fokussiert stärker die (kompetenzorientierte) Outputdiagnostik (vgl. Kapitel 5.1.3). Aus Letzterer geht hervor, dass LiV praxisorientiert und mit Problemsituationen arbeiten und entsprechend prüfen können sollen – die zentralen Prüfungsaufgaben wurden hingegen kritisiert. Die Interviews mit angehenden Lehrkräften zeigen, dass diese ihre Vorbereitung auf die Erstellung von problemhaltigen Prüfungs- bzw. Klausuraufgaben als unzureichend wahrnehmen. Weiter wurde die fehlende Praxisorientierung bzw. Unterstützung in der praktischen Umsetzung bemängelt (vgl. Kapitel 5.2.1). Auf die Limitationen der Teilstudien wurde in den Kapiteln 5.1.4 und 5.2.2 eingegangen, sodass in diesem Kapitel darauf verwiesen wird.

Dass auf der einen Seite das Prüfen der Problemlösekompetenz relevant ist, auf der anderen Seite bei der Erstellung von schriftlichen Prüfungsaufgaben (potentiell) beteiligte Lehrkräfte kaum auf diese Tätigkeit vorbereitet werden – oder dies so wahrnehmen – begründet ein entsprechendes Training. Für das hier vorgestellte Training wurde das Trainingskonzept des TeKoP-Projekts genutzt, da dieses durch ein multiprofessionelles Team zusammengestellt wurde. Zudem konnten zum Zeitpunkt des Einsatzes durch bisherige Trainingsdurchläufe im Projekt überwiegend (sehr) gute Ergebnisse erzielt werden (vgl. Kapitel 7.1). Die Trainingsevaluation erfolgte quantitativ und qualitativ (vgl. Kapitel 7.3 und 7.4). Zur quantitativen Evaluation

wurde ein quasi-experimentelles Design gewählt und – nach den biografischen Fragen im Pretest – das Wissen der Interventions- und Kontrollgruppe mittels Pre- und Posttests abgefragt. Zu Evaluationszwecken erhielten die Teilnehmenden der Interventionsgruppe im Posttest einen zusätzlichen Fragebogenteil zur Erfassung der Reaktion bezüglich der Instruktionsqualität, der Zufriedenheit, dem Nutzen sowie den subjektiven Lernzuwachs. Nach Abschluss der Intervention wurden zur qualitativen Evaluation freiwillige Teilnehmende interviewt. Die Ergebnisse sind als positiv zu bewerten, die die Trainings- und Instruktionsqualität sowie den (wahrgenommenen) Lernzuwachs bestätigten (vgl. Kapitel 8). Teilnehmende LiV und Masterstudierende zeigten eine weitestgehende Zufriedenheit mit dem Training und der Instruktion. Sie bestätigten die Relevanz der Inhalte des Trainings, lobten die begleiteten Anwendungsphasen und die bereitgestellten Materialien sowie das im Training erhaltene Feedback. Zudem zeigt sich bei den Teilnehmenden ein signifikanter Lernzuwachs in Bezug auf das Erstellen problemhaltiger Prüfungsaufgaben. Einige Teilnehmende wenden das Gelernte bereits in Praxis an (als LiV oder z. B. Vertretungslehrkraft an der Schule). Andere Teilnehmende berichteten von der Intention, das Gelernte und die Materialien künftig einzusetzen. Kritik zum Training bezieht sich etwa auf das Blockformat der Veranstaltung und auf den Trainingsbaustein zu diagnostischen Gütekriterien; dieser wurde als zu komplex wahrgenommen. Dass aus diesem Baustein kein signifikanter Lernzuwachs erreicht wurde, stützt auch die Analyse der quantitativ erhobenen Daten. Die Limitationen der Studien, u. a. die kleine Stichprobengröße der LiV, wurden in Kapitel 8.1.4 eingehend diskutiert, sodass in diesem Kapitel nicht darauf eingegangen wird.

Die überwiegend positiven Ergebnisse der Trainingsevaluation in Kombination mit der Relevanz der Thematik und der bisher nicht hinreichenden Behandlung des Themas zur Erstellung problemhaltiger Prüfungsaufgaben führen zur Empfehlung, (zumindest einzelne) Trainingsbausteine künftig für die Ausbildung von Lehrkräften beider Phasen zu nutzen. Der modulare Aufbau des Trainings ermöglicht es Lehrenden, einzelne Bausteine ggf. adaptiv und entsprechend der Bedingungen der Teilnehmenden (vgl. Kapitel 6.2) und der Rahmenbedingungen der jeweiligen Institution und des jeweiligen Standorts einzusetzen. Auch ist eine Erweiterung des Konzepts u. a. durch weitere Bausteine oder durch eine Vertiefung der Inhalte eines jeden Bausteins möglich. So kann etwa verstärkt auf das Thema der Bewertung und Beurteilung eingegangen werden. Laut den Ergebnissen der Dokumentenanalyse (vgl. Kapitel 5) sind Inhalte dazu bereits in den Modulen der Ausbildungsphasen zu finden. In Bezug auf die diagnostische Kompetenz von Lehrkräften wird häufig die Notengebung kritisiert (vgl. Schrader, 2013, S. 157). Dieses Thema könnte im Trainingsbaustein zu diagnostischen Gütekriterien integriert

werden, wodurch gleichzeitig die Thematik diagnostischer Gütekriterien näher am Alltag gekoppelt und weniger komplex gestaltet werden kann. Somit würde auf die Kritik der Teilnehmenden der Trainingsstudie reagiert (vgl. Kapitel 8.1.2 und 8.1.3) und eine Adaption vorgenommen werden.

Mit dem Konzept und den Evaluationsergebnissen soll ein Beitrag zur Optimierung der Lehrer*innenbildung an kaufmännisch-beruflichen Schulen geleistet werden. „Focus on competence development of teachers which will enable them to teach their students for an unknown future“ (Mulder, 2019, S. 56). Jedoch sind für Lehrkräfte nicht nur Kompetenzen im Hinblick auf die Gestaltung von Lerngelegenheiten zur Kompetenzförderung der Lernenden relevant, sondern auch für das adäquate Prüfen. Dadurch können auch die Kompetenzen angehender Fachkräfte gefördert und diese besser auf den Arbeitsmarkt vorbereitet werden. Auch, wenn nicht alle Lehrkräfte an beruflichen Schulen an der Erstellung von schriftlich-kaufmännischen Prüfungsaufgaben beteiligt sind/sein werden, besteht die Notwendigkeit, Assessments zu summativen Zwecken zu erstellen. Lehrkräfte erstellen Klausuren und erfassen damit die Ausprägung des interessierten Merkmals zum gegebenen Zeitpunkt. Die Ergebnisse werden dann in (Zeugnis-)Noten übertragen. Um aussagekräftige Ergebnisse gewinnen zu können, bedarf es (Klausur-)Aufgaben, die diagnostischen Gütekriterien entsprechen und die notwendige Kompetenz valide überprüfen. Zwar konzentriert sich das hier vorgestellte Trainingskonzept stark auf Prüfungsaufgaben, dennoch können Bausteine bzw. die Inhalte daraus genutzt werden.

Das Trainingskonzept lässt sich durch eine inhaltliche Adaption auch für andere Fachrichtungen nutzen und ist nicht nur für angehenden Lehrkräfte kaufmännisch-beruflicher Schulen relevant. Hierfür sprechen insbesondere auch die Möglichkeiten, die sich aus dem modularen Aufbau ergeben. Zu den Ausbildungsinhalten anderer Zielgruppen als Voraussetzungen können an dieser Stelle jedoch keine Forschungsergebnisse vorgestellt werden. Neben den Lehrkräften an beruflichen Schulen, sind auch Arbeitgeber- und Arbeitnehmervertreter*innen an der Konzeption von Prüfungsaufgaben beteiligt. Im Rahmen des Forschungsprojektes TeKoP wurde zu und mit dieser Gruppe jüngst geforscht und gearbeitet, sodass an dieser Stelle lediglich kurz auf die Studie eingegangen wird. Es wurden u. a. technologiebasierte problemhaltige Aufgaben bei Auszubildenden qualitativ und quantitativ erprobt (z. B. Wuttke et al., 2023). Die Ergebnisse zeigen, dass eine Evaluation des Trainings auf höheren Ebenen möglich ist, zumindest können dazu Hinweise generiert werden. Gleichzeitig können damit Potentiale technologiebasierten Arbeitens aufgezeigt werden. Dies kann das Auswerten der Auszubildendenantworten

vereinfachen, etwa durch (Teil-)Automatisierung. Weiterhin können Problemsituationen authentischer abgebildet werden, etwa mit E-Mail-Darstellungen und Videos (Wuttke et al., 2023; Wuttke et al., 2022, S. 47).

Neben der Problemlösekompetenz gewinnt – nicht unabhängig von der technologischen Entwicklung – das lebenslange Lernen an Bedeutung, etwa in Form von Weiterbildungen (Bauer & Hofmann, 2018, S. 11-12; Becker, 2015, S. 25; Bonekamp & Sure, 2015, S. 36; Lempert & Achtenhagen, 2000, S. 146; Weber, 2015, S. 723). Der Wandel der Arbeits- und Berufswelt erfordert zudem das „interdisziplinäre Denken und Handeln“ (Bauer & Hofmann, 2018, S. 12). So wird erwartet, dass künftig projektorientiert, hierarchieübergreifend und interdisziplinär vernetzt gearbeitet wird (Lewrick & Link, 2018, S. 51; Wilbers, 2019a, S. 21). Dies wiederum impliziert eine (funktionierende) Teamarbeit mit allen dafür notwendigen Fähigkeiten bzw. Kompetenzen wie Kommunikationsfähigkeit, Selbstorganisation und -verantwortung (Bauer & Hofmann, 2018, 8, 14). Nach Schrader (2013, S. 156) sind Instrumente, die nicht-kognitive Facetten erfassen, noch unzureichend in der Lehrkräfteausbildung implementiert. Es besteht daher weiterer Forschungsbedarf. Die daraus resultierenden Erkenntnisse würden einen weiteren Beitrag zur Optimierung der beruflichen Bildung insgesamt leisten, jedoch auch zur Verbesserung der Ausbildung von Lehrkräften, die sich an der beruflichen Bildung beteiligen.

Literaturverzeichnis

- Abele, S. (2017). Förderung berufsfachlicher Problemlösekompetenzen. *Lernen & Lehren*, 32 (126), 60-66.
- Abele, S., Deutscher, V., Nickolaus, R., Rausch, A., Seeber, S., Sembill, D. et al. (2021). Potenziale technologiebasierter Kompetenztests der ASCOT-Initiative aus der Perspektive der Curriculum-Instruction-Assessment-Triade. In K. Beck & F. Oser (Hrsg.), *Resultate und Probleme der Berufsbildungsforschung. Festschrift für Susanne Weber* (S. 13-41). Bielefeld: wbv.
- Abs, H. J. (2006). Zur Bildung diagnostischer Kompetenz in der zweiten Phase der Lehrerbildung. In C. Allemann-Ghionda & T. Ewald (Hrsg.) *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern. Zeitschrift für Pädagogik*. (51), 217-234 [Themenheft]. Weinheim u.a.: Beltz Juventa. Verfügbar unter https://www.pedocs.de/volltexte/2013/7379/pdf/Abs_Zur_Bildung_diagnostischer_Kompetenz.pdf
- Abs, H. J., Döbrich, P., Vögele, E. & Eckhard, K. (2005). *Skalen zur Qualität der Lehrerbildung - Dokumentation der Erhebungsinstrumente. Pädagogische Entwicklungsbilanzen an Studienseminaren (PEB-Sem)* (Materialien zur Bildungsforschung, Bd. 12, 2. Aufl.). Frankfurt am Main. Zugriff am 31.01.2023. Verfügbar unter https://www.pedocs.de/volltexte/2012/3105/pdf/MatBild_Bd12_2Aufl_D_A.pdf
- Achtenhagen, F. & Baethge, M. (2007). Kompetenzdiagnostik als Large-Scale-Assessment im Bereich der beruflichen Aus- und Weiterbildung. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.) *Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft. Sonderheft 8*, 51-70 [Themenheft]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Aebli, H. (1981). *Denken: Das Ordnen des Tuns* (Denkprozesse, Bd. 2). Stuttgart: Klett-Cotta.
- Ahmed, H. M. S. (2010). Hybrid E-Learning Acceptance Model: Learner Perceptions. *Decision Sciences Journal of Innovative Education*, 8 (2), 313-346. Empirical Research. Zugriff am 28.05.2021. Verfügbar unter <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1540-4609.2010.00259.x>
- Akreml, L. (2014). Stichprobenziehung in der qualitativen Sozialforschung. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (Springer VS Handbuch, S. 265-282). Wiesbaden: Springer VS.
- Amend, M. (29.10.2021). *Module Wirtschaftswissenschaften* (E-Mail).

- Amt für Lehrerbildung Hessen (Hrsg.). (2012). *Module für den Vorbereitungsdienst. Lehramt an Grundschulen, Haupt- und Realschulen, Förderschulen Lehramt an Gymnasien Lehramt an beruflichen Schulen Lehrbefähigung für arbeitstechnische Fächer*. Genehmigt vom Hessischen Kultusministerium durch Erlass 139.11 vom 16. November 2012.
- Artelt, C. & Kunter, M. (2019). Kompetenzen und berufliche Entwicklung von Lehrkräften. In D. Urhahne, M. Dresel & F. Fischer (Hrsg.), *Psychologie für den Lehrberuf* (S. 395-418). Berlin: Springer-Verlag.
- Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen (Hrsg.). (2022). *Fachausschüsse - hier entstehen Prüfungsaufgaben*. Zugriff am 07.03.2022. Verfügbar unter <https://www.ihk-aka.de/fachausschuesse>
- Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen (Hrsg.). (2023a). *Handlungsorientierte Prüfung*. Zugriff am 07.03.2022. Verfügbar unter <https://www.ihk-aka.de/handlungsorientiertepruefung>
- Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen (Hrsg.). (2023b). *Was tun wir? Wofür stehen wir?* Zugriff am 18.01.2023. Verfügbar unter <https://www.ihk-aka.de/wir-ueber-uns/qualitaetsmanagement>
- Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen (Hrsg.). (2023c). *Wie entsteht ein Aufgabensatz?* Zugriff am 02.03.2023. Verfügbar unter <https://www.ihk-aka.de/aufgabensatzentstehung>
- Aufschnaiter, C. von, Cappell, J., Dübbelde, G., Ennemoser, M., Mayer, J., Stiensmeier-Pelster, J. et al. (2015). Diagnostische Kompetenz. Theoretische Überlegungen zu einem zentralen Konstrukt der Lehrerbildung. *Zeitschrift für Pädagogik*, 6 (5), 738-758.
- Backhaus, K., Erichson, B., Gensler, S., Weiber, R. & Weiber, T. (2021). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (Lehrbuch, 16. Aufl.). Wiesbaden: Springer Gabler.
- Badura, J. (2015). *Handlungsorientierte Aufgaben für schriftliche Prüfungen in der kaufmännischen Berufsausbildung - Erstellung und Korrektur - Leitfaden für Aufgabenersteller/-innen und Korrektoren/-innen* (1. Aufl.) (Aka - Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen, Hrsg.), Nürnberg.
- Baltes-Götz, B. (2022). *Lineare Regressionsanalyse mit SPSS*. Trier: Universität Trier; Zentrum für Informations-, Medien- und Kommunikationstechnologie. Zugriff am 13.11.2022. Verfügbar unter <https://www.uni-trier.de/fileadmin/urt/doku/linreg/linreg.pdf>

- Bauer, G. (2020). Lineare Regression. In M. Tausendpfund (Hrsg.), *Fortgeschrittene Analyseverfahren in den Sozialwissenschaften. Ein Überblick* (Lehrbuch, S. 7-43). Wiesbaden: Springer VS.
- Bauer, W. & Hofmann, J. (2018). Arbeit, IT und Digitalisierung. In J. Hofmann (Hrsg.), *Arbeit 4.0 – Digitalisierung, IT und Arbeit* (S. 1-16). Springer Vieweg.
- Baumert, J. (2001). Vergleichende Leistungsmessung im Bildungsbereich. In J. Oelkers (Hrsg.) *Zukunftsfragen der Bildung. Zeitschrift für Pädagogik*. (43), 13-36 [Themenheft]. Weinheim: Beltz. Verfügbar unter https://www.pedocs.de/volltexte/2013/7912/pdf/Baumert_2001_Vergleichende_Leistungsmessung_im_Bildungsbereich.pdf
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520. Zugriff am 09.09.2021. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/s11618-006-0165-2.pdf>
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-53). Münster: Waxmann.
- Bayer, M. (2015). Bildung, Leistung und Kompetenz. In A. Lange, H. Reiter, S. Schutter & C. Steiner (Hrsg.), *Handbuch Kindheits- und Jugendsoziologie* (Springer NachschlageWissen, Living Reference Work, continuously updated edition, S. 1-13). Wiesbaden: Springer VS. Zugriff am 11.05.2020. Verfügbar unter https://link.springer.com/content/pdf/10.1007%2F978-3-658-05676-6_7-1.pdf
- Beck, K. (2002). Plädoyer für ein Kern-Curriculum zur Ausbildung von Lehrerinnen und Lehrern an berufsbildenden Schulen. In M. Eckert, M. Horlebein, I. Lisop, H. Reinisch & T. Tramm (Hrsg.), *Bilanzierungen. Schulentwicklung, Lehrerbildung und Wissenschaftsgeschichte im Feld der Wirtschafts- und Berufspädagogik* (S. 187-198). Frankfurt am Main: Verlag der Gesellschaft zur Förderung arbeitsorientierter Forschung und Bildung.
- Becker, K.-D. (2015). Arbeit in der Industrie 4.0 - Erwartung des Instituts für angewandte Arbeitswissenschaft e.V. In A. Botthof & E. A. Hartmann (Eds.), *Zukunft der Arbeit in Industrie 4.0* (pp. 23-29). o. A.: Springer Vieweg.
- Berendt, B., Voss, H.-P. & Wildt, J. (Hrsg.). (2002). *Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten*. [Loseblattsammlung]. Berlin: Raabe Verlag.

- Berge, B. von dem (2020). Teilstandardisierte Experteninterviews. In M. Tausendpfund (Hrsg.), *Fortgeschrittene Analyseverfahren in den Sozialwissenschaften. Ein Überblick* (Lehrbuch, S. 275-300). Wiesbaden: Springer VS.
- Besser, L., Rau, R. & Traum, A. (2022). Objektiv-bedingungsbezogene Tätigkeitsanalysen an Schulen der beruflichen Bildung. Welche Arbeitsinhalte und Ausführungsbedingungen prägen die Tätigkeit von Berufsschullehrkräften? In S. Mühlpfordt & G. Prodehl (Hrsg.), *Gesundheitsschutz und Gesundheitsförderung im Lehrberuf. multiprofessionelle Perspektiven* (S. 53-67). Lengerich: Pabst.
- Bethscheider, M., Höhns, G. & Münchhausen, G. (2011). Kompetenzorientierung in der beruflichen Bildung. In M. Bethscheider, G. Höhns & G. Münchhausen (Hrsg.), *Kompetenzorientierung in der beruflichen Bildung* (Berichte zur beruflichen Bildung, S. 9-18). Bielefeld: Bertelsmann, W.
- Betsch, T., Funke, J. & Plessner, H. (Hrsg.). (2011). *Denken - Urteilen, Entscheiden, Problemlösen. Allgemeine Psychologie für Bachelor* (Springer-Lehrbuch). Berlin: Springer.
- Bezirksregierung Köln (Hrsg.). (2021). *Der Regierungsbezirk*. Zugriff am 01.10.2021. Verfügbar unter https://www.bezreg-koeln.nrw.de/brk_internet/regierungsbezirk/index.html
- Blanz, M. (2021). *Forschungsmethoden und Statistik für die Soziale Arbeit. Grundlagen und Anwendungen* (2. Aufl.). Stuttgart: W. Kohlhammer Verlag.
- Blötz, U. (Hrsg.). (2008). *Planspiele in der beruflichen Bildung. Auswahl, Konzepte, Lernarrangements, Erfahrungen - Aktueller Planspielkatalog 2008* (Schriftenreihe des Bundesinstituts für Berufsbildung, 4. Aufl.). Multimedia-Publikation mit CD-Rom. Bielefeld: W. Bertelsmann Verlag.
- Bonekamp, L. & Sure, M. (2015). Consequences of Industry 4.0 on Human Labour and Work Organisation. *Journal of Business and Media Psychology*, 6 (1), 33-40. Zugriff am 27.10.2019. Verfügbar unter https://journal-bmp.de/wp-content/uploads/2015/12/04_Bonekamp-Sure_final.pdf
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation. für Human und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer Medizin Verlag.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (Springer-Lehrbuch, 7. Aufl.). Berlin: Springer.

- Bosse, D. (2012). Zur Situation der Lehrerbildung in Deutschland. In D. Bosse, L. Criblez & T. Hascher (Hrsg.), *Reform der Lehrerbildung in Deutschland, Österreich und der Schweiz. Teil 1. Analysen, Perspektiven und Forschung* (Theorie und Praxis der Schulpädagogik, Band 4, S. 11-28). Immenhausen bei Kassel: Prolog-Verlag.
- Brady, S., Gillis, M., Smith, T., Lavalette, M., Liss-Bronstein, L., Lowe, E. et al. (2009). First grade teachers' knowledge of phonological awareness and code concepts: Examining gains from an intensive form of professional development and corresponding teacher attitudes. *Reading and Writing*, 22 (4), 425-455.
- Bransford, J. & Stein, B. S. (1993). *The ideal problem solver. A guide for improving thinking, learning, and creativity* (2. Aufl.). New York: Freeman. Zugriff am 04.01.2020. Verfügbar unter https://www.tnitech.edu/cat/pdf/useful_links/idealproblemsolver.pdf
- Bremer, C. (2005). Handlungsorientiertes Lernen mit Neuen Medien. In B. Lehmann & E. Bloh (Hrsg.), *Online-Pädagogik* (Methodik und Content-Management, Bd. 2). Baltmannsweiler: Schneider-Verl. Hohengehren. Zugriff am 04.11.2020. Verfügbar unter https://www.e-teaching.org/praxis/themenspecials/quickstarter-online-lehre/kursprogramm/einfuehrungsmodul/bremer-2005_handlungsorientiert_lernen.pdf
- Brennan, R. L. & Prediger, D. J. (1981). Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement* (41), 687-699. Zugriff am 11.12.2020. Verfügbar unter <https://journals.sagepub.com/doi/pdf/10.1177/001316448104100307>
- Bresges, A., Harring, M., Kauertz, A., Nordmeier, V. & Parchmann, I. (2019) Die Theorie-Praxis-Verahnung in der Lehrerbildung - eine Einführung in die Thematik. In Bundesministerium für Bildung und Forschung (Hrsg.), *Verzahnung von Theorie und Praxis im Lehramtsstudium. Erkenntnisse aus Projekten der „Qualitätsoffensive Lehrerbildung“* (S. 4-7). Berlin.Referat Qualitätsförderung Schule.
- Bretschneider, M., Gutschow, K. & Lorig, B. (2014). Kompetenzorientiert prüfen - Prüfungspraxis und weiterer Entwicklungsbedarf. *Berufsbildung in Wissenschaft und Praxis*, 43 (3), 26-29.
- Bromme, R. (1995). Was ist „pedagogical content knowledge“? Kritische Anmerkungen zu einem fruchtbaren Forschungsprogramm. In S. Hopmann, K. Riquarts, W. Klafki & A. Krapp (Hrsg.) *Didaktik und/oder Curriculum. Grundprobleme einer international vergleichenden Didaktik*. (33), 105-113 [Themenheft]. Weinheim: Beltz Verlag. Verfügbar unter

https://www.pedocs.de/volltexte/2015/10002/pdf/Bromme_1995_Was_ist_pedagogical_content_knowledge.pdf

- Brunner, M., Anders, Y., Hochfeld, A. & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 215-234). Münster: Waxmann.
- Büchter, K., Hahn, C., Jastrzebski, A., Kräenbring, R. & Wölk, M. (2011). Qualitätsverständnisse in der betrieblichen Ausbildung von Kleinen und Mittleren Unternehmen (KMU) des Maler- und Lackiererhandwerks in Hamburg – Erste Ergebnisse des Modellversuches ML-QuES. *Berufs- und Wirtschaftspädagogik - online* (21). Zugriff am 03.11.2020. Verfügbar unter <https://www.bwpat.de/content/ausgabe/21/buechter-etal/index.html>
- Büchter, K. & Tramm, T. (2007). EDITORIAL zur Ausgabe 13: Selbstorganisiertes Lernen in der beruflichen Bildung. *bwp@ Berufs- und Wirtschaftspädagogik - online* (13), 1-5. Zugriff am 27.12.2019. Verfügbar unter http://www.bwpat.de/archiv/bwpat_ausgabe13.pdf
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (Studium Psychologie, 2. Aufl.). Hallbergmoos: Pearson Deutschland GmbH.
- Bundesagentur für Arbeit (Hrsg.). (2023). *Ausbildungswege im Überblick*. Zugriff am 09.04.2023. Verfügbar unter <https://www.arbeitsagentur.de/bildung/ausbildung/ausbildungswege-im-ueberblick>
- Bundesinstitut für Berufsbildung (Hrsg.). (o. A.). *Infos für Ausbildungssuchende und Auszubildende*. Zugriff am 09.04.2023. Verfügbar unter <https://www.bibb.de/de/757.php>
- Bundesinstitut für Berufsbildung (Hrsg.). (2002). *Verordnung über die Berufsausbildung zum Industriekaufmann/zur Industriekauffrau*. vom 23. Juli 2002 (Bundesgesetzblatt Jahrgang 2002 Teil I Nr. 51), Bonn. Zugriff am 08.02.2020. Verfügbar unter https://www.bibb.de/tools/berufesuche/index.php/regulation/industriekaufmann_2002.pdf
- Bundesinstitut für Berufsbildung. (2015). *Ausbildungsordnungen und wie sie entstehen* (7. Aufl.). Bonn: Bundesinstitut für Berufsbildung. Zugriff am 30.11.2019. Verfügbar unter <https://www.bibb.de/veroeffentlichungen/en/publication/download/2061>
- Bundesinstitut für Berufsbildung. (2019). *Datenreport zum Berufsbildungsbericht 2019. Informationen und Analysen zur Entwicklung der beruflichen Bildung*. Bonn: Bundesinstitut für Berufsbildung. Verfügbar unter https://www.bibb.de/dokumente/pdf/bibb_datenreport_2019.pdf

- Bundesministerium für Bildung und Forschung (Hrsg.). (2021). *Bildung und Forschung in Zahlen 2021. Ausgewählte Fakten aus dem Daten-Portal des BMBF*. datenportal.bmbf.de, Bonn.
- Bundesministerium für Bildung und Forschung (Hrsg.). (2022). *Teilnehmer/-innen an Maßnahmen zur beruflichen Weiterbildung nach Alter, Bildungsabschluss und Ländern 2020*. Datenportal. Zugriff am 02.11.2022. Verfügbar unter <https://www.datenportal.bmbf.de/portal/de/Tabelle-2.7.8.html>
- Bürger, N., Haselmann, S., Baumgart, J., Prinz, G., Girnat, B., Meisert, A. et al. (2021). Jenseits von Professionswissen: Eine systematische Überblicksarbeit zu einstellungs- und motivationsbezogenen Einflussfaktoren auf die Nutzung digitaler Technologien im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 24 (5), 1087-1112.
- Burkhardt, R. (23.02.2023). *Studienberatung-WiPäd* (E-Mail).
- Cademartori, I., Seifried, J., Wuttke, E., Krille, C. & Salge, B. (2017). Developing a Training Programme to Promote Professional Error Competence in Accounting. In E. Wuttke & J. Seifried (Hrsg.), *Professional Error Competence of Preservice Teachers. Evaluation and Support* (SpringerBriefs in Education, S. 29-46). Cham: Springer International Publishing; Imprint; Springer.
- Carnegie Forum on Education. (1986). *A Nation Prepared: Teachers for the 21st Century. The Report of the Task Force on Teaching as a Profession*. Hyattsville, Md.: Carnegie Forum on Education and the Economy.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Hillsdale, NJ u.a.: Lawrence Erlbaum Associates.
- Cook, D. R. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, 19 (1), 15-18. Zugriff am 13.11.2022. Verfügbar unter <https://www.jstor.org/stable/1268249>
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16 (3), 297-334.
- Darius, S., Bunzel, K., Ehms-Ciechanowicz, E. & Böckelmann, I. (2020). Psychische Gesundheit bei Referendaren. *Prävention und Gesundheitsförderung*, 16 (3), 215-224.

- Darling-Hammond, L. & Bransford, J. (Eds.). (2005). *Preparing teachers for a changing world. What teachers should learn and be able to do* (The Jossey-Bass education series). San Francisco, CA: Jossey-Bass.
- Davis, F. D., Bagozzi, R. P. & Warschaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of two theoretical Models. *Management Science*, 35 (8), 982-1003. Zugriff am 15.03.2020. Verfügbar unter <https://www.jstor.org/stable/pdf/2632151.pdf?refreqid=excelsior%3A616fcf49f3d27c28a503cfbc35583771>
- Dehnbostel, P. (2000). Erfahrungslernen in der beruflichen Bildung - Ansatzpunkte für eine neue Lernkultur? In P. Dehnbostel & H. Novak (Hrsg.), *Arbeits- und erfahrungsorientierte Lernkonzepte* (Berufsbildung zwischen innovativer Programmatik und offener Umsetzung, Bd. 18, S. 103-114). Bielefeld: Bertelsmann.
- Dengler, K. & Matthes, B. (2015). *Folgen der Digitalisierung für die Arbeitswelt. Substituierbarkeitspotenziale von Berufen in Deutschland* (IAB-Forschungsbericht 11/2015). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung. Zugriff am 25.10.2019. Verfügbar unter <http://doku.iab.de/forschungsbericht/2015/fb1115.pdf>
- Der Präsident der Humboldt-Universität zu Berlin (Hrsg.). (2015). Fachspezifische Studien- und Prüfungsordnung für das Bachelorstudium im Fach Wirtschaftspädagogik (Wirtschaft und Verwaltung). Kernfach im Kombinationsstudiengang mit Lehramtsoption. Überfachlicher Wahlpflichtbereich für andere Bachelorstudiengänge und -studienfächer (Kultur-, Sozial- und Bildungswissenschaftliche Fakultät) (125/2015), Berlin.
- Der Präsident der Humboldt Universität zu Berlin (Hrsg.). (2008). *Lehramtsmaster Wirtschaftspädagogik*. Amtliches Mitteilungsblatt (Nr. 33), Berlin.
- Deutscher, V. (26.04.2023). *Logic-Assessment-Modell* (E-Mail).
- Deutscher, V. & Winther, E. (2019). Zentrale Abschlussprüfungen in der Berufsausbildung - ein Plädoyer. *Berufsbildung in Wissenschaft und Praxis*, 48 (6), 11-15. Zugriff am 14.03.2020. Verfügbar unter https://www.uni-due.de/imperia/md/content/biwi/einrichtungen/vet/deutscher_winther_2019_zentrale_abschlusspr%C3%BCfungen.pdf
- Deutscher Qualifikationsrahmen für lebenslanges Lernen (Hrsg.). (2011). *Deutscher Qualifikationsrahmen für lebenslanges Lernen. verabschiedet vom Arbeitskreis Deutscher Quali-*

- fikationsrahmen (AK DQR) am 22. März 2011*, o. A. Zugriff am 09.02.2023. Verfügbar unter https://www.dqr.de/dqr/shareddocs/downloads/media/content/der_deutsche_qualifikationsrahmen_fue_lebenslanges_lernen.pdf?blob=publicationFile&v=2
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.) *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule. Zeitschrift für Pädagogik*. (41), 73-92 [Themenheft]. Weinheim: Beltz.
- Ditton, H. (2009). Evaluation und Qualitätssicherung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (SpringerLink Bücher, 2., überarbeitete und erweiterte Auflage, S. 607-623). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin, Heidelberg: Springer.
- Dörner, D. (1984). Denken, Problemlösen und Intelligenz. *Psychologische Rundschau* (1), 10-20.
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung* (Kohlhammer Standards Psychologie: Studententext: Teilgebiet Denkpsychologie, 3. Aufl.). Stuttgart: Kohlhammer.
- Drüge, M., Schleider, K. & Rosati, A.-S. (2014). Psychosoziale Belastungen im Referendariat - Merkmale, Ausprägungen, Folgen. *DDS - Die Deutsche Schule*, 106 (4), 358-372. Zugriff am 31.10.2022. Verfügbar unter https://www.waxmann.com/index.php?eID=download&id_artikel=ART101535&uid=frei
- Dubs, R. (2006). Entwicklung von Schlüsselqualifikationen in der Berufsschule. In R. Arnold & A. Lipsmeier (Hrsg.), *Handbuch der Berufsbildung* (2. Aufl., S. 191-203). Wiesbaden: VS Verl. für Sozialwiss.
- Dubs, R. (2008). Lehrerbildung zwischen Theorie und Praxis. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 11-28). Münster: Waxmann.
- Ebbinghaus, M. (2006). Stellenwert der Qualitätssicherung in der betrieblichen Berufsausbildung – Ergebnisse einer Betriebsbefragung. In Bundesinstitut für Berufsbildung (Hrsg.), *Qualitätssicherung beruflicher Aus- und Weiterbildung. Ergebnisse aus dem BIBB* (Wissenschaftliche Diskussionspapiere / Bundesinstitut für Berufsbildung, BIBB, Bd. 78, S. 31-52).

- Eid, M., Gollwitzer, M. & Schmitt, M. (2016). *Formelsammlung Statistik und Forschungsmethoden* (1. Aufl.). Weinheim: Beltz Verlag.
- (Zentrum für schulpraktische Lehrerbildung Krefeld, Hrsg.). (o. A.). *Entwurf Ausbildungsprogramm VD 2016*, Seminar für das Lehramt an Berufskollegs. Zugriff am 02.11.2021. Verfügbar unter https://www.zfsl.nrw.de/KRE/Seminar_BK/Download/Q1_FS_AusbildungsprogrammVD.pdf
- Ertl-Schmuck, R. (2019) Das Theorie-Praxis-Verhältnis als produktive Irritation in der Lehrerbildung für berufsbildende Schulen. In Bundesministerium für Bildung und Forschung (Hrsg.), *Verzahnung von Theorie und Praxis im Lehramtsstudium. Erkenntnisse aus Projekten der „Qualitätsoffensive Lehrerbildung“* (S. 48-53). Berlin.Referat Qualitätsförderung Schule.
- Euler, D. & Pätzold, G. (2004). *Selbst gesteuertes und kooperatives Lernen in der beruflichen Erstausbildung (SKOLA). Gutachten und Dossiers zum BLK-Programm* (Materialien zur Bildungsplanung und zur Forschungsförderung, Bd. 120). Bonn: BLK.
- Fachbereich Wirtschafts- und Sozialwissenschaften. (2019a). *Masterstudiengang Wirtschaftspädagogik. Modulhandbuch - Wintersemester 2019/2020*. Stand: 01. Oktober 2019 (Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Hrsg.).
- Fachbereich Wirtschafts- und Sozialwissenschaften. (2019b). *Wirtschaftswissenschaften, Sozialökonomik, Internatiol Business Studies, Wirtschaftsinformatik. Modulhandbuch - Wintersemester 2019/2020*. Bachelorstudiengänge (Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Hrsg.), Erlangen.
- Fasshauer, U., Wilbers, K. & Windelband, L. (2021). Lernfabriken: Ein Zukunftsmodell für die berufliche Bildung? In K. Wilbers & L. Windelband (Hrsg.), *Lernfabriken an beruflichen Schulen. Gewerblich-technische und kaufmännische Perspektiven* (Blaue Reihe, Bd. 26, S. 15-48). Berlin: epubli.
- Fenstermacher, G. D. (1994). The Knower and the Known: The Nature of Knowledge in Research on Teaching. *Review of Research in Education*, 20, 3-56.
- Field, A. (2009). *Discovering statistics using SPSS. (and sex and drugs and rock ,n‘ roll)* (3. Aufl.). Los Angeles: Sage Publications.
- Flick, U. (2006). Interviews in der qualitativen Evaluationsforschung. In U. Flick (Hrsg.), *Qualitative Evaluationsforschung. Konzepte - Methoden - Umsetzung* (Rororo Rowohlts

- Enzyklopädie, Bd. 55674, Orig.-Ausg, S. 214-232). Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Flick, U. (2010). Triangulation. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (1. Aufl., S. 278-289). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Flick, U. (2011). Das Episodische Interview. In G. Oelerich & H.-U. Otto (Hrsg.), *Empirische Forschung und Soziale Arbeit. Ein Studienbuch* (SpringerLink Bücher, S. 273-280). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Förster, N., Kuhn, J.-T. & Souvignier, E. (2017). Normierung von Verfahren zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik* (2), 116-122. Zugriff am 02.06.2020. Verfügbar unter https://www.pedocs.de/volltexte/2017/14998/pdf/ESP_2017_2_Foerster_et_al_Normierung_von_Verfahren.pdf
- Frackmann, M. & Tärre, M. (2009). *Lernen und Problemlösen in der beruflichen Bildung. Methodenhandbuch* (Berichte zur beruflichen Bildung). Bonn: Bundesinstitut für Berufsbildung.
- Frank, S. & Iller, C. (2013). Kompetenzorientierung - mehr als ein didaktisches Prinzip. *REPORT*, 36 (4), 32-41.
- Frey, A. (2014). Kompetenzmodelle und Standards in der Lehrerbildung und im Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2., überarb. und erw. Aufl., S. 712-744). Münster: Waxmann.
- Frey, C. B. & Osborne, M. A. (2013). *The Future of Employment: How Susceptible Are Jobs to Computerisation?* Zugriff am 28.10.2019. Verfügbar unter https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
- Frey, C. B. & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254-280.
- Friedrich-Schiller-Universität Jena. (2019). *Modulkatalog Master of Education 181 Wirtschaftspädagogik. PO-Version 2016*, Jena.
- Fritz, U. & Laueremann, K. (2019). Akzente für eine neue Lehr-, Lern- und Beurteilungskultur. In U. Fritz, K. Laueremann, M. Pächter, M. Stock & W. Weirer (Hrsg.), *Kompetenzorientierter Unterricht. Theoretische Grundlagen – erprobte Praxisbeispiele* (1. Aufl., S. 59-74). Stuttgart: UTB.

- Frommberger, D. (2013). Lernergebnisorientierung und Lernergebniseinheiten in der beruflichen Bildung. Eine theoretische und komparative Einordnung aktueller curricularer Gestaltungsansätze. *Berufs- und Wirtschaftspädagogik - online* (24).
- Frommberger, D., Held, G., Milolaza, A., Reinisch, H. & Steib, C. (2013). Zusammenfassung und Diskussion der didaktisch-curricularen Ansätze der DECVET-Projekte zur Förderung der Durchlässigkeit im Berufsbildungssystem. *Berufsbildung*, 67 (142), 131-147.
- Funke, J. (2003). *Problemlösendes Denken* (Einführungen und Allgemeine Psychologie, 1. Aufl.). Stuttgart: W. Kohlhammer.
- Funke, J. (2006). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen* (Bd. 8, S. 375-446). Göttingen u.a.
- Funke, J., Fischer, A. & Holt, D. V. (2018). Competencies for Complexity: Problem Solving in the Twenty-First Century. In E. Care, P. Griffin & M. Wilson (Eds.), *Assessment and Teaching of 21st Century Skills. Research and Applications* (Educational Assessment in an Information Age, pp. 41-53). Cham: Springer International Publishing.
- Gajek, S., Hewlett, C., Kaiser, F., Kahl-Andresen, A., Keup-Gottschalck, M. & Labusch-Schönwandt, G. (2021). *Handbuch für Prüfende in der beruflichen Bildung* (ver.di Bundesverwaltung Bildungspolitik, Hrsg.), Berlin.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F. & Yoon, K. S. (2001). What Makes Professional Development Effective? Results From a National Sample of Teachers. *American Education Research Journal*, 38 (4), 915-945. Zugriff am 19.03.2020. Verfügbar unter <https://journals.sagepub.com/doi/pdf/10.3102/00028312038004915>
- Geiser, C., Turhan, L., Wuttke, E., Seeber, S. & Meiners, H. (2021, September). *Entwicklung problemhaltiger technologiebasierter Prüfungsaufgaben: (Erste) Befunde aus einer Trainingsstudie mit Masterstudierenden der Wirtschaftspädagogik*, virtuell.
- Georg-August-Universität Göttingen. (o.A.). ‚Bachelor of Arts‘ (B.A.) in Wirtschaftspädagogik, Georg-August-Universität Göttingen. Zugriff am 21.12.2022. Verfügbar unter <https://www.uni-goettingen.de/de/81147.html>
- Georg-August-Universität Göttingen (Hrsg.). (2019). *Modulverzeichnis. konsekutiven Master-Studiengang „Wirtschaftspädagogik“*, Göttingen.

- Gläser, J. & Laudel, G. (2009). *Experteninterviews und qualitative Inhaltsanalyse. als Instrumente rekonstruierender Untersuchungen* (Lehrbuch, 3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gläser-Zikuda, M. (2013). Qualitative Inhaltsanalyse in der Bildungsforschung - Beispiele aus diversen Studien. In K. Aguado, L. Heine & K. Schramm (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung* (Kolloquium Fremdsprachenunterricht, Bd. 48, 1st, New ed., S. 136-159). Berlin: Peter Lang GmbH Internationaler Verlag der Wissenschaften.
- Gläser-Zikuda, M., Seidel, T., Rohlf, C., Gröschner, A. & Ziegelbauer, S. (2012). Mixed Methods in der empirischen Bildungsforschung - eine Einführung in die Thematik. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung* (S. 7-13). 74. Tagung der Arbeitsgruppe Empirische Pädagogische Forschung (AEPF) im September 2010 in Jena. Münster: Waxmann.
- Göb, N. (2017). Professionalisierung durch Lehrerfortbildung: Wie wird der Lernprozess der Teilnehmenden unterstützt. *DDS - Die Deutsche Schule*, 109 (1), 9-27.
- Goethe-Universität Frankfurt (Hrsg.). (2015a). *UniReport. Ordnung des Fachbereichs Wirtschaftswissenschaften der Johann Wolfgang Goethe-Universität für den Bachelorstudiengang "Wirtschaftspädagogik" mit dem Abschlussgrad "Bachelor of Science" vom 17. Oktober 2007 in der Fassung vom 17. Juni 2015*, Frankfurt am Main.
- Goethe-Universität Frankfurt (Hrsg.). (2015b). *UniReport. Ordnung des Fachbereichs Wirtschaftswissenschaften der Johann Wolfgang Goethe-Universität für den Masterstudiengang „Wirtschaftspädagogik“ mit dem Abschlussgrad „Master of Science“ vom 16. Juli 2014 in der Fassung vom 15. Juli 2015*. Satzungen und Ordnungen, Frankfurt am Main.
- Goethe-Universität Frankfurt; Georg-August-Universität Göttingen. (o. A.). *Handbuch zum Erstellen problemhaltiger technologiebasierter Prüfungsaufgaben in ILIAS. Forschungs- und Transferinitiative „ASCOT+ -Technologiebasierte Kompetenzmessung in der beruflichen Bildung“*. Förderkennzeichen 21AP001A, 21AP001B, o. A. Zugriff am 16.03.2023. Verfügbar unter https://ilias-tekop.as.wiwi.uni-goettingen.de/ilias.php?ref_id=5355&obj_id=5102&cmd=downloadFile&cmdClass=ilImpresentationgui&cmdNode=ek&baseClass=ilLMPresentationGUI&file_id=il_file_25434

- Goethe-Universität Frankfurt; Georg-August-Universität Göttingen. (2023). *Abschlussbericht TeKoP. Technologiebasiertes kompetenzorientiertes Prüfen*. [unveröffentlichtes Dokument], Frankfurt am Main.
- Goldschmidt, P. & Phelps, G. (2007). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review*, 29 (3), 432-439. CSE Technical Report 711.
- Goos, M. (2013). *How the world of work is changing: a review of the evidence* (International Symposium for Employers on the future of work). Zugriff am 28.10.2019. Verfügbar unter https://pdfs.semanticscholar.org/202a/53a7ba033440604fd7d8a0fa635f31b64069.pdf?_ga=2.98897624.772946124.1572240694-1198084783.1572240694
- Grabowski, F., Castello, A. & Brodersen, G. (2022). Diagnostische Gütekriterien bei Status-tests. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 247-258). Version 1.0.
- Gräsel, C. (2008). Einführung. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 201-205). Münster: Waxmann.
- Gräsel, C., Jäger, M. & Willke, H. (2005). *Konzeption einer übergreifenden Transferforschung unter Einbeziehung des internationalen Forschungsstandes*, Wuppertal. Zugriff am 18.12.2022. Verfügbar unter https://www.tib.eu/en/search?tx_tibsearch_search%5Baction%5D=download&tx_tibsearch_search%5Bcontroller%5D=Download&tx_tibsearch_search%5Bdocid%5D=TIB-KAT%3A510705960&cHash=aa3b3390b9034b29c4611151f641e69a#download-mark
- Gräsel, C., Parchmann, I., Puhl, T., Baer, A., Fey, A. & Demuth, R. (2004). Lehrerfortbildung und ihre Wirkung auf die Zusammenarbeit von Lehrkräften und die Unterrichtsqualität. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 133-151). Münster: Waxmann.
- Greiner, K. (2010). *Projektunterricht in der gewerblichen Berufsschule. Eine Fallstudie*. Dissertation, Universität Koblenz-Landau. Landau. Zugriff am 20.12.2019. Verfügbar unter https://kola.opus.hbz-nrw.de/opus45-kola/frontdoor/deliver/index/docId/454/file/Dissertation_Endfassung_gesamt_01_11_2010.pdf

- Greiwe, C. (2020). *Kompetenzen im Nachhaltigkeitsmanagement. Eine Interventionsstudie mit angehenden kaufmännischen Lehrkräften* (Berufsbildung, Arbeit und Innovation. Dissertationen/Habilitationen, Bd. 61). Bielefeld: wbv.
- Grohmann, A. & Kauffeld, S. (2013). Evaluating training programs: development and correlates of the Questionnaire for Professional Training Evaluation. *International Journal of Training and Development*, 17 (2), 135-155.
- Gröschner, A. (2012). Langzeitpraktika in der Lehrerinnen- und Lehrerausbildung –Für und wider ein innovatives Studienelement im Rahmen der Bologna-Reform. *Beiträge zur Lehrerbildung*, 30 (2), 200-208.
- Gröschner, A. & Hascher, T. (2019). Praxisphasen in der Lehrerinnen- und Lehrerbildung. In M. Haring, C. Rohlf & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (UTB Schulpädagogik, Bd. 8698, S. 652-664). Münster: Uni-Taschenbücher GmbH; Waxmann Verlag.
- Grosse-Beck, R. (1998). Was hat Innovation im Prüfungswesen mit den Ewiggestrigen bei der PAL gemein? *Gewerkschaftliche Bildungspolitik* (1-2), 3-5.
- Grotzer, T. A. (2002). Expanding Our Vision for Educational Technology: Procedural, Conceptual, and Structural Knowledge. *Educational Technology*, 42 (2), 52-59. Zugriff am 14.01.2020. Verfügbar unter <https://www.jstor.org/stable/pdf/44428737.pdf>
- Gruber, H., Mandl, H. & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze* (S. 139-156). Göttingen: Hogrefe.
- Halbheer, U. & Reusser, K. (2008). Outputsteuerung, Accountability, Educational Governance –Einführung in Geschichte, Begrifflichkeiten und Funktionen von Bildungsstandards. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 26 (3), 253-266. Zugriff am 24.04.2020. Verfügbar unter https://www.pedocs.de/volltexte/2017/13677/pdf/BZL_2008_3_253_266.pdf
- Hämäläinen, R., Cincinato, S., Malin, A. & Wever, B. de. (2014). VET workers' problem-solving skills in technology-rich environments: European approach. *International Journal for Research in Vocational Education and Training*, 1 (1), 57-80.
- Hanft, A. & Müskens, W. (o. A.). *Kompetenzbezogene Erfolgskontrollen internetgestützten Lernens. Gutachten des Projektes „Lernerfolgskontrolle und Zertifizierung des Lernens im*

- Netz und mit Multimedia unter dem Aspekt der Lernkulturveränderung und Kompetenzentwicklung“*. Zugriff am 17.07.2020. Verfügbar unter [https://uol.de/f/1/inst/paedagogik/ab/web/download/service/Endversion_Gutachten\(2\).pdf](https://uol.de/f/1/inst/paedagogik/ab/web/download/service/Endversion_Gutachten(2).pdf)
- Hartig, J. & Klieme, E. (2006a). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Mit 18 Tabellen. Heidelberg: Springer Medizin Verlag.
- Hartig, J. & Klieme, E. (2006b). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Mit 18 Tabellen. Heidelberg: Springer Medizin Verlag. Verfügbar unter https://link.springer.com/content/pdf/10.1007/3-540-33020-8_9.pdf
- Harvey, P. (2005). *Motivating factors influencing teachers' engagement in postgraduate study: The results of a study of five schools*. A paper presented at the Australian Association for Research in Education Conference 2005. Zugriff am 30.10.2022. Verfügbar unter <https://www.aare.edu.au/data/publications/2005/har05249.pdf>
- Hascher, T. (2008). Diagnostische Kompetenz im Lehrerberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung* (S. 71-86). Münster: Waxmann.
- Hassler, B. (2007). Anregungen zur Prüfungspraxis in der deutschen dualen Berufsausbildung aus der Perspektive der gewerblich-technischen Berufsausbildungsforschung. In P. Grollmann, K. Luomi-Messerer, M.-L. Stenström & H. Tutschner (Hrsg.), *Praxisbegleitende Prüfungen und Beurteilungen in der beruflichen Bildung in Europa* (Bildung und Arbeitswelt, Bd. 18, S. 193-220). Wien: Lit-Verl.
- Held, H. (2014). Ungleichheit und Leistung - ein Spannungsverhältnis? Eine ideologiekritische Betrachtung mit Blick auf Schule. In G. Birkgit & M. Demmer (Hrsg.), *Leistung zwischen Anspruch und Wirklichkeit. Wohin geht Schule?* (S. 131-148). Stuttgart: Raabe.
- Helmke, A. (2002). Kommentar: Unterrichtsqualität und Unterrichtsklima: Perspektiven und Sackgassen. *Zeitschrift für Lernforschung*, 30 (3), 261-277. Zugriff am 08.11.2022. Verfügbar unter https://www.pedocs.de/volltexte/2013/7689/pdf/UnterWiss_2002_3_Helmke_Kommentar.pdf

- Helmke, A. (2021). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (Schule weiterentwickeln, Unterricht verbessern Orientierungsband, Aktualisierte 8. Auflage berücksichtigt die Hattie-Studien). Hannover: Klett / Kallmeyer.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung. Voraussetzungen, Bedingungen, Erfahrungen* (S. 119-143). Baltmannsweiler: Schneider-Verl. Hohengehren.
- Helmke, A. & Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge zur Lehrerbildung*, 31 (2), 214-233.
- Hewlett, C. & Kahl-Andresen, A. (2014). Prüfungsökonomie statt Prüfungsqualität? *Berufsbildung in Wissenschaft und Praxis* (3), 6-9. Zugriff am 14.09.2022. Verfügbar unter <https://d-nb.info/1051616956/34>
- Hollmann, C., Kiedrowski, M. von, Lorig, B. & Schürger, B. (2021). 2.2.334 - *Das Prüfungswesen in der digitalen Transformation: Status quo und Entwicklungsperspektiven. Abschlussbericht*. Laufzeit III/2019 bis IV/2020. Bonn: Bundesinstitut für Berufsbildung.
- Holtsch, D. (2011). Fachdidaktische Kompetenz (künftiger) Lehrender im kaufmännischen Bereich. In U. Faßhauer, B. Fürstenau & E. Wuttke (Hrsg.), *Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung* (S. 21-34). Opladen: Verlag Barbara Budrich. Zugriff am 26.06.2022. Verfügbar unter https://www.pedocs.de/volltexte/2013/7065/pdf/Fasshauer_Grundlagenforschung_2011_Holtsch_Fachdidaktische_Kompetenz.pdf
- Hopf, C. (1995). Befragungsverfahren. In U. Flick, E. von Kardorff, H. Keupp, L. von Rosenstiel & S. Wolff (Hrsg.), *Handbuch Qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen* (2. Aufl., S. 177-182). Weinheim: Beltz.
- Horstkemper, M. (2004). Diagnosekompetenz als Teil pädagogischer Professionalität. *Neue Sammlung*, 44 (2), 201-214.
- Huber, P. J. (1981). *Robust statistics* (Wiley series in probability and mathematical statistics). New York, NY: John Wiley & Sons.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* (6. Aufl.). Weinheim: Beltz Verlagsgruppe.

- Ingvarson, L., Meiers, M., Beavis & Adrian. (2005). Factors Affecting the Impact of Professional Development Programs On Teachers' Knowledge, Practice, Student Outcomes & Efficacy. *Education Policy Analysis Archives*, 13 (10), 1-28.
- Johannes Gutenberg-Universität Mainz (Hrsg.). (2018). *Modulhandbuch. Master of Science Wirtschaftspädagogik*. gültig ab WiSe 2018/19, Mainz.
- Johannes Gutenberg-Universität Mainz (Hrsg.). (2019). *Modulhandbuch Master of Education Wirtschaftspädagogik*. (gültig am WiSe 2018/19). Stand: 30.10.2019, Mainz.
- Johannmeyer, K. & Cramer, C. (2021). Nachfrage und Auslastung von Fortbildungen für Lehrerinnen und Lehrer: Effekte struktureller und inhaltlicher Angebotsmerkmale. *Zeitschrift für Erziehungswissenschaft*, 24 (5), 1183-1204.
- Jonassen, D. H. (1997). Instructional Design Models for Well-Structured and Ill-Structured Problem-Solving Learning Outcomes. *Educational Technology Research and Development*, 45 (1), 65-94. Zugriff am 14.01.2020. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/BF02299613.pdf>
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48 (4), 63-85. Zugriff am 03.01.2020. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/BF02300500.pdf>
- Jonassen, D. H. (2011). *Learning to solve problems. A handbook for designing problem-solving learning environments*. New York: Routledge. Zugriff am 02.01.2020. Verfügbar unter <https://ebookcentral.proquest.com/lib/senc/reader.action?docID=574578>
- Jonassen, D. H. (2012). Problem Typology. In N. M. Seel (Hrsg.), *Encyclopedia of the sciences of learning* (Springer reference, S. 2683-2686). New York: Springer.
- Jonassen, D. H., Beissner, K. & Yacci, M. (1993). *Structural knowledge. Techniques for Representing, Conveying, and Acquiring Structural Knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Jonassen, D. H. & Hung, W. (2008). All Problems are Not Equal: Implications for Problem-Based Learning. *Interdisciplinary Journal of Problem-Based Learning*, 2 (2), 6-28. Zugriff am 11.01.2020. Verfügbar unter <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1080&context=ijpbl>
- Jonassen, D. H. & Hung, W. (2012). Problem Solving. In N. M. Seel (Hrsg.), *Encyclopedia of the sciences of learning* (Springer reference, S. 2680-2683). New York: Springer.

- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M. et al. (2006). *Klassifikationsschema für Mathematikaufgaben. Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt* (Materialien aus der Bildungsforschung, Bd. 81). Berlin: Max-Planck-Institut für Bildungsforschung.
- Jungjohann, J. & Gebhardt, M. (2022). Bezugsnormen im Unterricht. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 25-32). Version 1.0.
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule* (Pädagogik, Band 27). Weinheim und Basel: Beltz.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern* (Empirische Erziehungswissenschaft, Bd. 35). Münster: Waxmann.
- Kirkpatrick, D. L. (2007). *The Four Levels of Evaluation. Tips, Tools, and Intelligence for Trainers* (Infoline Ser). Measurement and Evaluation [Online-Ressource]. Alexandria: American Society for Training & Development.
- Kirkpatrick, D. L. & Kirkpatrick, J. D. (2005). *Transferring learning to behavior. Using the four levels to improve performance*. San Francisco: Berrett-Koehler Publishers.
- Kirkpatrick, D. L. & Kirkpatrick, J. D. (2006). *Evaluating training programs. The four levels* (3. Aufl.). San Francisco: Berrett-Koehler.
- Klein, H. (2009). *Transparente Leistungsbewertung und Zensurengebung in der Sekundarstufe I. Umfassend - praxisorientiert - fair*. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Klein, H. P. (2016). 10 Jahre Bildungsstandards und Kompetenzorientierung - Innovation oder Scheininnovation? In K. Zierer, J. Kahlert & M. Burchardt (Hrsg.), *Die pädagogische Mitte. Plädoyers für Vernunft und Augenmaß in der Bildung* (63-78). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* (2. Aufl.) (Bundesministerium für Bildung und Forschung, Hrsg.) (Bildungsreform Band 1), Berlin. Zugriff am 21.01.2023. Verfügbar unter http://sinus-transfer.uni-bayreuth.de/fileadmin/MaterialienBT/Expertise_Bildungsstandards.pdf

- Klieme, E., Funke, J., Leutner, D., Reimann, P. & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47 (2), 179-200. Zugriff am 28.05.2020. Verfügbar unter https://www.pedocs.de/volltexte/2012/5272/pdf/ZfPaed_2_2001_Klieme_Funke_Leutner_Reimann_Wirth_Problemloesen_D_A.pdf
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52 (6), 876-903.
- Klieme, E. & Tippelt, R. (Hrsg.). (2008a) Qualitätssicherung im Bildungswesen [Themenheft]. *Zeitschrift für Pädagogik* (53). Weinheim: Beltz.
- Klieme, E. & Tippelt, R. (2008b). Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz. In E. Klieme & R. Tippelt (Hrsg.) Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz. *Zeitschrift für Pädagogik*. (53), 7-13 [Themenheft]. Weinheim: Beltz.
- Klößner, J. & Friedrichs, J. (2014). Gesamtgestaltung des Fragebogens. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (Springer VS Handbuch, S. 675-685). Wiesbaden: Springer VS.
- Klotz, V. K. (2015). *Diagnostik beruflicher Kompetenzentwicklung. Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne* (Research). Wiesbaden: Springer.
- Klotz, V. K. & Winther, E. (2012). Kompetenzmessung in der kaufmännischen Berufsausbildung: Zwischen Prozessorientierung und Fachbezug. Eine Analyse der aktuellen Prüfungspraxis. *Berufs- und Wirtschaftspädagogik - online* (22). Zugriff am 10.09.2019. Verfügbar unter www.bwpat.de/ausgabe22/klotz_winther_bwpat22.pdf
- Klotz, V. K. & Winther, E. (2016). Kompetenzorientierte Assessments in der beruflichen Bildung - Zur Grundlegung eines kompetenzorientierten Prüfungswesens. In A. Dietzen, R. Nickolaus, B. Rammstedt & R. Weiß (Hrsg.), *Kompetenzorientierung. Berufliche Kompetenzen entwickeln, messen und anerkennen* (Berichte zur beruflichen Bildung, S. 173-188). Bielefeld: W. Bertelsmann Verlag GmbH & Co. KG.
- Knight, P. & Yorke, M. (2003). *Assessment, learning and employability*. Maidenhead: Society for Research into Higher Education & Open University Press.

- Koch-Priewe, B., Köker, A., Seifried, J. & Wuttke, E. (Hrsg.). (2015). *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Körper, A. (2010). Kompetenzorientierung versus Inhalte. Eine alte Debatte zu neuem Thema. *Schulmanagement* (6), 8-11. Zugriff am 16.07.2020. Verfügbar unter https://www.pedocs.de/volltexte/2014/9596/pdf/Schulmanagement_2010_6_Koerber_Kompetenzorientierung_vs_Inhalte.pdf
- Krainer, K. & Posch, P. (2010). Intensivierung der Nachfrage nach Lehrerfortbildung. Vorschläge für Bildungspraxis und Bildungspolitik. In F. H. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 479-495). Münster: Waxmann Verlag.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M. et al. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 31-53). Münster: Waxmann.
- Krille, C. (2017). Prospective Teachers' Training Motivation and Its Influence on Training Success. In E. Wuttke & J. Seifried (Hrsg.), *Professional Error Competence of Preservice Teachers. Evaluation and Support* (SpringerBriefs in Education, S. 99-114). Cham: Springer International Publishing; Imprint; Springer.
- Krille, C. (2019). *Fortbildungsmotivation von (wirtschaftspädagogischen) Lehrkräften*. Inaugural-Dissertation, Johann Wolfgang Goethe-Universität Frankfurt am Main. Frankfurt am Main.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (Grundlagentexte Methoden, 3. Aufl.). Weinheim: Beltz Juventa.
- Kultusministerkonferenz [KMK] (Hrsg.). (2002). *Rahmenlehrplan. für den Ausbildungsberuf Industriekaufmann/Industriekauffrau*. Beschluss der Kultusministerkonferenz vom 14.06.2002.

- Kultusministerkonferenz [KMK] (Hrsg.). (2013). *Rahmenlehrplan. für den Ausbildungsberuf Kaufmann für Büromanagement und Kauffrau für Büromanagement*. Beschluss der Kultusministerkonferenz vom 27.09.2013.
- Kunter, M. (2010). Modellierung von Lehrerkompetenzen. Kommentierung der Projektdarstellungen. In E. Klieme, D. Leutner & M. Kenk (Hrsg.) *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Zeitschrift für Pädagogik*. (56), 307-312 [Themenheft]. Weinheim: Beltz. Verfügbar unter https://www.pedocs.de/volltexte/2010/3442/pdf/Kunter_Review_D_A.pdf
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 55-68). Münster: Waxmann.
- Kunter, M., Klusmann, U. & Baumert, J. (2009). Professionelle Kompetenz von Mathematiklehrkräften: Das COACTIV-Modell. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. H. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (Beltz-Bibliothek, S. 153-165). Weinheim: Beltz Verlag.
- Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M. et al. (2007). Linking aspects of teacher competence to their instruction: Results from the COACTIV project. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG priority programme* (BIQUA, pp. 39-60). Münster: Waxmann.
- Kunter, M., Pohlmann, B. & Decker, A.-T. (2021). Lehrkräfte. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (3. Aufl., S. 269-288). Berlin: Springer.
- Küsters, I. (2009). *Narrative Interviews. Grundlagen und Anwendungen* (Hagener Studientexte zur Soziologie, 2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, Wiesbaden.
- Kuwan, H., Bilger, F., Gnahs, D. & Seidel, S. (2006). *Berichtssystem Weiterbildung IX. Integrierter Gesamtbericht zur Weiterbildungssituation in Deutschland* (Bundesministerium für Bildung und Forschung, Hrsg.), Berlin. Zugriff am 02.11.2022. Verfügbar unter <https://doku.iab.de/externe/2006/k060217f03.pdf>
- La Marca, P. M., Redfield, D. & Winter, P. C. (2000). *State Standards and State Assessment Systems: A Guide to Alignment. Series on Standards and Assessments*. Washington, DC.:

- Council of Chief State School Officers; Office of Elementary and Secondary Education (ED).
- Labusch, G. & Müller, H.-J. (2014). Prüfungen: Eine unerledigte Herausforderung. Veränderte Anforderungen an die Berufsbildung erfordern andere Prüfungen. *denk-doch-mal.de* (4). Zugriff am 24.01.2023. Verfügbar unter <https://bibb-dspace.bibb.de/rest/bitstreams/fffa8f51-8580-4e55-97d2-fba8df450fe4/retrieve>
- Lamnek, S. (2010). *Qualitative Sozialforschung. Lehrbuch; mit Online-Materialien* (Grundlagen Psychologie, 5. Aufl.). Weinheim: Beltz.
- Lamnek, S. & Krell, C. (2016). *Qualitative Sozialforschung. Mit Online-Materialien* (6. Aufl.). Weinheim: Beltz.
- Land Baden-Württemberg, vertreten durch das Institut für Bildungsanalysen Baden-Württemberg. (2022). *Leitfaden. zur Erstellung von Aufgaben- und Lösungsvorschlägen für die gemeinsame schriftliche Abschlussprüfung*. Berufsfeld Wirtschaft und Verwaltung (2. Aufl.), Stuttgart. Zugriff am 09.05.2022. Verfügbar unter <https://www.schule-bw.de/faecher-und-schularten/berufliche-schularten/berufsschule/lernfelder/wirtschaft-und-verwaltung/leitfaden-schriftliche-abschlusspruefung-berufsfeld-wuv.pdf>
- Landfried, K. & Senger, U. (2004). Neue Lehr- und Lernformen. Lehren und Lernen im Zeichen ganzheitlicher Persönlichkeitsentwicklung. In W. Benz, J. Kohler & K. Landfried (Hrsg.), *Handbuch Qualität in Studium, Lehre und Forschung*. Berlin: Raabe Verlag.
- Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33 (1), 159-174. Zugriff am 28.12.2020. Verfügbar unter https://www.jstor.org/stable/2529310?seq=7#metadata_info_tab_contents
- Lankes, E.-M. (2008). Einführung. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 41-45). Münster: Waxmann.
- Lempert, W. & Achtenhagen, F. (2000). Entwicklung eines Programmkonzepts „Lebenslanges Lernen“. *Unterrichtswissenschaft*, 28 (2), 144-159. Zugriff am 26.10.2019. Verfügbar unter https://www.pedocs.de/volltexte/2013/5468/pdf/UnterWiss_2000_2_Lempert_Achtenhagen_Entwicklung_Programmkonzept_D_A.pdf
- Leonard, J. (30.06.2021). *Referendariat in Niedersachsen* (E-Mail).
- Leuphana Universität Lüneburg (Hrsg.). (2019, 24. September). *Gazette. Amtliches Mitteilungsblatt der Körperschaft und der Stiftung* (51/19), Lüneburg.

- Leutner, D. (2010). Pädagogisch-psychologische Diagnostik. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 624-635). Weinheim: Beltz.
- Lewrick, M. & Link, P. (2018). Design Thinking: Mit dem richtigen Mindset die Zukunft gestalten. *Wirtschaftswissenschaftliches Studium* (7-8), 51-56. Zugriff am 26.10.2019. Verfügbar unter https://elibrary.vahlen.de/10.15358/0340-1650-2018-7-8-51.pdf?download_full_pdf=1
- Lipowsky, F. (2004). Was macht Fortbildungen für Lehrkräfte erfolgreich? Befunde der Forschung und mögliche Konsequenzen für die Praxis. *Die Deutsche Schule*, 96 (4), 462-479. Zugriff am 04.02.2020. Verfügbar unter http://www.digizeitschriften.de/download/PPN509092632_0096/PPN509092632_0096_log102.pdf
- Lipowsky, F. (2009). Unterrichtsentwicklung durch Fort- und Weiterbildungsmaßnahmen für Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 27 (3), 346-360. Zugriff am 05.02.2020. Verfügbar unter https://www.pedocs.de/volltexte/2017/13705/pdf/BZL_2009_3_346_360.pdf
- Lipowsky, F. (2010). Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildungen. In F. H. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 51-70). Münster: Waxmann Verlag.
- Lipowsky, F. & Rzejak, D. (2012). Lehrerinnen und Lehrer als Lerner - Wann gelingt der Rollentausch? - Merkmale und Wirkungen wirksamer Lehrerfortbildungen. In D. Bosse, L. Criblez & T. Hascher (Hrsg.), *Reform der Lehrerbildung in Deutschland, Österreich und der Schweiz. Teil 1. Analysen, Perspektiven und Forschung* (Theorie und Praxis der Schulpädagogik, Band 4, S. 235-253). Immenhausen bei Kassel: Prolog-Verlag.
- Lipowsky, F. & Rzejak, D. (2021). *Fortbildungen für Lehrpersonen wirksam gestalten. Ein praxisorientierter und forschungsgestützter Leitfaden*. Gütersloh: Bertelsmann Stiftung.
- Lissmann, U. (2010a). Leistungsbeurteilung gestern, heute, morgen. *EEO Enzyklopädie Erziehungswissenschaft Online*. Zugriff am 02.11.2020. Verfügbar unter https://content-select.com/media/moz_viewer/5282487a-1744-4c50-9d5f-11372efc1343/language:de
- Lissmann, U. (2010b). Schultests. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 737-750). Weinheim: Beltz.
- Lohmeyer, N. & Velten, S. (2019). Forschungs- und Transferinitiative ASCOT+ startet mit sechs Projekten. *Berufsbildung in Wissenschaft und Praxis*, 48 (6), 16-18.

- Lorig, B., Bretschneider, M., Gutschow, K. & Mpangara, Mirim, Weber-Höller, Robin. (2014). *Kompetenzbasierte Prüfungen im dualen System – Bestandsaufnahme und Gestaltungsperspektiven. Abschlussbericht*. Forschungsprojekt 4.2.333 (JFP 2010). Bonn: Bundesinstitut für Berufsbildung. Zugriff am 24.01.2023. Verfügbar unter https://www.bibb.de/dienst/dapro/daprodocs/pdf/eb_42333.pdf
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft* (13), 293-308. Zugriff am 02.05.2020. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/s11618-010-0124-9.pdf>
- Maier, U. (2012). *Lehr-Lernprozesse in der Schule: Studium. Allgemeindidaktische Kategorien für die Analyse und Gestaltung von Unterricht* (utb-studi-e-book, Bd. 3767, 1. Aufl.). Stuttgart: UTB; Klinkhardt.
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht. Schülerleistungen messen, bewerten und fördern* (utb-studi-e-book, Bd. 4178). Bad Heilbrunn: Verlag Julius Klinkhardt; UTB.
- Markowitsch, J., Messerer, K. & Prokopp, M. (2004). *Handbuch praxisorientierter Hochschulbildung* (Schriftenreihe des Fachhochschulrates, Bd. 10). Wien: WUV-Universitätsverlag.
- Mayring, P. (2012). Mixed Methods - ein Plädoyer für gemeinsame Forschungsstandards qualitativer und quantitativer Methoden. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung* (S. 287-300). 74. Tagung der Arbeitsgruppe Empirische Pädagogische Forschung (AEPF) im September 2010 in Jena. Münster: Waxmann.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Mayring, P. (2016). *Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken* (Beltz Studium, 6. Aufl.). Weinheim: Beltz, J.
- McCarthy, J. (1956). The Inversion of Functions Defined by Turing Machines. In C. E. Shannon & J. McCarthy (Hrsg.), *Atomata Studies* (Annals of Mathematical Studies, Bd. 34, S. 177-181). Princeton, NJ: Princeton University Press. Zugriff am 29.12.2019. Verfügbar unter <http://jmc.stanford.edu/articles/inversion/inversion.pdf>

- McMillan, J. H. & Schumacher, S. (2010). *Research in education. Evidence-Based Inquiry* (7. Aufl.). Boston, Mass.: Pearson Education.
- Mercator School of Management. (o. A.). *Studienverlaufsplan des Bachelorstudiengangs Wirtschaftspädagogik. inkl. der Wahlmöglichkeiten in den kleinen beruflichen Fachrichtungen*. Studienjahr 2019/20.
- Mercator School of Management. (2019a). Bachelor of Science in Wirtschaftspädagogik. Modulhandbuch. Version PO 17 (Universität Duisburg-Essen, Hrsg.), Duisburg.
- Mercator School of Management. (2019b). *Master of Education in Wirtschaftspädagogik - Lehramt an Berufskollegs. Modulhandbuch*. Stand: Wintersemester 2019/20 (Universität Duisburg-Essen, Hrsg.), Duisburg.
- Meuser, M. & Nagel, U. (2002). ExpertInneninterviews - vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In A. Bogner, B. Littig & W. Menz (Hrsg.), *Das Experteninterview. Theorie, Methode, Anwendung* (S. 71-93). Opladen: Leske + Budrich; Springer Fachmedien Wiesbaden GmbH.
- Ministerium der Justiz Rheinland-Pfalz (Hrsg.). (2012). *Landesverordnung über die Ausbildung und Zweite Staatsprüfung für das Lehramt an Grundschulen, an Realschulen plus, an Gymnasien, an berufsbildenden Schulen und an Förderschulen. Vom 3. Januar 2012*. Zugriff am 25.10.2021. Verfügbar unter <http://landesrecht.rlp.de/jportal/portal/t/ycl/page/bsrlpprod.psml?doc.hl=1&doc.id=jlr-SchulLehr2StPrVRPrahmen%3Ajuris-lr00&documentnumber=1&numberofresults=217&showdoccase=1&doc.part=X¶mfromHL=true>
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg (Hrsg.). (2016). *Leitlinien zur Pädagogischen Schulung. Technische Lehrerinnen und Lehrer (gewerblich, kaufmännisch, hauswirtschaftlich) im gehobenen Schuldienst*, Stuttgart.
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg (Hrsg.). (2018). *Fachdidaktikstandards. Vorbereitungsdienst und pädagogische Schulung für den Schuldienst an beruflichen Schulen*.
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg & Landesinstitut für Schulentwicklung (Hrsg.). (2016). *Leistungsfeststellung, Leistungsbeurteilung und Leistungsrückmeldung. an Gemeinschaftsschulen*. Zugriff am 26.05.2020. Verfügbar unter

https://www.km-bw.de/site/pbs-bw-new/get/documents/KULTUS.Dachmandant/KULTUS/KM-Homepage/Artikelseiten%20KP-KM/1_PDFS_2016/GMS/Leistungsbeurteilung%20GMS_13_06_16.pdf

- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (Hrsg.). (2021). *Kerncurriculum für die Lehrerausbildung im Vorbereitungsdienst. Verbindliche Zielvorgabe der schulpraktischen Lehrerausbildung in Nordrhein-Westfalen*, Düsseldorf.
- Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2003). *A brief introduction to Evidence-centered design*. Princeton, NJ: Educational Testing Service. Zugriff am 15.08.2021. Verfügbar unter <https://www.ets.org/Media/Research/pdf/RR-03-16.pdf>
- Mislevy, R. J. & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, 25 (4), 6-20. Zugriff am 15.08.2021. Verfügbar unter <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1745-3992.2006.00075.x>
- Mislevy, R. J. & Riconscente, M. M. (2005). *Evidence-Centered Assessment Design: Layers, Structures, and Terminology* (Principled Assessment Design for Inquiry Technical Report Nr. 9).
- Mislevy, R. J., Steinberg, L. S. & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspectives*, 1 (1), 3-62.
- Misoch, S. (2019). *Qualitative Interviews* (2. Aufl.). Berlin: De Gruyter Oldenbourg.
- Moosbrugger, H. & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion. Mit 79 Abbildungen und 43 Tabellen* (lehrbuch-psychologie.de, S. 7-26). Heidelberg: Springer Medizin Verlag.
- Morningstar, M. E. & Benitez, D. T. (2013). Teacher Training Matters. The Results of a Multi-State Survey of Secondary Special Educators Regarding Transition from School to Adulthood. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 36 (1), 51-64.
- Mulder, M. (2019). The Global Need for Competence: Competence-based VET and Implications for Policy and Practice. In A. Bahl & A. Dietzen (Hrsg.), *Work-based Learning as a Pathway to Competence-based Education. A UNEVOC Network Contribution* (Berichte zur beruflichen Bildung, Bd. 27, 1. Auflage, S. 44-65). Leverkusen: Verlag Barbara Budrich.

- Müller, F. H., Kemethofer, D., Andreitz, I., Nachbaur, G. & Soukup-Altrichter, K. (2019) Lehrerfortbildungen und Lehrerweiterbildungen. In *Nationaler Bildungsbericht Österreich 2019. Fokussierte Analysen und Zukunftsperspektiven für das Bildungswesen* (S. 99-142). Graz. Bundesministerium Bildung, Wissenschaft und Forschung; Bundesinstitut für Bildungsforschung, Innovation & Entwicklung.
- Müller, K. & Dieck, M. (2011). Schulpraxis als Lerngelegenheit? Mehrperspektivische empirische Befunde zu einem Langzeitpraktikum. *Journal für Lehrerinnen- und Lehrerbildung* (3), 46-50.
- Nagasaki, E. & Becker, J. P. (1993). Classroom assessment in Japanese mathematics education. In N. L. Webb (Hrsg.), *Assessment in the mathematics classroom* (Yearbook, 2. Aufl., S. 40-53). Reston, Va.: The National Council of Teachers of Mathematics.
- National Conference of State Legislatures. (2011). *National Board for Professional Teaching Standards Certification. What Legislators Need to Know*. Colorado. Zugriff am 06.02.2022. Verfügbar unter <https://www.ncsl.org/Portals/1/Documents/educ/NationalBoard.pdf>
- Neubrand, M., Jordan, A., Krauss, S., Blum, W. & Löwen, K. (2011). Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 112-132). Münster: Waxmann.
- Neuweg, G. H. (2010). Fortbildung im Kontext eines phasenübergreifenden Gesamtkonzepts der Lehrerbildung. In F. H. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 35-49). Münster: Waxmann Verlag.
- Nickolaus, R. (2011). Kompetenzmessung und Prüfungen in der beruflichen Bildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Band 107* (2), 161-173. Zugriff am 24.01.2023. Verfügbar unter <https://biblioscout.net/content/pdf/99.140005/zbw201102016101.pdf>
- Niedersächsisches Kultusministerium (Hrsg.). (2013a). *Seminarlehrpläne. für die Ausbildung von Lehrkräften im Vorbereitungsdienst an den niedersächsischen Studienseminaren für das Lehramt an berufsbildenden Schulen (Beschluss des Niedersächsischen Kultusministeriums vom ... 2012)* (Teil I).

- Niedersächsisches Kultusministerium (Hrsg.). (2013b). *Seminarlehrpläne für die Studienseminare für das Lehramt an berufsbildende Schulen* (Teil II).
- Niegemann, H. & Korbach, A. (2020). *Abschlussbericht zum Projekt LaSiDig*. Gefördert vom BMBF, FKZ 01PZ16007B, unveröffentlicht.
- Niermann, A. (2017). *Professionswissen von Lehrerinnen und Lehrern des Mathematik- und Sachunterrichts. „... man muss schon von der Sache wissen.“* (Klinkhardt forschung). Dissertation. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Nistor, N., Schnurer, K. & Mandl, H. (2005). *Akzeptanz, Lernprozess und Lernerfolg in virtuellen Seminaren – Wirkungsanalyse eines problemorientierten Seminarkonzepts*. Forschungsbericht, Ludwig-Maximilians-Universität München. München.
- Nitko, A. J. (2004). *Educational assessment of students* (4. Aufl.). Upper Saddle River, NJ: Pearson.
- Noe, R. A. & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, 39 (3), 497-523. Verfügbar unter <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1744-6570.1986.tb00950.x>
- Oehmann, K. & Blumschein, P. (2019). *Schluss mit der Donutpädagogik. Lebensnahe Lernaufgaben leicht gemacht* (1. Aufl.). Bern: hep der Bildungsverlag.
- Ohle, A., McElvany, N., Horz, H. & Ullrich, M. (2015). Text-picture integration - Teachers' attitudes, motivation and self-related cognitions in diagnostics. *Journal for educational research online*, 7 (2), 11-33.
- Oser, F. & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur* (Beltz-Pädagogik). Weinheim u.a.: Beltz.
- Otto von Guericke Universität Magdeburg (Hrsg.). (2019). *Modulhandbuch für den Bachelorstudiengang Beruf und Bildung*, Magdeburg.
- Pasternack, P., Baumgarth, B., Burkhardt, A., Paschke, S. & Thielemann, N. (2017). *Drei Phasen. Die Debatte zur Qualitätsentwicklung in der Lehrer_innenbildung*. Bielefeld: W. Bertelsmann Verlag.
- Pellegrino, J. W. (2002-03). Knowing What Students Know. *Issues in Science and Technology*, 19 (2 (Winter 2022-03)), 48-52. Zugriff am 22.02.2023. Verfügbar unter

https://www.jstor.org/stable/pdf/43312299.pdf?refreqid=excelsior%3A345499a9a4cf55039c991dd50c052ebd&ab_segments=&origin=&initiator=&acceptTC=1

- Pellegrino, J. W. (2010). *The Design of an Assessment System for the Race to the Top: A Learning Sciences Perspective on Issues of Growth and Measurement*. *Educational Testing Service*. Zugriff am 11.12.2019. Verfügbar unter <https://pdfs.semanticscholar.org/53b7/6668fec653df7db1261304bd43a4ce64e42d.pdf>
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, 20, 65-77.
- Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academy Press.
- Pereira Kastens, C., Döring-Seipel, E. & Nolle, T. (2020). Selbstwirksamkeit erlangen, Belastung reduzieren? Effekte des Feedbackverhaltens der Ausbilder/innen in Unterrichtsnachbesprechungen. *Jorunal for educational research online*, 12 (1), 67-90.
- Piowar, V., Thiel, F. & Ophardt, D. (2013). Training inservice teachers' competencies in classroom management. A quasi-experimental study with teachers of secondary schools. *Teaching and Teacher Education*, 30, 1-12.
- Qualitätsmanagement in der Lehrerbildung. (2018). *Modulbeschreibungen 1. Ausbildungsjahr*. 18. Juli 2018 (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, Hrsg.), München.
- Qualitätsmanagement in der Lehrerbildung. (2019). *Modulbeschreibungen 2. Ausbildungsjahr*. 22. März 2019 (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, Hrsg.), München.
- Qualitätsmanagement in der Lehrerbildung. (2020a). *Modulbeschreibungen 1. Ausbildungsjahr*. 09. November 2020 (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, Hrsg.), München. Zugriff am 31.12.2020. Verfügbar unter https://studien-seminar.de/inhalte/downloads/Hauptseminar-%20und%20Deutschmodule/Modulbeschreibungen_AJ_1.pdf
- Qualitätsmanagement in der Lehrerbildung. (2020b). *Modulbeschreibungen 2. Ausbildungsjahr*. 09. November 2020 (Staatliches Studienseminar für das Lehramt an beruflichen

- Schulen, Hrsg.), München. Zugriff am 31.12.2020. Verfügbar unter https://studien-seminar.de/inhalte/downloads/Hauptseminar-%20und%20Deutschmodule/Modulbeschreibungen_AJ_2.pdf
- Rasch, B., Friese, M., Hofmann, W. & Naumann (2014). Zweifaktorielle Varianzanalyse. In B. Rasch, M. Friese, W. Hofmann & E. Naumann (Hrsg.), *Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (Springer-Lehrbuch, 4. Aufl., S. 35-63). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2021). *Quantitative Methoden 1. Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl.). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Rausch, A. & Wuttke, E. (2016). Development of a multi-faceted model of domain-specific problem-solving competence and its acceptance by different stakeholders in the business domain. *Unterrichtswissenschaft*, 44 (2), 164-189. Zugriff am 13.01.2020. Verfügbar unter https://www.researchgate.net/publication/299746973_Development_of_a_Multi-Faceted_Model_of_Domain-Specific_Problem-Solving_Competence_and_its_Acceptance_by_Different_Stakeholders_in_the_Business_Domain
- Redenius, N.-H. (30.09.2021). *Ausbildungsverordnung Vorbereitungsdienst* (E-Mail).
- Reetz, L. (2005). Situierete Prüfungsaufgaben. Die Funktion von Situationsaufgaben in Abschlussprüfungen des Dualen Systems der Berufsausbildung. *Berufs- und Wirtschaftspädagogik - online* (8).
- Reichardt, C. S. (2019). *Quasi-Experimentation. A Guide to Design and Analysis* (Methodology in the social sciences). New York: The Guilford Press.
- Reinmann-Rothmeier, G., Mandl, H. & Prenzel, M. (1997). Qualitätssicherung bei multimedialen Lernumgebungen. In H. F. Friedrich, G. Eigler, H. Mandl, W. Schnotz, F. Schott & N. M. Seel (Hrsg.), *Multimediale Lernumgebungen in der betrieblichen Weiterbildung. Gestaltung, Lernstrategien und Qualitätssicherung* (Grundlagen der Weiterbildung, S. 267-333). Neuwied: Luchterhand.
- Renkl, A. (1994). *Die „unerklärliche“ Kluft zwischen Wissen und Handeln*. Forschungsbericht Nr. 41. München: Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie.

- Ressel, T. & Kiesecker, D. (2019). *Die duale Kompetenzprüfung. Konzept zur Weiterentwicklung der Abschlussprüfung zu einem Kompetenznachweis für die Lernorte Schule und Betrieb.* : IG Metall. Zugriff am 11.12.2020. Verfügbar unter https://wap.igmetall.de/docs_Duale_Kompetenzpruefung_Langfassung_c0a0d2a45a0e13c11d72ca159d7ba82d771dcff6.pdf
- Reusser, K. (2014). Kompetenzorientierung als Leitbegriff der Didaktik. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32 (3), 325-339. Zugriff am 22.10.2019. Verfügbar unter https://www.pedocs.de/volltexte/2017/13873/pdf/BZL_2014_3_325_339.pdf
- Reynolds, A. (1995). The Knowledge Base for Beginning Teachers: Education Professionals' Expectations versus Research Findings on Learning to Teach. *The Elementary School Journal*, 95 (3), 199-221. Zugriff am 05.02.2022. Verfügbar unter <https://www.jstor.org/stable/pdf/1001931.pdf>
- Reynolds, A., Tannenbaum, R. J. & Rosenfeld, M. (1992). *Beginning Teacher Knowledge of General Principles of Teaching and Learning: A National Survey* (ETS Research Report Series). Zugriff am 08.02.2022. Verfügbar unter <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/j.2333-8504.1992.tb01491.x>
- Richter, A. & Vigerske, S. (2011). Fort- und Weiterbildung von Lehrerinnen und Lehrern – Die Bedeutung der dritten Phase am Beispiel einer Evaluation der Lehrer/-innenfortbildung im Land Baden-Württemberg. *bwp@ Berufs- und Wirtschaftspädagogik - online* (Spezial 5), 1-15.
- Richter, D., Engelbert, M., Weirich, S. & Pant, H. A. (2013). Differentielle Teilnahme an Lehrerfortbildungen und deren Zusammenhang mit professionsbezogenen Merkmalen von Lehrkräften // Differentielle Teilnahme an Lehrerfortbildungen und deren Zusammenhang mit professionsbezogenen Merkmalen von Lehrkräften*. *Zeitschrift für Pädagogische Psychologie*, 27 (3), 193-207.
- Richter, D., Kunter, M., Klusmann, U., Lüdtke, O. & Baumert, J. (2011). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education*, 27 (1), 116-126.
- Riedl, A. & Schelten, A. (2006, 14. Februar). *Handlungsorientiertes Lernen. Aktuelle Entwicklungen aus der Lehr-Lern-Forschung und deren Anwendung im Unterricht*. Unterlagen für die Teilnehmer der Fortbildung (Technische Universität München, Hrsg.). München: Lehrstuhl für Pädagogik.

- Rosenfeld, M. & Tannenbaum, R. J. (1991). *Identification of Core of Important Enabling Skills for the NTE Successor Stage I Examination* (Educational Testing Service), Princeton. Zugriff am 08.02.2022. Verfügbar unter <https://files.eric.ed.gov/fulltext/ED384661.pdf>
- Rost, D. H. (2007). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung* (Beltz Bibliothek, 2. Aufl.). Weinheim: Beltz PVU.
- Rzejak, D., Lipowsky, F. & Bleck, V. (2020). Synopse bedeutsamer Merkmale von Lehrkräftefortbildungen. Wirkungsebenen und konzeptionelle Aspekte. *journal für lehrerInnenbildung*, 20 (4). Zugriff am 05.11.2022. Verfügbar unter https://www.pedocs.de/volltexte/2021/21372/pdf/jlb_2020_4_Rzejak_et_al_Synopse_bedeutsamer_Merkmale.pdf
- Sälzer, C., Reiss, K., Schiepe-Tiska, A., Prenzel, M. & Heinze, A. (2013). Zwischen Grundwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 47-97). Münster: Waxmann.
- Sauer, J. & Gamsjäger, E. (2010). Prognose von Schulerfolg. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 651-663). Weinheim: Beltz.
- Schaefers, C. (2002). Forschung zur Lehrerausbildung in Deutschland – eine bilanzierende Übersicht der neueren empirischen Studien. *Schweizerische Zeitschrift für Bildungswissenschaften*, 24 (1), 65-90. Zugriff am 21.11.2022. Verfügbar unter https://www.pedocs.de/volltexte/2011/3785/pdf/SZBW_2002_H1_S65_Schaefers_D_A.pdf
- Schaper, N. (2012). *Fachgutachten zur Kompetenzorientierung in Studium und Lehre. HRK-Fachgutachten ausgearbeitet für die HRK von Niclas Schaper unter Mitwirkung von Oliver Reis und Johannes Wildt sowie Eva Horvath und Elena Bender* (nexus). o. A.: Bundesministerium für Bildung und Forschung; Hochschulrektorenkonferenz. Zugriff am 22.10.2019. Verfügbar unter https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/fachgutachten_kompetenzorientierung.pdf
- Schlicht, J. (2012). *Kosten-Nutzen-Analyse von beruflicher Weiterbildung. Untersuchungen zur Wirtschaftlichkeit und pädagogischen Wirksamkeit* (Gabler Research, 1. Aufl.). Wiesbaden: Gabler Verlag / Springer Fachmedien Wiesbaden GmbH Wiesbaden.
- Schlömerkemper, J. (2014). Leistungsmessung und die Professionalität des Lehrerberufs. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3. Aufl., S. 311-321). Weinheim: Beltz.

- Schmidt, H. (2010) Zur Reform der Abschlussprüfungen in der kaufmännischen Berufsausbildung am Beispiel des Ausbildungsberufs Kaufmann/ Kauffrau für Spedition und Logistikdienstleistung. In D. Frommberger (Hrsg.), *Magdeburger Schriften zur Berufs- und Wirtschaftspädagogik* Otto-von-Guericke-Universität Magdeburg. Verfügbar unter https://www.bwp.ovgu.de/bwp_media/Neu_Magdeburger+Schriften/Jahrgang+2010/Heft+7+_+2010.pdf
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 102-108). Weinheim: Beltz.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31 (2), 154-165. Verfügbar unter https://www.pedocs.de/volltexte/2017/13843/pdf/BZL_2013_2_154_165.pdf
- Schrader, F.-W. & Helmke, A. (2014). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3. Aufl., S. 45-58). Weinheim: Beltz.
- Schröder, H., Thienen, L. von, Müller, A. & Homann-Vorderbrück, S. (2020). Arbeitskreise im IT- und Prozessmanagement: Erfolgsfaktoren eines gelungenen Theorie-Praxis-Transfers. *HMD Praxis der Wirtschaftsinformatik*, 57 (2), 244-256.
- Schütze, B., Souvignier, E. & Hasselhorn, M. (2008). Stichwort - formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21 (4), 697-715. Zugriff am 02.06.2020. Verfügbar unter https://www.pedocs.de/volltexte/2019/16754/pdf/Hasselhorn_2018_Formatives_Assessment_A.pdf
- Schwaighofer, M., Heene, M. & Bühner, M. (2019). Grundlagen und Kriterien der Diagnostik. In D. Urhahne, M. Dresel & F. Fischer (Hrsg.), *Psychologie für den Lehrberuf* (S. 471-491). Berlin: Springer-Verlag.
- Seeber, S. & Nickolaus, R. (2010). Kompetenzmessung in der beruflichen Bildung. *Berufsbildung in Wissenschaft und Praxis*, 30 (1), 10-13.
- Seeber, S., Weber, S., Geiser, P., Zarnow, S., Hackenberg, T. & Hiller, F. (2019). Effekte der Digitalisierung auf kaufmännische Tätigkeiten und Sichtweisen ausgewählter Akteure. *Berufsbildung*, 73 (176), 2-7.
- Seel, N. M. (2012). Problems: Definition, Types, and Evidence. In N. M. Seel (Hrsg.), *Encyclopedia of the sciences of learning* (Springer reference). New York: Springer.

- Seifried, J. & Wuttke, E. (2017). Teacher Training as a Contribution to Teachers' Professional Development: Conclusions from a Research Programme to Foster Professional Error Competence in Accounting. In E. Wuttke & J. Seifried (Hrsg.), *Professional Error Competence of Preservice Teachers. Evaluation and Support* (SpringerBriefs in Education, S. 115-120). Cham: Springer International Publishing; Imprint; Springer.
- Seifried, J., Wuttke, E., Kögler, K., Rausch, A., Eigenmann, R., Küster, J. et al. (2015). *Abchlusspublikation ASCOT „Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse und Bedeutung für Politik und Praxis*.
- Seifried, J., Wuttke, E., Türling, J., Krille, C. & Paul, O. (2015). Teachers' strategies for handling student errors. The contribution of teacher training programs. In M. Gartmeier, H. Gruber, T. Hascher & H. Heid (Hrsg.), *Fehler. Ihre Funktionen im Kontext individueller und gesellschaftlicher Entwicklung = Errors : their functions in context of individual and societal development* (S. 177-188). Münster: Waxmann.
- Sekretariat der Kultusministerkonferenz (Hrsg.). (2007). *Handreichung. für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe*, Bonn. Zugriff am 08.09.2019. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2007/2007_09_01-Handreich-Rlpl-Berufsschule.pdf
- Sekretariat der Kultusministerkonferenz (Hrsg.). (2018). *Handreichung. für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe*, Berlin. Zugriff am 24.12.2019. Verfügbar unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2011/2011_09_23-GEP-Handreichung.pdf
- Sekretariat der Kultusministerkonferenz (Hrsg.). (2019). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019, Berlin. Zugriff am 04.03.2022. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf
- Sekretariat der Kultusministerkonferenz (Hrsg.). (2021a). *Handreichung. für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz für den berufsbezogenen Unterricht in*

der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe, Berlin. Zugriff am 09.02.2023. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_06_17-GEP-Handreichung.pdf

Sekretariat der Kultusministerkonferenz (Hrsg.). (2021b). *Positionspapier der Kultusministerkonferenz zur Weiterentwicklung der Abschlussprüfung in der dualen Berufsausbildung*. (Beschluss der Kultusministerkonferenz vom 09.09.2021), Berlin. Zugriff am 25.01.2023. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_09_09-Abschlusspruefung-duale-Berufsausbildung.pdf

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (1998). *Überlegungen der Kultusministerkonferenz zur Weiterentwicklung der Berufsbildung*. (verabschiedet von der Kultusministerkonferenz am 23.10.1998). Zugriff am 09.06.2021. Verfügbar unter https://www.kmk.org/fileadmin/Daten/pdf/PresseUndAktuelles/Beschluesse_Veroeffentlichungen/uebb98-10-23.pdf

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. (Beschluss der Kultusministerkonferenz vom 16.12.2004), o. A. Zugriff am 30.10.2020. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2009). *Erklärung der Kultusministerkonferenz zur zukünftigen Stellung der Berufsschule in der dualen Berufsausbildung*. (Beschluss der Kultusministerkonferenz vom 10.12.2009). Zugriff am 25.01.2023. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2009/2009_12_10-Berufsschule-in-der-dualen_Berufsausbildung.pdf

Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (Hrsg.). (2003). *Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik*. Beschluss der Mitgliederversammlung, Jena. Zugriff am 15.01.2020. Verfügbar unter https://www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2003_Basiscurriculum_BWP.pdf

Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (Hrsg.). (2014). *Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik*. Beschluss der Mitgliederversammlung, Jena. Zugriff am 15.01.2020. Verfügbar unter https://www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf

- schaftspädagogik im Rahmen berufs- und wirtschaftspädagogischer Studiengänge*. Beschluss der Mitgliederversammlung in Schwäbisch-Gmünd am 25.09.2014, Schwäbisch-Gmünd. Verfügbar unter https://www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf
- Sektion Politik - Recht - Wirtschaft; Fachbereich Wirtschaftswissenschaften. (o. A.). *Modulhandbuch. Masterstudiengang Wirtschaftspädagogik* (Universität Konstanz, Hrsg.), Konstanz. Zugriff am 06.01.2023. Verfügbar unter https://www.uni-konstanz.de/uploads/tx_studiengang/en_modulebook_186_1506066866.pdf
- Sembill, D. (1992). *Problemlösefähigkeit, Handlungskompetenz und emotionale Befindlichkeit. Zielgrößen forschenden Lernens* (Ergebnisse der Pädagogischen Psychologie, Bd. 10). Göttingen: Hogrefe.
- Seminar für das Lehramt an Berufskollegs. (2020). *Ausbildungsprogramm. Jahrgang November 2020* (zentrum für schulpraktische Lehrerausbildung Köln, Hrsg.), Köln.
- Severing, E. (2011). Prüfungen und Zertifikate in der beruflichen Bildung: eine Einführung. In E. Severing & R. Weiß (Hrsg.), *Prüfungen und Zertifizierungen in der beruflichen Bildung. Anforderungen - Instrumente - Forschungsbedarf* (S. 15-36). Bielefeld: W. Bertelsmann Verlag.
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15 (2), 4-14.
- Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, 57 (1), 1-23.
- Siegfried, C. (2017). *Professionelles ökonomisches Fachwissen und ökonomische Einstellungen von angehenden Lehrpersonen*. Inaugural-Dissertation, Johann Wolfgang Goethe-Universität Frankfurt am Main. Frankfurt am Main. Zugriff am 08.10.2022. Verfügbar unter https://www.wiwi.uni-frankfurt.de/fileadmin/user_upload/dateien_wirtschaftspaedagogik/Unveroeffentlichte_Manuskripte/Christin_Doktorarbeit_Mantel.pdf
- Siegfried, C., Kögler, K., Rausch, A., Seifried, J., Wuttke, E. & Eigenmann, R. (2019). Individuelle und kontextuelle Einflussfaktoren domänenspezifischer Problemlösekompetenz in der kaufmännischen Ausbildung. *Zeitschrift für Erziehungswissenschaft* (22), 989-1017. Zugriff am 01.02.2020. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/s11618-019-00892-2.pdf>

- Speth, H. (2007). *Theorie und Praxis des Wirtschaftslehreunterrichts. Eine Fachdidaktik* (Wirtschaftswissenschaftliche Bücherei für Schule und Praxis, 9. Aufl.). Rinteln: Merkur Verlag.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1/2), 85-95.
- Staatliches Studienseminar für das Lehramt an beruflichen Schulen (Hrsg.). (o. A.). *Ausbildungsinhalte*. Zugriff am 01.11.2021. Verfügbar unter <https://www.studien-seminar.de/index.php/hauptmenue-vorbereitungsdienst/hauptmenue-festlegungen/hauptmenue-hauptseminar>
- Staatliches Studienseminar für das Lehramt an berufsbildenden Schulen Speyer / Kaiserslautern (Hrsg.). (2018, 09. Mai). *Didaktische Abschnittsplanung. Leitfaden* (1. Aufl.), Speyer. Zugriff am 16.01.2020. Verfügbar unter https://studienseminar.rlp.de/fileadmin/user_upload/studienseminar.rlp.de/bb-sp/Downloadbereich/leitf/Leitfaden_Didaktische_Abschnittsplanung_1.1.pdf
- Stahel, W. (2017). *Statistische Regressionsmodelle. Teil I: Lineare Regression*. Unterlagen zum Teil 1 der Vorlesung/ des Kurses in Angewandter Regression, ETH Zürich. Zürich. Zugriff am 13.11.2022. Verfügbar unter <https://ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/Advanced%20Studies%20in%20Applied%20Statistics/course-material-1921/Regression/reg-script-full.pdf>
- Strübing, J., Hirschauer, S., Ayaß, R., Krähnke, U. & Scheffer, T. (2018). Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Zeitschrift für Soziologie*, 47 (2), 83-100.
- Südkamp, A., Möller, J. & Pohlmann, B. (2008). Der Simulierte Klassenraum. Ein Instrument zur Untersuchung von diagnostischer Kompetenz. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 87-97). Münster: Waxmann.
- Technische Universität Chemnitz. (2018). *Prüfungsordnung für den konsekutiven Studiengang Berufs- und Wirtschaftspädagogik (Vocationomics) mit dem Abschluss Master of Science (M.Sc.) an der Technischen Universität Chemnitz*. Vom 31. Juli 2018 (Nr. 32).
- Technische Universität Dresden. (2016, 11. August). *Anlage 1a SO Bachelor Wipäd: Studienablaufplan Bachelor Wirtschaftspädagogik, Studienrichtung I (Vollzeit, Beginn Wintersemester)*, Technische Universität Dresden. Zugriff am 21.12.2022. Verfügbar unter

https://tu-dresden.de/bu/wirtschaft/ressourcen/dateien/studium/studienangebot/Studienangebot/wipaed_ba/ordnungen/stand-20161025/SO_1_BA-Wipaed-I_VZ_20160908.pdf?lang=de

- Technische Universität Dresden (Hrsg.). (2020). *Modulhandbuch Master-Studiengang Wirtschaftspädagogik*. Fakultätsratsbeschluss vom 19.06.2019 Stand: 04.03.2020.
- Tenorth, H.-E. (2006). Professionalität im Lehrerberuf. Ratlosigkeit der Theorie, gelingende Praxis. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 580-597. Zugriff am 05.02.2022. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/s11618-006-0169-y.pdf>
- Tent, L. & Birkel, P. (2010). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 949-958). Weinheim: Beltz.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz*. : Westfälische Wilhelms-Universität Münster. Zugriff am 08.02.2022. Verfügbar unter <https://d-nb.info/1141683334/34>
- Tharenou, P. (2001). The relationship of training motivation to participation in training and development. *Journal of Occupational and Organizational Psychology*, 74 (5), 599-621.
- Tramm, T. & Naeve, N. (2007). Auf dem Weg zum selbstorganisierten Lernen – Die systematische Förderung der Selbstorganisationsfähigkeit über die curriculare Gestaltung komplexer Lehr-Lern-Arrangements. *Berufs- und Wirtschaftspädagogik - online* (13). Zugriff am 15.01.2020. Verfügbar unter http://www.bwpat.de/ausgabe13/tramm_naeve_bwpat13.pdf
- Tribelhorn, T. (2014). „Toolbox Assessment“ – ein hochschuldidaktischer Service im virtuellen Raum (Praxis- und Werkstattbericht). In K. Rummler (Hrsg.), *Lernräume gestalten - Bildungskontexte vielfältig denken* (Medien in der Wissenschaft, Bd. 67, S. 207-212). Münster: Waxmann.
- Universität Kassel (Hrsg.). (2014). *Fachprüfungsordnung für den Bachelorstudiengang Wirtschaftspädagogik des Fachbereichs Wirtschaftswissenschaften der Universität Kassel vom 11. Dezember 2013* (10/2014), Kassel.
- Universität Mannheim (Hrsg.). (2018). *Master of Science (M.Sc.) „Wirtschaftspädagogik“ der Universität Mannheim*. Modulkatalog, Mannheim.
- Universität Mannheim (Hrsg.). (2019). *Bachelor of Science (B.Sc.) „Wirtschaftspädagogik“ der Universität Mannheim*. Modulkatalog, Mannheim.

- Universität Rostock (Hrsg.). (2014, 30. Juli). *Amtliche Bekanntmachung. Studiengangsspezifische Prüfungs- und Studienordnung für den Masterstudiengang Wirtschaftspädagogik der Universität Rostock vom 5. Juli 2014* (Nr. 34), Rostock. Zugriff am 06.01.2023. Verfügbar unter https://www.uni-rostock.de/storages/uni-rostock/UniHome/Gremien/Rechtsgrundlagen/Amtliche_Bekanntmachungen/2014/NR_34_2014.pdf
- Universität zu Köln (Hrsg.). (2019a). *Modulhandbuch. Studienbereich berufliche Fachrichtung Wirtschaftswissenschaft Bachelor mit bildungswissenschaftlichem Anteil - Bachelor of Arts*. Nach dem Anhang der Wirtschafts- und Sozialwissenschaftlichen Fakultät und der gemeinsamen Prüfungsordnung der Universität zu Köln für den Studiengang Bachelor of Arts im Lehramt an Berufskolleg, Köln.
- Universität zu Köln (Hrsg.). (2019b). *WIRTSCHAFTSPÄDAGOGIK / Lehramt an Berufskollegs (Studienrichtung I und II). Master of Education/ Master of Science. NACH DEN ANHÄNGEN DER WIRTSCHAFTS- UND SOZIALWISSENSCHAFTLICHEN FAKULTÄT*.
- Van Ophuysen, S. & Behrmann, L. (2015). Die Qualität pädagogischer Diagnostik im Lehrerberuf - Anmerkungen zum Themenheft „Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz“. *Journal for educational research online*, 7 (2), 82-98.
- Vigerske, S. (2016). *Transfer von Lehrerfortbildungsinhalten in die Praxis. Eine empirische Untersuchung zur Transferqualität und zu Einflussfaktoren* (Research). Dissertation. Wiesbaden: Springer VS.
- Vogel, W. (2014). Bundeseinheitliche IHK-Prüfungen am Beispiel der zentralen Aufgabenerstellung bei der Aka. *Berufsbildung in Wissenschaft und Praxis*, 43 (3), 16-19.
- Vogel, D. & Funck, B. J. (2018). Immer nur die zweitbeste Lösung? Protokolle als Dokumentationsmethode für qualitative Interviews. *Forum: Qualitative Sozialforschung (FQS)*, 19 (1).
- Vonken, M. (2005). *Handlung und Kompetenz. Theoretische Perspektiven für die Erwachsenen- und Berufspädagogik* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften (GWV).
- Vroom, V. H. (1964). *Work and motivation*. Oxford: Wiley.
- Vuori, V., Helander, N. & Okkonen, J. (2019). Digitalization in knowledge work: the dream of enhanced performance. *Cognition, Technology & Work* (21), 237-252. Zugriff am

30.06.2021. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/s10111-018-0501-3.pdf>

- Wagner, P., Lüftenegger, M., Finsterwald, M., Schober, B. & Spiel, C. (2012). „Multi-method“-Evaluation am Beispiel des Lehrkräftetrainingsprogramms TALK. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung* (S. 243-258). 74. Tagung der Arbeitsgruppe Empirische Pädagogische Forschung (AEPF) im September 2010 in Jena. Münster: Waxmann.
- Wanitschek, I., Katschnig, T., Auferbauer, M. & Prorok, J. (2020). Die Wirksamkeit von Lehrer*innenfortbildung aus der Sicht der Fortbilder*innen. *Herausforderung Lehrer*innenbildung Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 3 (1), 739-760.
- Weber, A. (2005). Problem-Based Learning. - Ansatz zur Verknüpfung von Theorie und Praxis. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 1 (23), 94-104. Zugriff am 24.12.2019. Verfügbar unter https://www.pedocs.de/volltexte/2017/13566/pdf/BZL_2005_1_94_104.pdf
- Weber, E. (2015). Industrie 4.0 - Wirkungen auf Wirtschaft und Arbeitsmarkt. *Wirtschaftsdienst*, 95 (11), 722-723. Zugriff am 03.12.2019. Verfügbar unter <https://www.econstor.eu/bitstream/10419/125203/1/839444966.pdf>
- Weinert, F. E. (2000). *Lehren und Lernen für die Zukunft - Ansprüche an das Lernen in der Schule*. Vortragsveranstaltungen mit Prof. Dr. Franz E. Weinert, Max-Planck-Institut für psychologische Forschung, gehalten am 29. März 2000 im Pädagogischen Zentrum in Bad Kreuznach. Bad Kreuznach: Max-Planck-Institut für psychologische Forschung.
- Weinert, F. E. (2014). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3. Aufl., S. 17-32). Weinheim: Beltz.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Educational Expertise. Closing the Gap between Educational Research and Classroom Practice. *School Psychology International*, 11 (3), 163-180.
- Weiß, R. (2011). Prüfungen in der beruflichen Bildung – ein vernachlässigter Forschungsgegenstand. In E. Severing & R. Weiß (Hrsg.), *Prüfungen und Zertifizierungen in der beruflichen Bildung. Anforderungen - Instrumente - Forschungsbedarf* (S. 37-52). Bielefeld: W. Bertelsmann Verlag. Zugriff am 15.09.2019. Verfügbar unter https://www.agbfn.de/dokumente/pdf/a12_voevz_agbfn_10_weiss_1.pdf

- Weiß, R. (2014). Prüfungen - Instrument der Qualitätsentwicklung. *Berufsbildung in Wissenschaft und Praxis*, 43 (3), 3.
- Weyer, E., Wachendorf, N. M. & Mört, A. (2017). Kompetenzorientierung, was ist gemeint? In Cendon, Eva, Donner, Noëmi, U. Elsholz, A. Jandrich, A. Mört, N. M. Wachendorf & E. Weyer (Hrsg.), *Die kompetenzorientierte Hochschule. Kompetenzorientierung als Mainstreaming-Ansatz in der Hochschule* (Handreichung der wissenschaftlichen Begleitung des Bund-Länder-Wettbewerbs „Aufstieg durch Bildung: offene Hochschulen“, S. 6-12). Bonn: Bundesministerium für Bildung und Forschung.
- White, R. & Gunstone, R. (1992). *Probing Understand*. London: Routledge Taylor and Francis.
- Whitehead, A. N. (1929). *The Aims of Education*. New York: Macmillan.
- Wijnen, W. (2008). Problem-based learning (PBL) improves pedagogical professionalism. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 29-40). Münster: Waxmann.
- Wilbers, K. (2014a). *Wirtschaftsunterricht gestalten. Eine traditionelle und handlungsorientierte Didaktik für kaufmännische Bildungsgänge* (2. Aufl.). Lehrbuch. Berlin: epubli GmbH.
- Wilbers, K. (2014b). *Wirtschaftsunterricht gestalten. Toolbox* (2. Aufl.). Berlin: epubli GmbH. Zugriff am 16.01.2020. Verfügbar unter https://www.pedocs.de/volltexte/2013/8421/pdf/Wilbers_2014_Wirtschaftsunterricht_Toolbox_Aufl_2.pdf
- Wilbers, K. (2018). *Wirtschaftsunterricht gestalten* (3. Aufl.). Berlin: epubli GmbH.
- Wilbers, K. (Hrsg.). (2019a). *Digitale Transformation kaufmännischer Bildung. Ausbildung in Industrie und Handel hinterfragt* (1. Aufl.). Berlin: epubli. Zugriff am 26.10.2019. Verfügbar unter https://www.pedocs.de/volltexte/2019/17968/pdf/Wilbers_2019_Digitale_Transformation_kaufmaennischer_Bildung.pdf
- Wilbers, K. (2019b). *Wirtschaftsunterricht gestalten* (4. Aufl.). Berlin: epubli. Zugriff am 23.01.2020. Verfügbar unter https://www.pedocs.de/volltexte/2019/17949/pdf/Wilbers_2019_Wirtschaftsunterricht_gestalten.pdf
- Wilbers, K. (2020) Leitfaden Unterrichtsvorbereitung. In *Wirtschaftsunterricht gestalten. Toolbox* (5. Aufl.). Berlin: epubli. Verfügbar unter <http://www.wirtschaftsunterricht-gestalten.de/luv>

- Wilcox, R. R. (2005). Robust Testing Procedures. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1768-1769). Chichester: John Wiley & Sons.
- Wilcox, S. K. & Lanier, P. E. (2000). *Using assessment to reshape mathematics teaching. A casebook for teachers and teacher educators, curriculum and staff development specialists* (Studies in Mathematical Thinking and Learning Series). Mahwah, N.J: Lawrence Erlbaum Associates.
- Wild, K.-P. & Krapp, A. (2006). Pädagogisch-psychologische Diagnostik. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (Anwendung Psychologie, 5. Aufl., S. 525-574). Weinheim: Beltz PVU.
- Wilhelm, O. & Kunina-Habenicht, O. (2021). Pädagogisch-psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (3. Aufl., S. 311-334). Berlin: Springer.
- Windelband, L., Spöttl, G. & Becker, M. (2014). Qualität in der Berufsbildung – Chancen und Gefahren einer Output-/Outcome-Orientierung. In M. Fischer (Hrsg.), *Qualität in der Berufsausbildung. Anspruch und Wirklichkeit* (Schriftenreihe des Bundesinstituts für Berufsbildung, Bonn, S. 297-317). Bielefeld: Bertelsmann.
- Winter, F. (2018). *Lerndialog statt Noten. Neue Formen der Leistungsbeurteilung* (2. Aufl.). Weinheim: Beltz. Zugriff am 17.07.2020. Verfügbar unter https://content-select.com/media/moz_viewer/5aa7b79d-8a00-4d09-b0eb-6955b0dd2d03/language:de
- Winther, E. (2006). *Motivation in Lernprozessen. Konzepte in der Unterrichtspraxis von Wirtschaftsgymnasien* (Wirtschaftswissenschaft). Wiesbaden: Deutscher Universitäts-Verlag.
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. Bielefeld: W. Bertelsmann Verlag.
- Winther, E. (2011). Kompetenzorientierte Assessments in der beruflichen Bildung - Am Beispiel der Ausbildung von Industriekaufleuten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 107 (1), 33-54.
- Winther, E., Klotz, V. K. & Sangmeister, J. (2015). *Kompetenzorientiertes Prüfen in der kaufmännischen Berufsbildung. Handlungsempfehlungen für die Prüfungspraxis und Entwicklungsperspektiven*. Bonn: DIE Bonn. Zugriff am 11.05.2020. Verfügbar unter <http://www.die-bonn.de/doks/2015-kompetenzentwicklung-01.pdf>

- Wirtschafts- und Sozialwissenschaftliche Fakultät. (2019c). *WIRTSCHAFTSPÄDAGOGIK / Lehramt an Berufskollegs (Studienrichtungen I und II). MASTER OF EDUCATION / MASTER OF SCIENCE*. 2019/20 (Universität zu Köln, Hrsg.), Köln.
- Wirtschafts- und Sozialwissenschaftliche Fakultät; Philosophische Fakultät. (2018). *Modulhandbuch für den 2-Fächer-Bachelor-Studiengang und den 2-Fächer-Master-Studiengang mit dem 1. Fach ‚Wirtschaftswissenschaft‘ und dem ‚Profil Wirtschaftspädagogik‘*. Stand: 26.03.2018 (Christian-Albrechts-Universität zu Kiel (CAU), Hrsg.), Kiel.
- Wolter, F. (2012). *Heikle Fragen in Interviews. Eine Validierung der Randomized Response-Technik*. Zugl. Dissertation an der Johannes Gutenberg-Universität Mainz, 2011. Wiesbaden: Springer VS. Zugriff am 04.12.2022. Verfügbar unter <https://link.springer.com/content/pdf/10.1007/978-3-531-19371-7.pdf?pdf=button>
- Wottawa, H. (2006). Evaluation. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (Anwendung Psychologie, 5. Aufl., S. 659-687). Weinheim: Beltz PVU.
- Wuttke, E., Seeber, S., Geiser, C. & Turhan, L. (2022). Zur Problemhaltigkeit von Aufgaben in kaufmännischen Abschluss- und Zwischenprüfungen – Ergebnisse aus Aufgabenanalysen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 118 (1), 25-52.
- Wuttke, E., Seeber, S., Meiners, H. & Turhan, L. (2023). Entwicklung und Evaluation eines Trainingskonzeptes für die Gestaltung problemhaltiger technologiebasierter Lern- und Prüfungsaufgaben. In M. Hommel, C. Aprea & K. Heinrichs (Hrsg.), *bwp@ Profil 8: Netzwerke - Strukturen von Wissen, Akteuren und Prozessen in der beruflichen Bildung*. Digitale Festschrift für Bärbel Fürstenau zum 60. Geburtstag (S. 1-20). Zugriff am 15.10.2023. Verfügbar unter https://www.bwpat.de/profil8_fuerstenau/wuttke_etal_profil8.pdf
- Wuttke, E., Seeber, S. & Schumann, M. (o. A.). *Projektkonzept zur Forschungs- und Transferinitiative „ASCOT+ - Technologiebasierte Kompetenzmessung in der beruflichen Bildung“*. [unveröffentlichtes Dokument].
- Wuttke, E., Seeber, S., Schumann, M., Niegemann, H., Turhan, L., Meiners, H. et al. (2021, 20. September). *Technologiebasiertes kompetenzorientiertes Prüfen (TeKoP)* (3. Projekt-treffen ASCOT+).
- Wuttke, E., Seeber, S., Schumann, M., Turhan, L., Geiser, C., Hartmann, P. et al. (2022). *Handbuch für die Fortbildung von Prüfenden und Lehrenden zur Erstellung problemhalti-*

- ger technologiebasierter Prüfungsaufgaben und Aufgaben zur Lernprozessdiagnostik. Forschungs- und Transferinitiative „ASCOT+ - Technologiebasierte Kompetenzmessung in der beruflichen Bildung“*. Förderkennzeichen: 21AP001A, 21AP001B. Frankfurt: Goethe-Universität Frankfurt; Georg-August-Universität Göttingen.
- Wuttke, E. & Seifried, J. (2017). Competence, Teacher Competence and Professional Error Competence: An Introduction. In E. Wuttke & J. Seifried (Hrsg.), *Professional Error Competence of Preservice Teachers. Evaluation and Support* (SpringerBriefs in Education, S. 1-14). Cham: Springer International Publishing; Imprint; Springer.
- Wuttke, E., Seifried, J., Brandt, S., Rausch, A., Sembill, D., Martens, T. et al. (2015). Modellierung und Messung domänenspezifischer Problemlösekompetenz bei angehenden Industriekaufleuten. Entwicklung eines Testinstruments und erste Befunde zu kognitiven Kompetenzfacetten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111 (2), 189-207.
- Wuttke, E. & Wolf, K. D. (2007). Entwicklung eines Instrumentes zur Erfassung von Problemlösefähigkeit – Ergebnisse einer Pilotstudie. *Europäische Zeitschrift für Berufsbildung*, 2 (41), 99-118.
- Yoon, K. S., Duncan, T., Wen-Yu Lee, S., Scarloss, B. & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report REL 2007 - No. 033). Washington DC: Institute of Education Sciences. Zugriff am 05.11.2022. Verfügbar unter https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033.pdf
- Zaniboni, S., Fraccaroli, F., Truxillo, D. M., Bertolino, M. & Bauer, T. N. (2011). Training valence, instrumentality, and expectancy scale (T-VIES-it): Factor structure and nomological network in an Italian sample. *Journal of Workplace Learning*, 23 (2), 133-151.
- Zentrum für schulpraktische Lehrerbildung Aachen (Hrsg.). (2021). *Fachseminare*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/AAC/Seminar_BK/Fachseminare/index.html
- Zentrum für schulpraktische Lehrerbildung Bielefeld (Hrsg.). (2021). *Seminar BK*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/BIE/Seminar_BK/index.html
- Zentrum für schulpraktische Lehrerbildung Dortmund (Hrsg.). (2021). *Ausbildungsprogramm*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/DOR/Seminar_BK/Ausbildungsprogramm/Dortmunder_Plattform1/index.html

- Zentrum für schulpraktische Lehrerbildung Duisburg (Hrsg.). (2021). *Regelausbildung*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/DUI/Seminar_BK/Ausbildung/Regelausbildung/index.html
- Zentrum für schulpraktische Lehrerbildung Düsseldorf (Hrsg.). (2021). *Kerncurriculum*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/DUS/Seminar_BK/Ausbildung/Kerncurriculum/index.html
- Zentrum für schulpraktische Lehrerbildung Gelsenkirchen (Hrsg.). (2021). *Seminarprogramm*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/GEL/Seminar_BK/Seminarprogramm/index.html
- Zentrum für schulpraktische Lehrerbildung Hagen (Hrsg.). (2021). *Seminarprogramm*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/HAG/Seminar_BK/Seminarprogramm/index.html
- Zentrum für schulpraktische Lehrerbildung Krefeld (Hrsg.). (2021). *Fachseminare*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/KRE/Seminar_BK/Ausbildung/Fachseminare/index.html
- Zentrum für schulpraktische Lehrerbildung Leverkusen (Hrsg.). (o. A.a). *Fachcurriculum Spezielle Wirtschaftswissenschaften*. Zugriff am 07.12.2021. Verfügbar unter https://www.zfsl.nrw.de/LEV/Seminar_BK/Fachseminare/FS-Curriculum-spezelle_WiWi-Wocke.pdf
- Zentrum für schulpraktische Lehrerbildung Leverkusen (Hrsg.). (o.A.b). *Fachcurriculum Wirtschaftswissenschaften*, Leverkusen. Zugriff am 24.11.2021. Verfügbar unter https://www.zfsl.nrw.de/LEV/Seminar_BK/Fachseminare/Fachcurriculum-Wirtschaftswissenschaften-.pdf
- Zentrum für schulpraktische Lehrerbildung Leverkusen (Hrsg.). (2021). *Fachseminare*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/LEV/Seminar_BK/Fachseminare/index.html
- Zentrum für schulpraktische Lehrerbildung Münster (Hrsg.). (2021). *Ausbildungsprogramm*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/MUE/Seminar_BK/Ausbildungsprogramm/index.html
- Zentrum für schulpraktische Lehrerbildung Paderborn (Hrsg.). (2021). *Ausbildung am Seminar*. Zugriff am 01.11.2021. Verfügbar unter https://www.zfsl.nrw.de/PBN/Seminar_BK/Ausbildung_am_Seminar/index.html

Zentrum für schulpraktische Lehrerbildung Solingen (Hrsg.). (2021). *Seminarprogramm*.

Verfügbar unter https://www.zfsl.nrw.de/SOL/Seminar_BK/Seminarprogramm/index.html

Zimmer, G. (2009). Prüferhandbuch. *BWP* (3), 61.

ZPA Nord-West (Hrsg.). (2021). *Bundeseinheitliche Prüfungsaufgabenerstellung. Die IHK-Organisation gewährleistet zuverlässige Prüfungsergebnisse*. Zugriff am 18.01.2023. Verfügbar unter <http://www.ihk-zpa.de/opencms/pages/pruefungen/Aufgaben/Aufgabenerstellung>

Rechtsquellenverzeichnis

Berufsbildungsgesetz in der Fassung der Bekanntmachung vom 4. Mai 2020 (BGBl. I S. 920), das zuletzt durch Artikel 2 des Gesetzes vom 20. Juli 2022 (BGBl. I S. 1174) geändert worden ist.

Büromanagementkaufleute-Ausbildungsverordnung vom 11. Dezember 2013 (BGBl. I S. 4125), die durch Artikel 1 der Verordnung vom 16. Juni 2014 (BGBl. I S. 791) geändert worden ist.

Verordnung über die Berufsausbildung zum Industriekaufmann/zur Industriekauffrau vom 23. Juli 2002 (BGBl. I S. 2764), die durch Artikel 1 der Verordnung vom 20. Juli 2007 (BGBl. I S. 1518) geändert worden ist.

ANHANG

Anhang A: Kodierleitfaden zu den Ausbildungsinhalten aus der Angebotsperspektive

Kategorie (Kappa-Wert)	Beschreibung und Kodierregeln Darunter sind Passagen zu verstehen, ...	Ankerbeispiel(e) ¹⁷⁶
1. INPUT (.79)		
1.1 Vorgaben, Richtlinien und Regelungen (.75)	<p>... in denen organisatorische und rechtliche Vorgaben zur beruflichen Bildung behandelt werden. Kodiert wird immer dann, wenn von der Steuerung, Organisation, rechtlichen Regelungen und Regelwerken (z. B. Lehrpläne und Curricula) und/oder Standards zur beruflichen Bildung gesprochen wird. Diese müssen sich nicht explizit auf die Prüfung beziehen.</p> <p>Lehrerbildungsstandards und Werke, die sich auf die rechtliche, organisatorische Gestaltung der Lehrerbildung beziehen, z. B. LPO, werden nicht in diese Kategorie kodiert.</p>	<p>„Spezielle rechtliche ... Aspekte der Diagnostik im betrieblichen Kontext“ (Georg-August-Universität Göttingen, 2019, S. 150)</p> <p>„Lehrpläne, Ausbildungsordnung“ (Amt für Lehrerbildung Hessen, 2012, S. 67)</p> <p>„Also für die Pädagogik ist es so ... schon zu Beginn der Ausbildung stellen wir ja dar, dass Lernziele aus den Lehrplänen abgeleitet werden und das beispielsweise auch die Prüfungsaufgaben einen wichtigen Standard darstellen“ (Lara, Pos. 45)</p>
1.2 Voraussetzungen der Lehrenden (.90)		
1.2.1 Unspezifische Formulierungen zum Lehrerhandeln und -tätigkeiten (.81)	<p>... in denen Anforderungen an die Bildungsarbeit bzw. das -personal genannt werden, die im weitesten Sinne Diagnostik und Prüfen umfassen. Kodiert wird immer dann, wenn in einer Passage Anforderungen an das Bildungspersonal formuliert werden, die sich auf die Bildungsarbeit beziehen, jedoch allgemein formuliert sind und sich nicht explizit auf die diagnostische Tätigkeit beziehen.</p> <p>Begriffe/ Passagen zum Lehrere ethos, zur Lehrerrolle etc. werden nicht kodiert.</p>	<p>„Die Studierenden ... kennen den Arbeitsplatz ‚Schule‘ sowie die damit verbundenen Anforderungen“ (Universität Mannheim, 2019, S. 37)</p> <p>„Für die Fachdidaktiken sind vor allem jene beruflichen Aufgaben von Interesse, die sich unmittelbar auf die Lernprozesse von Schülerinnen und Schülern beziehen, auch wenn es für Lehrkräfte noch viele andere Aufgaben und Herausforderungen gibt, wie beispielsweise die Beteiligung an Schulentwicklungsprozessen, Teambildung</p>

¹⁷⁶ Hier werden nur die relevanten Passagen der jeweiligen Modulbeschreibung genannt. Entsprechend der Kodierregel (1) zur Kodiereinheit wurde in der Analyse das gesamte Modul kodiert, in dem die Passage(n) enthalten ist (sind).

im Kollegium, Kontakt mit Eltern und Ausbildungsbetrieben, lebenslange Weiterbildung in fachlicher, fachdidaktischer und pädagogischer Hinsicht“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 13)

„Also, wir haben natürlich eine Vorstellung, wie wir uns vorstellen was sollen Lehrkräfte am Ende können. Dazu gibt es ein Kompetenzprofil und dazu gibt es ... Grundsätze.“ (Levin, Pos. 13)

1.2.2 Diagnostische Kompetenz
(1.00)

... in denen Diagnostik als ein Kompetenzbereich bzw. als Anforderung an Personen, die am Bildungsprozess beteiligt sind, aufgeführt oder als Zieldimension für künftige Lehrkräfte/Bildungspersonal benannt wird. Kodiert wird immer dann, wenn diagnostische Kompetenz als konkrete Anforderung an das Bildungspersonal genannt und/oder thematisiert wird bzw. den Studierenden als solche vermittelt wird. Es wird auch dann kodiert, wenn statt von „diagnostischer Kompetenz“ von „pädagogisch-psychologische Kompetenz“ geschrieben wird.

Passagen, in denen die Diagnostik als ein Thema bzw. Themenfeld in dem Modul behandelt wird, in der sie nicht als direkte Anforderung oder Kompetenz genannt werden, werden nicht kodiert.

„Diskussion empirischer Studien zu diagnostischen Kompetenzen und Praktiken von Lehrkräften“ (Georg-August-Universität Göttingen, 2019, S. 150)

„Sie begreifen die pädagogische Diagnostik als kontinuierliche Aufgabe und diskutieren die Möglichkeiten und Grenzen der Anwendbarkeit sowohl im Hinblick auf die Schüler als auch für sich selbst und im Lehrerteam“ (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, o. A.)

„Wenn ... Sie sich dieses Kompetenzraster mal anschauen. ... und da sehen Sie ... diese Struktur und auch dann eben diesen Bereich Diagnose, Beratung und Bewertung ... das Thema halt jetzt auch Leistungsmessungen“ (Leon, Pos. 17)

2. EINGANGSDIAGNOSTIK (.91)

2.1 Lern- und Leistungsvoraussetzungen diagnostizieren
(.81)

... in denen die Messung, Erfassung, Bestimmung o. Ä. des Lern- bzw. Leistungsstands der Lernenden als Voraussetzung für den Lehr-Lern-Prozess beschrieben wird. Kodiert wird immer dann, wenn eine Beschreibung eingesetzter Instrumente bzw. genutzter Verfahren erfolgt, die zur Bestimmung der Voraussetzungen der Lernenden in Bezug auf deren Lernstand und Leistungen dienen. Ebenso wird

„Diagnostik: Erfassung von Schüler(wohl)vorstellungen, Vorwissen und Kompetenzen, Einstellungen etc.“ (Goethe-Universität Frankfurt, 2015, S. 41)

	<p>dann kodiert, wenn Diagnostik als solches für die Lern- und Leistungsvoraussetzungen der Lernenden beschrieben bzw. eingesetzt wird.</p> <p>Es wird nur dann in diese Kategorie kodiert, wenn das Wort „Diagnose“ oder „Diagnostik“ als Inhaltspunkt und/oder ein (Teil des) Lernziels im Zusammenhang mit der Eingangsdiagnostik genannt wird, aber keine Konkretisierung (z. B. Bewertung und Beurteilung) erfolgt. Wird beispielsweise nur die Motivation als eine Lernvoraussetzung beschrieben, nicht aber die Diagnostik hierfür genannt, wird nicht kodiert. Hierunter fallen auch nicht die in Kapitel 5.5 genannten Diagnoseinstrumente und -verfahren, wenn sie sich nicht direkt auf die Eingangsdiagnostik beziehen.</p>	<p>„Die Lehrkräfte im Vorbereitungsdienst diagnostizieren Lernvoraussetzungen“ (Amt für Lehrerbildung Hessen, 2012, S. 84)</p>
<p>2.2 Aufgabenerstellung zur Eingangsdiagnose (k. K.)</p>	<p>... bei denen es um inhaltliche Konzeption von Instrumenten zur Eingangsdiagnostik geht. Kodiert wird immer dann, wenn sich die Konzeption von Aufgaben zur Eignungsdiagnostik bezieht.</p>	<p>„Die Studierenden erstellen adäquate Aufgaben zur Erfassung der Lern- und Leistungsvoraussetzungen der Lernenden“ (fiktiv)</p>
<p>2.3 Bewertung und Beurteilung der Lern- und Leistungsvoraussetzungen (1.00)</p>	<p>... in denen der ermittelte Lern- bzw. Leistungsstand der Lernenden als Voraussetzung für den Lehr-Lern-Prozess beschrieben, beurteilt und/oder bewertet wird. Kodiert wird immer dann, wenn eine Bewertung und/oder eine Beurteilung der Voraussetzungen der Lernenden erfolgt.</p>	<p>„Sie können Gegenstände, zentrale Verfahren und konkrete Instrumente der Beurteilung individueller Lernvoraussetzungen, ... anhand ausgewählter Beispiele erläutern und begründet bewerten“ (Georg-August-Universität Göttingen, 2019, S. 149)</p> <p>„Die Lehrkräfte im Vorbereitungsdienst ... Lernvoraussetzungen ... und üben ihre Beurteilungsaufgabe transparent und verantwortungsbewusst aus“ (Amt für Lehrerbildung Hessen, 2012, S. 84)</p>
<p>2.4 Gütekriterien in der Eignungsdiagnostik (1.00)</p>	<p>Darunter sind Passagen zu den Kriterien Validität, Objektivität, Reliabilität sowie Nebengütekriterien zu verstehen, die im Input bzw. in der Diagnostik dessen in Bezug auf die Lernenden zum Einsatz kommen und/oder bedacht werden. Kodiert wird immer dann, wenn die Gütekriterien in Bezug auf das Input im Sinne der Lernenden bzw. die Eingangsdiagnose formuliert werden.</p>	<p>„Diagnostik der Lernvoraussetzungen der Lernenden unter Beachtung der Kriterien Reliabilität, Objektivität und Validität“ (fiktiv)</p>

2.5 Handlungsorientierte Eingangsdiagnostik (k. K.)	... in denen Inhalte zu handlungsorientierten Ansätzen in der Diagnostik der Lern- und Leistungsvoraussetzungen der Lernenden behandelt werden. Kodiert wird immer dann, wenn beschrieben wird, dass die Eingangsidagnostik handlungsorientiert gestaltet werden soll.	„Wissen, wie handlungsorientierte Diagnostik zur Erfassung der Lernvoraussetzungen der Lernenden funktioniert“ (fiktiv)
2.6 Problemorientierte Eingangsdiagnostik (k. K.)	... in denen Inhalte zu den problemorientierten Ansätzen in der Diagnostik der Lern- und Leistungsvoraussetzungen der Lernenden behandelt werden. Kodiert wird immer dann, wenn beschrieben wird, dass die Eingangsidagnostik problemorientiert gestaltet werden soll.	„Wissen, wie problemorientierte Diagnostik zur Erfassung der Lernvoraussetzungen der Lernenden funktioniert“ (fiktiv)
3. PROZESSDIAGNOSTIK (.85)		
3.1 Diagnostik im Lernprozess der Lernenden (.73)	<p>... in denen der Stand der Lernenden im Lehr-Lern-Prozess gemessen, erfasst und bestimmt wird. Kodiert wird immer dann, wenn aus der Passage deutlich wird, dass die Diagnostik im Lehr-Lern-Prozess geschieht. Es wird dann in diese Kategorie kodiert, wenn das Wort „Diagnose“ oder „Diagnostik“ als Inhaltspunkt und/oder ein (Teil des) Lernziels im Zusammenhang mit der Prozessdiagnostik genannt wird, aber keine Konkretisierung (z. B. Bewertung und Beurteilung) erfolgt.</p> <p>Passagen, die sich ausschließlich auf die Evaluation des Lehr-Lern-Prozesses beziehen, werden nicht in diese Kategorie kodiert, da die Aussagen zu unkonkret sind. Ebenso werden Passagen zur Unterrichtsevaluation nicht kodiert. Auch wird nicht kodiert, wenn es um die Förderung von Lernprozessen geht. Anders ist es, wenn die Förderung von Lernprozessen gemeinsam mit der bzw. dessen Diagnostik erwähnt wird. Hierunter fallen auch nicht die in Kapitel 5.5 genannten Diagnoseinstrumente und -verfahren, wenn sie sich nicht direkt auf die Prozessdiagnostik beziehen.</p>	<p>„Lernprozesse ... von Schüler/-innen in ihrer Unterschiedlichkeit erkennen und diagnostizieren“ (Universität Kassel, 2014, S. 1024)</p> <p>„Die Lehrkraft kennt ... auch Möglichkeiten der Leistungsfeststellung in offenen Unterrichtsformen und setzt diese zielgerichtet ein“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26)</p> <p>„Also hier geht es dann tatsächlich um formative ... Formen der Leistungsbewertung“ (Luisa, Pos. 16)</p>
3.2 Aufgabenerstellung zur Prozessdiagnostik (1.00)	... bei denen es um die inhaltliche Konzeption von Diagnoseinstrumenten geht, die zur Erhebung der Leistungen der Lernenden im Lehr-Lern-Prozess genutzt werden. Kodiert wird immer dann, wenn sich die Konstruktion von Aufgaben auf die Prozessanalyse bezieht und in dies in der Passage deutlich angegeben wird.	„Die Studierenden erwerben Kenntnisse über die lernzielorientierte Entwicklung von Aufgabenstellungen und können diese anhand ausgewählter wirtschaftlicher und kaufmännischer Lerninhalte umsetzen“ (Georg-August-Universität Göttingen, 2019, S. 149)

„Meistens ist es im Bereich Wirtschaft und Verwaltung zeitlich homogen und im pädagogischen Seminar kommt das Thema Aufgabenerstellung, dann themengebunden wieder vor, wenn zum Beispiel das Thema Binnendifferenzierung auftaucht, dann schauen wir mit dieser Perspektive noch einmal auf die Aufgabenstellung“ (Laurentia, Pos. 13)

3.3 Bewertung und Beurteilung der Lernenden im Lehr-Lern-Prozess
(1.00)

... in denen ermittelte Leistungen der Lernenden beschrieben, beurteilt und/oder bewertet werden. Kodiert wird immer dann, wenn eine Bewertung und/oder eine Beurteilung im Lehr-Lern-Prozess der Lernenden (formatives Assessment) geschieht.

„Sie können Gegenstände, zentrale Verfahren und konkrete Instrumente der Beurteilung individueller Lernvoraussetzungen, -verläufe und -leistungen ... begründet bewerten“ (Georg-August-Universität Göttingen, 2019, S. 149)

„Die Studienreferendarinnen und Studienreferendare kennen Beurteilungsbögen [...] zur Leistungsfeststellung im kompetenzorientierten Unterricht“ (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, o. A.)

„Also wenn jemand eine Lernaufgabe, die wirklich sehr problemorientiert und sehr offen ist, dann frage ich ihn der Nachbesprechung schon mal was machst du denn mit Bewertungen dazu?“ (Liah, Pos. 59)

3.4 Gütekriterien in der Prozessdiagnostik
(k. K.)

Darunter sind Passagen zu den Kriterien Validität, Objektivität, Reliabilität sowie Nebengütekriterien zu verstehen, die im Lehr-Lern-Prozess bzw. in der Diagnostik dessen zum Einsatz kommen. Kodiert wird immer dann, wenn die Gütekriterien in Bezug auf den Lehr-Lern-Prozess bzw. die Prozessdiagnostik formuliert werden.

„Sie [die Lehrkraft] stellt eine Klassenarbeit unter Beachtung der Gütekriterien auf und korrigiert sie“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26)

„Da ... thematisieren wir Gütekriterien von Unterricht“ (Laurentia, Pos. 20)

3.5 Handlungsorientierung im Prozess (.73)	... in denen Inhalte zu handlungsorientierten Ansätzen in der Prozessdiagnostik behandelt werden. Die Prozessdiagnostik kann sich auf die Gestaltung des Lehr-Lern-Prozesses beziehen.	„Handlungsorientierte (komplexe) Lehr- Lernarrangements“ (Universität Duisburg-Essen, 2019a, S. 84) „Unterricht gestalten und steuern (vorwiegend situativ-handlungspraktisch)“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 14) „Also deswegen also Grundlage sind die die Phasen der vollständigen Handlung, die idealtypisch sechs Phasen beinhalten, aber nicht jede Phase entsprechend durchlaufen werden kann.“ (Lorenz, Pos. 32)
3.6 Problemorientierung im Prozess (.81)	... in denen Inhalte zu problemorientierten Ansätzen in der Prozessdiagnostik behandelt werden. Die Prozessdiagnostik kann sich auf die Gestaltung des Lehr-Lern-Prozesses beziehen.	„Problemorientierte Lernsituationen und komplexe Lehr-Lernarrangements“ (CAU, 2018, S. 7) „Die Studienreferendarinnen und Studienreferendare klären die theoretischen Grundlagen zur Konzeption sowie zur Planung und Durchführung von kompetenz- und problemorientiertem Unterricht im beruflichen Schulwesen“ (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, o. A.) „Und genauso ist es eigentlich umgekehrt, aber eher möglich, dass man zum Beispiel Problemlösungskontext als Lernsituation hat und daraus dann halt reorganisativ ein reproduktives Lernziel noch erreichen kann.“ (Lukas, Pos. 24)
4. OUTPUTDIAGNOSTIK (1.00)		
4.1 Diagnostik des Outputs (1.00)	... bei denen es um diagnostische Verfahren zur Erfassung des Outputs zum Ende eines Bildungsprozesses, bspw. in Form von Prüfungen, geht. Kodiert wird immer dann, wenn von einer Prüfung(-sform) geschrieben wird, die zum Ende eines Lernprozesses oder konkret zum Abschluss der beruflichen Erstausbildung stattfindet. Hierzu zählt auch die Selektionsdiagnostik. Es wird nur dann in diese Kategorie kodiert, wenn das Wort „Diagnose“ oder „Diagnostik“ als Inhaltspunkt und/oder	„Evaluation, dabei insbesondere: Outputanalysen in Form von Prüfungen“ (Goethe-Universität Frankfurt, 2015, S. 41) „Die Lehrkraft kennt unterschiedliche Aufgabentypen, die in den kaufmännischen Prüfungen angewandt werden und kann diese anforderungsgerecht in Klassenarbeiten und sonstigen Leistungsüberprüfungen

	<p>ein (Teil des) Lernziels im Zusammenhang mit der Outputdiagnostik genannt wird, aber keine Konkretisierung (z. B. Bewertung und Beurteilung) erfolgt. Hierunter fallen auch nicht die in Kapitel 5.5 genannten Diagnoseinstrumente und -verfahren, wenn sie sich nicht direkt auf die Outputdiagnostik beziehen.</p>	<p>einsetzen“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26)</p> <p>„Und ... in dem Moment, wo wir mit Referendaren Lernziele formulieren, vorformulieren wir, also denken wir ... mit ihnen gemeinsam, in welcher Form können wir überprüfen, inwieweit die Schülerinnen und Schüler die Lernziele erreichen?“ (Lars, Lara und Lamia, Pos. 44)</p>
<p>4.2 Aufgabenerstellung zur Outputdiagnostik (1.00)</p>	<p>... bei denen es um die inhaltliche Konzeption von Instrumenten zur Outputdiagnostik geht. Kodiert wird immer dann, wenn sich die Konstruktion von Aufgaben auf die Outputdiagnostik, bspw. Prüfungen oder Klassenarbeiten, bezieht.</p>	<p>„Die Studierenden ... entwickeln Aufgabendesigns für spezifische Kompetenzdimensionen und vergleichen diese mit Ansprüchen typischen Aufgaben abschlussbezogener Prüfungen“ (Universität zu Köln, 2019b, S. 52)</p> <p>„beachten bei den Aufgabenstellungen sachliche Zusammenhänge mit anderen, bereits behandelten Themenbereichen (z. B. System der Besteuerung von Einkommen und die sich daraus ergebenden Aspekte der Kaufkraft)“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 123)</p> <p>„Nichtsdestotrotz ist auch ein Thema kompetenzorientierte Klassenarbeiten erstellen, weil Klassenarbeiten erstellen, ist gar nicht so einfach am Anfang. Und das behandeln wir schon auch, wie man daran geht, damit auch die Klausuren, Klassenarbeiten kompetenzorientiert formuliert sind“ (Levin und Liah, Pos. 23)</p>
<p>4.3 Bewertung und Beurteilung des Outputs der Lernenden (k. K.)</p>	<p>... in denen der ermittelte Output der Lernenden beschrieben, beurteilt und/oder bewertet wird. Kodiert wird immer dann, wenn eine Bewertung und/oder eine Beurteilung im Zusammenhang mit der Outputdiagnostik (summatives Assessment) geschieht.</p>	<p>„Die Studierenden ... erfassen und bewerten den spezifischen Kompetenzerwerb“ (Universität zu Köln, 2019b, S. 52)</p> <p>„Und auch bei Klassenarbeiten ... ist immer ein heikles Thema. Notengebung, dann ... wie kann ich das werten? ... Wie geht man mit Folgefehlern um und so weiter,</p>

also das sind Themen, die immer wieder aufs Tablett kommen.“ (Lorenz, Pos. 14)

4.4 Gütekriterien in der Outputdiagnostik
(1.00)

Darunter sind Passagen zu den Kriterien Validität, Objektivität, Reliabilität sowie Nebengütekriterien zu verstehen, die im Output bzw. in der Diagnostik dessen zum Einsatz kommen und/oder bedacht werden. Kodiert wird immer dann, wenn die Gütekriterien in Bezug auf das Output bzw. die Outputdiagnostik formuliert werden.

„können Kritik an traditionellen Prüfungsverfahren begründen, kennen reformierte Prüfungsverfahren und können sie hinsichtlich der Reformintentionen und im Hinblick auf die Gütekriterien der Testtheorie diskutieren“ (FAU, 2019a, S. 85)

„Sie stellt eine Klassenarbeit unter Beachtung der Gütekriterien auf und korrigiert sie“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26)

„Ich nehme an (unv.) Theorie, wir beschäftigen uns natürlich mit der Validität, Reliabilität und Objektivität von diesen entsprechenden Kriterien und wir hinterfragen natürlich im Kontext sowohl mündlicher als auch andererseits schriftlicher Prüfungen die Kontexte entsprechend dieser Prüfungssituationen und analysieren selbstverständlich auch dann zum Teil anhand von Videovignetten, zum Teil anhand von Fallbeispielen in Form von Skriptbestandteilen, welche entsprechenden Fehlerquellen da auch letztendlich daraus entstehen können. Also subjektive Fehlerquellen oder auch Fehler, die dann zu Einbußen der Objektivität, Reliabilität und Validität führen.“ (Lukas, Pos. 12)

4.5 Handlungsorientierung in der Outputdiagnostik
(k. K.)

... in denen Inhalte bzw. Ansätze zur handlungsorientierter Outputdiagnostik behandelt werden. Kodiert wird immer dann, wenn Handlungsorientierung im (summativen) Assessment angewandt wird.

„Handlungsorientierte Prüfungen“ (FAU, 2019a, S. 84)

„Da geht's auch ganz konkret und auch um Kompetenzen und kompetenzorientiertes Prüfen. Und dann bekommen die Referendare von mir eine quasi sehr kleinschrittige Prüfungsaufgabe ‘Nennen sie...‘ so sehr kleinschrittig aufgebaut und dann bekommen sie zugleich auch noch eine, sagen wir mal mit einem Fallbeispiel situativ, und dann müssen sie sich in die Situation rein

und müssen dann entsprechend auch ein betriebliches Handlungsergebnis erstellen.“ (Lorenz, Pos. 10)

4.6 Problemorientierung in der Outputdiagnostik
(k. K.)

... in denen Inhalte bzw. Ansätze zur problemorientierter Outputdiagnostik behandelt werden. Kodiert wird immer dann, wenn Problemorientierung im (summativen) Assessment angewandt wird.

„Problemorientierte Prüfungsgestaltung“ (fiktiv)

„Und wenn wir uns den anschauen, dann können wir aus der empirischen Forschung eben ableiten, dass es darum geht, authentische Aufgaben, problemhaltige Aufgaben und Aufgaben aus dem späteren Beruf zu integrieren in die Prüfungsaufgaben und die Formate der Prüfungsaufgaben. Also das wäre so der Anteil der allgemeinen Pädagogik zu sagen, es soll möglichst authentisch, lebensnah, berufsnah sein und da möglichst auch wenn es im Beruf situiert ist, auch mit authentischen Belegen“ (Lara, Pos. 10)

5. DIAGNOSTIK IN DER PÄDAGOGIK (.83)

5.1 Diagnostik
(.88)

... bei denen Aussagen zur Diagnostik getroffen werden, die allgemein bzw. unkonkret gehalten sind. Kodiert wird immer dann, wenn unkonkrete Aussagen zur Diagnostik gemacht werden, die keiner spezifischeren Subkategorie (z. B. Gütekriterien) bzw. keiner der Phasen (Input/Eingang-, Prozess- oder Outputdiagnostik) zugeordnet werden können. Geht es um die *Förderung* als einzelnes Thema, wird nicht kodiert.

„Die Studierenden besitzen umfassende Kenntnisse zu Aufgaben und Funktionen pädagogisch-psychologischer Diagnostik in schulischen und betrieblichen Anwendungsfeldern der beruflichen Bildung“ (Georg-August-Universität Göttingen, 2019, S. 149)

„Fachbezogene Diagnose- u. Förderkonzepte“ (Amt für Lehrerbildung Hessen, 2012, S. 7)

„Ein Modul hat sieben Veranstaltungen a drei Stunden, und in dem Rahmen beschäftigen wir uns natürlich mit Diagnostik und mit Aufgabenerstellung.“ (Liah, Pos. 14)

5.2 Aufgabenerstellung zur Diagnostik
(1.00)

... bei denen es um die inhaltliche Konzeption von Diagnoseinstrumenten in Form von Aufgaben und Situation geht, wobei nicht beschrieben wird, zu welcher Phase (Input/Eingang, Prozess oder Output) diese eingesetzt werden.

„diagnostische Instrumente ... selbst zu entwickeln“ (Universität Duisburg-Essen, 2019b, S. 24)

„Ein Modul hat sieben Veranstaltungen a drei Stunden, und in dem Rahmen beschäftigen wir uns natürlich mit Diagnostik und mit Aufgabenerstellung.“ (Liah, Pos. 14)

5.3 Bewertung und Beurteilung (.73)	<p>Darunter sind Passagen zu Verfahren zum Abschluss der Erhebung von Leistungen, Kompetenzen o. Ä. zu verstehen, die dazu dienen, die individuellen Ergebnisse Einzelner als Wert wiederzugeben und/oder zu interpretieren. Kodiert wird immer dann, wenn es um die Beurteilung und/oder Bewertung von Leistungen, Kompetenzen o. Ä. geht, nicht aber um das Instrument oder zum Einsatz dessen. Hierzu zählen Bezugsnormen und Fehlerquellen. Wenn sich die Bewertung und Beurteilung auf die Instrumente bzw. Verfahren, die unter der Kategorie 5.5 aufgeführt sind, beziehen, so ist die Passage bzw. die Modulbeschreibung in diese Kategorie zu kodieren, es sei denn, aus der Passage geht hervor, auf welche Phase (Input/Eingang, Prozess oder Output) sich die Passage bezieht.</p>	<p>„Sie kennen die Grundkonzepte pädagogisch-psychologischer Leistungsmessung, können mögliche Fehler bei der Beurteilung benennen und kennen Möglichkeiten, diese zu vermeiden.“ (Universität Mannheim, 2019, S. 33)</p> <p>„praktizieren geeignete Verfahren der Leistungsmessung und -bewertung transparent und begründen Beurteilungen und Bewertungen adressatengerecht“ (Amt für Lehrerbildung Hessen, 2012, S. 6)</p> <p>„Ja, wir haben ... fachdidaktische Seminare, bezogen auf die berufliche Fachrichtung Wirtschaft und Verwaltung. Und die machen das Thema Leistungsbewertung in der Regel relativ kurz, also über meistens nur eine Sitzung und dann vertiefend, wenn es als Bedarf der teilnehmenden wiederauftaucht.“ (Laurentia, Pos. 16)</p>
5.4 Gütekriterien (1.00)	<p>Darunter sind Passagen zu den Kriterien Validität, Objektivität, Reliabilität sowie Nebengütekriterien zu verstehen. Kodiert wird immer dann, wenn die Gütekriterien in Bezug auf die pädagogisch-psychologische Diagnostik formuliert werden. Beispielsweise werden Passagen zu Gütekriterien einer empirischen Forschungsarbeit nicht kodiert.</p>	<p>„Gütekriterien pädagogisch-psychologischer Diagnostik“ (Georg-August-Universität Göttingen, 2019, S. 149)</p> <p>„beurteilt Leistung anhand fachdidaktisch reflektierter Gütekriterien“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 16)</p> <p>„Das ist ja, steht in engem Zusammenhang mit den Gütekriterien, besteht denn Objektivität, Reliabilität, Validität überhaupt hier, und das behandeln wir ganz allgemein in Pädagogik, in einem Modul, das da heißt Bewertung von Kompetenz.“ (Lara, Pos. 41)</p>
5.5 Diagnoseinstrumente und –verfahren (.78)		

5.5.1 Diagnoseinstrumente und -verfahren allgemein (.73)	Darunter sind Passagen zu Verfahren und Instrumenten zu verstehen, die in zur Diagnose eingesetzt werden, allerdings nicht konkreter benannt und/oder beschrieben werden. Kodiert wird immer dann, wenn nicht von einer bestimmten Assessmentform und einem bestimmten Verfahren der Diagnostik geschrieben wird, wohl aber mit dem Instrument eine pädagogisch-psychologische Diagnostik bestrebt wird, ist die Passage in diese Kategorie zu kodieren. Ist aus der Passage ersichtlich, in welcher Phase (Input/Eingang, Prozess oder Output) das unspezifisch genannte bzw. beschriebene Instrument oder Verfahren eingesetzt wird, ist die Passage in die entsprechende Kategorie der Phase (2.1, 3.1 oder 4.1) zu kodieren.	<p>„... wählen für die Zielgruppen und die Profile der Bildungsgänge des Berufskollegs geeignete Diagnoseansätze, -verfahren und -instrumente aus“ (Universität zu Köln, 2019b, S. 52)</p> <p>„Instrumente, z. B. Reflexionsbogen, Kompetenzraster“ (Staatliches Studienseminar für das Lehramt an beruflichen Schulen, 2018, o. A.)</p> <p>„Letztendlich Leistungsfeststellung in Form von schriftlichen Arbeiten ist nur eine Form. Wir haben natürlich noch einen ein Potpourri von anderen. Ja, das werden halt Lerntagebücher oder Präsentation beziehungsweise Vorträge, Plakatgestaltungen und Bewertung von Handlungsprodukten.“ (Luisa, Pos. 16)</p>
5.5.2 Kompetenzmessung und –diagnostik (.67)	Darunter sind Passagen zu Verfahren und Instrumenten der Messung, Testung, Operationalisierung bzw. Erhebung zu verstehen, die der Bestimmung der Kompetenz dienen. Kodiert wird immer dann, wenn in der Passage allgemein das Wort „Kompetenzmessung“ oder „Messung der Kompetenz“ fällt, oder aber von Operationalisierung, Testung bzw. Messung der Kompetenzen geschrieben wird, wird die Passage zu dieser Kategorie kodiert. Ist aus der Passage ersichtlich, in welcher Phase das Instrument bzw. das Verfahren zur Kompetenzmessung bzw. -diagnostik eingesetzt wird, ist die Passage in die entsprechende Kategorie der Phase (2.1, 3.1 oder 4.1) zu kodieren.	<p>„[Bullet-Point] kennen verschiedene Ansätze zur Messung von Kompetenzen sowie einschlägige empirische Befunde aus der beruflichen Bildung, [Bullet-Point] verstehen die Herausforderungen im Kontext der Messung von Kompetenzen.“ (Universität Mannheim, 2019, S. 6)</p> <p>„Wie evaluiere ich die Entwicklung der Fachkompetenz/ Handlungskompetenz meiner Schülerinnen und Schüler?“ (Zentrum für schulpraktische Lehrerbildung Leverkusen, o.A.b, S. 6)</p>
5.5.3 Leistungsmessung und –diagnostik (.81)	Darunter sind Passagen zu Verfahren und Instrumenten der Messung, Testung, Operationalisierung bzw. Erhebung zu verstehen, die der Bestimmung der Leistung dienen. Kodiert wird immer dann, wenn in der Passage allgemein das Wort „Leistungsmessung“ oder „Messung von Leistung(en)“ fällt, oder aber von Operationalisierung, Testung bzw. Messung der Leistungen geschrieben wird, wird die Passage zu dieser Kategorie zugeordnet. Ist aus der Passage ersichtlich, in welcher Phase das Instrument bzw. das Verfahren zur Leistungsmessung bzw. -diagnostik	<p>„Sie kennen die Grundkonzepte pädagogisch-psychologischer Leistungsmessung“ (Universität Mannheim, 2019, S. 33)</p> <p>„Fachdidaktische Konzepte zur Leistungsmessung“ (Amt für Lehrerbildung Hessen, 2012, S. 6)</p>

	eingesetzt wird, ist die Passage bzw. das Modul in die entsprechende Kategorie der Phase (2.1, 3.1 oder 4.1) zu kodieren.	
5.5.4 Lernerfolgskontrolle und -diagnostik (.81)	Darunter sind Passagen zu Verfahren der Messung, Testung, Operationalisierung bzw. Erhebung zu verstehen, die der Bestimmung und/oder Kontrolle des Lernerfolgs dienen. Ist aus der Passage ersichtlich, in welcher Phase das Instrument bzw. das Verfahren zur Lernerfolgsmessung bzw. -diagnostik eingesetzt wird, ist die Passage bzw. das Modul in die entsprechende Kategorie der Phase (2.1, 3.1 oder 4.1) zu kodieren. Kodiert wird immer dann, wenn über die Kontrolle, Evaluation und/oder Messung von Lernergebnissen oder dem Lernerfolg geschrieben wird.	„Lernerfolgskontrollen“ (Goethe-Universität Frankfurt, 2015a, S. 59) „Sie stellt auch den individuellen Lernfortschritt fest.“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26)
6. HANDLUNGS- UND PROBLEMORIENTIERUNG (1.00)		
6.1 Handlungsorientierung (1.00)	... in denen Inhalte zu Handlungsorientierung zu finden sind, diese aber nicht dem Input/Eingang, Prozess und/oder Output zugeordnet werden können.	“Handlungsorientierung” (Der Präsident der Humboldt-Universität zu Berlin, 2015, S. 5) “Problemorientierung ja nicht losgekoppelt ist von dem [1]Thema oder der übergeordneten Kompetenz der Handlungsorientierung. Also, das muss man ja in einer Verbindung miteinander betrachten.” (Luisa, Pos. 35)
6.2 Problemorientierung (1.00)	... in denen Inhalte zur Problemorientierung zu finden sind, diese aber nicht dem Input/Eingang, Prozess und/oder Output zugeordnet werden können.	„Problem based Learning“ (Der Präsident der Humboldt-Universität zu Berlin, 2015, S. 5) “Problemorientierung als fachdidaktisches Prinzip ist sehr früh immer Thema am fachdidaktischen Seminar, also die fachdidaktischen Prinzipien Problemorientierung, Handlungsorientierung im doppelten Sinne.” (Laurentia, Pos. 18)
7. AKTUELLE FRAGESTELLUNGEN (1.00)		
7. Aktuelle Fragestellungen (1.00)	... in denen aktuelle Fragen der Berufs- und Wirtschaftspädagogik behandelt werden. Diese können variieren. Kodiert wird immer dann, wenn als Inhalt eines Moduls „aktuelle Fragen“, „aktuelle Fragestellungen“, „aktuelle Probleme“, „aktuelle	„Das Modul greift aktuelle Diskussionslinien der Wirtschaftspädagogik auf, wobei wissenschaftliche Bedeutung mit bildungspraktischer und bildungspolitischer Relevanz verknüpft werden.“ (Universität Jena, 2019, S. 224)

Problemstellungen“ o. Ä. beschrieben wird. Stehen die Inhalte bzw. Ziele der Veranstaltung zwar unter der Bezeichnung „aktuell“, wird jedoch nicht der Aspekt der beruflichen Erstausbildung erwähnt, ist diese Kategorie nicht zu wählen.

„Die Lehrkraft verfügt über Strategien, um wirtschaftsethische Fragen an Hand [sic!] aktueller betriebs- und gesamtwirtschaftlicher Problemstellungen zu reflektieren.“ (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2018, S. 26)

Anhang B: Leitfaden für das Interview mit Fachleiter*innen an Studienseminaren

(1) Informationsphase

Guten Tag Frau*Herr [Name]. Herzlichen Dank dafür, dass Sie sich die Zeit nehmen, um ein Gespräch zur Ausbildung von Referendar*innen an kaufmännisch-beruflichen Schulen zu führen. Wie bereits angekündigt, wird das Gespräch ca. 45 Minuten dauern. Damit die Erkenntnisse aus dem Interview genutzt werden können, würde ich das Gespräch gerne aufzeichnen. Sind Sie damit einverstanden? [Wenn ja, Tonaufnahme starten und Frage für das Band wiederholen.]

Im Rahmen meiner Dissertation interessiere ich mich dafür, ob und wie angehende Lehrkräfte im Verlauf ihres Referendariats auf die Erstellung von problemhaltigen (Prüfungs-)Aufgaben sowie deren Beurteilung und Bewertung vorbereitet werden. [...]Außerdem würde ich gerne die Frage klären, ob und wie sie darauf vorbereitet werden, solche Aufgaben technologiegestützt umzusetzen. Hierzu habe ich einige Fragen vorbereitet und bitte Sie, sich bei der Beantwortung der Fragen auf die Inhalte in dem letzten und aktuellen Ausbildungsjahr zu beziehen.

(2) Aufwärm- und Einstiegsphase

1. Für einen Einstieg würde ich Sie bitten, sich kurz vorzustellen. Dabei brauchen Sie nicht Ihren Namen zu sagen oder Ihren Standort zu nennen. Mich würde insbesondere interessieren, welche Funktion Sie genau am Studienseminar haben und wie Sie sich an der Ausbildung angehender Lehrkräfte beteiligen.

(3) Hauptphase

2. Wie werden in der Fachrichtung Wirtschaft und Verwaltung sowie im allgemein-pädagogischen Seminar [entsprechend dem*r Gesprächspartner*in anpassen] Inhalte zur **Aufgabenerstellung** behandelt?

Ggf. hinweisen: Beziehen Sie sich dabei auf die Inhalte im letzten und aktuellen Ausbildungsjahr.

Ggf. Nachfassen:

- Wie lernen Ihre LiV/Referendar*innen beispielsweise das Erstellen von Arbeitsaufträgen im Unterricht, bspw. auf Arbeitsblättern?
- Wie lernen Ihre LiV/Referendar*innen das Erstellen von Klausuraufgaben?

3. [Nur, wenn in 2 gesagt wurde, dass Aufgabenerstellung behandelt wird:] Können Sie bitte genauer erklären, wie das Thema behandelt wird?

Folgende Inhalte sind besonders interessant [ggf. nachfassen, wenn nicht angesprochen]:

- Welche Inhalte bearbeiten Sie genau zum Thema?
- Wie bringen Sie den Teilnehmenden die Aufgabenerstellung bei?
- Behandeln Sie dabei auch das Thema der Gütekriterien?
- Welche Methoden wenden Sie an?
- Wie schätzen Sie den zeitlichen Umfang, der für das Thema investiert wird?
- Wird dabei auch technologiegestützt gearbeitet?

4. [Nur, wenn in 2 gesagt wurde, dass Aufgabenerstellung behandelt wird:] Zu welchem Zweck, etwa für die Unterrichtsgestaltung oder Gestaltung von Klausuraufgaben, werden die Aufgaben erstellt?

Folgende Inhalte sind besonders interessant [ggf. nachfassen, wenn nicht angesprochen]:

- Aufgabenerstellung zu Zwecken der Eingangs-/Statusdiagnostik
- Aufgabenerstellung für den Unterricht/ Lehr-Lern-Prozess
- Aufgabenerstellung für Tests (auf Klassen- oder Schulebene)
- Aufgabenerstellung für (Abschluss-)Prüfungen

5. Werden in der Fachrichtung Wirtschaft und Verwaltung sowie im allgemein-pädagogischen Seminar [entsprechend dem Gesprächspartner anpassen] Inhalte zur **Bewertung und Beurteilung** behandelt?

Ggf. hinweisen: Beziehen Sie sich dabei auf die Inhalte im letzten und aktuellen Ausbildungsjahr.

6. [Nur, wenn 5 bejaht:] Können Sie bitte genauer erklären, wie das Thema behandelt wird?

Folgende Inhalte sind besonders interessant [ggf. nachfassen, wenn nicht angesprochen]:

- Welche Inhalte bearbeiten Sie genau zum Thema?
- Wie bringen Sie den Teilnehmenden die Bewertung und Beurteilung bei?
- Behandeln Sie dabei auch das Thema der Gütekriterien?
- Welche Methoden wenden Sie an?
- Wie schätzen Sie den zeitlichen Umfang, der für das Thema investiert wird?
- Wird dabei auch technologiegestützt gearbeitet?

7. Werden in der Fachrichtung Wirtschaft und Verwaltung sowie im allgemein-pädagogischen Seminar [entsprechend dem Gesprächspartner anpassen] Inhalte zur **Problemorientierung (in Instruktion und/oder Assessment)** behandelt? Mit Problemorientierung meine ich, dass beispielsweise in Aufgaben oder auch in der Unterrichtsgestaltung darauf geachtet wird, dass diese Probleme abbilden, die von den Lernenden zu lösen sind.

Ggf. hinweisen: Beziehen Sie sich dabei auf die Inhalte im letzten und aktuellen Ausbildungsjahr.

8. [Nur, wenn 7 bejaht:] Können Sie bitte genauer erklären, wie das Thema behandelt wird?

Folgende Inhalte sind besonders interessant [ggf. nachfassen, wenn nicht angesprochen]:

- Wie bringen Sie den Teilnehmenden das Thema der Problemorientierung bei?
- Welche Methoden wenden Sie an?
- Wie schätzen Sie den zeitlichen Umfang, der für das Thema investiert wird?
- Wird dabei auch technologiegestützt gearbeitet?

9. [Nur, wenn 7 bejaht:] In welchem Zusammenhang wird die Problemorientierung behandelt?

Nachfragen/Checkbox [ggf. nachfassen, wenn nicht angesprochen]:

- Problemhaltige Aufgabenerstellung zu Zwecken der Eingangs-/Statusdiagnostik
- Problemorientierte Unterrichtsgestaltung
- Problemhaltige Aufgabenerstellung für Tests (auf Klassen- oder Schulebene)
- Problemhaltige Aufgabenerstellung für (Abschluss-)Prüfungen

...

(4) *Ausklang- und Abschlussphase*

Ich danke Ihnen sehr für das Gespräch, Sie haben damit einen wichtigen Beitrag zur Beantwortung meiner Forschungsfrage beigetragen. Welche Themen möchten Sie diesbezüglich noch ansprechen? Welche Fragen haben Sie?

[Wenn Themen/Fragen: Klären.]

[Nur dann, wenn sich die Fragen nicht schon aus der Vorbereitung und/oder dem Gespräch geklärt haben/nachträglich beantwortet werden können:] Zur Beschreibung Ihres Standorts möchte ich gerne weitere Fragen zum Standort klären.

12. Wie viele Absolvent*innen haben Sie durchschnittlich pro Jahr?

13. Wie häufig wird Ihr Studienseminar wöchentlich von Refendar*innen besucht?

14. Welche Funktion haben Sie an Ihrem Studienseminar?

- | | |
|--|---|
| <input type="checkbox"/> Ausbildungsbeauftragte*r | <input type="checkbox"/> Direktor*in |
| <input type="checkbox"/> Bereichsleiter*in Wirtschaftswissenschaften | <input type="checkbox"/> Seminarleiter*in |
| <input type="checkbox"/> stellv. Seminarleiter*in | |
| <input type="checkbox"/> Andere Funktion, nämlich _____ | |

[Wenn keine weiteren Themen/Fragen, dann nochmals danken und Tonaufnahme beenden.]

Anhang C: Zuordnung der Fragen des teilstandardisierten Interviewleitfadens mit angehenden Lehrkräften (Ausbildungsinhalte aus der Nachfragerperspektive) zu den Forschungsfragen und Kategorien des Kodierleitfadens

Frage aus dem Leitfaden ¹⁷⁷	Forschungsfrage	Kategorie-kürzel ¹⁷⁸
Für einen Einstieg würde ich Sie bitten, sich kurz vorzustellen. Dabei brauchen Sie nicht Ihren Namen zu sagen. Bitte beschreiben Sie kurz, welche Vorbildung Sie haben, ob Sie Lehrerfahrung haben und was Sie aktuell beruflich tun.	---	D
<u>LiV und Studierende</u> (die vor der Intervention bereits Erfahrung im Klausurerstellen hatten): Sie erinnern sich sicherlich an die erste Klausur ¹⁷⁹ , die Sie erstellt und eingesetzt haben. Wie sind Sie vorgegangen?		
<ul style="list-style-type: none"> • Welche Hindernisse bzw. Schwierigkeiten hatten Sie und wie sind Sie damit umgegangen? 	F_{1-3}	E
<ul style="list-style-type: none"> • Haben Sie dabei an folgende Punkte gedacht (optional): <ul style="list-style-type: none"> ○ Bewertung, Beurteilung und Auswertung ○ Gütekriterien ○ Aufgaben-/Itemformulierung ○ ... 		
<u>LiV</u> : Wo und wie haben Sie Ihr damaliges Vorgehen gelernt?	F_{1-3}	S REF
<u>LiV</u> : Sie haben erklärt, wie Sie bei Ihrer ersten Klausurerstellung vorgegangen sind. Gehen Sie aktuell immer noch so vor oder hat sich Ihr Vorgehen verändert? Wenn ja, warum?	F_{1-3}	E
<u>LiV</u> : Wo und wie haben Sie Ihr jetziges Vorgehen gelernt?	F_{1-3}	S
<u>Studierende</u> : Wo und wie haben Sie das gelernt?		REF
Wenn Sie sich an Ihr Studium zurückdenken, wie fühlen Sie sich auf allgemein auf Aufgaben der pädagogischen Diagnostik vorbereitet?		
<ul style="list-style-type: none"> • Wie wurde darin die Erstellung von Aufgaben/Items thematisiert? • Inwiefern wurde dabei auf verschiedene Kompetenzdimensionen eingegangen? Wie haben Sie gelernt, Problemlösekompetenz zu messen? • Können Sie genauer beschreiben, wie Ihnen das gelehrt wurde? <ul style="list-style-type: none"> ○ Bewertung, Beurteilung und Auswertung ○ Gütekriterien ○ Aufgaben-/Itemformulierung 	F_{1-3}	S

¹⁷⁷ Zum Teil wurden die Fragen im Interviewleitfaden an die Zielgruppe der Masterstudierenden bzw. der LiV angepasst. Wenn dies in der vorgestellten Frage des Leitfadens getan wurde, wird dies in dieser Tabelle dargestellt. Wenn in der Zelle keine Unterscheidung gemacht wird, wurde für beide Gruppen die gleiche Frage genutzt.

¹⁷⁸ Zu den Kategorie-kürzeln: vgl. Kodierleitfaden Anhang D.

¹⁷⁹ In Kapitel 4.3.1 wurde begründet, warum die Frage nicht nach dem Prüfungsaufgabenerstellen, sondern Klausurerstellen gestellt wurde.

○ ...

Was könnte man Ihrer Meinung nach im Studium mit Blick auf Inhalte zur pädagogischen Diagnostik, insb. auf die Erstellung von (Prüfungs-) Aufgaben, verbessern?

F₁₋₃

S

LiV: Wie fühlen Sie sich durch das Referendariat/Vorbereitungsdienst allgemein auf Aufgaben der pädagogischen Diagnostik vorbereitet?

- Wie wurde darin die Erstellung von Aufgaben/Items thematisiert?
- Inwiefern wurde dabei auf verschiedene Kompetenzdimensionen eingegangen? Wie haben Sie gelernt, Problemlösekompetenz zu messen?
- Können Sie genauer beschreiben, wie Ihnen das gelehrt wurde?
 - Bewertung und Beurteilung
 - Gütekriterien
 - ...

F₁₋₃

REF

LiV: Was könnte man Ihrer Meinung nach im Seminar mit Blick auf Inhalte zur pädagogischen Diagnostik, insb. auf die Erstellung von (Prüfungs-) Aufgaben, verbessern?

F₁₋₃

REF

Ich danke Ihnen sehr für das Gespräch, Sie haben damit einen wichtigen Beitrag zur Beantwortung meiner Forschungsfragen beigetragen. Welche Themen möchten Sie diesbezüglich noch ansprechen? Welche Fragen haben Sie?

Anhang D: Kodierleitfaden zur Interviewstudie mit angehenden Lehrkräften zur Ausbildung aus der Nachfrageperspektive

Kategoriekürzel – Kategorie (Kappa-Wert)	Beschreibung und Kodierregeln Darunter sind Passagen zu verstehen, aus denen hervorgeht, ...	Ankerbeispiel ¹⁸⁰
D - DESKRIPTIVE DATEN (.83)		
D1 – Berufsausbildung absolviert (.88)	... ob und ggf. welche Ausbildung die Proband*innen absolviert haben.	„... davor eine Ausbildung zur Steuerfachangestellten“ (Seila, Pos. 6)
D2 – Wunsch ins Referendariat (nur Studierende) (.73)	... ob studentische Proband*innen den Wunsch haben, nach Abschluss ihres Studiums ins Referendariat zu gehen. In diese Kategorie sind nur Passagen aus Transkripten von Interviews mit Studierenden zu kodieren.	[Auf die Frage, ob die Person ins Referendariat/ Vorbereitungsdienst gehen möchte:] „Ziemlich sicher, ja, also zu 80 Prozent, will ich sagen.“ (Samuel, Pos. 21)
D3 – Erfahrung in Unterricht/Schule (.86)	... ob und wie studentische Proband*innen und/oder LiV praktische Erfahrungen an Schulen bzw. im Unterricht an anderen Instituten sammeln konnten. In die Kategorie D3.1.1 sind nur Passagen aus Transkripten von Interviews mit Studierenden zu kodieren, da davon auszugehen ist, dass alle LiV bereits Unterrichtserfahrung gesammelt haben. Beim Kodieren von Passagen wird darauf geachtet, dass Praktika im Rahmen des Studiums nicht zu der hier gemeinten Erfahrung gehören. Die Kategorie unterteilt sich in weitere Subkategorien.	
D3.1 – Erfahrung besteht (.82)	... dass Proband*innen Erfahrung im Unterrichten an der Schule oder anderen Instituten sammeln konnten. Diese Kategorie unterteilt sich in zwei weitere Subkategorien.	
D3.1.1 – Erfahrung besteht im Unterrichten (nur Studierende) (.85)	... dass studentische Proband*innen Erfahrung im Unterrichten haben (bspw. an einer Schule oder einem anderen Institut). Die Erfahrung über universitäre Lehrveranstaltungen wird nicht kodiert.	„Ich habe einen TVH-Vertrag und unterrichte ... die sechs Stunden die Woche an einer Berufsschule Mathematik.“ (Sarah, Pos. 6)
D3.1.2 – Erfahrung besteht auch im Klausurerstellen (.73)	... dass studentische Proband*innen und/oder LiV auch Erfahrungen im Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments sammeln konnten. Darunter fällt nicht das Erstellen von beispielsweise Arbeitsblättern.	„Ha, ich habe ja vor kurzem noch Klausuren vorbereitet.“ (Sarah, Pos. 17)

¹⁸⁰ Hier werden nur die relevanten Passagen der jeweiligen Modulbeschreibung genannt. Entsprechend der Kodierregel (1) zur Kodiereinheit wurde in der Analyse das gesamte Modul kodiert, in dem die Passage(n) enthalten ist (sind).

D3.2 – Nein, keine Erfahrung im Unterricht/Schule (nur Studierende) (1.00)	... dass studentische Proband*innen keine Erfahrung im Unterrichten (bspw. an einer Schule oder einem anderen Institut) bzw. im Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments gesammelt haben.	„Damit habe ich aber tatsächlich, also ich habe einige Erfahrungen tatsächlich mit ... der Betreuung von Auszubildenden, dann im Betrieb für verschiedene Projektarbeiten, Abschlussarbeiten und Ähnliches. Aber keine Lehrererfahrung.“ (Shirin, Pos. 6)
D4 – Ausbildungsfortschritt (.67)	... in welchem Masterjahr Studierende bzw. in welchem Ausbildungsjahr LiV sind.	Diese Kategorie unterteilt sich in zwei Subkategorien.
D4.1 – Erstes Masterjahr (Studierende)/ Ausbildungsjahr (LiV) (.67)	... dass studentische Proband*innen im ersten Masterjahr (höchstens zweites Semester) sind bzw. LiV im ersten Ausbildungsjahr. In diese Passage wird auch dann kodiert, wenn Sachverhalte/Äußerungen einschränkend wiedergegeben oder Limitationen in den Aussagen getätigt werden, die mit dem Ausbildungsstadium begründet werden.	„Also ich bin im Master Wirtschaftspädagogik jetzt im zweiten Fachsemester, dass ich bald abschließe.“ (Samira, Pos. 8) „Und dann habe ich im November 2019 den Vorbereitungsdienst ... angefangen“ (Rebecca, Pos. 7)
D4.2 – Zweites Masterjahr (Studierende)/ Ausbildungsjahr (LiV) (--) ¹⁸¹	... dass studentische Proband*innen (mindestens) im zweiten Masterjahr (mindestens 3. Semester) sind bzw. LiV im zweiten Ausbildungsjahr.	„Ich bin jetzt im dritten Mastersemester.“ (Sabrina, Pos. 48)
E - ERFAHRUNG PROBLEMHALTIGE (KLAUSUR-) AUFGABENERSTELLUNG (.77)		
E1 – Bisheriges Vorgehen Klausurerstellung (.77) wie Proband*innen bisher Klausuren oder andere schriftliche (summativ) Assessments erstellt haben. Dabei werden auch Passagen kodiert, aus denen hervorgeht, welche Hilfsmittel oder Wege sie genutzt haben. Die Kategorie besteht aus mehreren Subkategorien. In diese werden nur Passagen aus Transkripten kodiert, die aus Interviews mit erfahrenen Proband*innen sind, die bereits solche Assessments erstellt haben.	
E1.1 –Eigene Erfahrungen (1.00)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments auf eigene Erfahrungen zurückgreifen. Hierzu zählt auch die Kreativität.	„Man wird halt geübter ... also ich habe natürlich selber jetzt auch mehr Expertise seit dem gewonnen und unterrichte auch eher in Bereichen, wo ich meine Fachexpertise mit einbringen kann, da ist es natürlich einfacher“ (Rebecca, Pos. 19)

¹⁸¹ Es gibt keine Kodierung in den Transkripten, die in die Doppelkodierung gingen. Das Ankerbeispiel ist aus einem der anderen Transkripte.

E1.2 – Bewertung und Beurteilung (.64)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments an die Bewertung und Beurteilung dessen denken und berücksichtigen. Darunter zählen beispielsweise der Erwartungshorizont oder die Gedanken um die Bepunktung der Lösungen von Schüler*innen. Des Weiteren zählen auch Gütekriterien wie Objektivität, die in Bezug dazu genannt werden.	„... desto einfacher die Auswertung, desto besser für die Schüler und für uns. ... ähm als Lehrer gerade was die Prüfung anbetrifft, soll, muss man objektiv sein.“ (Sarah, Pos. 30)
E1.3 – Implementierung von Reproduktionsaufgaben (.73)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments auch Aufgaben einbauen, die deklaratives Wissen abfragen.	„Ansonsten habe ich natürlich auch so ein bisschen diese Standardaufgaben gemacht, ... aber es war schon sehr klar in meinen Aufgaben verständlich, dass ich jetzt, was mein Ziel war, also nenne mir Vor- und Nachteile kam es auch mal vor.“ (Seila, Pos. 22)
E1.4 – Abbildung realistischer beruflicher Probleme, Problemgehalt (--) ¹⁸²	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments darauf achten, ein realistisches berufliches Problem abzubilden. Hierzu zählen die Merkmale problemhaltiger Aufgaben: - Klarheit/Eindeutigkeit (Formulierung des Ist- und Zielzustands, Qualität und Quantität von Informationen, Lösungsoffenheit), - Dynamik, - Vernetztheit, - Interessensberücksichtigung, - Handlungskontrolle/ Reflexion sowie - Berufsrelevanz und Authentizität.	„Das heißt, da musste ich auch erst mal fragen: was ist dann vielleicht auch wichtig, was für Probleme können die auch in der Realität haben, mit denen sie umgehen müssen.“ (Rebecca, Pos. 11)
E1.5 – Abstimmung Instruktion und Assessment (1.00)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments darauf achten, dass diese mit der Instruktion im Einklang sind. Hierzu zählt unter anderem, dass ein Bezug zu den Inhalten aus dem Unterricht besteht.	„Und dann habe ich mir angeschaut: was habe ich mit denen im Unterricht gemacht? Und welche Prüfungsaufgaben kann ich daraus generieren, wie auch dann den entsprechenden Leistungsstand wirklich gut abbilden.“ (Rebecca, Pos. 11)
E1.6 – Komplexität (.64)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments auf die Komplexität der Aufgaben achten. Darunter fallen etwa die Bearbeitungsdauer, der Schwierigkeitsgrad (nicht unabhängig des Leistungs-/Lernstands der Schüler*innen) sowie die zugelassenen Hilfsmittel.	„Ich habe halt versucht, dass ich die Punkte irgendwie so anpasse, dass das Zeitmanagement von denen klappt, also dass sie sich nicht über- und unterfordert

¹⁸² Hierzu wurde kein Kappa-Wert berechnet, da im doppeltkodierten Material nur eine Passage von einer Kodiererin gefunden wurde.

fühlen, dass sie, ich glaube, ich habe 60 Minuten damals mit denen geschrieben, dass dann halt auch so 60 Punkte ungefähr gab und genau, habe halt versucht, dass sie die volle Zeit ausschreiben, aber auch jetzt nicht irgendwie in Zeitdruck kommen. Das war so was ich damals so berücksichtigt habe, dann natürlich auch die Hilfsmittel, die sie benutzen dürfen. Also die durften da Taschenrechner und das Umsatzsteuergesetz benutzen. Dass sie das auf, dass sie, also das wäre also meine Aufgaben so konstruiert, dass es notwendig war, diese dieses Buch zu benutzen und einen Taschenrechner und ja.“ (Seila, Pos. 22)

E1.7 – Hilfe über Dritte, Recherchearbeit
(.67)

... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments über Kolleg*innen, Mentor*innen, Familienmitgliedern oder anderen Personen Hilfe einholen. Hierzu zählt auch beispielsweise die Nutzung von Altklausuren als Arbeiten von Dritten. Zudem sind zu dieser Kategorie Passagen zu kodieren, aus denen hervorgeht, dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments Recherche betreiben, etwa im Internet und/oder durch die Nutzung von Schulbüchern. Die Berücksichtigung curricularer Anforderung wird nicht in diese Kategorie kodiert, sondern in die Kategorie zur Abstimmung des Curriculums mit Assessment bzw. Instruktion (E1.5).

„Also, ich bin so vorgegangen, wie gesagt halt nur Mathematik, dass ich ähm meinen Mathe-Mentor gefragt habe, er hat parallel zu meiner Klasse dieselbe Klasse oder in der selben Schulart unterrichtet und glaube alles BBV. Und der hat mir dann seine Klausuren gegeben. Und ich habe sie dann modifiziert, ... Und dann habe ich mir Bücher ausgeliehen und mir Aufgaben angeguckt und bisschen im Internet geforscht und ja und dann ein Mix gemacht aus den Klausuren, die mir vorlagen und das, was ich dann so ein bisschen recherchiert und eigene Aufgaben erstellt habe“ (Sarah, Pos. 17)

S – AUSBILDUNGSINHALTE IM STUDIUM (.75)

S1 – Vorhandene Inhalte im Studium
(.75)

... welche Inhalte im Sinne der pädagogischen Diagnostik allgemein und der problemhaltigen (Prüfungs-) Aufgabenerstellung im Speziellen im Studium behandelt werden. Dabei geht es um die Erfahrungen der Proband*innen bzw. deren Wahrnehmungen, was nicht unabhängig deren Nutzung von Lehr-Lernangeboten ist. Es wird dann kodiert, wenn von der Erfahrung in einem Modul gesprochen, das nicht die ist, in der die Intervention durchgeführt wurde.

S1.1 – (problemhaltige) Aufgabenerstellung/ -formulierung (.69)	... dass im Studium Inhalte zum Erstellen und/oder Formulieren von Aufgaben behandelt werden. In diese Kategorie wird auch dann kodiert, wenn von der Erstellung problemhaltiger Aufgaben gesprochen wird und auch, wenn Einschränkungen hierzu folgen.	„... ne, in schulpraktische Übung hatten wir auch, aber das war eine ... Stunde, das war wirklich nicht ein vertieftes Modul. Das war einmal, ein Block, hatten wir auch was gute Aufgaben ausmacht.“ (Samuel, Pos. 52)
S1.2 – Kompetenzorientierung (1.00)	... dass im Studium das Thema der Kompetenzen bzw. der Orientierung an diese behandelt wird. Es wird auch dann kodiert, wenn Einschränkungen hierzu folgen.	„Ja, ja. Doch, doch, darüber haben wir gesprochen und die Betonung liegt auf gesprochen. So, aber diese richtige Anwendung, die hattest du halt nicht. Also was Kompetenz bedeutet, was Handlungsorientierung bedeutet. Äh, vollständige Handlung.“ (Raphael, Pos. 47)
S1.3 – Praxisbezug (durch Praktika) (.68)	... dass im Studium praktische Erfahrungen, etwa durch Schulpraktika, ermöglicht werden. Diese wiederum werden als relevant für die pädagogische Diagnostik bzw. die problemhaltige (Prüfungs-) Aufgabenerstellung erachtet. Hierin werden auch Aussagen zur DBWA/VFD kodiert. Es wird auch dann kodiert, wenn Einschränkungen hierzu folgen.	„... für mich ... hat tatsächlich so diese diese Vorgehensweise auch viel Nutzen gestiftet, also wirklich dann doch erstmal theoretisch dieses Grundwissen aufzubauen und also im Bachelor-Nachstudiums-Auflagen und das dann letztendlich in der Praxis im Master anzuwenden.“ (Shirin, Pos. 65)
S1.4 – Bewertung und Beurteilung (.73)	... dass im Studium das Thema der Bewertung und/oder Beurteilung behandelt wird. Es wird auch dann kodiert, wenn Einschränkungen hierzu folgen.	„Ansonsten hatten wir auch was zu, ich weiß nicht mehr welches Modul es war, zu Noten und zum ... um Einfluss, ja, von verschiedenen Aspekten auf ... Noten also wie allein Namen zum Beispiel die Notengebung. So Aspekte. Aber um jetzt große Themengebiete zu sagen, weiß ich jetzt gar nicht mehr.“ (Samuel, Pos. 48)
S1.5 – Gütekriterien (.67)	... dass im Studium das Thema der Gütekriterien (im Kontext mit problemhaltigen (Prüfungs-) Aufgaben) behandelt wird. Es wird auch dann kodiert, wenn Einschränkungen hierzu folgen.	„Das war ähm in der OWIP zum Beispiel der Fall ... da sind wir das theoretisch durchgegangen und in der Verantwortung und in einer weiteren Veranstaltung der WPF, auch dieses Semester, da sind wir auch auf Gütekriterien eingegangen, aber nicht so detailliert tatsächlich wie in der DBWA.“ (Samira, Pos. 82)
S2 – Fehlende Inhalte im Studium & Wünsche zur Verbesserung	... welche Inhalte im Sinne der pädagogischen Diagnostik allgemein und der problemhaltigen (Prüfungs-) Aufgabenerstellung im Spezialen im Studium nicht (ausreichend) behandelt werden. Dabei ist zu beachten, dass es um die Wahrnehmungen der Proband*innen geht.	

(.72)	Zudem fallen in diese Kategorie auch die Wünsche, die Proband*innen zur Verbesserung des Studiums äußern. Diese Kategorie hat mehrere Subkategorien.	
S2.1 – nicht/kaum (problemhaltige) Aufgabenerstellung/ -formulierung (.78)	... dass Proband*innen wahrnehmen, dass im Studium Inhalte zur Abbildung realistischer beruflicher Probleme fehlen/ nicht (ausreichend) behandelt werden. Zu dieser Kategorie wird auch insgesamt auf das Thema der Erstellung von Aufgaben eingegangen, auch mit Blick auf deren Komplexität.	[Auf die Frage, ob im Studium Inhalte zur Aufgabenerstellung eingegangen wurde:] „Nein.“ (Sarah, Pos. 65)
S2.2 – Fehlender Praxisbezug/Anwendung (.65)	... dass Proband*innen wahrnehmen, dass im Studium der Praxisbezug bzw. die Anwendung der gelernten Inhalte fehlt/ nicht (ausreichend) vorhanden ist. Zu dieser Kategorie wird auch dann kodiert, wenn Aussagen getätigt werden, aus denen hervorgeht, dass zu viel/nur theoretische Inhalte behandelt werden.	„Und jetzt dieser Umschwung zu dieser Output-Orientierung, sage ich jetzt mal, dass das irgendwie auch im Referendariat auf dem Niveau ein bisschen bleibt, also dass, man merkt halt schon so ein bisschen, ... dass sie immer sagen, ja die aus Frankfurt, die sind so hochgebildet manchmal, also, so dass die so viel theoretisches Wissen haben, aber so praktisch sind sie dann eher hinten dran. Und ich finde das halt das halt das kein Nachteil sein soll, dass wir dann halt praktisch hintendran bisschen sind, sondern dass wir das dann halt im Ref sozusagen aufholen können.“ (Seila, Pos. 65)
S2.3 – Nicht gut/ unvorbereitet auf pädagogische Diagnostik (.64)	... dass Proband*innen wahrnehmen, im Studium nicht oder nicht ausreichend auf die Aufgabe der pädagogischen Diagnostik (im Allgemeinen) vorbereitet werden. In diese Kategorie wird auch dann kodiert, wenn Proband*innen ihre Wahrnehmung schildern, dass im Studium Inhalte zur pädagogischen Diagnostik im Allgemeinen fehlen/ nicht (ausreichend) behandelt werden. Ebenso werden Passagen in diese Kategorie aufgenommen, in denen nach mehr pädagogischen Modulen gefordert wird. Ebenso wird in diese Kategorie kodiert, wenn es um die Bewertung und Beurteilung geht, die den Teilnehmenden thematisch zu kurz kam/fehlte.	„Ähm einzelne Module definitiv ... Aber das sind so die Grundlagen, die geschaffen wurde, die sehr wichtig sind, gar keine Frage, um auch ein Verständnis für die Pädagogik zu bekommen. Ähm aber so richtig vorbereitet, ... aber abgesehen von ... unserem Modul jetzt, hatte ich kein Modul, das mich auf die Praxis wirklich vorbereitet hat.“ (Sarah, Pos. 51)
S2.4 – (Noch) nicht viele Pädagogik-Module (.73)	... dass Proband*innen noch nicht viele (wirtschafts-) pädagogische-Module belegt haben.	„um ehrlich zu sein, ... ich muss aber auch dazu sagen, dass ich noch einige Module vor mir habe, was das angeht.“ (Samuel, Pos. 37)
REF – AUSBILDUNGSINHALTE IM REFERENDARIAT (.85)		

REF1 – Vorhandene Inhalte im Referendariat (.88)	... welche Inhalte im Sinne der pädagogischen Diagnostik allgemein und der problemhaltigen (Prüfungs-) Aufgabenerstellung im Speziellen im Referendariat behandelt werden. Dabei geht es um die Erfahrungen der Proband*innen bzw. deren Wahrnehmungen, was nicht unabhängig deren Nutzung von Lehr-Lernangeboten ist.	
REF1.1 – (problemhaltige) Aufgabenerstellung/ -formulierung (1.00)	... dass im Referendariat das Thema der Erstellung von Aufgaben (wenn auch zu Unterrichtszwecken) behandelt wird. In diese Kategorie wird auch dann kodiert, wenn davon gesprochen wird, dass die Abbildung realistischer beruflicher Probleme für formative und/oder summative Zwecke behandelt wird.	„Bei der Erstellung problemhaltiger Aufgaben, hm. .. Ich habe tatsächlich, also mein Fachleiter in ... BWL, jetzt im Vorbereitungsdienst, der guckt auch ganz, ganz arg auf die Problemhaltigkeit von Lernsituationen zum Beispiel, das ist ja dasselbe Prinzip eigentlich wie bei Aufgaben.“ (Rebecca, Pos. 99)
REF1.2 – Kompetenzorientierung (.73)	... dass im Referendariat das Thema der Kompetenzen bzw. der Orientierung an diese behandelt wird.	„Also, grundsätzlich liegt bei uns am Seminar der Fokus immer auf den Kompetenzen, also auf der Kompetenz Anbahnung und der Förderung.“ (Rebecca, Pos. 39)
REF1.3 – Praxisbezug/ Anwendung (1.00)	... dass im Referendariat praktische Erfahrungen ermöglicht werden. Diese wiederum werden als relevant für die pädagogische Diagnostik bzw. die problemhaltige (Prüfungs-) Aufgabenerstellung erachtet. In diese Kategorie wird auch dann kodiert, wenn über Gruppenarbeiten oder aber auch den Austausch auf verschiedenen Ebenen gesprochen wird, die zur Anwendung beitragen.	„Also das kommt tatsächlich erst durch den Austausch jetzt im Studienseminar. Das ist auch das, was mir mehr hilft. Ich ... brauche nicht unbedingt diese klassische[n] Aufgaben, die wir erarbeiten uns jetzt dies, sondern ich habe das eher dann, wenn wir im Austausch sind und wir sprechen darüber und der eine bringt mir das Beispiel, der andere bringt mir das Beispiel, man spricht darüber, entwickelt vielleicht noch mal so ein paar Situationen, wie man das lösen hätte können und so weiter, das das ist deutlich fruchtbarer, wie wenn man das aus dem Buch zum Beispiel erfährt.“ (Raphael, Pos. 43)
REF1.4 – Bewertung und Beurteilung (1.00)	... dass im Referendariat das Thema der Bewertung und/oder Beurteilung behandelt wird.	„Da haben wir tatsächlich explizit auch Seminare zur Leistungsbewertung, Leistungsrückmeldung“ (Rebecca, Pos. 31)
REF1.5 – Diagnose/ pädagogische Diagnostik	... dass im Referendariat das Thema der pädagogischen Diagnostik (im Allgemeinen) behandelt wird.	„Und das ist auch ein Modul, das extra zur Diagnose und Bewertung dient, also die Bereitschaft fokussiert.

(.67)		Also da wird schon mehr Fokus daraufgelegt.“ (Rebecca, Pos. 31)
REF1.6 – Gütekriterien (1.00)	... dass im Referendariat das Thema der Gütekriterien (im Kontext mit problemhaltigen (Prüfungs-) Aufgaben) behandelt wird.	„Doch es wird thematisiert ... also bisher im pädagogischen Kontext und im Wirtschaftskontext. ... Ja, es wurde aber auch nur angerissen, dass wir da verschiedene Dinge zu beachten hätten und (unv.) wurde da auf Problemstellungen schon eingegangen, also in den Pädagogik bei uns. Dass wir jetzt aber konkret uns Aufgabenstellungen genommen haben, diese daraufhin überprüft haben und verändert haben und das mal im Unterricht durchgeführt haben, nein.“ (Ricardo, Pos. 65)
REF2 – Fehlende Inhalte im Referendariat & Wünsche zur Verbesserung (.78)	... welche Inhalte im Sinne der pädagogischen Diagnostik allgemein und der problemhaltigen (Prüfungs-) Aufgabenerstellung im Speziellen im Referendariat nicht (ausreichend) behandelt werden. Dabei ist zu beachten, dass es um die Wahrnehmung bzw. um die Gefühle der Proband*innen geht. Zudem fallen in diese Kategorie auch die Wünsche, a*innen zur Verbesserung des Referendariats äußern. Diese Kategorie hat mehrere Subkategorien.	
REF2.1 – nicht/kaum (problemhaltige) Aufgabenerstellung/ -formulierung (.73)	... dass Proband*innen wahrnehmen, dass im Referendariat Inhalte zur Erstellung von Aufgaben nicht, kaum oder nicht ausreichend behandelt werden. In diese Kategorie wird auch dann kodiert, wenn davon gesprochen wird, dass die Abbildung realistischer beruflicher Probleme fehlen/ nicht (ausreichend) behandelt werden.	„Es gibt dabei immer noch wenig Seminare, die über die Aufgabenerstellung gehen.“ (Rebecca, Pos. 35)
REF2.2 – Fehlender Praxisbezug/Anwendung (.73)	... dass Proband*innen wahrnehmen, dass im Referendariat der Praxisbezug bzw. die Anwendung der gelernten Inhalte fehlt/ nicht (ausreichend) vorhanden ist. Zu dieser Kategorie wird auch dann kodiert, wenn Aussagen getätigt werden, aus denen hervorgeht, dass zu viel/nur theoretische Inhalte besprochen bzw. behandelt werden.	„Also diese Verknüpfung von Theorie und Praxis fehlt mir dann oft, damit ich wirklich weiß, was gemeint ist.“ (Rebecca, Pos. 39)
REF2.3 – Nicht (ausreichend) vorbereitet auf pädagogische Diagnostik (.81)	... dass Proband*innen wahrnehmen, im Referendariat nicht oder nicht ausreichend auf die Aufgabe der pädagogischen Diagnostik (im Allgemeinen) vorbereitet worden zu sein.	„Da haben wir tatsächlich explizit auch Seminare zur Leistungsbewertung, Leistungsrückmeldung Und das ist auch ein Modul, das extra zur Diagnose und Bewertung dient, also die Bereitschaft fokussiert. Also da wird schon mehr Fokus daraufgelegt. Aber auch, finde ich, immer noch nicht genug.“ (Rebecca, Pos. 31)

Anhang E: Raster zu den Merkmalen problemhaltiger Aufgaben

Mit angepasstem Format aus Wuttke et al. (2022)

Merkmal/Frage		nicht/wenig ausgeprägt	Mittelmäßig ausgeprägt	stark ausgeprägt
Klarheit/Eindeutigkeit	Muss der Ist-Zustand identifiziert werden?	alle relevanten Informationen sind gegeben		Informationen vage gegeben
	Müssen Handlungsziele/ muss das Handlungsziel identifiziert und/oder präzisiert werden?	ist konkret gegeben		ist vage/unkonkret gegeben
	Müssen verschiedene Informationen erschlossen und abgeglichen werden?	wenig Informationen, strukturiert	viele Informationen, strukturiert	viele Informationen, nicht strukturiert
	Sind irrelevante Informationen vorhanden?	nur relevant		irrelevant
	Wie eindeutig/offen ist die zu entwickelnde Lösung?	Lösung eindeutig	Lösung aus fest definiertem begrenzten Lösungsraum	lösungsoffen
Dynamik	Ist die Situation dynamisch?	nein, statische Situation		dynamische Situation
Vernetztheit	Müssen Nebenwirkungen bei der Entscheidungsfindung bzw. Lösungsentwicklung berücksichtigt werden?	keine Wirkungen	Nebenwirkungen, ohne negative Konsequenzen (Synergieeffekte)	Nebenwirkungen mit möglichen negativen Konsequenzen
Interessens-berücksichtigung	Müssen Interessen verschiedener Stakeholder*innen (Geschäftspartner*innen, Kund*innen, Geldgeber*innen etc.) für die Entscheidungsfindung/Lösungsentwicklung berücksichtigt werden und sind diese widersprüchlich?	nur betriebliche Perspektive	betriebliche Perspektive und Perspektive eines Stakeholders	betriebliche Perspektive und Perspektive von mindestens zwei Stakeholdern (Widersprüchliche Interessen)
Handlungskontrolle/ Reflexion	Ist die Entscheidung bzw. Lösung zu begründen?	nein	ja, aus einer Perspektive	Ja, mehrperspektivisch

Anhang F: Zuordnung der Forschungsfragen zu den Fragen des teilstandardisierten Interviewleitfadens zur Trainingsevaluation und den Kategorien des Kodierleitfadens

Frage aus dem Leitfaden ¹⁸³	For- schungs- frage	Kategorie- kürzel ¹⁸⁴
Für einen Einstieg würde ich Sie bitten, sich kurz vorzustellen. Dabei brauchen Sie nicht Ihren Namen zu sagen. Bitte beschreiben Sie kurz, welche Vorbildung Sie haben, ob Sie Lehrerfahrung haben und was Sie aktuell beruflich tun.	---	D
<u>Studierende</u> (die vor der Intervention keine Erfahrung im Klausurerstellen hatten): Stellen Sie sich vor, Sie sind im Referendariat/Vorbereitungsdienst und haben von Ihrem*r Mentor*in die Aufgabe bekommen, eine Klausur zu erstellen. Wie würden Sie vorgehen? (Bei Studierenden mit Lehrerfahrung): Waren Sie schon mit einer solchen Situation konfrontiert? Wie sind Sie vorgegangen?	<i>F₂₋₂</i>	LA
<ul style="list-style-type: none"> • Welche Hindernisse bzw. Schwierigkeiten hatten Sie und wie sind Sie damit umgegangen? • Würden Sie dabei an folgende Punkte denken (optional): <ul style="list-style-type: none"> ○ Bewertung, Beurteilung und Auswertung ○ Gütekriterien ○ Aufgaben-/Itemformulierung ○ ... 		
Das Training bestand aus mehreren Teilen, die auf der Tischvorlage dargestellt sind.		
[Die Tischvorlage mit den Trainingsbausteinen – ähnlich zu Abbildung 14 und farbig– wurde im Querformat auf Vollbildschirmgröße eingeblendet]		
Wie fanden Sie die einzelnen Bausteine der Veranstaltung? Welchen Nutzen sehen Sie für sich aus den einzelnen Bausteinen? Bitte stellen Sie das mit den Punkten dar und begründen Ihre Einschätzung. Damit das übersichtlicher für Sie wird, habe ich eine Vorlage hierfür bereitgestellt [siehe Tischvorlage].	<i>F₂₋₁</i> <i>F₃</i>	R LA
Wie zufrieden waren Sie mit der Instruktion im Training?	<i>F₂₋₁</i> <i>F₃</i>	R
Welche Verbesserungen würden Sie insgesamt für das Training vorschlagen?	<i>F₃</i>	R LA

¹⁸³ Zum Teil wurden die Fragen im Interviewleitfaden an die Zielgruppe der Masterstudierenden bzw. der LiV/Referendar*innen angepasst. Wenn dies in der vorgestellten Frage des Leitfadens getan wurde, wird dies in dieser Tabelle dargestellt. Wenn in der Zelle keine Unterscheidung gemacht wird, kann davon ausgegangen werden, dass die Frage für beide Gruppen gleich lautete.

¹⁸⁴ Zu den Kategoriekürzeln: vgl. Kodierleitfaden Anhang G.

Denken Sie, dass Sie die Inhalte des Trainings anwenden werden? Warum?	<i>F</i> ₂₋₂ <i>F</i> ₃	LA
Würden Sie das Training weiterempfehlen? Warum?	<i>F</i> ₂₋₁ <i>F</i> ₃	R LA 1.
Ich danke Ihnen sehr für das Gespräch, Sie haben damit einen wichtigen Beitrag zur Beantwortung meiner Forschungsfragen beigetragen. Welche Themen möchten Sie diesbezüglich noch ansprechen? Welche Fragen haben Sie?	---	

Anhang G: Kodierleitfaden zur Interviewstudie mit angehenden Lehrkräften zur Trainingsevaluation

Kategoriekürzel – Kategorie (Kappa-Wert)	Beschreibung und Kodierregeln Darunter sind Passagen zu verstehen, aus denen hervorgeht, ...	Ankerbeispiel
D – DESKRIPTIVE DATEN		
Zur Ausführung der Subkategorien zu den deskriptiven Daten wird auf den Kodierleitfaden zur Interviewstudie mit Masterstudierenden der Wirtschaftspädagogik und LiV zu Ausbildungsinhalten verwiesen (vgl. Anhang D). Dadurch sollen Wiederholungen vermieden und Parallelen genutzt werden, die sich daraus ergaben, dass die Interviews zur Beantwortung der Forschungsfragen mit denselben Personen geführt wurden.		
E – ERFAHRUNG (KLAUSUR-) AUFGABENERSTELLUNG		
Siehe Hinweis zu „D – Deskriptive Daten“.		
R – REAKTION ZUR INTERVENTION (.80)		
R1 – Reaktion zur Gestaltung einzelner Bausteine (.84)	... wie Proband*innen die Bausteine der Interventionsstudie empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben. Die Kategorie unterteilt sich in mehrere Subkategorien, aufgeteilt nach den Bausteinen.	
R1.1 – Baustein Problemhaltige Aufgabenerstellung (.82)	... wie Proband*innen den Baustein zur <i>Erstellung problemhaltiger Aufgaben</i> empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	
R1.1.1 – Positives (.81)	... dass Proband*innen den Baustein zur <i>Erstellung problemhaltiger Aufgaben</i> (ACHTUNG: Keine Eigenentwicklung der Aufgaben, Eigenentwicklung findet in der Selbstlernphase und nach dem Transfer bei Studierenden statt) positiv empfunden und entsprechende Rückmeldung formuliert haben. Hierzu zählen unter anderen Aussagen wie <ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung 	„Also der erste Teil war der die Erstellung problemhaltiger Aufgaben. Und ich glaube, das habe ich eben schon mal kurz anklingen lassen. Das war für mich noch mal sehr erkenntnisreich, weil die Sachen, die wir dort gelernt oder besprochen haben, waren mir zwar schon bekannt, also gerade Problemorientierung, Handlungsorientierung. Aber das noch mal in einem anderen Kontext bei euch jetzt zu Hören und vielleicht nochmal mit einer anderen Unterfütterung von Beispielen auch, hat mir noch mal einige Erkenntnisse

	- Anwendung des Gelernten	über problemhaltige Aufgaben gebracht, was auch jetzt für mein Vorbereitungsdienst tatsächlich noch sehr hilfreich war und ist, das würde ich sagen, dann noch so die grundlegenden Erkenntnisse indem Bereich, das fand ich sehr gut, auch gerade so als ersten Schritt, damit man noch mal die Theorie wirklich präsent hat im Kopf.“ (Rebecca, Pos. 89)
R1.1.2 – Mehr Übung (.81)	... dass Proband*innen zum Baustein <i>Erstellung problemhaltiger Aufgaben</i> mehr Übung bzw. Anwendung gewünscht hätten und/oder die vorhandenen Einheiten als nicht ausreichend empfunden haben.	„Beim ersten Block, ... vielleicht wäre es besser gewesen, noch mehr selbst wirklich diese Beispielaufgaben zu machen und mehr Gruppenarbeiten und mehr, ja, das Umformulieren und wie man es problemhaltiger machen ... vielleicht diese, wenn ich das richtig im Kopf habe, war das erst noch so ein theoretischer Teil, sage ich, sage ich mal, ja vielleicht da die die Anteile bisschen verschieben.“ (Samuel, Pos. 157)
R1.2 – Baustein Mediendidaktik und Technologiebasiertes Arbeiten (.72)	... wie Proband*innen die Bausteine zur <i>Mediendidaktik</i> und/oder zum technologiebasierten Arbeiten/ <i>Umgang mit ILIAS</i> empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	
R1.2.1 – Positives (.73)	... dass Proband*innen die Bausteine zur <i>Mediendidaktik</i> und/oder zum technologiebasierten Arbeiten/ <i>Umgang mit ILIAS</i> positiv empfunden und entsprechende Rückmeldung formuliert haben. Hierzu zählen unter anderen Aussagen wie <ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung - Anwendung des Gelernten. 	„... digitale Medien zum Lernen und Prüfen fand ich einen ... guten Misch ... dadurch, dass Herr Niegemann ja auch vorgestellt hat, wie man auf was man auch medial achten sollte, kann die Vorlesung auch, welche Schrift und welche Schriftarten und was es bei einem bewirkt gerade auch bei meiner Präsentation mit ... diesem Bild, das eigentlich nicht so viel ausgesagt hat. Was bewirkt es eigentlich bei demjenigen, der, der das anschaut? Fande ich also, darüber habe ich mir so nie Gedanken gemacht. Also auch über die Schriftarten nicht, war schon immer klar, man nimmt halt die,

		die einem gefällt ja, und ... das hat ... auch neue Denkanstöße äh gegeben.“ (Sarah, Pos. 117)
R1.2.2 – Mehr Inhalt/Übung gewünscht (.69)	... dass Proband*innen zu mindestens einem der Bausteine <i>Mediendidaktik</i> und/oder Technologiebasiertes Arbeiten/ <i>Umgang mit ILLIAS</i> mehr Inhalt und Übung bzw. Anwendung gewünscht hätten und/oder die vorhandenen Einheiten als nicht ausreichend empfunden haben.	„Ja, oder halt, was mir auch überlegt hatte. Eine Aufgabe zum interaktiven Bild zu machen. Das habe mir so überlegt. Das wäre perfekt gewesen.“ (Seila, Pos. 88)
R1.3 – Baustein Selbstlernphase (.87)	... wie Proband*innen tutoriell begleitete <i>Selbstlernphase</i> empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	Rückmeldungen sie zu diesen äußern und welche Verbesserungs-
R1.3.1 – Positives (.78)	... dass Proband*innen die <i>Selbstlernphase</i> positiv empfunden und entsprechende Rückmeldung formuliert haben. Hierzu zählen unter anderen Aussagen wie <ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung - Anwendung des Gelernten. 	„Die Selbstlernphase. Dadurch, dass wir die Möglichkeit hatten, immer wieder in Kontakt mit ihnen zu treten, fand ich die auch sehr gut. Und man muss ja auch sagen, generell ist es ja am Berufsleben auch so, dass ich nicht dann jede Woche was habe, wo ich dann noch mal mit dem Dozenten oder nicht jede Woche mit meinem Vorgesetzten treffe oder mit der Schulleiterin sage, ja können wir jetzt noch mal kurz den Lehrplan durchgehen, sondern dann, wenn ich Fragen habe. Deswegen finde ich so eine Selbstlernphase eigentlich voll wichtig“ (Seila, Pos. 115)
R1.3.2 – Mehr Unterstützung gebraucht (1.00)	... dass Proband*innen in der <i>Selbstlernphase</i> mehr Unterstützung gebraucht hätten, feststellten, dass Ihnen gewisse Inhalte und/oder das Wissen gefehlt hat oder ähnliches. Auch wird in diese Kategorie kodiert, wenn eine Gruppenarbeit gewünscht wäre.	„Auch mit der Aufgabenstellung an sich fand ich es sehr schwierig, die alleine zu lösen. Mir hat das total gut gefallen, als wir das im Team gemacht haben. Da fiel es mir wesentlich einfacher, weil ich mich da auf meine Kompetenzen verlassen konnte, oder oder die die stärker fokussieren konnte und meine Mitstreiter in der Gruppe sich auf ihre Kompetenzen und so hatten wir ein ganz hervorragendes Team und konnten schöner und schneller und gefühlt auch besser Aufgaben erstellen, weil wir auch viel über Kleinigkeiten diskutiert haben.“ (Ricardo, Pos. 73)

R1.4 – Baustein Feedback/Reflexion (.85)	... wie Proband*innen den Baustein zum <i>Feedback/Reflexion</i> empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	
R1.4.1 – Positives (.85)	<p>Darunter sind Passagen zu verstehen, aus denen hervorgeht, dass Proband*innen den Baustein zum <i>Feedback/</i> zur <u>Reflexion</u> positiv empfunden und entsprechende Rückmeldung formuliert haben. Hierzu zählen unter anderen Aussagen wie</p> <ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung - Anwendung des Gelernten 	„Feedback nach der Präsentation hat sehr viel geholfen. Weil ähm das habe ich bisher in keinem Modul so erlebt, dass so intensiv eingegangen worden ist. ... fande diesen Feedbackbogen mega. Also da habe ich mir gedacht, ah ja, das war es und es hat nochmal, natürlich hat man sich, habe ich mir Notizen gemacht, als nach der Präsentation Feedback gegeben worden ist. ... was hatte ich in meiner Problemlöseaufgabe denn für Punkte, die ich nicht, ja, so nicht intensiv genug oder anders bedacht habe. Und was könnte ich noch mal anders machen? Also so als Reflektion nochmal finde ich sehr gut.“ (Sarah, Pos. 121)
R1.5 – Baustein Diagnostik/Gütekriterien (.86)	... wie Proband*innen den Baustein zur <i>Diagnostik, insb. Gütekriterien</i> , empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	
R1.5.1 – Positives (1.00)	<p>... dass Proband*innen den Baustein zur <i>Diagnostik, insb. Gütekriterien</i>, positiv empfunden und entsprechende Rückmeldung formuliert haben. Hierzu zählen unter anderen Aussagen wie</p> <ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung - Anwendung des Gelernten 	„... fande ich auch gut, fande ich auch gut erklärt, weil gerade, ja zum Beispiel das Scoring war für mich auch komplett neu, fande ich auch gut, dass wir da auch Gruppenarbeitsphasen hatten. Mit die die anderen Gütekriterien hatte ich schon in anderen Kursen. Aber ... ich muss trotzdem irgendwie immer noch mal nachschauen. Und da kann auch eigentlich Wiederholung nie schaden, finde ich. Und ja vor allen Dingen auf Aufgaben bezogen war das, denke ich, ja, war das doch auch noch mal ein anderer Kontext.“ (Samuel, Pos. 139)
R1.5.2 – Komplexität (.73)	... dass Proband*innen den Baustein <i>Diagnostik, insb. Gütekriterien</i> , inhaltlich überladen, zu schwierig, kompliziert, anspruchsvoll und/oder komplex empfunden haben. Auch Aussagen dazu, dass sie kognitiv ausgestiegen sind, werden zu dieser Kategorie kodiert.	„Der letzte Teil Diagnostik und Gütekriterien ... Wobei das noch mal sehr in die Tiefe gang, gegangen ist, fande ich. ... Wo ich dann, ehrlich gesagt, teilweise

		manchmal ein bisschen ausgestiegen bin. [lacht] So kognitiv.“ (Rebecca, Pos. 93)
R1.5.3 – Mehr Inhalt/ Übung gewünscht (1.00)	... dass Proband*innen zum Baustein <i>Diagnostik</i> , insb. <i>Gütekriterien</i> , mehr inhaltlichen Input, Übung bzw. Anwendung gewünscht hätten und/oder die vorhandenen Einheiten als nicht ausreichend empfunden haben.	„... nur bei der Diagnostik der Gütekriterien, klar, wir haben das gefühlt schon hundertmal gehört, aber genau bezogen auf die Prüfungsaufgabe, wäre es vielleicht besser gewesen, da noch ein bisschen mehr zu machen. ... ja man hört es zwar immer wieder, und irgendwie weiß man auch so man kann es auch noch mal googeln zur Not, was so sich dahinter verbirgt. Aber das so vielleicht noch mal speziell auf die auf die Prüfungsaufgabe anzuwenden oder auf einem Beispiel anzuwenden wäre irgendwie glaube ich noch besser gewesen“ (Seila, Pos. 117)
R1.6 – Baustein Transfer (.73)	... wie Proband*innen den Baustein zum <i>Transfer</i> der erstellten Aufgabe für Industriekaufleute auf den Beruf Kaufmann*frau für Büromanagement empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	
R1.6.1 – Positives (.73)	... dass Proband*innen den Baustein zum <i>Transfer</i> positiv empfunden und entsprechende Rückmeldung formuliert haben. Hierzu zählen unter anderen Aussagen wie <ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung - Anwendung des Gelernten 	„Der Transfer zu den Bürokaufleuten fand ich auch noch einmal sehr spannend, weil das auch für unseren Berufsalltag noch mal gezeigt hat, wie viel man aus einer Aufgabe auch für andere Bereiche machen kann. Wenn deine Verknüpfungen irgendwie herstellbar ist, das auch eben didaktisch zu reduzieren oder zu erweitern, dass das eigentlich gar nicht so ein, so eine Hürde ist, die ... übersprungen werden kann, so, ja. Und wie viele Parallelen da auch im kaufmännischen Bereich oft hergestellt werden können.“ (Rebecca, Pos. 91-93)
R1.7 – Sonstige Rückmeldung (.73)	... wie Proband*innen die Intervention empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben. Der Unterschied zu den anderen Subkategorien ist hier allerdings, dass Proband*innen diese nicht auf einen bestimmten Baustein bezogen formulieren.	
R1.7.1 – Positives (.73)	... dass Proband*innen die Intervention positiv empfunden und entsprechende Rückmeldung formuliert haben, ohne sich dabei auf einen konkreten Baustein zu beziehen. Hierzu zählen unter anderen Aussagen wie	„... ich finde einfach das Modul toll nur wie erwähnt. Also ich hoffe, dass Sie auch weiterhin so machen werden. Also es ist es wäre es war sehr bereichernd.

	<ul style="list-style-type: none"> - Interessant - Neue Inhalte - Verständlich - Praxisnähe - Prozessorientierung - Anwendung des Gelernten 	Und ja, und ich hoffe, für meine, für nachkommenden Studenten, dass die auch so ein Modul belegen dürfen, dass die Forschung weiter beibehalten wird oder in irgendeiner anderen Form wieder so, ja.“ (Sarah, Pos. 143)
R2 – Reaktion zum Aufbau der Intervention (.74)	... wie Proband*innen das Format, den Aufbau, die Anordnung und das (Nicht-)Vorhandensein einzelner Inhalte empfunden haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.	
R2.1 – Positives (.60)	... dass Proband*innen das Format, den Aufbau, die Anordnung und das (Nicht-)Vorhandensein einzelner Inhalte positiv bewertet haben. Hierzu zählen auch die positive Rückmeldung und Bewertung des Blockformats der Veranstaltung.	„Erstens von also vom Zeitmanagement finde ich es eigentlich ganz cool, dass man halt so mal fünf Stunden hatte und das dann alles so ein bisschen geballter war. ... Ansonsten finde ich es aber auch gut, dass es halt irgendwie Zeitmanagement mäßig stringent war. Also wir hatten Block, man wusste auch an dem Tag habe ich die DBWA, dann wird es mein Hauptaugenmerk. Und danach habe ich aber auch Zeit, wenn ich will, mich mit den Inhalten auseinanderzusetzen, ohne zu denken, okay es könnten jetzt noch 18 andere Inhalte kommen, das finde ich halt das Positive daran.“ (Seila, Pos. 113)
R2.2 – Umpositionierung der Bausteine (.81)	... dass Proband*innen einzelne Bausteine vorher oder nachher gewünscht bzw. an einer anderen Stelle als besser/sinnvoller empfunden hätten.	„Was man also, was man vielleicht noch weiter vorne machen könnte, wäre die Diagnostik. Also der letzte Baustein.“ (Rebecca, Pos. 107)
R2.3 – Fehlende Bausteine (1.00)	... dass Proband*innen einzelne Bausteine/ Inhalte gefehlt haben bzw. sie sich gewünscht hätten, weitere Bausteine/ Inhalte in der Intervention zu behandeln.	„Also das Einzige, was mir gefehlt hat, war wirklich ist, testen zu können.“ (Sarah, Pos. 130)
R2.4 – Contra Blockformat (.69)	... dass Proband*innen das Blockformat der Veranstaltung negativ bewerten bzw. negative Rückmeldung hierzu formulieren.	„... aber es war relativ viel dann für mich, weil ähm ja, vielleicht lag das auch einfach am am ... Block Termin und dass irgendwann, ja, dass irgendwann der Kopf auch ... Und irgendwann wird es dann viel. ... Aber ja, ich glaube, ich hätte mehr mitgenommen aus, aus

kleineren Teilen, also, wenn das irgendwie so 20 Minuten Blöcke oder so was wären.“ (Samuel, Pos. 133-135)

R3 – Reaktion zum Material der Intervention
(.64)

... wie Proband*innen das zur Verfügung gestellte Material im Rahmen der Interventionsstudie (bspw. Folien, Aufnahmen, Handbücher) empfunden und ggf. genutzt haben, welche Rückmeldungen sie zu diesen äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben. Das Lernmanagementsystem ILIAS ist nicht als Material zu verstehen.

„... gerade dieses Raster zur Problemhaltigkeit, was ihr im Training uns auch zur Hand gegeben habt, finde ich sehr hilfreich. Und das habe ich tatsächlich jetzt auch schon zur Überprüfung von meinen eigenen Lernsituationen eingesetzt, um zu gucken, was was an welchen Stellschrauben könnte ich da vielleicht noch drehen. Und das ist, glaube ich so eine Grunderkenntnis, die ich auch aus dem Training mitgenommen habe.“ (Rebecca, Pos. 99)

R4 – Reaktion zur Instruktion/Hilfestellung
(.73)

... wie Proband*innen die Instruktion in der Interventionsstudie sowie die Rückmeldungen und Hilfestellungen der Dozent*innen empfunden haben, welche Rückmeldungen sie dazu äußern und welche Verbesserungswünsche/-vorschläge sie ggf. haben.

„... dadurch, dass ich aber von Ihnen auch noch so viel Feedback bekommen habe, war das schon sehr hilfreich ... Und dann fand ich es halt gut, dass Sie dann kamen und gesagt haben, okay die Ideen sind gut, aber sie müssen da irgendwie gucken, dass Sie eine Struktur reinbekommen oder ein bisschen weniger werden oder halt gucken, wenn Sie so viel machen, ob Sie es vielleicht aufteilen und so weiter. Und genau, das war mir dann zum Beispiel sehr hilfreich, dass man halt Feedback bekommt.“ (Seila, Pos. 80)

LA –LERNZUWACHS DURCH INTERVENTION UND ANWENDUNG (.78)

L1 – Nutzen/Lernzuwachs durch einzelne Bausteine
(.83)

... dass Proband*innen einen Nutzen aus der Intervention ziehen konnten. Darunter fällt auch der Lernzuwachs. Dieser wird häufig verbal als „etwas mitgenommen“, „etwas gebracht“ formuliert. Diese Kategorie unterteilt sich in Subkategorien.

L1.1 – Unspezifisch
(.68)

... dass Proband*innen einen Nutzen bzw. einen Lernzuwachs haben. Worauf dieser sich bezieht, ist allerdings unklar formuliert.

„Also ... also ich fand das Training für mich sehr gewinnbringend und ich glaube, auch die anderen Teilnehmer, mit denen ich mich so ausgetauscht habe, fanden das auch sehr gewinnbringend für sich und fast alle haben gesagt, gerade im Vorbereitungsdienst

kommt dieses Thema einfach zu kurz. Und ich könnte mir gut vorstellen, dass man wirklich als Referendarin oder Referendar besser auf die Berufspraxis vorbereitet werden würde, wenn so ein Training, wie ihr das anbietet, auch mal im Vorbereitungsdienst integriert wäre, also verpflichtend integriert wäre, zum Beispiel.“ (Rebecca, Pos. 142)

L1.2 – Problemhaltige Aufgabenerstellung
(.78)

... dass Proband*innen einen Nutzen bzw. einen Lernzuwachs haben. Dieser bezieht sich auf den Baustein *zum Erstellen problemhaltiger Aufgaben* bzw. um den Nutzen und/oder das Lernen in Bezug auf das *Erstellen problemhaltiger Aufgaben*.

“Erstellung problemhaltiger Aufgaben ..., auf jeden Fall das mit einem Raster, das war mir gar nicht so bewusst, welche einzelne Aspekte man da drauf achten muss. Also man denkt immer so, ja, wenn man ein Aspekt hat, also so eine offene Aufgabe einfach zu erstellen, dann ist es schon problemhaltig genügt genug. Aber dass da noch andere Faktoren reinspielen, das hat mir halt auch auf jeden Fall ein Lernzuwachs gegeben.” (Seila, Pos. 128)

L1.3 – Mediendidaktik und Technologiebasiertes Arbeiten
(.90)

... dass Proband*innen einen Nutzen bzw. einen Lernzuwachs haben. Dieser bezieht sich auf den Baustein *Mediendidaktik* und/oder *Technologiebasiertes Arbeiten/ Umgang mit ILIAS* bzw. um die mediendidaktisch ansprechende Gestaltung von Aufgaben und/oder dessen Umsetzung in ILIAS.

„Also auf jeden Fall auch das mit den digitalen Medien, also mit diesem interaktiven Bild, also das man, oder auch mit dem Video. Da hat ich ja auch ein Lernzuwachs, da ich halt jetzt mittlerweile ein bisschen mich besser auskenne mit Video schneiden, dann auch dieses Urheberrechts, also ich wusste, dass es da was gibt und dass man nicht einfach irgendwie Random irgendetwas aus dem Internet ziehen darf. Aber wie sehr man das doch beachten muss und so weiter, hat mir das noch mal bisschen so verdeutlicht, da hatte ich auch einen kleinen Lernzuwachs und natürlich irgend-eine technologiebasierte Anwendung mal zu testen.“ (Seila, Pos. 128)

L1.4 – Selbstlernphase
(1.00)

... dass Proband*innen einen Nutzen bzw. einen Lernzuwachs haben. Dieser bezieht sich auf die *Selbstlernphase*.

„Selbstlernphase ist immer in irgendeiner Weise Lernzuwachs, weil man halt sich selbst damit beschäftigen muss, ohne irgendwie nochmal von also, klar man

kann sich Hilfe holen, aber dass man halt trotzdem erst mal probiert, es selbst zu machen” (Seila, Pos. 128)

L1.5 – Feedback, Reflexion
(1.00)

... dass Proband*innen einen Nutzen bzw. einen Lernzuwachs haben. Dieser bezieht sich auf den Baustein *Feedback und Reflexion*.

„Ich glaube, wenn ich an einem früheren Punkt gestoppt hätte nach der Präsentation ... beziehungsweise wenn da überhaupt keine Präsentation stattgefunden hätte, sondern man gesagt hätte, an der Stelle, das ist das Ende ... und man gibt es ab und diskutiert es nicht irgendwie in der Gruppe und erhält nicht Ihr Feedback, Herr Niegemanns Feedback, dann hätte mir wirklich viel von meiner Lerngruppe gefehlt, würde ich sagen, ja.“ (Shirin, Pos. 73)

L1.6 – Diagnostik
(1.00)

... dass Proband*innen einen Nutzen bzw. Lernzuwachs haben. Dieser bezieht sich auf den Baustein *Diagnostik, insb. Gütekriterien* bzw. um eine entsprechende Gestaltung und Prüfung einer Aufgabe. Auch fällt unter diese Kategorie, wenn Proband*innen auf den Nutzen bzw. Lernzuwachs in Bezug auf die Bewertung und Beurteilung von Aufgaben aber auch Schüler*innenleistungen eingehen, ebenso wie auf Gütekriterien.

„Ja, das habe ich jetzt tatsächlich auch intensiv gemerkt. Also dafür war der Kurs tatsächlich auch sehr gut. Ich bin es tatsächlich auch gewohnt, hauptsächlich Klausuren, Prüfungen, Arbeiten zu schreiben, die wirklich sehr sehr viel mit Freitextarbeiten. Und habe aber jetzt auch gesehen, dass tatsächlich auch verschiedene andere Varianten wie zum Beispiel Multiple Choice dann auch große Chancen bieten und gleichzeitig auch nicht sehr sehr viel an Mehrwert verlieren. ... es gibt natürlich große Chancen ... Frage-Typen auszuwählen oder wo im Nachhinein die Auswertung leichter funktioniert. Und das habe ich jetzt auch bei dieser Hausarbeit wieder gemerkt, dass natürlich, ... dass letztendlich der Erwartungshorizont viel viel aufwendiger in der Erstellung war, auch das Scoring dann nochmal aufwendiger war und es auch mir gar nicht so leichtgefallen ist.“ (Shirin, Pos. 30)

L1.7 – Transfer
(--)¹⁸⁵

... dass Proband*innen einen Nutzen bzw. einen Lernzuwachs aus dem Trainingsbaustein zum *Transfer* haben.

„Den Transfer, wie gesagt, habe ich sehr viel mitgenommen.“ (Sami, Pos. 136)

¹⁸⁵ Es gab keine Kodierung in den Transkripten, die in die Doppelkodierung gingen. Das Ankerbeispiel ist aus einem der anderen Transkripte.

L2 – Kein Nutzen/Lernzuwachs durch einzelne Bausteine (.88)	... dass Proband*innen keinen Nutzen aus der Intervention ziehen konnten. Darunter fällt auch der fehlende Lernzuwachs. Dieser wird häufig verbal als „nicht/wenig mitgenommen“, „nichts/wenig gebracht“ formuliert. Diese Kategorie unterteilt sich in Subkategorien.	
L2.1 – Mediendidaktik und Technologiebasiertes Arbeiten (.85)	... dass Proband*innen keinen Nutzen bzw. Lernzuwachs haben. Dieser bezieht sich auf den Baustein <i>Mediendidaktik</i> und/oder Technologiebasiertes Arbeiten/ <i>Umgang mit ILIAS</i> bzw. um die mediendidaktisch ansprechende Gestaltung von Aufgaben und/oder dessen Umsetzung in ILIAS.	„Aber teilweise war es auch etwas trocken Aber ja, ich glaube, da habe ich wahrscheinlich weniger von mitgenommen, was ich jetzt noch im Kopf habe, als aus anderen Sachen.“ (Samuel, Pos. 133)
L2.2 – Diagnostik (1.00)	... dass Proband*innen keinen Nutzen bzw. keinen Lernzuwachs haben. Dieser bezieht sich auf den Baustein <i>Diagnostik</i> , insb. <i>Gütekriterien</i> , bzw. um eine entsprechende Gestaltung und Prüfung einer Aufgabe.	„... einen Lernzuwachs bei Diagnostik der Gütekriterien war es halt so, okay, ich setze mich jetzt noch einmal mit dem Inhalt auseinander, aber das ist so, hört sich jetzt auch ein bisschen Lump an, aber ich habe es halt nur gemacht, weil ich es weiß, ich muss es jetzt in der Hausarbeit machen. Und ich versuche mich da jetzt kurz reinzufuchsen und irgendwie zu verstehen, was dahinter steckt. Aber das ist so eine Sache, die werde ich wahrscheinlich in drei Wochen wieder vergessen haben.“ (Seila, Pos. 128)
A1 – Anwendbarkeit/Transfer in Praxis (.72)	... (1) welchen Transfer die Proband*innen aus der Interventionsstudie in die Praxis leisten würden bzw. wie und wo sie die Anwendbarkeit des Gelernten sehen. Zudem wird in diese Kategorie auch kodiert (2) wie Proband*innen Klausuren oder andere schriftliche (summative) Assessments erstellen würden. Dabei werden auch Passagen kodiert, aus denen hervorgeht, welche Hilfsmittel oder Wege sie nutzen würden.	
A1.1 – unspezifisch (1.00)	... dass Proband*innen aus der Interventionsmaßnahme Transfer leisten bzw. die Inhalte anwenden können. Was genau, wird allerdings in diesen Passagen nicht ausgeführt.	[Auf die Frage, ob aus dem Training etwas mitgenommen wurde, was angewendet werden kann:] „Ja, auf jeden Fall.“ (Rebecca, Pos. 133)
A1.2 – Problemhaltige Aufgabengestaltung (.63)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments darauf achten werden/würden, ein realistisches berufliches Problem abzubilden. Hierzu zählen die Merkmale problemhaltiger Aufgaben:	„Ah ja, ... da würde ich mir konkret natürlich das das zu dem Thema entsprechend konkret überlegen, was ist der Ist-Zustand, was ist der Soll-Zustand? Und wo kann ich dazwischen mixen, dass das Problem zwar erkannt wird, aber nicht so eindeutig erkannt wird. Ja?

	<ul style="list-style-type: none"> - Klarheit/Eindeutigkeit (Formulierung des Ist- und Zielzustands, Qualität und Quantität von Informationen, Lösungsoffenheit), - Dynamik, - Vernetztheit, - Interessensberücksichtigung, - Handlungskontrolle/ Reflexion sowie - Berufsrelevanz und Authentizität. <p>Auch Realitätsnähe wird in diese Kategorie kodiert.</p>	Vielleicht so noch ein bisschen versteckt für die guten Schüler, so ein Teil eindeutschen, ein Teil ein bisschen versteckt. Und über diesen Weg würde ich mir genauer Gedanken machen.“ (Sarah, Pos. 21)
A1.3 – (keine) Implementierung von Reproduktionsaufgaben (.64)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments sich Gedanken darüber machen werden/würden, ob sie Aufgaben einbauen, die deklaratives Wissen abfragen. In diese Kategorie wird dann kodiert, wenn sie davon berichten, auf die Implementierung solcher Aufgaben zu verzichten.	„Ich würde mich aber definitiv nicht mit hundert Prozent auf die Aufgaben stürzen, sage ich mal.“ (Samuel, Pos. 33)
A1.4 – Abstimmung Curriculum-Instruktion-Assessment (.61)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments auf die Abstimmung zwischen Curriculum-Instruktion-Assessment als Triade oder zumindest des Curriculums und des Assessments oder der Instruktion und des Assessments achten werden/würden. Hierzu zählt auch das kompetenzorientierte Arbeiten, wobei mit Kompetenzen, die curricular formulierten gemeint sind. Ebenso wird auch dann in diese Kategorie kodiert, wenn das Thema „teaching-to-test“ implizit oder explizit kritisch betrachtet wird.	„Als Erstes würde ich berücksichtigen wie intensiv wir das Thema besprochen haben, dann, ähm welche Themen wie intensiv besprochen wurde. Dann würde ich mir ein Thema rauspicken, das in der Prüfungsordnung oder im ja Ausbildungsordnung, im Rahmenlehrplan, wie auch immer, explizit genauer genannt ist, weil meistens wird es ja grob genannt, und es gibt so ... Unterteilungen, da würde ich mir eins zwei rauspicken und ähm, je nachdem, welche Ausbildung oder Schulart es ist, würde ich mich auch so ein bisschen an die Prüfungen orientieren, welche Themen schon vorkamen, das wäre das das erste, dass ich mir über das Thema Gedanken mache.“ (Sarah, Pos. 25)
A1.5 – Komplexität (.67)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments auf die Komplexität der Aufgaben achten werden/würden. Darunter fallen etwa die Bearbeitungsdauer, der Schwierigkeitsgrad (nicht unabhängig des Leistungs-/Lernstands der Schüler*innen) sowie die zugelassenen Hilfsmittel.	„Und dann ähm würde ich mir überlegen, welcher Schwierigkeitsgrad, also ich würde die Aufgaben ähm so konzipieren, dass die erst halt vielleicht ein bisschen einfacher sind, oder die Erste, und dann mit den folgenden Aufgaben es einfach schwieriger wird oder herausfordernder, sagen wir es so.“ (Sarah, Pos. 25)

A1.6 – Bewertung und Beurteilung (.73)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments sich Gedanken um die Bewertung und Beurteilung der Schüler*innenlösungen zur Aufgabe machen werden/würden. Darunter zählen beispielsweise der Erwartungshorizont oder die Gedanken um die Bepunktung der Lösungen von Schüler*innen. Des Weiteren zählen auch Gütekriterien wie Objektivität, die in Bezug dazu genannt werden.	„Weil, wenn im Vorhinein ein Erwartungshorizont oder eine Musterlösung erstellt wird, dann weiß ich ja die und die Aspekte erwarte ich von den Schülerinnen und Schülern, dass sie in der Aufgabe enthalten sind, um so einen für mich dann einen Leitfaden zu haben oder eine Richtlinie zu haben. Ob das in etwa dem entspricht, was dem Inhalt und dann den Anforderungen des Themas gleichkommt.“ (Samira, Pos. 48)
A1.7 – Technologiebasiertes Arbeiten (.62)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments oder aber auch im Unterricht auf das technologiebasierte (und dadurch authentische) Abbildung bzw. Umsetzung der Aufgabe darauf achten werden/würden. In diese Kategorie wird auch dann kodiert, wenn sie sich dabei nicht auf ILIAS beziehen und/oder sagen, dass sie nicht mit ILIAS arbeiten werden.	„Ich glaube, dass E-Learning insgesamt auf jeden Fall sehr wichtig ist. Und man ... das auf jeden Fall anwenden kann, das dann letztendlich Ilias ist oder ein anderes Programm.“ (Samuel, Pos. 147)
A1.8 – Transfer einer Aufgabe (1.00)	... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments diese so transferieren werden/würden, dass sie auch bei anderen Berufsgruppen eingesetzt werden können.	„Ja, und der Transfer, ... was ich mitnehmen kann, dass man halt nicht so denkt, ... oh ne jetzt muss ich noch eine andere Aufgabe, ... wie soll ich das jetzt hibekommen und so, aber dass man halt weiß, okay, das ist aber Alltag, das muss man einfach machen und ich kann jetzt nicht nur, weil es jetzt einmal bei einer Aufgabe hibekommen habe, kann ich jetzt nicht sagen na ja, jetzt Scheuklappen auf, ich mache es jetzt nur noch. Ich mach's übernehm es jetzt überall da so, sondern ... man ... muss wissen, okay, man kann es aber nicht, auch wenn es auch das gleiche Fach ist, einfach so blind übernehmen.“ (Seila, Pos. 135)
A2 – Schwierigkeiten und Hilfe in Anwendbarkeit/ Transfer in Praxis (.84)	... welche Schwierigkeiten bzw. Herausforderungen bezüglich des Transfers/ die Anwendbarkeit aus der Interventionsstudie in die Praxis aus Sicht der Proband*innen bestehen (würden). Zudem wird in diese Kategorie auch kodiert welche Schwierigkeiten Proband*innen beim Erstellen von Klausuren oder andere schriftliche (summativ) Assessments sehen würden. Dabei werden auch Passagen kodiert, aus denen hervorgeht, welche Hilfsmittelgebraucht werden (würden). Die Kategorie unterteilt sich in mehrere Subkategorien.	

A2.1 – Hilfe über Dritte
(1.00)

... dass Proband*innen beim Erstellen von Klausuren oder anderen schriftlichen (summativen) Assessments Hilfe über etwa Kolleg*innen, Mentor*innen bzw. deren vorherigen Arbeiten in Form von etwa Altklausuren einholen werden/würden.

„Ähm, ich ... glaube, da auch an den Kurs direkt anzuschließen, ist es ziemlich schwer, inwieweit man auf Prüfungsvorbereitung für die Auszubildenden sich da fokussiert und quasi aus sich an den alten ja, Abschlussprüfungen et cetera orientiert oder hier eben selbst anders vorgeht, weil wir haben ja gesehen, dass das nicht unbedingt die die besten Aufgaben sind ... Ich glaube, ich würde mich auch ein bisschen an ... den Mentoren et cetera orientieren. Was, was die mir vorschlagen.“ (Samuel, Pos. 25)

A2.2 – Eingeschränkte Möglichkeiten in Bezug auf die Prüfungspraxis
(.73)

... dass Proband*innen nur eingeschränkte Möglichkeiten für den Transfer bzw. die Anwendung der gelernten Inhalte in der Prüfungspraxis sehen, da beispielsweise der Einfluss einer einzelnen Person nicht ausreichend scheint. Ebenso wird in die Kategorie kodiert, wenn Teilnehmende aufgrund der Prüfung(-vorbereitung) eingeschränkte Möglichkeiten zum Transfer bzw. zur Anwendung des Gelernten sehen.

„Ja, was die Prüfungsaufgaben oder Aufgaben insgesamt angeht ... sind da einem eben auch die Hände gebunden als ... Lehrer. ... ich glaube, im Unterricht kann man das sehr gut anwenden .. Aber letztendlich muss man ja auch auf die Prüfungen vorbereiten, es sollte nicht so sein. Aber letztendlich ist das ja die Realität Und da muss man halt auch ein Stück weit dann pragmatisch sein und gucken, was dann am Ende abgefragt wird und hat ja da irgendwie auch eine soziale Verantwortung, nicht nur eine Verantwortung, den Schülern was beizubringen, sondern eben auch, ja, es ist vielleicht bisschen demotivierend, aber ich glaube, bei zumindest bei einigen Schülern ist wichtiger, dass sie letztendlich eine gute Note am Ende dann haben, als wenn der Unterricht noch toll ist und noch problemhaltiger und so weiter. Es ist deprimierend das so zuzusagen.“ (Samuel, Pos. 147)
