



TECHNISCHE UNIVERSITÄT MÜNCHEN



TUM School of Education

Susanne Klatten-Stiftungslehrstuhl für Empirische Bildungsforschung

Herausforderung Prüfen:

Eine fallbasierte Untersuchung der Prüfungspraxis von Hochschullehrenden im
Rahmen eines Qualitätsentwicklungsprogramms

Christoph Josef Schindler

Vollständiger Abdruck der von der Fakultät *TUM School of Education*
der Technischen Universität München zur Erlangung des akademischen Grades
eines

Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitzender:	apl. Prof. Dr. Alfred Riedl
	1. Univ.- Prof. Dr. Manfred Prenzel
Prüfer der Dissertation:	2. Univ.- Prof. Dr. Christina Seidel

Die Dissertation wurde am 02.07.2015 bei der Technischen Universität München eingereicht und von der Fakultät *TUM School of Education* am 30.07.2015 angenommen.

Danksagung

Das Leben besteht aus kleineren und größeren Prüfungen. Rückblickend war das Anfertigen dieser Dissertation für mich bisher wohl die anspruchsvollste, aber auch mit die lehrreichste Prüfungsphase. Ich habe von vielen Seiten Unterstützung erhalten, um die damit verbundenen Prüfungsanforderungen bewältigen zu können. An erster Stelle ist hier Prof. Dr. Manfred Prenzel zu nennen, der mir das notwendige Vertrauen zur Durchführung des mit der Dissertation verbundenen Forschungsprojekts entgegenbrachte und mir vielfältige Gestaltungsspielräume einräumte. Ich danke ihm für die vielen anregenden und konstruktiven Gespräche, aus denen ich viele Ideen und Hinweise für meine Arbeit mitnehmen konnte. Meiner Zweitbetreuerin, Prof. Dr. Tina Seidel, möchte ich an dieser Stelle dafür danken, dass sie mir beigebracht hat, meine Ideen zu strukturieren und „beforschbar“ zu machen. Außerdem danke ich ihr für das immer wertschätzende Feedback, das mich auf der einen Seite in meinem Tun bekräftigt hat, mir aber auch stets Verbesserungspotenziale aufzeigte.

Bei meinen beiden Kolleginnen von ProLehre, Dr. Alexandra Strasser und Dr. Nadine Schlomske-Bodenstein, möchte ich mich für die Zusammenarbeit bedanken. Sie haben wesentlich dazu beigetragen, das Programm „Herausforderung Prüfen“ in der Praxis umzusetzen. An dieser Stelle gilt mein Dank auch den Teilnehmenden, die sich mit uns auf den Weg gemacht haben, Prüfungen an der TU München zu verbessern.

Außerdem danke ich PD Dr. Johannes Bauer, der mir stets als Berater zur Seite stand. Im Prozess konnte ich viel von ihm lernen und von seiner Erfahrung profitieren. Viel Dank geht auch an meinen geschätzten Kollegen Florian Schulz, mit dem ich mich in den letzten Jahren in das Thema „Prüfungen“ einarbeiten durfte. Weiterhin danke ich meinen Kolleginnen und Kollegen, sowie Hilfskräften am Lehrstuhl für ihre Unterstützung und die stets angenehme Arbeitsatmosphäre.

Ganz besonders möchte an dieser Stelle Prof. Dr. Richard Shavelson danken, der mir ein unverzichtbarer Mentor war und so einige gedankliche Knoten bei mir lösen konnte.

Mein tief verbundener Dank gilt meiner Lebensgefährtin Ann-Kathrin Pehmer für ihre unabdingbare Unterstützung. Von ihrer Leidenschaft für die Forschung konnte ich fachlich und persönlich viel lernen. Besonders bedanke ich mich bei meinen Eltern - Angela und Josef Schindler - und meinen Geschwistern - Christina, Johannes und Sebastian - für Ihre lebenslange Unterstützung bei der Bewältigung von Prüfungssituationen. Meiner Cousine Cornelia Weber und ihrem Ehemann Dr. Norbert Huber danke ich ebenfalls für ihre Unterstützung in vielerlei Hinsicht und dafür, dass sie mir in allen Lebenslagen mit Rat und Tat zur Seite stehen. Zum Abschluss ein besonderer Gruß an mein Patenkind Johann und seine Geschwister Valentin und Viola.

Kurzzusammenfassung

Prüfungen nehmen eine zentrale Rolle im Lehr-Lernalltag an Hochschulen ein. Prüfungsanforderungen steuern das Lernverhalten von Studierenden und beeinflussen so die Qualität und nachhaltige Verfügbarkeit der im Studium erworbenen Qualifikationen und Kompetenzen (Cilliers, Schuwirth, Adendorff, Herman, & Van der Vleuten, 2010; Segers, Nijhuis, & Gijsselaers, 2006). Unabdingbar ist dabei, dass Prüfungen bestimmte Standards und Kriterien erfüllen (Rost, 2004; Wilson, 2005). Das Prüfen an Hochschulen liegt im Verantwortungsbereich der dort Lehrenden. Diese werden oftmals ohne besondere Vorkenntnisse mit dem Thema konfrontiert. Für Programme zur Unterstützung der Lehrenden bei der Prüfungserstellung fehlte es bisher an einer Ausgangsbasis in Form verlässlicher Daten. In der vorliegenden Arbeit werden deshalb zunächst Ergebnisse aus einer fallbasierten Untersuchung der Hochschulpraxis von Lehrenden vorgestellt. Die Untersuchungen wurden entlang eines Modells, das im Rahmen dieser Arbeit auf Basis des aktuellen Kenntnisstands zur Kompetenzmessung und zur pädagogischen Diagnostik entwickelt wurde, vorgenommen. Die Ergebnisse der Analysen bestehender Prüfungsaufgaben zeigen diverse Problemstellen auf: So bezogen sich die Lehrenden bei der Erstellung von Prüfungsaufgaben nicht – wie bei der Testerstellung üblich - auf vorab definierte Lehrziele. Die Ergebnisse zeigen weiter, dass die Prüfungsaufgaben großteils entkoppelt von den deklarierten Zielen der Veranstaltung konzipiert wurden. Ebenso wurden Merkmale von Kompetenzorientierung, wie sie immer stärker auch von Hochschulen eingefordert werden, in den betrachteten Prüfungen nur unzureichend umgesetzt. Das weitere Anliegen dieser Arbeit war deshalb die Entwicklung eines Qualitätsentwicklungsprogramms, das unter anderem die genannten Schwachstellen in der bestehenden Prüfungspraxis aufgreift und sich dabei aber gleichermaßen an den Zielvorstellungen der Teilnehmenden orientiert. Untersucht wurde dabei das Potenzial zur nachhaltigen Veränderung der Prüfungspraxis durch die Maßnahme. Ein Ergebnis der Analysen war, dass die Lehrenden schwerpunktmäßig an der Bewertung von Prüfungsantworten arbeiten wollten, da sie hierin eine Möglichkeit sahen, möglichen Beschwerden von Seiten der Studierenden zu begegnen. Wie bereits in den bestehenden Prüfungen und Prüfungsaufgaben war auch in den Zielsetzungen kein Bewusstsein hinsichtlich einer Kompetenz- und Lehrzielorientierung erkennbar. Im Verlauf des Qualitätsentwicklungsprogramms entwickelten die Teilnehmenden eine solche Lehrzielorientierung und konnten dadurch zahlreiche Probleme in ihren Prüfungen beheben. Ebenso konnte die Kompetenzorientierung in den Prüfungsaufgaben gesteigert werden. Die Ergebnisse der Studie zeigen auf, inwieweit Lehrende durch ein auf sie zugeschnittenes Qualitätsentwicklungsprogramm in der Verbesserung ihrer Prüfungspraxis unterstützt werden können. Auf dieser Grundlage werden abschließend Handlungsempfehlungen für die Hochschuldidaktik im Sinne eines Best-Practices gegeben und weiterführende Forschungsperspektiven aufgezeigt.

Inhaltsverzeichnis

Inhaltsverzeichnis	4
1 Einleitung	1
2 Theoretischer Hintergrund	5
2.1 Entwicklung eines analytischen Rahmens zur Untersuchung der Prüfungspraxis von Hochschullehrenden	5
2.1.1 Kompetenzen als Gegenstand von Hochschulprüfungen	6
2.1.1.1 Der Kompetenzbegriff	7
2.1.1.2 Theoretische Modellierung von Kompetenzen	9
2.1.1.3 Operationalisierung von Kompetenzen durch Aufgaben.....	12
2.1.1.4 Der Kompetenzbegriff im Qualifikationsrahmen für deutsche Hochschulabschlüsse	15
2.1.1.5 Indikatoren für eine Kompetenzorientierung in Prüfungen	18
2.1.2 Gütekriterien der pädagogischen Diagnostik im Kontext der Prüfungspraxis an Hochschulen	27
2.1.2.1 Validität	27
2.1.2.2 Reliabilität.....	32
2.1.2.3 Objektivität	36
2.1.3 Vorgehensmodell zur Erstellung und Auswertung von Prüfungen.....	38
2.1.3.1 Festlegung der Prüfungsgegenstandes	40
2.1.3.2 Erstellung von Prüfungsaufgaben	47
2.1.3.3 Rückbindung der Aufgaben an den Prüfungsgegenstand	48
2.1.3.4 Bewertung von Prüfungsantworten.....	49
2.1.3.5 Punkte- und Notenvergabe.....	52
2.1.3.6 Rückschlüsse aus den Prüfungsdaten.....	53
2.1.4 Kurzzusammenfassung.....	57
2.2 Erkenntnisse zur Prüfungspraxis von Hochschullehrenden	59
2.2.1 Forschungsergebnisse zur Prüfungspraxis an deutschen Hochschulen.....	59
2.2.2 Erkenntnisse zur Prüfungspraxis an der Technischen Universität München	65
2.2.3 Kurzzusammenfassung.....	68
2.3 Theoretische Grundlagen zur Entwicklung eines Qualitätsentwicklungsprogramms.....	70
2.3.1 Prinzipien effektiver Professionalisierung	70
2.3.2 Ziele des Qualitätsentwicklungsprogramm	72
2.3.3 Ablauf der Fortbildung.....	73
2.3.4 Ablauf der Arbeitstreffen	74
2.3.5 Kurzzusammenfassung.....	74

3	Forschungsfragen	76
4	Methoden	79
4.1	Untersuchungsdesign	79
4.2	Stichprobe und Fallbeschreibungen	80
4.2.1	Akquirierung von Teilnehmerinnen und Teilnehmern	80
4.2.2	Bildung der Fälle und Zusammensetzung der Teams	81
4.3	Datenquellen	84
4.3.1	Materialbasierte Daten	84
4.3.2	Verbaldaten	86
4.4	Datenauswertung	87
4.4.1	Auswertung der materialbasierten Daten	88
4.4.1.1	Deduktive Entwicklung der Kategoriensysteme	89
4.4.1.2	Ablauf der Kodierungen	91
4.4.1.3	Erzielte Übereinstimmungswerte	94
4.4.2	Auswertung der Verbaldaten	96
4.4.2.1	Vorstrukturierung der Textpassagen	97
4.4.2.2	Induktive Bildung von Kategorien	99
5	Ergebnisse	103
5.1	Festlegung des Prüfungsgegenstandes	103
5.1.1	Zielsetzungen der Hochschullehrenden	104
5.1.2	Erkenntnisse zur bestehenden Prüfungspraxis	104
5.1.2.1	Konzeptionelle Analyse des Prüfungsgegenstands	105
5.1.2.2	Zerlegung von Kompetenzen in Teilkompetenzen und Festlegung von Bearbeitungsniveaus	113
5.1.2.3	Formulierung von Lehrzielen	113
5.1.2.4	Verwendung einer Lehrzielmatrix	117
5.1.3	Erkenntnisse zur veränderten Prüfungspraxis	117
5.1.3.1	Anzahl der Lehrziele in der veränderten Praxis	118
5.1.3.2	Spezifität der neu erstellten Lehrziele	120
5.1.3.3	Wissensarten in den neu erstellten Lehrzielen	122
5.1.3.4	Verwendung einer Lehrzielmatrix	124
5.1.4	Kurzzusammenfassung	127
5.2	Erstellung von Prüfungsaufgaben	129
5.2.1	Zielsetzungen der Hochschullehrenden	129
5.2.2	Erkenntnisse zur bestehenden Prüfungspraxis	130
5.2.2.1	Formale Qualitätsindikatoren von Prüfungsaufgaben	131

5.2.2.2	Passung zwischen Lehrzielen und Prüfungsaufgaben	132
5.2.2.3	Differenzierung von Aufgabentypen	134
5.2.2.4	Merkmale einer Kompetenzorientierung in Aufgaben	138
5.2.3	Erkenntnisse zur veränderten Prüfungspraxis	143
5.2.3.1	Ausgangspunkt für die Erstellung von Aufgaben	143
5.2.3.2	Formale Gestaltungsmerkmale von Aufgaben.....	146
5.2.3.3	Passung zwischen Lehrzielen und Aufgaben.....	146
5.2.3.4	Differenzierung von Aufgabentypen	148
5.2.3.5	Merkmale einer Kompetenzorientierung in den veränderten Aufgaben.....	149
5.2.4	Kurzzusammenfassung.....	154
5.3	Rückbindung der Aufgaben an den Prüfungsgegenstand	157
5.3.1	Zielsetzungen der Hochschullehrenden.....	157
5.3.2	Erkenntnisse zur bestehenden Prüfungspraxis	157
5.3.3	Erkenntnisse zur veränderten Prüfungspraxis	159
5.3.4	Kurzzusammenfassung.....	163
5.4	Bewertung von Prüfungsantworten.....	165
5.4.1	Zielsetzungen der Hochschullehrenden.....	165
5.4.2	Erkenntnisse zur bestehenden Prüfungspraxis	166
5.4.3	Erkenntnisse zur veränderten Prüfungspraxis	168
5.4.4	Kurzzusammenfassung.....	169
5.5	Punkte- und Notenvergabe	171
5.5.1	Vergabe von Punkten	171
5.5.2	Notenvergabe	173
5.5.3	Kurzzusammenfassung.....	175
5.6	Rückschlüsse aus den Prüfungsdaten.....	177
5.6.1	Zielsetzungen der Hochschullehrenden.....	177
5.6.2	Erkenntnisse zur bestehenden Prüfungspraxis	177
5.6.2.1	Aufbereitung von Datensätzen.....	178
5.6.2.2	Rückschlüsse aus den Prüfungsdaten.....	178
5.6.3	Erkenntnisse zur veränderten Prüfungspraxis	180
5.6.4	Kurzzusammenfassung.....	182
6	Diskussion	185
6.1	Diskussion zentraler Befunde	188
6.1.1	Diskussion der Befunde zu den Zielsetzungen der Teilnehmenden für das Qualitätsentwicklungsprogramm	188
6.1.2	Diskussion der Befunde zur bestehenden Prüfungspraxis	190
6.1.3	Diskussion der Befunde zur veränderten Praxis	198

6.2	Methodische Reflexion und Limitationen.....	204
6.3	Implikationen	208
6.3.1	Praktische Implikationen.....	208
6.3.2	Implikationen für weitere Forschungsvorhaben.....	210
7	Schlusswort.....	213
	Literatur	214
	Abbildungsverzeichnis	221
	Tabellenverzeichnis.....	223
	Anhang	225
A.	Kategoriensysteme	226
a.	Kategoriensystem zur Analyse der Lehrziele	226
b.	Kategoriensystem zur Analyse der Prüfung im Gesamten	233
c.	Kategoriensystem zur Analyse der Einzelaufgaben.....	235
d.	Kategoriensystem zur Strukturierung der Verbaldaten.....	259
B.	Verwendete Zitate zur Kategorienbildung	264
C.	Zitate zur Festlegung des Prüfungsgegenstands bestehende Praxis.....	264
D.	Zitate zur Erstellung von Prüfungsaufgaben.....	281
E.	Zitate zur Rückbindung von Aufgaben an den Prüfungsgegenstand	282
F.	Zitate zur Bewertung von Prüfungsantworten	283
G.	Zitate zur Punkte- und Notenvergabe	287
H.	Zitate zu Rückschlüssen aus der Empirie.....	292

1 Einleitung

Prüfungen nehmen eine zentrale Rolle im Lehr-Lernalltag an Hochschulen ein. Hochschulen können auf Grundlage von Prüfungsergebnissen offen legen, inwieweit es mit dem bestehenden Studienangebot gelingt, Studierende zu qualifizieren, sodass diese die für einen Studiengang festgelegten Lehrziele erreichen, um damit der Forderung nach Rechenschaftslegung über den Studienerfolg nachzukommen (Wissenschaftsrat, 2012). Die auf Grundlage von Prüfungsergebnissen vergebenen Noten dienen zukünftigen Abnehmern, den Studierenden, sowie den Hochschulen als Indikator für die Leistungsfähigkeit und das Beherrschen von Fachinhalten. Dabei ist es die Aufgabe der Hochschulen, die Qualität von Prüfungen zu sichern. Werden Prüfungen eingesetzt, mit denen die ausgewiesenen Fähigkeiten und Kompetenzen nicht zuverlässig erfasst werden, verlieren die Einschätzungen der Hochschulen zu den Fähigkeiten und Potenzialen ihrer Studierenden an Kreditabilität (Braun, Weiss, & Seidel, 2014).

Auf Ebene von Studiengängen liefern Prüfungsergebnisse wichtige Steuerungsinformationen und können zur Qualitätssicherung von Studiengängen verwendet werden. Sind die Module eines Studiengangs kumulativ angelegt, insofern die Inhalte und Kompetenzen der einzelnen Module aufeinander aufbauen und sukzessive zum Erreichen der Studiengangziele hinführen, kann anhand von Prüfungsergebnissen nachvollzogen werden, inwieweit die Hierarchisierung und Sequenzierung von Studienhalten lernwirksam gelungen ist. Inhalte und Anforderungen von Lehrveranstaltungen können effektiv aufeinander abgestimmt werden, sofern die vorangegangenen Prüfungen belastbare Informationen zum Leistungsstand von Studierenden bereitstellen. Die Prüfungsanforderungen der jeweils nachgelagerten Module, die zu einem späteren Zeitpunkt im Studienverlauf vorgesehen sind, können betrachtet werden, um gezielt Grundlagen für weiterführende Veranstaltungen zu legen. Außerdem können lerndiagnostische Informationen darüber gewonnen werden, ob gewisse Teilkompetenzen für Studierende bzw. bestimmte Studierendengruppen schwieriger zu erlangen sind. Ausgehend davon können Probleme identifiziert und Unterstützungsmaßnahmen zur Erhöhung des Studienerfolgs bedarfsorientiert konzipiert und angeboten werden. Prüfungen liefern damit formative Informationen auf Ebene der Institution (Shavelson, 2010a) und ermöglichen es, datengestützte Entscheidungen über die Gestaltung von Studiengängen oder die Konzeption von Maßnahmen für Studierende und Lehrende zu treffen (Secolsky & Denison, 2012).

Studierende verwenden im Rahmen eines Studiums üblicherweise viel Zeit, um sich auf Prüfungen vorzubereiten. Dabei setzen sie sich intensiv mit den antizipierten Prüfungsinhalten und Anforderungen auseinander, vermutlich in stärkerem Ausmaß als durch den Besuch von Lehrveranstaltungen. Daher sollte es ein Anliegen von Lehrenden sein, dieses Engagement möglichst optimal auszunutzen, indem in den Prüfungen die Kernanforderungen der jeweiligen Veranstaltungen

abgebildet sind (Schulz, Zehner, Schindler, & Prenzel, 2014). Außerdem haben Prüfungen für die Studierenden eine herausragende Bedeutung, da anhand von Prüfungsergebnissen über den Studienfortschritt entschieden wird. Für die Studierenden stellen Prüfungen somit Hürden dar, die es zu überwinden gilt, um das Studium erfolgreich zu absolvieren. Die Prüfungsinhalte signalisieren den Studierenden, welche Inhalte wichtig und welche weniger wichtig sind. Studierende erhalten eine Rückmeldung zu ihrem Leistungsstand und erfahren mehr darüber, in welchen Teilbereichen ihr Lernen erfolgreich war und an welchen Stellen sie noch Schwächen haben. Die Prüfungsergebnisse werden herangezogen, um sich mit anderen Studierenden zu vergleichen und das eigene Lernen und die gewählten Lernstrategien zu bewerten und gegebenenfalls anzupassen. Die Prüfungsanforderungen steuern damit das Lernverhalten von Studierenden und beeinflussen so die Qualität und nachhaltige Verfügbarkeit der im Studium erworbenen Qualifikationen und Kompetenzen (Cilliers et al., 2010; Gijbels, Segers, & Struyf, 2008; Segers et al., 2006).

Die Lehrenden als Prüfungserstellende erhalten mit den Prüfungsergebnissen wiederum Informationen darüber, inwieweit die Studierenden die Lehrziele einer Veranstaltung erreicht haben. Das Abschneiden in Teilbereichen kann Aufschluss darüber geben, ob bestimmte Inhalte besser oder schlechter beherrscht werden. Aus diesen Informationen können Rückschlüsse gezogen werden, ob die entsprechenden Inhalte in der Lehrveranstaltung adäquat vermittelt wurden. Lehrende können die Lehrveranstaltungsinhalte und Lehr-Lernaktivitäten an den Prüfungsanforderungen ausrichten und das Lernen von Studierenden mit geeigneten Übungsaufgaben unterstützen (Shaeiwitz, 1996). Unterschiede zwischen Studierendengruppen liefern gegebenenfalls Hinweise, ob das Anforderungsniveau und die Eingangsvoraussetzungen zielgruppengerecht festgelegt wurden (Klauer & Leutner, 2007). Die gewonnenen Daten ermöglichen die Qualität von Prüfungsaufgaben einzuschätzen (Brown & Wilson, 2011).

Anhand der obigen Ausführungen wird deutlich, dass Prüfungen ein wichtiger Hebel sind, um die Lehr- und Lernqualität zu sichern. Das Prüfen an deutschen Hochschulen liegt in der Verantwortung der dort Lehrenden. Soll die Qualität von Prüfungen entwickelt werden, muss daher an der Prüfungspraxis von Lehrenden angesetzt werden, die eine zentrale Rolle bei der Gestaltung und Durchführung von Prüfungen einnehmen. Vor diesem Hintergrund erscheint es vielversprechend zu sein, eine Professionalisierung der Prüfungspraxis von Hochschullehrenden anzustreben. Um mit Maßnahmen zur Professionalisierung eine nachhaltige Wirkung im Handeln von Lehrenden zu erzielen, müssen entsprechende Programme die Probleme in der täglichen Praxis aufgreifen und kohärent auf dem Wissens- und Erfahrungsstand von Lehrenden aufbauen (Desimone, 2009). Trotz der zentralen Rolle, die Lehrende bezüglich des Prüfens an Hochschulen einnehmen, existieren kaum empirische Befunde über die Qualität von Prüfungen, und darüber, wie Lehrende bei der Erstellung, Durchführung und Auswertung von Prüfungen vorgehen und wie sie dabei unter-

stützt werden können, die Qualität von Prüfungen zu entwickeln. Damit fehlt auch eine Ausgangsbasis, um Programme zur Professionalisierung der Prüfungspraxis von Lehrenden effektiv anzulegen.

Das Anliegen der Dissertation ist es daher, die Prüfungspraxis von Hochschullehrenden zu untersuchen, mit einem Fokus darauf, inwieweit Maßnahmen getroffen und berücksichtigt werden, die zu der Einhaltung zentraler Gütekriterien von Prüfungen beitragen und welche Prüfungsqualität derzeit von den Lehrenden letztlich erreicht wird. Ein weiteres Anliegen besteht darin abzuklären, welche Herausforderungen durch die im Rahmen der Bologna Reform geforderte Kompetenzorientierung von Studium und Lehre auf Ebene des Prüfens für die Hochschullehrenden entstehen und inwieweit es den Lehrenden gelingt, die entsprechenden Anforderungen umzusetzen. Schließlich soll überprüft werden, inwieweit Lehrende im Rahmen eines Qualitätsentwicklungsprogramms dazu angeregt und angeleitet werden können, bestehende Problemlagen in ihren Prüfungen zu identifizieren und aufzugreifen, um entsprechende Lösungs- und Entwicklungsmöglichkeiten zu erarbeiten und gemeinsame Qualitätsstandards für das Prüfen zu entwickeln.

Bei der Auseinandersetzung mit der Prüfungspraxis an Hochschulen müssen die Besonderheiten akademischer Lehre berücksichtigt werden, die geprägt ist durch die Vermittlung spezialisierter Inhalte auf einem hohen Niveau (Braun et al., 2014). Ein weiteres Spezifikum stellt die Freiheit der Lehre dar, die den Lehrenden viele Gestaltungsmöglichkeiten eröffnet (ebd.). Außerdem ist zu beachten, dass Lehrende an Hochschulen mit einer Doppelbelastung von Forschung und Lehre umgehen müssen (Johannes & Seidel, 2012). Vor dem Hintergrund dieser hohen Fachlichkeit sowie der Autonomie in der Gestaltung von Lehre und Prüfungen, schien es zielführend zu sein, die bestehende Prüfungspraxis an einer kleinen Stichprobe, im Sinne einer fallbasierten Aufklärung an verschiedenen Lehrstühlen aus der Fakultät für Maschinenwesen und der Fakultät für Mathematik der Technischen Universität München, zu betrachten, und dann im Rahmen einer kooperativen Qualitätsentwicklung, deren Kern Arbeitstreffen zwischen Lehrenden und Prüfungsexperten bilden, zielorientiert und systematisch zu verbessern. Betrachtet werden sollen schriftliche Prüfungen mit hohen Studierendenzahlen, da dieses Prüfungsformat an Hochschulen die größte Reichweite aufweist. In die Analysen werden Lehrziele, Prüfungen und Prüfungsaufgaben (materialbasierte Daten) sowie Verbaldaten aus den Arbeitstreffen mit den Lehrenden einbezogen. Mit der Analyse der Prüfungsaufgaben werden die Qualität von Prüfungen sowie die in Prüfungen angelegten Anforderungen untersucht. Anhand der Verbaldaten soll beschrieben werden, wie Lehrende bei der Prüfungserstellung und -auswertung vorgehen. Mit der Betrachtung dieser prozessbezogenen Daten können Gründe und Ursachen für identifizierte Probleme in den Prüfungsaufgaben herausgearbeitet werden, die wiederum als Ansatzpunkte für Verbesserungen der Prüfungspraxis im Rahmen des geplanten Qualitätsentwicklungsprogramms dienen. Die empirischen Erhe-

bungen finden im Rahmen des Qualitätsentwicklungsprogramms statt. Dadurch entsteht für die Lehrenden ein praktischer Nutzen, da im Rahmen des Programms an den Materialien und Problemen der Lehrenden gearbeitet wird. Aus wissenschaftlicher Perspektive bietet dieses Vorgehen den Vorteil, dass Veränderungen der Prüfungspraxis und deren Bedingungen bzw. Voraussetzungen in ihrer Komplexität prozessnah und damit authentisch betrachtet und analysiert werden können.

Um die skizzierten Fragestellungen zur Prüfungspraxis von Hochschullehrenden und deren Entwicklung durch eine Qualitätsentwicklungsmaßnahme anhand des gewählten Designs beantworten zu können, wird ein analytischer Rahmen benötigt, der im ersten Teil des Theoriekapitels erarbeitet wird (Abschnitt 2.1). Im zweiten Teil des Theoriekapitels (Abschnitt 2.2) erfolgt eine Betrachtung des nationalen Forschungsstands zur Prüfungspraxis von Hochschullehrenden, um systematisch Anknüpfungspunkte zum bestehenden Erkenntnisstand zu identifizieren. Für die Konzeption und Durchführung des geplanten Qualitätsentwicklungsprogramms müssen unter Rückgriff auf Erkenntnisse zu Prinzipien effektiver Professionalisierung entsprechende Gestaltungsaspekte erarbeitet werden (Abschnitt 2.3). Basierend auf den Ausführungen im Theorieteil wird das Erkenntnisinteresse der Arbeit in Forschungsfragen konkretisiert (Kapitel 3). Im Methodenteil (Kapitel 4) werden das Untersuchungsdesign, die Stichprobe, die Datengrundlage sowie das methodische Vorgehen bei der Datenanalyse beschrieben. Anschließend folgt die Darstellung der Ergebnisse (Kapitel 5), um die Forschungsfragen zu beantworten. In der Diskussion (Kapitel 6) werden die Ergebnisse in Bezug auf die Fragestellungen interpretiert. Die Arbeit endet mit Schlussfolgerungen (Kapitel 1), in denen die Bedeutung der Ergebnisse für die Gestaltung von Maßnahmen zur Entwicklung der Prüfungspraxis an Hochschulen aufgezeigt wird.

2 Theoretischer Hintergrund

Der Theorieteil gliedert sich in drei Abschnitte. Im ersten Schritt wird ein analytischer Rahmen erarbeitet, der es ermöglicht, die Prüfungspraxis von Hochschullehrenden zu charakterisieren und die Qualität von Prüfungen einzuschätzen und zu bewerten (Abschnitt 2.1). Am aktuellen Wissensstand zu Prinzipien und Erkenntnissen der pädagogischen Diagnostik sollen dazu Kriterien erarbeitet werden, die von den Lehrenden bei der Erstellung und Auswertung von Prüfungen berücksichtigt werden müssen, um eine hohe Prüfungsqualität zu erreichen. Zwei Aspekte können dabei als wesentlich erachtet werden, die als Bezugspunkte für die Erstellung des analytischen Rahmens herangezogen werden sollen: (1) der aktuelle Trend zur Entwicklung und Überprüfung von Kompetenzen, dessen Umsetzung zudem mit den Bologna Vorgaben explizit von den Hochschulen eingefordert wird, bei gleichzeitiger (2) Einhaltung zentraler Gütekriterien der pädagogischen Diagnostik.

Im zweiten Teil (Abschnitt 2.2) folgt eine Betrachtung des nationalen Forschungsstands zur Prüfungspraxis von Hochschullehrenden. Der analytische Rahmen bildet dafür eine Ausgangsbasis. Auf Grundlage von existierenden Befunden zur Prüfungspraxis an deutschen Hochschulen soll zum einen das Erkenntnisinteresse der Arbeit konkretisiert werden. Zum anderen wurden in vorhandenen Studien möglicherweise bereits Problemlagen in Prüfungen identifiziert, die bei der Konzeption und inhaltlichen Ausgestaltung des geplanten Qualitätsentwicklungsprogramms aufgegriffen werden können.

Im dritten Teil des Theoriekapitels (Abschnitt 2.3) werden Gestaltungsaspekte und Kriterien für die Konzeption und Durchführung des angesprochenen Qualitätsentwicklungsprogramms erarbeitet. Das Programm soll nach erprobten Prinzipien effektiver Professionalisierung angelegt werden, um die Wahrscheinlichkeit zu erhöhen, dass mit der Maßnahme positive Resultate erzielt werden können. Dazu wird der aktuelle Forschungsstand zur Professionalisierung von Lehrpersonen betrachtet.

2.1 Entwicklung eines analytischen Rahmens zur Untersuchung der Prüfungspraxis von Hochschullehrenden

Aktuell wird die Vermittlung und Überprüfung von fachlichen und überfachlichen Kompetenzen gefordert und als Kriterium für die Bewertung der Qualität von Studiengängen herangezogen (Kultusministerkonferenz, 2011). Der Kompetenzerwerb bei Studierenden muss an den Hochschulen mit den dort eingesetzten Prüfungsformen zuverlässig erfasst werden. Bislang existieren nur wenige gesicherte Erkenntnisse darüber, wie Merkmale der Prüfungs- und Aufgabengestal-

tung die Studierenden bei der Bearbeitung von Prüfungen beeinflussen (Bücker et al., 2015). In diesem Kapitel wird daher ein analytischer Rahmen entwickelt, anhand dessen die Prüfungspraxis hinsichtlich der geforderten Kompetenzorientierung charakterisiert und die Qualität von Prüfungen eingeschätzt und bewertet werden kann. Dazu wird der aktuelle Forschungsstand zur Kompetenzmessung (Abschnitt 2.1.1) und zur pädagogischen Diagnostik (Abschnitt 2.1.2) betrachtet, um Kriterien zu erarbeiten, die zu einer hohen Prüfungsqualität beitragen. Diese Kriterien werden in handlungsnahen Indikatoren übersetzt und in ein Modell integriert, das den Prozess zur Erstellung und Auswertung einer Prüfung in verschiedene Schritte strukturiert (Abschnitt 2.1.3). Dieses Modell bildet den analytischen Rahmen der Arbeit. Die Unterteilung des Prüfungsprozesses in ein Vorgehensmodell mit mehreren Schritten soll eine möglichst handlungs- und prozessnahe Untersuchung der Prüfungspraxis von Hochschullehrenden ermöglichen. Entlang der einzelnen Schritte des Modells sollen Herausforderungen und Probleme in der Prüfungspraxis von Hochschullehrenden identifiziert werden. Gleichzeitig dient das Modell als Grundlage für das angesprochene Qualitätsentwicklungsprogramm. Dadurch können die identifizierten Probleme direkt im Rahmen des Programms aufgegriffen und gelöst werden.

2.1.1 Kompetenzen als Gegenstand von Hochschulprüfungen

Wie einleitend bereits erläutert, sollen mit Hochschulprüfungen fachliche sowie überfachliche Kompetenzen erfasst werden. Bislang besteht jedoch kein einheitliches Verständnis über den Kompetenzbegriff (Schaper, 2012). Entsprechend wenig ist darüber bekannt, ob die Kompetenzorientierung in der Prüfungspraxis von Hochschullehrenden umgesetzt wird. Ein wesentliches Ziel dieser Arbeit wird daher sein, basierend auf dem aktuellen Forschungsstand zur Erfassung von Kompetenzen, Indikatoren zu erarbeiten, anhand derer der Grad der Kompetenzorientierung in der Prüfungspraxis von Hochschullehrenden eingeschätzt werden kann.

In den letzten 15 Jahren konnten erhebliche Fortschritte in der Modellierung und Erfassung von Kompetenzen erzielt werden. Um das dabei entstandene Erfahrungswissen in dieser Arbeit nutzen zu können, werden zunächst zwei bedeutsame Ansätze zur Modellierung und Erfassung von Kompetenzen betrachtet, um die Bedeutung des Konstrukts Kompetenz begrifflich zu klären (Abschnitt 2.1.1.1). Anschließend wird dargelegt, welche Schritte durchgeführt werden müssen, um Kompetenzen mit Prüfungen erfassbar zu machen und welche Testformen dafür in Frage kommen (Abschnitte 2.1.1.2 und 2.1.1.3). Das dabei entwickelte Kompetenzrahmenmodell wird dahingehend diskutiert, ob es sich für eine Anwendung auf den Hochschulbereich eignet, indem betrachtet wird, ob die im Qualifikationsrahmen für deutsche Hochschulabschlüsse (HQR) beschriebenen Qualifikationsstufen abgebildet werden können (Abschnitt 2.1.1.4). Der HQR erscheint dafür als geeigneter Bezugspunkt, da er für die Hochschulen als verpflichtender Referenzrahmen bei der

Festlegung von fachlichen und überfachlichen Kompetenzen von Studiengängen gilt (Kultusministerkonferenz, 2005, 2011). Auf Grundlage der begrifflichen Klärung von Kompetenz und der Unterscheidung von möglichen Testvarianten zur Erfassung von Kompetenzen werden Indikatoren erarbeitet, anhand derer die Umsetzung einer Kompetenzorientierung in Hochschulprüfungen eingeschätzt werden kann (Abschnitt 2.1.1.5).

2.1.1.1 Der Kompetenzbegriff

Bei der Definition und Modellierung von Kompetenzen sowie in der Entwicklung von geeigneten Messinstrumenten haben sich in den letzten Jahrzehnten zwei bedeutende Perspektiven und Forschungsstränge herausgebildet - ein *holistischer Ansatz* und ein *analytischer Ansatz* zur Erfassung von Kompetenzen - denen jedoch ein unterschiedliches Begriffsverständnis von Kompetenz zugrunde liegt (Blömeke, Gustafsson, & Shavelson, 2015).

Im Rahmen des holistischen Forschungsstrangs, wie er in der Arbeits- und Organisationspsychologie weit verbreitet ist, wird Kompetenz als erfolgreiches Handeln einer Person in realen Situationen verstanden (Blömeke et al., 2015). Beim Bewältigen von Anforderungssituationen greifen Personen auf ihre Potenziale zurück, die als ein komplexes Konglomerat von Fertigkeiten und Fähigkeiten, im Sinne kognitiver, motivationaler und emotionaler Voraussetzungen und Einstellungen aufgefasst werden (Shavelson, 2010b). Die Art von Personenmerkmalen und deren Zusammenspiel während der Bewältigung von Handlungssituationen, ist bei dieser Form der Kompetenzerfassung nicht von Interesse. Diese Konzeption von Kompetenz wird von Blömeke et al. (2015) als holistischer Ansatz bezeichnet, da das gezeigte Verhalten von Personen in realen Anforderungssituationen in seiner Gesamtheit betrachtet wird, ohne die dahinterliegenden Fähigkeits- und Fertigkeitsbündel analytisch zu trennen (Blömeke et al., 2015).

Beim zweiten Ansatz, wie er verstärkt in der empirischen Bildungsforschung verfolgt wird, werden Kompetenzen als die bei Personen verfügbaren kognitiven, motivationalen und affektiven Ressourcen betrachtet, die analytisch in verschiedene Dimensionen zerlegt und gesondert voneinander erfasst werden (Blömeke et al., 2015). Nach dieser Konzeption beziehen sich Kompetenzen auf die Voraussetzungen, die für erfolgreiches Handeln in realen Anforderungssituationen bei Personen vorhanden sein müssen. Blömeke et al. (2015) bezeichnen diesen Ansatz als analytisch, da eine Kompetenz in zugrunde liegende latente Merkmale zerlegt wird. Beim analytischen Ansatz wird nicht das Verhalten von Personen in realen Situationen betrachtet, sondern ob eine Person die kognitiven, affektiven und motivationalen Fähigkeiten und Ressourcen (z.B. relevantes Wissen, Wissen über Problemlösestrategien) besitzt, die als Voraussetzung für erfolgreiches Handeln in realen Situationen betrachtet werden. Auf Grundlage des Abschneidens bei einem entsprechenden Test wird versucht, Verhalten in realen Situationen vorherzusagen (Blömeke et al.,

2015). Für den Bereich der Leistungsmessung (insbesondere im large-scale Kontext) wurde eine Fokussierung auf die kognitive Dimension von Kompetenz vorgeschlagen. Kompetenzen werden dabei bezeichnet als „*kontextspezifische kognitive Leistungsdispositionen*, die sich funktional auf Situationen und Anforderungen in bestimmten *Domänen* beziehen“ (Klieme & Leutner, 2006, Seite 879). Die beiden beschriebenen Ansätze wurden bislang vorwiegend nebeneinander betrieben. Da beide Ansätze mit ihren unterschiedlichen Konzeptionen den Kompetenzbegriff für sich beanspruchen, soll an dieser Stelle der Versuch einer begrifflichen Klärung vorgenommen werden.

Eine Gemeinsamkeit zwischen den beiden Ansätzen besteht darin, dass reale Anforderungssituationen aus einer (Lern-) Domäne als zentraler Referenzpunkt für die Modellierung von Kompetenzen herangezogen werden (Blömeke et al., 2015). Ebenfalls übereinstimmend wird die Handlungsfähigkeit in realen Situationen als zentrales Kriterium für eine Kompetenz angenommen (ebd.). Eine Kompetenz definiert sich demnach über Anforderungssituationen und deren erfolgreichen Bewältigung. Um Kompetenzen bei Personen zu erfassen, muss also eine direkte Observation von Verhalten in realen oder möglichst realitätsnahen Situationen (umsetzbar z.B. durch Simulationen) erfolgen, wie es beim holistischen Ansatz angestrebt wird. Bei den kognitiven Fähigkeiten und Ressourcen handelt es sich daher nicht um Kompetenzen sondern um einen Teilaspekt von Kompetenz. Mit dem analytischen Ansatz, wie er oben beschrieben wurde, werden demnach keine Kompetenzen erfasst, sondern inwieweit die Voraussetzungen für erfolgreiches Handeln in realen Situationen bei einer Person ausgeprägt sind. Auf der kognitiven Dimension werden diese Voraussetzungen verstärkt in Form von kontextbezogenem Wissen modelliert und getestet (Blömeke et al., 2015). Auf Grundlage dieser Ausführungen soll der Begriff Kompetenz in dieser Arbeit definiert werden als *die Fähigkeit einer Person, in realen Situationen erfolgreich zu handeln*.

Nachfolgend wird auf Verwendungsmöglichkeiten beider Ansätze eingegangen, um herauszuarbeiten, wie sie sich bei einer Anwendung auf den Hochschulbereich gegenseitig ergänzen können. Mit dem holistischen Ansatz wird versucht, Verhalten in realen oder realitätsnahen Situationen zu observieren. Realitätsnahe Situationen werden verwendet, wenn eine Observation von Verhalten in realen Situationen aus testökonomischen Gründen nicht durchführbar ist oder die Komplexität realer Situation reduziert werden soll. Eine realitätsnahe Situation kann durch Simulationen echter Situationen hergestellt werden. Der holistische Ansatz kann beispielsweise bei der Auswahl von geeigneten Bewerbern für eine Arbeitsstelle oder der Zulassung von Personen zur Ausübung kritischer Tätigkeiten (Chirurgin, Pilotin, Feuerwehrmann) relevant sein. Dabei ist es nicht von Interesse, auf welche Potenziale und Ressourcen eine Person bei der Bewältigung der gestellten Situationen zurückgreift, sondern ob die Anforderungssituationen zufriedenstellend gelöst werden

(Blömeke et al., 2015). Bei Nichtbewältigung einer Testsituation kann jedoch nicht aufgeklärt werden, in welchen Bereichen Defizite bestehen, die ursächlich für das Versagen in der Testsituation sind. Mit dem analytischen Ansatz können diese Informationen gezielt gewonnen werden, da die kognitiven Potenziale und Ressourcen, die bei der Bewältigung von Anforderungssituationen zum Tragen kommen, getrennt voneinander erfasst werden (Klieme & Leutner, 2006). Dadurch kann analysiert werden, auf welchen Fähigkeiten und Fertigkeiten Kompetenzen beruhen, wie die verschiedenen Aspekte aufeinander aufbauen und sich gegenseitig bedingen und welche Entwicklungsverläufe Personen beim Erwerb von Kompetenzen durchlaufen. Basierend auf derartigen Erkenntnissen können Curricula lernwirksam und kumulativ angelegt werden. Außerdem können zielgerichtet Maßnahmen entwickelt werden, um die Voraussetzungen für kompetentes Handeln effektiv zu vermitteln (Blömeke et al., 2015).

Für den Bildungsbereich respektive die Gestaltung von Bildungsangeboten (z.B. Studiengänge) erscheint eine Kombination der beiden Ansätze als vielversprechend. Mit der Kompetenzmessung nach dem holistischen Ansatz können Kernkompetenzen zuverlässig erfasst werden, bevor diese zertifiziert werden. Mit dem analytischen Ansatz kann aufgeklärt werden, an welchen Stellen bei Lernenden Defizite bestehen. Lernende können dann systematisch an die Handlungsfähigkeit in realen Situationen herangeführt werden, indem zunächst modelliert wird, welche Voraussetzungen und Ressourcen für erfolgreiches Handeln benötigt werden, wie diese Wissens- und Könnensaspekte aufeinander aufbauen und miteinander interagieren und wie sie lernwirksam vermittelt werden können (Blömeke et al., 2015). Nach der begrifflichen Klärung des Kompetenzbegriffs wird in den nächsten beiden Abschnitten dargelegt, welche Schritte durchgeführt werden müssen, um Kompetenzen erfassbar zu machen.

2.1.1.2 Theoretische Modellierung von Kompetenzen

Kompetenzen definieren sich über Anforderungssituationen. Bevor Prüfungsaufgaben entwickelt werden können, muss festgelegt werden, was die Prüflinge können sollen oder anders ausgedrückt, welche Anforderungssituationen sie bewältigen können müssen, um die angestrebten Kompetenzen unter Beweis zu stellen. Außerdem muss geklärt werden, ob eine direkte Messung von Kompetenzen durch Observation von Verhalten in realen Situationen durchgeführt werden soll, die Observation von Verhalten in realitätsnahen Testsituationen erfolgt, oder ob die Voraussetzungen für kompetentes Handeln erfasst werden sollen. Bei allen drei Ansätzen sind mehrere Modellierungsschritte notwendig, um Kompetenzen erfassbar zu machen:

Im ersten Schritt werden sowohl Anforderungssituationen identifiziert, die für eine bestimmte Kompetenz relevant sind, als auch das Verhalten, das zu einer erfolgreichen Bewältigung der Situationen führt (Shavelson, 2012). Der oben eingeführten Definition von Kompetenz folgend,

müssen die Anforderungssituationen einem realen und für die Zielgruppe relevanten Kontext entstammen. Diese Vorgabe schränkt das Suchfeld zur Identifikation entsprechender Situationen ein. Bei der Identifikation von Anforderungssituationen und entsprechenden Verhaltensweisen sollen Erkenntnisse aus verschiedenen Informationsquellen integriert werden (Klauer & Leutner, 2007). Dazu gehören bereits existierende Studien, die sich mit Kompetenzniveaus und –strukturen sowie Bedingungen für einen Kompetenzzuwachs bei Lernenden in einem Lernbereich beschäftigen (Pellegrino, Chudowsky, & Glaser, 2001). Darüber hinaus können je nach Bereich weitere Informationen, beispielsweise durch empirische Analyse von Arbeitsplätzen, Analysen von Curricula oder der Orientierung an Kompetenzstandards, gewonnen werden (Bühner, 2011; Klauer, 1974).

In einem nächsten Schritt werden die identifizierten Anforderungssituationen in verschiedene Teilaspekte oder Teilkompetenzen unterteilt (Blömeke et al., 2015). Dabei muss definiert werden, wie viele oder welche Teilkompetenzen ausgeprägt sein müssen, um als kompetent im Hinblick auf eine übergeordnete Kompetenz zu gelten (ebd.). Die Teilkompetenzen können miteinander interagieren. Hier gilt es zu klären, ob eine Schwäche hinsichtlich einer Teilkompetenz mit einer besonderen Stärke bei einer anderen Teilkompetenz kompensiert werden kann und wo das Mindestniveau liegt, um noch als kompetent zu gelten (ebd.). Legt man zusätzlich eine Lern- oder Entwicklungsperspektive zugrunde, wird modelliert, welche Teilkompetenzen zuerst erlernt bzw. beherrscht werden müssen, bevor weiterführende Kompetenzbereiche vermittelt werden (typischerweise beginnen viele Studiengänge mit der Vermittlung von Grundlagenwissen, bevor anwendungsorientierte Aspekte aufgegriffen werden). Eine andere Entwicklungsperspektive könnte sein, dass Lern- oder Könnenszuwächse hinsichtlich aller Kompetenzbereiche gleichermaßen oder nur miteinander zustande kommen. Hier stellen sich Fragen dazu, ob Kompetenzzuwächse (oder –verluste) linear erfolgen, von allgemein nach speziell und umgekehrt, oder durch qualitative Sprünge, wie es im Experte-Novize Paradigma angenommen wird (z.B. Berliner, 2001)

Innerhalb von Kompetenzen bzw. Teilkompetenzen können verschiedene Kompetenzniveaus unterschieden werden (Wilson, 2005). Mit den Niveaus kann ausgedrückt werden, ob eine Person mehr oder weniger kompetent ist (Blömeke et al., 2015). Mögliche Anforderungsniveaus können sich darauf beziehen, ob die Anforderungssituationen mehr oder weniger gut, zufriedenstellend, genau, schnell, nachhaltig, mit angemessener Fachsprache, oder ähnlichem, bewältigt werden (ebd.). Ausgehend von diesen Überlegungen kann im nächsten Schritt bestimmt werden, welches Verhalten Studierende zeigen müssen, um die Anforderungssituationen zufriedenstellend zu lösen, beziehungsweise welcher Zielzustand in einer Situation hergestellt werden muss, damit sie als gelöst gilt. Mit der Durchführung der beschriebenen Schritte ist das latente Konstrukt „Kompetenz“ anhand von observierbaren Handlungen definiert und für eine Erfassung zugänglich gemacht.

Für eine direkte Messung von Kompetenzen durch die Observation von Verhalten in realen Situationen sind mit den beschriebenen Schritten die wesentliche Aspekte und Merkmale einer Kompetenz geklärt. Ein weiterer Modellierungsschritt wird nötig, wenn das Verhalten nicht in realen Situationen, sondern in realitätsnahen Testsituationen observiert wird. Je authentischer die realen Situationen in realitätsnahe Testsituationen übersetzt werden, desto belastbarer sind Prognosen, die auf Grundlage des Abschneidens in den simulierten Situationen hinsichtlich der tatsächlichen Handlungsfähigkeit in realen Situationen getroffen werden können. Um eine möglichst hohe Authentizität zu erreichen, ist es erforderlich, anknüpfend an die bereits beschriebenen Modellierungsschritte, die in den realen Situationen auftretenden Kontextvariablen systematisch zu identifizieren und zu beschreiben. Außerdem müssen Annahmen darüber getroffen werden, durch welche Stimuli und Repräsentationsformen in den Testsituationen genau die Reaktionen bei den Testpersonen hervorgerufen werden können (Blömeke et al., 2015), die auch in der korrespondierenden realen Situation auftreten und im Hinblick auf das Vorhandensein einer Kompetenz interpretiert werden können. Die Observation von Verhalten in realitätsnahen Situationen bietet den Vorteil, dass die Komplexität und Dynamik realer Situationen durch die Kontrolle von Kontextvariablen systematisch reduziert werden kann. Dadurch kann aufgeklärt werden, welche Merkmale in diesen Situationen schwierigkeitsgenerierend wirken.

Bei der Erfassung von Voraussetzungen für kompetentes Handeln müssen Kompetenzen in dahinterliegende latente Fähigkeitsfacetten zerlegt werden. Neben der Identifikation von Anforderungssituation und der Festlegung von Kompetenzdimensionen und –niveaus, müssen in einem weiteren Modellierungsschritt deshalb latente (kognitive) Fähigkeiten identifiziert werden, die als Voraussetzung für kompetentes Handeln betrachtet werden. Hinsichtlich der kognitiven Dimension von Kompetenzen wird angenommen, dass Kompetenzen auf kontext- oder domänenbezogenen Wissensbeständen beruhen, die bei der Bewältigung von Anforderungssituationen abgerufen und mittels bestimmter kognitiver Prozesse verarbeitet werden (Shavelson, 2010b). Bei der theoretischen Modellierung steht die Frage im Vordergrund, auf welche latenten Fähigkeiten (Wissensbestände und kognitive Prozesse) im Einzelnen bei der Bewältigung von Anforderungssituationen zurückgegriffen wird, wie sie miteinander interagieren, sich gegenseitig bedingen oder aufeinander aufbauen, und ob sich diese Interaktion während der Bearbeitung einer Aufgabe verändern (Snow & Swanson, 1992). Außerdem muss beachtet werden, dass sich diese Vorgänge zwischen Personen unterscheiden können (ebd.). Ausgehend von diesen Überlegungen müssen belastbare fachdidaktische Theorien zu Bestandteilen und der Struktur der entsprechenden kognitiven Voraussetzungen entwickelt werden, die bislang jedoch nur vereinzelt und auf eingeschränkte Bereiche bezogen existieren (Leuders, 2014). Hinzu kommt die Modellierung entsprechender Kontexte, auf die das Wissen angewendet werden soll. Die Kontexte lassen sich aus den zuvor bereits identi-

fizierten realen Anforderungssituationen erschließen. Mit der theoretischen Modellierung wird die Grundlage für die Erstellung von Prüfungsaufgaben gelegt (siehe Abbildung 1). In der vorliegenden Arbeit wird untersucht, inwieweit die Lehrenden die beschriebenen Schritte zur theoretischen Modellierung von Kompetenzen berücksichtigen. Dabei wird auf die Verbaldaten als Datengrundlage zurückgegriffen.

2.1.1.3 Operationalisierung von Kompetenzen durch Aufgaben

Im Anschluss an die theoretische Modellierung werden Test- beziehungsweise Prüfungsaufgaben entwickelt. In welcher Form die theoretisch beschriebenen Anforderungssituationen in Testsituationen übersetzt werden, bestimmt welche Art von Kompetenztest vorliegt. Hierzu wird folgende begriffliche Unterscheidung vorgeschlagen: Findet die Observation des Verhaltens in realen Situationen und unter realen Bedingungen statt, kann dies als *Kompetenztest* bezeichnet werden. Bei einer Observation von Verhalten in realitätsnahen Testsituationen, liegt ein *kompetenzorientierter Test* vor. In einigen Anwendungsbereichen werden über Simulationen vielversprechende Annäherungen an reale Situationen erreicht. Werden latente Fähigkeiten gemessen, die als Voraussetzung für kompetentes Handeln betrachtet werden, wird diese Art von Test als *kompetenzorientierter Wissenstest* bezeichnet. In den Testaufgaben werden Kontextbezüge hergestellt, auf die das entsprechende Wissen angewendet werden muss. Eine Unterscheidung zwischen einem Kompetenztest, einem kompetenzorientierten Test oder einem kompetenzorientierten Wissenstests ist vermutlich nicht immer eindeutig möglich und kann sich nach Anwendungsbereichen unterscheiden. Die hier getroffene Unterscheidung ist deshalb als heuristische Kategorisierungshilfe zur Einteilung verschiedener Ansätze und Instrumente zur Kompetenzmessung zu verstehen. In Abbildung 1 sind die drei Testansätze schematisch dargestellt. Aus der Abbildung gehen außerdem die der Aufgabenerstellung vorgelagerten Schritte bei der theoretischen Modellierung hervor, die in Kapitel 2.1.1.2 ausführlich beschrieben sind.

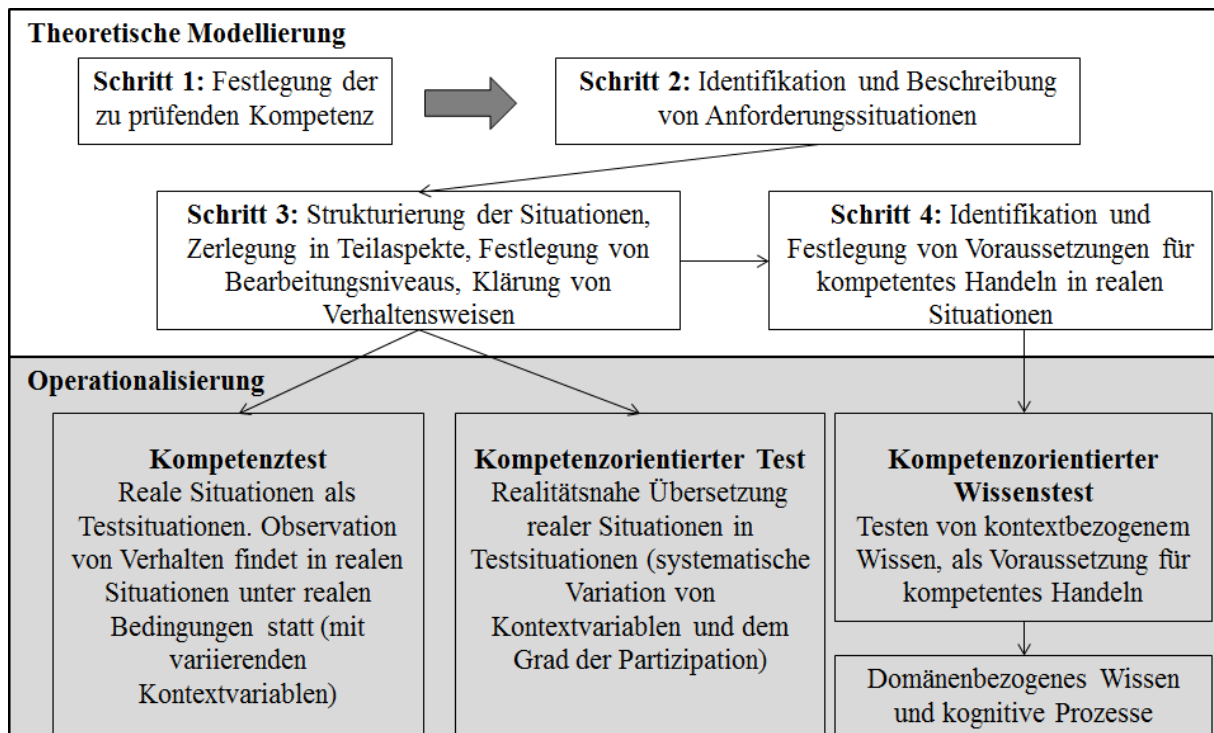


Abbildung 1: Heuristische Kategorisierungshilfe zur Einteilung verschiedener Ansätze und Instrumente zur Kompetenzmessung¹

Im Folgenden werden Besonderheiten bei der Erstellung entsprechender Aufgabentypen anhand eines Beispiels dargestellt. Bei einem Kompetenztest entsprechen die Testaufgaben den realen Anforderungssituationen bzw. der Test wird in realen Situationen, unter realen Bedingungen, durchgeführt. Möchte man beispielsweise überprüfen, ob eine angehende Lehrkraft kompetent unterrichten kann, muss, nachdem festgelegt wurde, was es bedeutet, kompetent zu unterrichten, die Observation des Testverhaltens in einer realen Klassenzimmersituation unter realen Bedingungen erfolgen. Nur unter realen Bedingungen, wenn also alle Anforderungsfacetten einer Unterrichtssituation zum Tragen kommen, kann valide schlussgefolgert werden, ob eine Lehrkraft in der Lage ist, kompetent zu unterrichten. Neben der Vermittlung der Unterrichtsinhalte muss eine Lehrkraft in realen Klassenzimmersituationen parallel auf weitere Einflüsse, wie Fragen von Schülerinnen und Schülern oder disziplinarische Aspekte, eingehen. In der zweiten Phase der Lehramtsausbildung (dem sogenannten Vorbereitungsdienst) wird ein Kompetenztest eingesetzt. Angehende Lehrkräfte müssen im Rahmen des zweiten Staatsexamens sogenannte Lehrproben absolvieren. Dabei wird der Unterricht einer Lehrkraft in einer realen Klassenzimmersituation bewertet (Strietholt & Terhart, 2009).

Die Fähigkeit in solchen realen Situationen angemessen zu handeln erfordert vielfältige Voraussetzungen auf Seiten der Lehrkräfte, die im Rahmen einer Ausbildung jedoch erst angebahnt wer-

¹ Abbildungen, unter denen keine Quelle ausgewiesen ist, wurden eigens vom Autor erstellt.

den müssen, um sich schrittweise an das Handeln in realen Situationen anzunähern (Grossman et al., 2009). Mit einem kompetenzorientierten Test werden reale Situationen in realitätsnahe Testaufgaben übersetzt. Dabei sind Abstufungen hinsichtlich der Realitätsnähe zu beachten, die auf einem Kontinuum angeordnet werden können. Testsituationen, in denen die Facetten einer realen Situation annähernd originalgetreu abgebildet sind, indem die betreffenden Kontextvariablen möglichst vollständig wirksam werden und die Partizipation der Testperson alle bzw. sehr viele Anforderungsfacetten umfasst, die auch in den realen Situationen zu beherrschen sind, weisen eine hohe Authentizität auf. Eine geringere Authentizität ist in Testsituationen gegeben, bei denen Kontextvariablen reduziert oder vereinfacht werden und sich die Partizipation der Testperson auf wenige - oder eine einzelne - Anforderungsfacette(n) beschränkt (Grossman et al., 2009). Simulationen stellen eine Möglichkeit dar, um reale Anforderungssituationen in realitätsnahe Testsituationen zu übersetzen, in denen mehr oder weniger viele Aspekte der realen Situationen modelliert werden können (z.B. Nickolaus, Behrendt, Dammann, Stefanica, & Heinze, 2013). Je stärker die Testsituationen von den echten Situationen abweichen, indem Kontextvariablen vernachlässigt oder vereinfacht werden, oder der Grad der Partizipation der Testperson verringert wird, desto weniger belastbar sind die Prognosen, die auf Grundlage der erzielten Testergebnisse das Verhalten einer Person in den betreffenden realen Situationen vorhersagen (ebd.).

Betrachtet man wieder das Beispiel einer Lehrkraft, so können in einem kompetenzorientierten Test die Komplexität von Unterrichtssituationen und der Grad der Partizipation der Testpersonen reduziert werden, indem Kontextvariablen ausgeblendet oder vereinfacht werden. Beispielsweise kann der zu vermittelnde Inhalt vereinfacht oder die Klassengröße reduziert werden. Schülerinnen und Schüler können durch Schauspieler ersetzt werden, um Komplexität, die aufgrund des Verhaltens von Schülerinnen und Schülern erzeugt wird, zu kontrollieren, oder um gezielt auf den Umgang mit bestimmten Schülertypen vorzubereiten. Die Schauspieler können beispielsweise verhaltensauffällige, unterforderte oder überforderte Schülerinnen und Schüler simulieren. Seidel, Stürmer, Schäfer und Jahn (2015) setzen dieses Konzept um, indem sie Lehr-Lernsettings schaffen, in denen Lehramtsstudierende einer kleinen Gruppe von Schülerinnen und Schülern, die von Schauspielern simuliert werden, einen einfachen Inhalt vermitteln. Dabei werden verschiedene Kompetenzfacetten hinsichtlich der Unterrichtsgestaltung bewertet, wie die Transparentmachung von Lehrzielen, einer Darlegung des Unterrichtsablaufs, die Art der eingesetzten Fragen und das Feedback, das an die Schülerinnen und Schüler gegeben wurde (ebd.). Dadurch können Lehramtsstudierende schrittweise an die Komplexität realer Situationen herangeführt werden.

Bei einem kompetenzorientierten Wissenstest weisen die Testaufgaben einen Kontextbezug auf, z.B. durch Fallbeispiele, und gehen damit über das isolierte Abfragen von Wissensseinheiten hinaus. Das dabei abgefragte Wissen und die kognitiven Prozesse, die zum Lösen der Aufgaben not-

wendig sind, werden als Voraussetzung für kompetentes Handeln in realen Situationen betrachtet. Hierbei geht es jedoch nicht mehr um die Observation von Verhalten in realen oder realitätsnahen Situationen, sondern um das Abfragen von latenten Fähigkeiten, die als Voraussetzung für eine Kompetenz betrachtet werden. Ein entsprechender Test, der sich auf das eingeführte Beispiel über die Kompetenz einer Lehrkraft zum Unterrichten bezieht, könnte Fallbeispiele beinhalten, die eine Unterrichtssituation beschreiben oder darstellen (z.B. in einem Unterrichtsvideo), um einen Kontextbezug herzustellen, auf den das Wissen über effektives Unterrichten bezogen werden muss. Eine Handlungsaufforderung in der Aufgabenstellung könnte lauten, Aspekte zu benennen, die laut der Forschung zu effektivem Unterrichten in den dargestellten Situationen vorkommen oder eben auch fehlen. Hier kann ebenfalls ein Beispiel aus der Lehramtsausbildung herangezogen werden: Der Observer, als Instrument zur Erfassung professioneller Unterrichtswahrnehmung, arbeitet mit Unterrichtsvideos, die als eine effektive Möglichkeit betrachtet werden, um Unterrichtssituationen authentisch darzustellen (Sherin, 2004). Lehramtsstudierende müssen die in den Videos dargestellten Unterrichtssituationen unter Rückgriff auf ihr Wissen über Komponenten effektiven Unterrichtens einschätzen. Damit wird festgestellt, ob Studierende in der Lage sind, Unterrichtssituationen professionell zu beobachten und zu interpretieren (Jahn, Stürmer, Seidel, & Prenzel, 2014). Der Kontextbezug wird durch die Unterrichtsvideos geschaffen. Die Studierenden müssen ihr Wissen über effektives Unterrichten direkt auf den Kontext Unterricht anwenden, anstatt die Komponenten beispielsweise nur zu nennen oder zu beschreiben.

2.1.1.4 Der Kompetenzbegriff im Qualifikationsrahmen für deutsche Hochschulabschlüsse

In diesem Abschnitt soll geklärt werden, ob sich der oben eingeführte Kompetenzbegriff, sowie das Kompetenzframework mit der Unterscheidung von drei Testvarianten, für die Anwendung auf den Hochschulbereich und damit für das Anliegen der vorliegenden Dissertation, nämlich die Prüfungspraxis hinsichtlich Merkmalen einer Kompetenzorientierung zu untersuchen, eignen. Dazu wird betrachtet, welches Kompetenzverständnis dem Qualifikationsrahmen für deutsche Hochschulabschlüsse (im Folgenden mit HQR abgekürzt) zugrunde liegt, der für die Hochschulen als verpflichtender Referenzrahmen bei der Festlegung von fachlichen und überfachlichen Kompetenzen von Studiengängen gilt. Anschließend wird dargelegt, wie die Zieldimensionen des HQR in das oben eingeführte Kompetenzframework eingeordnet werden können.

Im Rahmen der Bologna Reform wurden Hochschulen dazu verpflichtet, für ihre Studiengänge Qualifikationsziele beziehungsweise Lehrziele² festzulegen (Kultusministerkonferenz, 2010). Mit den Lehrzielen soll definiert werden, welche Kompetenzen Studierende mit Durchlaufen eines Studiengangs erlangen sollen (Terhart, Baumgart, Meder, & Sychowski, 2009). Diese Änderung wird häufig als eine Veränderung der Perspektive vom Input (was wurde gelehrt) zum Outcome (was wurde tatsächlich von den Studierenden gelernt) charakterisiert. Die Festlegung der fachspezifischen Kompetenzen liegt dabei bei den Hochschulen. Als Referenzrahmen für die Auswahl und Festlegung der Kompetenzen dient der „Qualifikationsrahmen für deutsche Hochschulabschlüsse“ (Kultusministerkonferenz, 2005). Bei der Akkreditierung von Studiengängen wird der HQR als Maßstab für die Bewertung der für Studiengänge festgelegten Qualifikationsziele herangezogen und sollte daher bei der Gestaltung von Studiengängen berücksichtigt werden. Daher sollen nachfolgend die inhaltlichen Dimensionen des HQR betrachtet werden, um zu erarbeiten, wie diese in das oben eingeführte Kompetenzmodell integriert werden können.

Im HQR wird unterschieden zwischen den Kategorien „Wissen und Verstehen“ und „Können“. Die Kategorien werden im HQR wie folgt näher bestimmt: Die Kategorie *Wissen und Verstehen* ist unterteilt in die Unterkategorien *Wissensverbreiterung* und *Wissensvertiefung*. Für Bachelor-Abschlüsse ist für die Wissensverbreiterung der Erwerb einer breiten und integrierten Wissensbasis und ein Verständnis der wissenschaftlichen Grundlagen des entsprechenden Lerngebiets anzustreben. Unter Wissensvertiefung werden Lehrziele verstanden, die ein kritisches Verständnis der wichtigsten Theorien, Methoden und Prinzipien und eine vertikale, horizontale und laterale Vertiefung des Wissens auf dem Stand der Fachliteratur anstreben. Hinzu kommen vertiefte Wissensbestände zu aktueller Forschung. Im Master-Studium soll diese Fachwissensbasis erweitert werden. Die Studierenden sollen nach einem Master-Studium außerdem in der Lage sein, eigenständige Ideen zu entwickeln und anzuwenden. Hinzu kommt ein breites, detailliertes und kritisches Verständnis in Spezialbereichen, basierend auf dem aktuellen Wissens- und Forschungsstand. Die Kategorie *Können* ist unterteilt in *Instrumentale Kompetenzen*, *Systemische Kompetenzen* und *Kommunikative Kompetenzen*. Mit den instrumentalen Kompetenzen sollen Bachelor-Studierende dazu befähigt werden, ihr Wissen und Verstehen auf ihre Tätigkeit oder ihren Beruf anzuwenden und Problemlösungen und Argumente in ihrem Fachgebiet zu erarbeiten und weiterzuentwickeln. Als Zielsetzung von Master-Studiengängen steht hier die Erweiterung dieses Problemlösevermö-

² Die KMK verwendet die Begriffe „Qualifikationsziele“, „Lernergebnisse“ und Lernziele ohne deren Bedeutungen zu unterscheiden. In dieser Arbeit wird der Begriff „Lehrziel“ verwendet, für Ziele, die von Lehrenden für Veranstaltungen, Module oder Studiengänge festgelegt werden (Klauer, 1974). Als Lernziel werden Ziele betrachtet, die sich die Studierenden als Ziel für ihr Lernen setzen (ebd.). Ein Lernergebnis ist das Resultat, dass von einem Studierenden durch Lernanstrengungen erzielt wurde. Lernergebnisse können mit einer Prüfung festgestellt werden.

gens auf unbekannte Situationen mit multidisziplinären Zusammenhängen im Vordergrund. Die systemischen Kompetenzen umfassen für Bachelor-Studierende die Fähigkeit, relevante Informationen zu sammeln, zu interpretieren, um daraus wissenschaftlich fundierte Urteile abzuleiten, sowie das selbstständige Weiterführen von Lernprozessen. Bei Master-Studierenden kommen der Umgang mit Komplexität, das Treffen von Entscheidungen auf unvollständiger Informationsbasis, das selbstständige Aneignen von Wissen und die weitgehend selbstständige und autonome Durchführung von Forschungs- und Anwendungsprojekten hinzu. Mit den kommunikativen Kompetenzen sollen Bachelor-Studierende fachbezogene Probleme und Positionen formulieren und argumentativ verteidigen können. Sie sollen in der Lage sein, sich mit Fachvertretern und Laien über ihr Fachgebiet auszutauschen. Außerdem sollen sie Verantwortung in einem Team übernehmen können. Bei Master-Studierenden sollen diese kommunikativen Fähigkeiten auf einem höheren Niveau ausgebildet werden (Kultusministerkonferenz, 2005).

Die Kategorien des HQR sind domänenunabhängig beschrieben und sollen von den Hochschulen fachspezifisch ausgestaltet werden. Die Dimension *Wissen und Verstehen* bildet die Grundlage für die Dimension des *Könnens*. Die Dimension Können kann dem oben eingeführten Kompetenzbegriff als entsprechend interpretiert werden, wonach Kompetenzen zu erfolgreichem Handeln in realen Situationen befähigen. In der Könnensdimension des HQR müssen demnach Kompetenztests (Observation von Verhalten in realen Situationen) oder kompetenzorientierte Tests (Observation von Verhalten in realitätsnahen Situationen) eingesetzt werden. Mit der Dimension *Wissen und Verstehen* werden die Voraussetzungen für kompetentes Handeln beschrieben. Als Testverfahren können kompetenzorientierte Wissenstests verwendet werden. Die Qualifikationsziele des HQR können demnach mit den drei Varianten zur Erfassung von Kompetenzen abgedeckt werden. Diese Einschätzung soll nachfolgend genauer erläutert werden, indem exemplarisch dargelegt wird, wie das eingeführte Kompetenzkonzept auf Studiengänge übertragen werden kann.

Auf Ebene von Studiengängen oder für bestimmte Module kann mittels Kompetenztests die Erreichung zentraler Kompetenzen besonders zuverlässig und mit hoher Gültigkeit festgestellt werden. Das Bestehen des zweiten Staatsexamens bei Lehrkräften und damit die Zulassung zum Lehrerberuf ist beispielsweise an drei Testsituationen (sogenannte Lehrproben) gekoppelt, in denen die angehende Lehrkraft eine Unterrichtsstunde in einer realen Klassenzimmersituation hält (Strietholt & Terhart, 2009). Über die Zulassung zu einem Kompetenztest kann mit kompetenzorientierten Tests entschieden werden, die beispielsweise auf Ebene von Modulen angesiedelt werden können. Aufgrund des zum Teil sehr hohen Entwicklungsaufwands für kompetenzorientierte Tests findet die Entwicklung solcher Testverfahren bislang vorwiegend im Rahmen von Forschungsprojekten statt (für den Bereich der Lehrerbildung siehe beispielsweise Blomberg,

Stürmer, & Seidel, 2009; Gartmeier et al., 2015; Seidel & Stürmer, 2014; Stürmer & Seidel, 2015). Der Fachbereich der Medizin bildet hier eine Ausnahme. Im Medizinstudium werden mit den sogenannten „Objective Structured Clinical Examinations“ schon seit Längerem situationsnahe Testverfahren eingesetzt, bei denen die Studierenden an verschiedenen Stationen praktisch relevante Handlungen (z.B. Anamnese oder körperliche Untersuchungen) an simulierten Patienten ausführen (Rushforth, 2007). Welche Ressourcen und externer Input nötig sind, damit Lehrende an Hochschulen kompetenzorientierte Tests umsetzen können, gilt es noch zu klären.

Auf Ebene von Lehrveranstaltungen können kompetenzorientierte Wissenstests eingesetzt werden, um zu überprüfen, ob Studierende die Voraussetzungen für kompetentes Handeln, im Sinne von kontextbezogenen Wissensseinheiten und kognitiven Fähigkeiten, besitzen. Ausgehend von den auf Studiengangs- oder Modulebene definierten Anforderungssituationen legen Lehrende fest, welches Wissen als Grundlage für die Bewältigung von Anforderungssituationen dient, und wie dieses Wissen durchdrungen und aufbereitet werden muss, um es kontextspezifisch und flexibel zur Lösung der für eine Kompetenz relevanten Anforderungssituationen einsetzen zu können (Klauer & Leutner, 2007). Aus diesen Festlegungen werden entsprechende Aufgaben abgeleitet, in denen dieses Wissen und dessen kontextbezogene Anwendung entsprechend abgeprüft wird. Bereits in schriftlichen Prüfungen, sowohl mit offenen als auch Wahlantwortformaten, ist eine Vielzahl derartiger Anforderungsfacetten umsetzbar (Büchter & Leuders, 2005; Reisse, 2008; Shavelson, 2010a; Shavelson & Ruiz-Primo, 1999a). Die Prognosefähigkeit auf VerhaltensPotenziale in realen Situationen ist bei einem kompetenzorientierten Wissenstest jedoch nicht oder nur eingeschränkt gegeben (Shavelson, 2010b). Die Auswahl einer Prüfungsvariante sollte daher stets zielbezogen erfolgen.

Wie die genannten Beispiele zeigen, können die drei Testvarianten auf Studiengänge angewendet werden. Das eingeführte Kompetenzframework erscheint daher als analytischer Rahmen geeignet, um die Umsetzung einer kompetenzbasierten beziehungsweise -orientierten Lehr- und Prüfungspraxis an Hochschulen einzuschätzen und soll in dieser Arbeit verwendet werden. Aus diesen Überlegungen wird außerdem deutlich, dass die Etablierung einer kompetenzbasierten Lehr- und Prüfungspraxis an Hochschulen inhaltliche Abstimmungen zwischen mehreren Ebenen sowie entsprechende Ressourcen und Kenntnisse zur Prüfungsgestaltung erfordert.

2.1.1.5 Indikatoren für eine Kompetenzorientierung in Prüfungen

Nachdem geklärt wurde, dass das eingeführte Kompetenzframework auf den Hochschulbereich angewendet werden kann, werden in diesem Abschnitt Indikatoren entwickelt, anhand derer eingeschätzt werden soll, ob in der Prüfungspraxis von Hochschullehrenden Merkmale einer Kompetenzorientierung umgesetzt werden. Aufgrund des hohen Aufwands zur Entwicklung und Durch-

führung von Kompetenztests und kompetenzorientierten Tests soll in dieser Arbeit der Fokus auf einer Analyse von kompetenzorientierten Wissenstests liegen. Für diese Fokussierung spricht zusätzlich, dass schriftliche Prüfungsformate an der Technischen Universität München, an der diese Arbeit durchgeführt wird, am weitesten verbreitet sind und eine entsprechend hohe Reichweite aufweisen (Prenzel, Schindler, & Schulz, 2012).

Die Umsetzung einer Kompetenzorientierung erfolgt sowohl auf Ebene der theoretischen Modellierung von Prüfungsanforderungen, als auch bei der Operationalisierung der theoretisch beschriebenen Anforderungssituationen in Testaufgaben. Wie in Abschnitt 2.1.1.2 dargestellt, legen die Lehrenden auf Ebene der Modellierung zunächst fest, welche Kompetenzen geprüft werden sollen und über welche Anforderungssituationen und Verhaltensweisen zur Bewältigung dieser Situationen sich eine Kompetenz definiert (siehe dazu Modellierungsschritt 1 in Abbildung 1). Dazu sollen verschiedene Informationsquellen und Referenzpunkte herangezogen werden, wie beispielsweise die im HQR beschriebenen Wissens- und Könnensaspekte, oder die Betrachtung von Anforderungssituationen in typischen Berufsfeldern, auf die das jeweilige Studium vorbereiten soll. Anschließend müssen die Lehrenden bestimmen, welche kognitiven Ressourcen für die Bewältigung der Anforderungssituationen notwendig sind.

Auf Ebene der theoretischen Modellierung wird in dieser Arbeit daher der Frage nachgegangen welche Überlegungen die Lehrenden bei der theoretischen Modellierung von Prüfungsinhalten und -anforderungen anstellen und welche Referenzpunkte und Informationsquellen sie dabei nutzen. Weiterhin soll betrachtet werden, ob die Lehrenden Überlegungen dazu anstellen, warum das in der Prüfung abgefragte Wissen als Voraussetzung für kompetentes Handeln angenommen wird. Weiterhin wird betrachtet inwieweit die Lehrenden Überlegungen zu Strukturen und Niveaus von Kompetenzen in die Modellierung mit einfließen lassen. Die entsprechenden Informationen werden aus den Verbaldaten gewonnen.

Auf Ebene der Operationalisierung der theoretisch modellierten Aspekte in einen kompetenzorientierten Wissenstest, stellt sich zunächst die Frage, welches Verhalten in einem Test Aufschluss über dahinterliegende latente Fähigkeiten zulässt und durch welche Aufgabenmerkmale und –anforderungen dieses Verhalten evoziert wird (Leuders, 2014). Daher gilt es zu klären, welche Merkmale von Aufgaben über eine Kompetenzorientierung in Wissensaufgaben Aufschluss geben können. Im Folgenden sollen Indikatoren entwickelt werden, anhand derer der Grad der Kompetenzorientierung in Prüfungsaufgaben eingeschätzt werden kann. Diese Indikatoren dienen als Grundlage für die Entwicklung eines Kodiermanuals, um die in dieser Arbeit betrachteten Prüfungsaufgaben auszuwerten.

Wie bereits dargelegt, basieren Kompetenzen auf domänenspezifischem Wissen, das durch bestimmte kognitive Prozesse abgerufen, integriert und verarbeitet wird, um es in realen Situationen zur Lösung von Problemen anzuwenden. In verschiedenen Studien konnte gezeigt werden, dass dieses Wissen durch eine Einteilung in folgende vier Wissensarten konzeptionalisiert werden kann: (1) deklaratives Wissen, (2) prozedurales Wissen, (3) konzeptionelles Wissen und (4) strategisches Wissen (Li, Ruiz-Primo, & Shavelson, 2006; Shavelson, 2010a; Shavelson, Ruiz-Primo, Li, & Ayala, 2003). In dieser Arbeit werden die Prüfungsaufgaben deshalb dahingehend analysiert, welche Wissensarten darin aufgerufen werden. Dabei ist zu berücksichtigen, dass eine reine Betrachtung von Aufgabentypen in der Regel nichts darüber aussagt, welche Wissensart mit einer Aufgabe angesprochen wird (beispielsweise ist es möglich, mit Multiple-Choice Aufgaben sowohl deklaratives als auch konzeptionelles Wissen abzufragen). Daher werden weitere Indikatoren benötigt, um einschätzen zu können, welche Wissensart in einer Aufgabe angelegt ist. Nachfolgend werden zunächst die vier Wissensarten genauer beschrieben, um zu erarbeiten, wie damit die Voraussetzungen für kompetentes Handeln modelliert werden können. Anschließend wird auf weitere Indikatoren eingegangen, die erwiesenermaßen dazu verwendet werden können, um Wissensarten in Aufgaben zu bestimmen (Li et al., 2006) und darüber hinaus auch im Hinblick auf eine Kompetenzorientierung in Aufgaben interpretiert werden können. Konkret werden im Rahmen dieser Arbeit die Offenheit von Aufgaben, die Funktion von Aufgabenkontexten und die Komplexität von Aufgaben als entsprechende Indikatoren betrachtet und nachfolgend beschrieben. Die folgenden Betrachtungen der genannten Indikatoren von Kompetenzorientierung dienen als Grundlage für die Kodierung der in dieser Arbeit betrachteten Prüfungsaufgaben.

Wissensarten als Indikatoren für Kompetenzorientierung in Wissenstests

- **Deklaratives Wissen** umfasst Definitionen, gesicherte Fakten und Prinzipien zu einem Lernbereich (Beispiel: Schmelzpunkte verschiedener Materialien bei verschiedenen Temperaturen). Die Aneignung und das Abspeichern von deklarativem Wissen erfolgt in der Regel in Form von Begriffen, Bildern und Graphen, Zahlen und deren sinngemäßen Repräsentationen (Li et al., 2006). Die Repräsentation von Konzepten anhand verschiedener Darstellungsformen wird als essentiell für das Erlernen von deklarativem Wissen angesehen (Bransford, Brown, & Cocking, 2000).
- **Prozedurales Wissen** ist Wissen darüber, wie etwas aus- oder durchgeführt werden kann. Es enthält sogenannte „Wenn-dann“-Produktionsregeln oder eine Sequenz an Schritten, die ausgeführt werden müssen (Shavelson et al., 2003). Durch zielgerichtetes Üben kann prozedurales Wissen in automatisierte Abläufe überführt werden (Ericsson, Krampe, &

Tesch-Römer, 1993). Unter prozedurales Wissen fallen typischerweise Berechnungsaufgaben, sofern nicht auch ein Verständnis dahinterliegender Konzepte angesprochen wird.

- **Konzeptionelles Wissen** befähigt dazu, wissenschaftliche Prinzipien und Erklärungsmodelle zu integrieren und anzuwenden, um bestimmte Phänomene, Funktionsweisen oder Zusammenhänge zwischen Konzepten und Prinzipien zu erklären. Konzeptionelles Wissen kommt auch bei Fehleranalysen oder der Entwicklung von Problemlösestrategien für bekannte Anwendungsbereiche zum Tragen, um Probleme zu strukturieren und Teilprobleme auszuweisen (Li et al., 2006). Konzeptionelles Wissen baut auf Fakten- und Prozesswissen auf und stellt Verbindungen zwischen Wissens-elementen her (Shavelson et al., 2003).
- **Strategisches Wissen** kommt in unbekanntem Situationen zum Einsatz (ebd.). Es befähigt zu entscheiden, wann, wo und wie vorhandenes Wissen eingesetzt und integriert werden muss, um unbekannt Problemstellungen zu lösen. Strategisches Wissen umfasst domänenspezifische Strategien und ermöglicht zu entscheiden, wann beispielsweise eine bestimmte Prozedur angewendet werden kann oder wie konzeptionelles Wissen genutzt werden kann, um ein Problem zu strukturieren, Ziele zu setzen oder Fortschritte zu überwachen (ebd.).

In Abbildung 2 ist in einem heuristischen Modell der Zusammenhang zwischen den Wissensarten dargestellt.

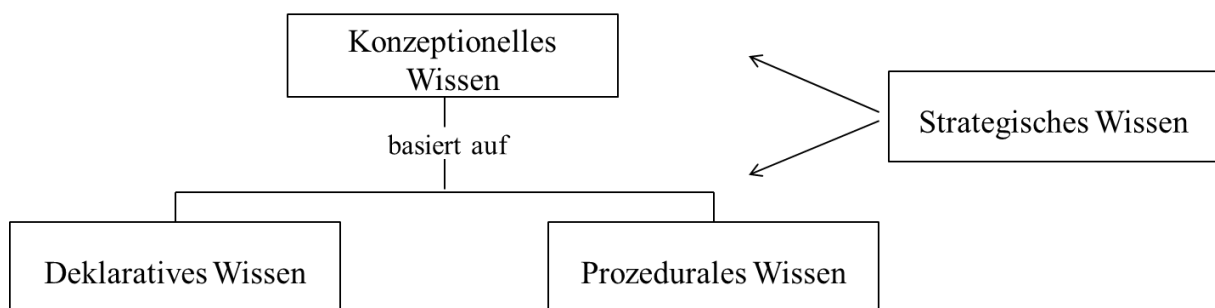


Abbildung 2: Konzeptionelle Darstellung der Wissensarten zur Einteilung von Anforderungsdimensionen in Prüfungen; Quelle: Shavelson et al. (2003)

Betrachtet man die Eigenschaften der verschiedenen Wissensarten, kann zunächst angenommen werden, dass bei der Bewältigung komplexer realer Anforderungssituationen insbesondere konzeptionelles und strategisches Wissen zum Tragen kommen. Mit konzeptionellem Wissen können die Situationen strukturiert und mit bekannten Lösungsmöglichkeiten in Verbindung gebracht werden. Außerdem können Bezüge zwischen Fakten oder Konzepten hergestellt werden, die es erlauben, bestimmte Entscheidungen, beispielsweise zum optimalen Vorgehen in einer Situation,

zu treffen. Strategisches Wissen wird benötigt, um in unbekanntem Situationen eine Problemlösestrategie zu entwickeln oder um zu entscheiden, welches Verfahren oder Wissen hilfreich zur Bewältigung der Situation ist. Bei derartigen Überlegungen greifen Personen auf eine deklarative Wissensbasis zurück. Auf prozeduraler Ebene kann es erforderlich sein, dass bestimmte Techniken oder Verfahren von einer Person auch tatsächlich ausgeführt werden können, um eine Situation zu bewältigen. Überträgt man diese Überlegungen auf die Voraussetzungen für kompetentes Handeln, sind zunächst zwei Typen von Situationen zu unterscheiden: Konzeptionelles Wissen kommt bei Situationen zum Tragen, zu deren Bewältigung Konzepte, Wissen, Strategien und Verfahren eingesetzt werden können, die einer Person bekannt sind. In unbekanntem Situationen wird zusätzlich strategisches Wissen erforderlich, um zu entscheiden, welches Wissen überhaupt auf die Situation anwendbar ist (und welches nicht). Bei einem kompetenzorientierten Wissenstest müssen Aufgaben enthalten sein, die insbesondere auf konzeptionelles und unter Umständen auf strategisches Wissen abzielen. Die Einteilung von Prüfungsanforderungen in die vier Wissensarten erscheint damit als geeignet, um abzuschätzen, ob kompetenzorientierte Anforderungen in Prüfungen gestellt werden. Da in diesem Abschnitt relativ viele Annahmen darüber getroffen wurden, ob und wie die beschriebenen Wissensarten bei der Bewältigung von realen Anforderungssituationen zum Tragen kommen, soll überprüft werden, ob durch diese Konzeptionalisierung die Qualifikationsstufen des HQR, die konkrete Anforderungssituationen benennen, entsprechend abgebildet werden können.

Laut HQR (siehe Kapitel 2.1.1.4) sollen Studierende im Verlauf eines Studiums eine breite und integrierte Wissensbasis sowie ein Verständnis wesentlicher wissenschaftlicher Prinzipien und Zusammenhänge in ihrem Studienfach erlangen. Dieses Wissen sollen sie anwenden können auf bekannte und unbekanntem Problemstellungen, auf Basis von Wissen eigene Ideen entwickeln und diese umsetzen. Außerdem sollen sie in der Lage sein relevante Informationen zu sammeln und zu interpretieren, um auf dieser Basis Urteile für weiteres Handeln abzuleiten. Eine deklarative und prozedurale Wissensbasis bilden dafür die Grundlage. Als Voraussetzung für kompetentes Handeln in realen Situationen können nach der Konzeptionalisierung im HQR die übergeordneten Wissensarten des Konzept- und Strategiewissen betrachtet werden und sollten daher über den Studienverlauf in Prüfungen verstärkt gewichtet werden. Die eingeführte Klassifikation nach Wissensarten erscheint daher auch aufgrund der Passung mit Dimensionen und Niveaus im HQR als geeignet, um die in Lehrzielen und Prüfungen abgebildeten Anforderungen zu differenzieren und im Hinblick auf eine Kompetenzorientierung zu interpretieren.

Kognitive Prozesse

Eine Betrachtung der kognitiven Anforderungen von Aufgaben ist ein relevanter Aspekt, um Hinweise auf die mit einer Aufgabe angesprochene Wissensart zu erhalten. Jedoch ist zur Feststellung der kognitiven Prozesse für Kompetenzen grundsätzlich eine domänenspezifische Festlegung vorzusehen, da allgemeingültige Bezeichnungen nicht zwangsläufig in verschiedenen Domänen auftreten und relevant sind (Leuders, 2014). Außerdem ist es möglich, dass gängige domänenunabhängige Bezeichnungen wie verstehen, anwenden, analysieren, bewerten, erschaffen (siehe Anderson & Krathwohl, 2001) in verschiedenen Domänen unterschiedliche Bedeutungen haben. In empirischen Studien zeigt sich außerdem, dass diese Kategorien nicht trennscharf unterschieden werden können (z.B. Dubs, 1978). Daher wird in dieser Arbeit auf eine Analyse kognitiver Aufgabenpotenziale verzichtet.

Aufgabenoffenheiten

Hinsichtlich der Aufgabenoffenheit wird eingeschätzt, welche Freiheitsgrade Studierende bei der Bearbeitung von Prüfungsaufgaben haben. Im Sinne der Kompetenzorientierung sollen Studierende Wissen flexibel auf unterschiedliche Problemstellungen anwenden beziehungsweise ihr Wissen dazu nutzen, um Situationen zu analysieren und zu strukturieren. Gleiches gilt für Aufgaben, die auf konzeptionelles oder strategisches Wissen abzielen. Diese Möglichkeiten müssen in Aufgaben systematisch eröffnet werden. Hinsichtlich der Offenheit einer Aufgabe kommen zwei Merkmale zum Tragen. Zum einen, ob der Anfangszustand des zu lösenden Problems klar definiert ist oder nicht. Hinzu kommt die Unterscheidung zwischen konvergenten Aufgaben (eine Lösung) und divergenten Aufgaben (mehrere Lösungen) (Kleinknecht, Maier, Metz, & Bohl, 2011). Aufgaben werden dabei nach Klarheit von Anfangs- und Zielzustand, sowie nach der Art der bei der Aufgabenbearbeitung vorzunehmenden Transformation eingeteilt. Dadurch ergeben sich drei Stufen der Offenheit (nach Kleinknecht et al., 2011):

- **Definierte und konvergente Aufgaben:** Die Aufgabe umfasst einen eindeutigen Arbeitsauftrag beziehungsweise eine klar identifizierbare Fragestellung. Die Ausgangsbedingungen in der Aufgabe sind vorgegeben. Es ist genau eine Lösung richtig. Die Gliederung der Aufgabe ist typischerweise kleinschrittig. Es ist genau eine Lösung gesucht beziehungsweise richtig. Die Lösung der Aufgabe ist daher konvergent.
- **Definierte und divergente Aufgaben:** Die Aufgabe umfasst einen eindeutigen Arbeitsauftrag beziehungsweise Fragestellung, der Ausgangszustand ist klar definiert. Es sind jedoch mehrere richtige Lösungen und/oder Lösungswege möglich.
- **undefinierte und divergente Aufgaben:** Weder Ausgangszustand noch Zielzustand sind klar definiert. Es sind mehrere Bearbeitungswege möglich. Der Aufgabentext besteht aus

einer Problemstellung oder einer Situation, jedoch ohne konkrete Handlungsaufforderungen. Es können unterschiedliche Lösungen richtig sein.

Aufgabenkontexte

Für kompetenzorientierte Wissenstests spielen insbesondere Kontext- und Situationsbezüge in Aufgaben eine tragende Rolle (Leuders, 2014). Für die Bewertung von Aufgaben aus dem schulischen Bereich wird häufig deren Lebensweltbezug betrachtet (Maier, Bohl, Kleinknecht, & Metz, 2013). Dabei wird typischerweise unterschieden, ob Aufgaben entweder keinen Lebensweltbezug, einen konstruierten Lebensweltbezug, einen konstruierten, aber authentisch wirkenden Lebensweltbezug, oder einen realen Lebensweltbezug aufweisen. Die Unterschiede zwischen den Kategorien beziehen sich darauf, inwieweit es gelingt, eine möglichst authentische Verbindung zwischen Fachwissen und der aktuellen Lebenswelt von Schülerinnen und Schülern herzustellen. Diese Unterteilung wird für die vorliegende Arbeit grundsätzlich beibehalten. Das entscheidende Kriterium bei der Analyse der Aufgaben ist jedoch die Funktion des Kontexts für Bearbeitung von Aufgaben. Bei kompetenzorientierten Wissenstests geht es darum, Wissen kontextbezogen anzuwenden. Daher wird unterschieden, ob Aufgaben (1) keinen Kontextbezug aufweisen, (2) der Kontextbezug zwar vorhanden, jedoch schwach ausgeprägt ist, (3) der Kontextbezug der Problemstellung vorgeschaltet ist, oder (4) der Kontext als Teil der Problemstellung fungiert:

- **kein Kontextbezug:** Es gibt keine Bezüge zwischen Fachwissen und Kontexten oder Situationen. Die Probleme sind rein innerfachlich.
- **Kontextbezug schwach ausgeprägt:** In der Aufgabenstellung wird keine Situation oder Kontext beschrieben, jedoch wird in einem Wort oder durch einen Beisatz eine mögliche Praxisanwendung angedeutet. Die Aufgabenstellung ist aber unabhängig davon und rein innerfachlich zu lösen.
- **Kontextbezug vorgeschaltet:** Es wird ein Kontext oder eine Situation beschrieben. Die Aufgabenstellung kann jedoch losgelöst von dem konstruierten Kontext gelöst werden. Der Kontext beziehungsweise die Situation spielt keine Rolle für die Lösung der Aufgabe.
- **Kontext als Teil der Problemstellung:** Die Aufgabenstellung ist direkt in einer Situation oder einem Kontext verortet. Die Situationen beziehungsweise Kontexte und darin enthaltene Informationen spielen bei der Bearbeitung beziehungsweise Lösung der Aufgabe eine Rolle insofern bestimmte Bedingungen daraus hervorgehen oder gestellt werden, die bei der Lösung der Aufgabe berücksichtigt werden müssen.

Sprachlogische Komplexität von Aufgaben

Als weiterer Indikator für Kompetenzorientierung in Aufgaben soll die sprachlogische Komplexität von Prüfungsaufgaben, im Sinne des Umfangs der Informationen, die bei der Aufgabebearbeitung verarbeitet werden müssen, verwendet werden (Maier et al., 2013). Eine Betrachtung der sprachlogischen Komplexität von Aufgaben ermöglicht es unter anderem abzuschätzen, inwieweit die im HQR vorgesehene Zieldimension, relevante Informationen zu sammeln und zu interpretieren, um daraus wissenschaftlich fundierte Urteile abzuleiten oder Probleme zu lösen, in Prüfungen angelegt wird. Für ein Master Studium wird der Anspruch diesbezüglich im HQR erweitert um den Umgang mit Komplexität und dem Treffen von Entscheidungen auf unvollständiger Informationsbasis (siehe Kapitel 2.1.1.4). Dazu sind Testaufgaben notwendig, in denen der Umfang der bei der Aufgabebearbeitung zu beachtenden Informationen systematisch variiert wird. Zur Einschätzung der sprachlogischen Komplexität von Aufgaben wird zwischen drei Stufen der Komplexität unterschieden: (1) einfache Darstellung, (2) wenig komplex und (3) komplex.

- **Einfache Darstellung:** Neben der Handlungsaufforderung werden keine oder nur wenige, einfach abzulesende Informationen gegeben, wie zum Beispiel für eine Berechnung benötigte Werte wie die Raumtemperatur. Die Informationen sind übersichtlich angeordnet, beispielsweise in einer Tabelle oder Grafik, die zusammen mit der Handlungsaufforderung positioniert sind. Es müssen keine Informationen gefiltert werden.
- **Weniger komplex:** Es ist möglich, sich einen schnellen Überblick über die bereitgestellten Informationen zu verschaffen, jedoch muss genauer beziehungsweise gezielter in den Informationen gelesen werden, um diese zu sichten und zu verarbeiten als es bei einer einfachen Darstellung der Fall ist. Ein weiterer Indikator auf die Kategorie *weniger komplex* kann sein, dass die Anordnung der Informationen in etwa der Chronologie der Aufgabebearbeitung entspricht. Es sind vorwiegend nur solche Informationen gegeben, die auch benötigt werden. Die Informationen sind übersichtlich dargestellt und dadurch gut erfassbar. Es müssen keine irrelevanten Informationen gefiltert werden.
- **Komplex:** Die Aufgabe enthält typischerweise mehrere Informationsquellen. Es sind zu viele Informationen verfügbar, aus denen die relevanten Aspekte gefiltert werden müssen. Ein weiterer Hinweis kann sein, dass die Anordnung der Informationen nicht der Chronologie der Aufgabebearbeitung entspricht. In der Aufgabenstellung wird nicht konkret auf die zu nutzenden Informationsquellen verwiesen.

Zwischenfazit

Ziel dieses Abschnitts war es Indikatoren zu erarbeiten, anhand derer eingeschätzt werden kann, inwieweit eine kompetenzorientierte Prüfungspraxis an Hochschulen umgesetzt wird. Dazu wurde zunächst der Versuch einer begrifflichen Klärung des Konstrukts der Kompetenz unternommen. Kompetenzen wurden definiert als die Fähigkeit einer Person, in realen Situationen erfolgreich zu handeln. Diese Konzeption von Kompetenz entspricht einem holistischen Kompetenzverständnis, bei dem das Verhalten von Testpersonen in realen oder realitätsnahen Situationen observiert wird, um festzustellen, ob die Situationen erfolgreich bewältigt werden. Ergänzend dazu wurde ein Ansatz beschrieben, bei dem die Voraussetzungen für kompetentes Handeln im Sinne kognitiver Fähigkeiten und Ressourcen erfasst werden. Für die Gestaltung von Studiengängen konnte aufgezeigt werden, dass eine Kombination der beiden Ansätze anzustreben ist. Zur Erfassung von Kernkompetenzen bietet es sich an, diese situationsnah zu erfassen, um belastbare Aussagen über die Handlungsfähigkeit in den relevanten Anforderungssituationen zu erhalten. Um die Studierenden sukzessive an die Fähigkeit zum erfolgreichen Handeln in realen Situationen heranzuführen, sollen im Studienverlauf die Voraussetzungen für kompetentes Handeln entwickelt werden. Inwieweit die Studierenden diese Voraussetzungen erfüllen kann mit kompetenzorientierten Wissenstests festgestellt werden. Das hier eingeführte Kompetenzkonzept erscheint für eine Anwendung auf den Hochschulbereich außerdem als geeignet, da damit die Qualifikationsstufen des HQR abgedeckt werden können.

Auf Ebene der theoretischen Modellierung wird betrachtet, ob die Lehrenden reale Anforderungssituationen als Referenzpunkte und Quellen heranziehen, um Prüfungsinhalte und –anforderungen festzulegen. Zudem wird untersucht, ob die Lehrenden Teilkompetenzen ausweisen und lernwirksam sequenzieren und ob zwischen Kompetenzniveaus unterschieden wird. Diese Erkenntnisse werden aus den Verbaldaten gewonnen.

Als Indikatoren für eine Kompetenzorientierung in schriftlichen Prüfungen werden die in Aufgaben umgesetzten *Wissensarten*, die *Aufgabenoffenheit*, die Verwendung von *Aufgabenkontexten* und die *sprachlogische Komplexität* von Aufgaben verwendet. Der Schwerpunkt der Aufgaben in einer kompetenzorientierten Prüfung muss auf konzeptionellem oder strategischem Wissen liegen, da diese beiden Wissensarten als direkte Voraussetzung für erfolgreiches Handeln in realen Situationen gelten. In Bezug auf die Offenheit sind Aufgaben vorzusehen, in denen von den Studierenden eigene Lösungswege entwickelt werden müssen, die unter Umständen zu unterschiedlichen, jedoch gleichermaßen richtigen, Lösungen führen können. Kompetenzorientierte Aufgaben enthalten zudem einen Kontext, auf den das Wissen bezogen werden muss. Bezüglich der sprachlogischen Komplexität sind Aufgaben vorzusehen, bei denen Informationen in dem Umfang und in

der Darbietungsform verarbeitet werden müssen, wie in den entsprechenden realen Situationen. Die genannten Indikatoren werden in einem Kodiermanual weiter spezifiziert, um damit Prüfungsaufgaben zu analysieren (siehe Tabelle 37).

2.1.2 Gütekriterien der pädagogischen Diagnostik im Kontext der Prüfungspraxis an Hochschulen

Neben der Umsetzung von Aspekten einer Kompetenzorientierung müssen weitere Gestaltungsprinzipien in Prüfungen beachtet und umgesetzt werden, damit anhand von Prüfungsergebnissen gültige und zuverlässige Rückschlüsse auf den Wissens- und Könnensstand von Studierenden gezogen werden können. Eine Prüfung muss dazu die drei Gütekriterien der Validität, Reliabilität und Objektivität erfüllen (Ingenkamp & Lissmann, 2007). Um aufklären zu können, ob diese Kriterien von den Lehrenden beachtet und in Prüfungen entsprechend umgesetzt werden, wird im Folgenden, entlang des aktuellen Wissens- und Forschungsstand der pädagogischen Diagnostik, geklärt, durch welche Maßnahmen und Gestaltungsprinzipien die Gütekriterien bei der Erstellung und Auswertung von Hochschulprüfungen erreicht werden können.

2.1.2.1 Validität

Mit dem Konzept der Validität wird bestimmt, ob mit einer Prüfung der Wissens- und Könnensstand von Studierenden zutreffend eingeschätzt werden kann. Kompetenzen und Wissen sind theoretische Konstrukte, die nicht direkt beobachtbar sind (man sieht es einer Person nicht an, ob sie über Wissen oder bestimmte Kompetenzen verfügt oder nicht) (Li et al., 2006). Um Wissens- und Könnensaspekte erfassbar zu machen, werden Prüfungen, beziehungsweise die darin enthaltenen Prüfungsaufgaben, eingesetzt, die von Studierenden bearbeitet werden. Die Lehrenden versuchen die Aufgabenanforderungen so zu gestalten, dass sie nur von den Studierenden richtig beantwortet werden können, die über die fraglichen Wissens- und Könnensaspekte verfügen. Das Konzept der Validität bezieht sich darauf, ob mit den erstellten Aufgaben gültige Rückschlüsse auf das Vorhandensein von Kompetenzen oder Wissen bei Studierenden zutreffend gezogen werden können. In dieser Arbeit wird eine Konzeptionalisierung von Validität zugrunde gelegt, die als Standard für alle Arten psychologischer und pädagogischer Leistungsmessung betrachtet wird und sich sowohl für praktische, als auch forschungsorientierte Anwendungen eignet (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Messick, 1995). Dabei steht die Konstruktvalidität als übergreifendes Konzept im Zentrum, die sich wiederum aus folgenden sechs Validitätsfacetten zusammensetzt (American Educational Research Association et al., 2014; Bühner, 2011; Messick, 1995): (1) Inhaltliche Validität (content validity), (2) kognitive Validität (substantive validity oder

response processes bei den AERA Standards), (3) strukturelle Validität (structural validity), (4) Generalisierbarkeit (generalizability), (5) externe Validität (external validity) und (6) konsequentielle Validität (consequential validity).

Die Unterteilung in sechs verschiedene Validitätsfacetten liefert ein hilfreiches und differenziertes Gerüst, um die Validität von Prüfungen umfassend einzuschätzen, ohne dabei relevante Kriterien zu ignorieren. Nachfolgend werden die einzelnen Facetten genauer beschrieben, um herauszuarbeiten, durch welche Maßnahmen Lehrende die Validität von Rückschlüssen, die auf Grundlage von Prüfungsergebnissen gezogen werden, sicherstellen können. Anhand dieser Maßnahmen werden Indikatoren erarbeitet, mit denen im Rahmen dieser Arbeit festgestellt werden kann, inwieweit die Lehrenden Maßnahmen zur Sicherstellung der Validität einsetzen.

(1) Inhaltliche Validität: Die inhaltliche Validität einer Prüfung ist gewährleistet, wenn in den Prüfungsaufgaben ausschließlich die für einen Lernbereich relevanten Inhalte und Anforderungssituationen enthalten sind. Um eine inhaltliche Validität zu erreichen, müssen von den Lehrenden Vorkehrungen auf drei Ebenen getroffen werden. Bei der theoriegeleiteten Festlegung von Prüfungsinhalten muss ein inhaltlicher Rahmen gesetzt werden, durch den festgelegt wird, welche Aspekte eines Lernbereichs relevant sind und überprüft werden sollen (Messick, 1995). Die Identifikation von Inhaltsbereichen und Anforderungssituationen, die als repräsentativ für eine Lerndomäne und die zugehörigen Kompetenzen anzusehen sind, ist dabei zentral (Klauer & Leutner, 2012; Klieme & Leutner, 2006; Pellegrino et al., 2001). Um Inhalte und Anforderungssituationen zu identifizieren, sollten mehrere Referenzpunkte und Informationsquellen verwendet werden, wie bereits bestehende Studien zum Kompetenzerwerb in einer Lerndomäne, Ergebnisse von Lehrplananalysen, Arbeitsplatzanalysen oder Kompetenzstandards (Klauer & Leutner, 2012) Auf Ebene von Prüfungen müssen die Lehrenden sicherstellen, dass die theoretisch festgelegten Inhalte und Anforderungssituationen repräsentativ durch Prüfungsaufgaben abgedeckt sind. Ist dies nicht der Fall, ist das zu messende Konstrukt in der Prüfung inhaltlich unterrepräsentiert. Eine inhaltliche Überrepräsentation eines Konstrukts liegt vor, wenn Prüfungsaufgaben zu Inhaltsbereichen enthalten sind, die im Konstrukt nicht genannt werden (American Educational Research Association et al., 2014). Letztlich müssen die Lehrenden sicherstellen, dass die identifizierten Inhalte auch auf Ebene der Aufgaben angelegt sind.

Im Rahmen der empirischen Erhebungen wird im Hinblick auf die inhaltlichen Validität untersucht, wie Lehrende bei der Auswahl von Prüfungsinhalten und -anforderungen vorgehen und welche Referenzpunkte und Informationsquellen sie dabei heranziehen. Auf Ebene der Prüfung wird der Frage nachgegangen, ob und durch welche Maßnahmen die Lehrenden die

Repräsentativität von Prüfungsinhalten und -anforderungen sicherstellen (beispielsweise durch ein zufälliges Ziehen von Aufgaben aus einem Aufgabenpool). Auf Ebene der Aufgaben wird betrachtet, inwieweit die Lehrenden überprüfen, ob in den Aufgaben die relevanten Inhalte abgebildet sind. Die von den Lehrenden formulierten Lehrziele und die darin beschriebenen Anforderungen können dabei als Referenzpunkt dienen.

(2) **Kognitive Validität:** Hinsichtlich der kognitiven Validität stellt sich die Frage, ob die Studierenden zur Lösung der Prüfungsaufgaben tatsächlich auf die für eine Kompetenz relevanten kognitiven Prozesse zurückgreifen müssen (Leuders, 2014). Um dies zu gewährleisten ist bereits bei der theoretischen Modellierung von den Lehrenden sicherzustellen, dass die ausgewählten kognitiven Prozesse (z.B. Analysieren oder Bewerten von Sachverhalten) zur Lösung der identifizierten Anforderungssituationen relevant und repräsentativ für die abzurufenden Kompetenzen sind. Auf Ebene von Prüfungsaufgaben ist die zentrale Frage, ob die Lösung von Aufgaben tatsächlich nur unter Rückgriff auf die identifizierten kognitiven Prozesse möglich ist, oder ob durch den Einsatz konstruktferner Strategien eine richtige Lösung gelingen kann, also Studierende eine Aufgabe lösen können, obwohl sie die fragliche Kompetenz gar nicht besitzen (z.B. durch test-wisness, dem Erkennen ungewollter Lösungshinweise in der Aufgabenstellung bei einem hohen Bekanntheitsgrad der verwendeten Aufgabentypen, da diese zuvor in ähnlicher Form eingeübt werden konnten, oder durch Raten oder Plausibilitätsannahmen bei Wahlantwortaufgaben) (Messick, 1995). Andererseits kann es vorkommen, dass Aufgaben für Personen oder Gruppen von Personen nicht beantwortbar sind, obwohl sie die Kompetenzen besitzen, die vermeintlich für die Lösung benötigt werden. Diese Personen würden eine zu niedrige Punktzahl bei einer entsprechenden Prüfung erzielen (Messick, 1995). Das kann dazu führen, dass diesen Personen bestimmte Kompetenzen nicht zertifiziert werden, obwohl sie diese möglicherweise besitzen. Gründe dafür können beispielsweise begründet liegen in einer hohen Leselast bei Mathematikaufgaben, sprachlichen Schwierigkeiten, die durch umständliche Aufgabenformulierung entstehen oder bei Nicht-Muttersprachlern auftreten, unklaren Handlungsanweisungen sowie einer falschen Einschätzung beziehungsweise nicht zielgruppengerechten Festlegung von Aufgabenschwierigkeiten (ebd.).

Die empirischen Analysen in dieser Arbeit beziehen sich darauf, ob und durch welche Maßnahmen die Lehrenden auf Ebene der theoretischen Modellierung absichern, dass die vorgesehenen Prüfungsanforderungen tatsächlich Aufschluss über das abzurufende Konstrukt geben. Es soll ermittelt werden, ob die Lehrenden Bezüge zwischen den Theorieelementen und der Festlegung von Aufgabenanforderungen herstellen und wie sie diese begründen. Inwieweit in den Aufgaben die relevanten Anforderungen umgesetzt sind, wird anhand der Lehrziele eingeschätzt.

- (3) Strukturelle Validität:** Mit der strukturellen Validität wird überprüft, ob die theoretisch angenommenen Strukturen und die Niveaus einer Kompetenz auch empirisch durch ein Messmodell bestätigt werden können. Eine Person, die eine bestimmte Kompetenz besitzt, sollte in der Lage sein, alle Aufgaben zu lösen, die diese Kompetenz repräsentieren. Umgekehrt sollte eine Person, die diese Kompetenz nicht besitzt die entsprechenden Aufgaben nicht lösen können. Diese Zusammenhänge können anhand von psychometrischen Messmodellen statistisch überprüft werden. Die Anwendung solcher Messmodelle erfordert ein spezialisiertes Fachwissen und kann nur von Testexperten durchgeführt werden (Leuders, 2014). Für Lehrende an Hochschulen erscheint ein Einsatz daher als nicht durchführbar. Dennoch können Lehrende bereits anhand einfacher statistischer Kennzahlen erste Hinweise auf die strukturelle Validität erhalten, indem sie beispielsweise Zusammenhänge zwischen theoretisch postulierten Anforderungsstrukturen und –niveaus und dem Abschneiden bei einzelnen Aufgaben herstellen. Fällt eine Aufgabe unerwartet schwierig aus, werden damit möglicherweise andere Kompetenzen oder Kompetenzniveaus angesprochen, wie ursprünglich intendiert. In der vorliegenden Arbeit soll daher geklärt werden, ob Lehrende statistische Kennzahlen zu ihren Prüfungen bestimmen und welche Rückschlüsse sie daraus ableiten.
- (4) Generalisierbarkeit:** Der Aspekt der Generalisierbarkeit bezieht sich darauf, ob anhand von Prüfungsergebnissen generalisierbare Aussagen über das Leistungsvermögen der geprüften Personen hinsichtlich des betreffenden Kompetenzbereichs getroffen werden können. In einer Prüfung wird aus testökonomischen Gründen im Normalfall lediglich eine Auswahl aus der Menge aller denkbaren Aufgaben gezogen, die zur Erfassung eines Kompetenzbereichs in Frage kommen. Mögliche Einschränkungen der Generalisierbarkeit können beispielsweise aus einer nicht-repräsentativen Auswahl von Aufgaben resultieren (American Educational Research Association et al., 2014; Fricke, 1974). Insbesondere bei kontextualisierten Aufgaben, also Aufgaben die einen Anwendungsbezug zu einem außerfachlichen Kontext herstellen, stellt sich die Frage, inwieweit das Abschneiden bei diesen Aufgaben eine generalisierbare Aussage über das Abschneiden bei anderen Aufgaben mit zwar gleicher Anforderungsdimension jedoch anderem Kontext (bzw. eine Aussage über alle relevanten Kontexten hinsichtlich der Lerndomäne) erlauben (Shavelson & Webb, 1991). Neben der Aufgabenauswahl kann die Generalisierbarkeit von Prüfungsergebnissen durch Messfehler, durch zum Beispiel eine inkonsistente Bewertung, eingeschränkt sein. Diese Aspekte sind konzeptionell jedoch dem Aspekt der Reliabilität zugeordnet und werden deshalb in Abschnitt 2.1.2.2 ausführlicher betrachtet. Auf Hochschulprüfungen bezogen bedeutet die Generalisierbarkeit, dass Studierende die gestellten Anforderungssituationen auch in anderen Prüfungen mit dem gleichen Ergebnis bewältigen können. Hinsichtlich des Aspekts der Generalisierbarkeit soll überprüft werden, ob

die Lehrenden die Reichweite der Aussagen, die auf Grundlage des Abschneidens bei den Aufgaben getroffen werden können, realistisch einschätzen.

(5) Externe Validität: Die Überprüfung des Aspekts der externen (oder prognostischen) Validität ist besonders dann relevant, wenn auf Grundlage von Prüfungsergebnissen, das Verhalten von Personen in anderen (realen) Situationen vorhergesagt werden soll (z.B. Gschwendtner, Geißel, & Nickolaus, 2010). Beispielsweise kann die externe Validität eines Eignungstests für ein Studium überprüft werden, indem die erreichten Testergebnisse mit den Noten, die im Verlauf des Studiums erzielt werden, korreliert werden. Ist der Zusammenhang hoch, erlaubt der Test eine valide Vorhersage der Studierfähigkeit. Besteht kein statistisch bedeutsamer Zusammenhang, ist der Test nicht geeignet, um die Studiereignung vorherzusagen. Dabei wird unterschieden zwischen diskriminanten und konvergenten Zusammenhängen (Bühner, 2011). Bei einem diskriminanten Zusammenhang besteht eine negative Korrelation zwischen den Messwerten zweier Konstrukte, bei konvergenten Zusammenhängen ist diese Korrelation positiv (ebd.). Die Gründe für eine fehlende externe Validität können möglicherweise über die anderen Validitätsaspekte aufgeklärt werden. Für einen Studieneingangstest ist bei fehlender externer Validität beispielsweise denkbar, dass die getesteten Inhalte und kognitiven Anforderungen für das betreffende Studium nicht relevant sind. In Bezug auf die externe Validität soll im Rahmen dieser Arbeit aufgeklärt werden, inwieweit die Lehrenden die in den Prüfungen abgebildeten Anforderungen als Voraussetzung für weiterführende Situationen (z.B. weiterer Studienverlauf oder Arbeitsmarktanforderungen) betrachten und welche Maßnahmen getroffen werden, um die externe Validität zu gewährleisten.

(6) Konsequentielle Validität: Prüfungsergebnisse werden als Grundlage und Legitimation für verschiedene Entscheidungen verwendet. Mit dem Aspekt der konsequenten Validität wird überprüft, ob der Einsatz von Prüfungen und die Entscheidungen, die auf Grundlage von Prüfungsergebnissen getroffen werden, die erwarteten Auswirkungen haben (Messick, 1995). Dazu muss in erster Linie die Reichweite der Prüfungsergebnisse eingeschätzt werden. Beispielsweise könnte man diesbezüglich die Frage stellen, ob die Exmatrikulation von Studierenden über das (mehrmalige) Nicht-Bestehen einer einzelnen Prüfung möglich sein sollte. Eine weitere Facette konsequentieller Validität bezieht sich auf die Auswirkungen, die durch den Einsatz einer Prüfung erzeugt werden. Studierende können durch Prüfungen zum Lernen motiviert werden, Prüfungsergebnisse können den Studierenden wichtige Rückmeldungen liefern und das weitere Lernen unterstützen. Es könnte jedoch auch zu Prüfungsangst, Stigmatisierung durch schlechte Prüfungsergebnisse oder generell Demotivation kommen. Diese und weitere mögliche Auswirkungen von Prüfungen (auch auf andere Ebenen, z.B. auf die Hoch-

schule, die Lehrenden, etc.) gilt es laut Messick (1995) abzuklären, um bei negativen Konsequenzen entgegenzusteuern.

2.1.2.2 Reliabilität

Mit dem Konzept der Reliabilität wird die Zuverlässigkeit einer Messung eingeschätzt (Bühner, 2011). Eine hohe Reliabilität liegt vor, wenn bei mehreren Messungen eines bestimmten Merkmals bei einer Person keine Unterschiede in den Messwerten auftreten, die durch die Messung bedingt sind. Dabei ist vorauszusetzen, dass das gemessene Merkmal stabil ist und damit keinen zeitlichen Schwankungen unterliegt (Shavelson & Webb, 1991). Treten Unterschiede bei derartigen Messungen auf, sind diese auf Messfehler beziehungsweise Messungenauigkeiten zurückzuführen (ebd.). In diesem Abschnitt wird auf mögliche Ursachen für Reliabilitätseinschränkungen eingegangen. Anschließend werden Aspekte und Maßnahmen beschrieben, die bei der Erstellung und Durchführung von Prüfungen beachtet werden sollten, um Messfehler zu reduzieren.

Das Prüfungsergebnis einer Person sollte im Idealfall ausschließlich über den Wissens- und Könnensstand dieser Person zustande kommen. Die Unterschiede (Varianz) in Prüfungsergebnissen zwischen Personen dürfen nur über das zu messende Konstrukt hervorgerufen werden, also durch tatsächlich vorhandene Wissens- und Könnensunterschiede bedingt sein (konstruktrelevante Unterschiede). Unterschiede in Prüfungsergebnissen, die nicht über das Leistungsniveau von Personen, sondern durch Messfehler zustande kommen, erzeugen ungewollte Varianz (konstruktferne Unterschiede). Die Ursachen für Messfehler bei Prüfungen können vielfältig sein. Nachfolgend werden in Tabelle 1, basierend auf Untersuchungen zur Reliabilität von Leistungstests (Ingenkamp & Lissmann, 2007; Shavelson, Baxter, & Gao, 1993; Shavelson & Ruiz-Primo, 1999b; Shavelson, Ruiz-Primo, & Wiley, 1999; Shavelson & Webb, 1991; Webb, Shavelson, Kim, & Chen, 1989), verschiedene Ursachen und Facetten von Messfehlern erarbeitet und beschrieben und anschließend auf die Durchführung von Hochschulprüfungen übertragen.

Tabelle 1³: Ursachen und Art der dadurch erzeugten Varianz bei Leistungstests

Ursachen für Varianz	Art der Varianz	Beschreibung
Personen	Konstruktbedingte (gewollte) Varianz	Die beobachtete Varianz in den Messergebnissen ist auf Unterschiede in den geprüften Personenmerkmalen zurückzuführen (Kompetenz- oder Wissensunterschiede). Diese Art von Varianz entsteht auf Grundlage der wahren Messwerte und ist gewollt.
Systematische Messfehler, mit gleicher Wirkung auf alle Testpersonen		
Aufgaben	Ungewollte Unterschiede in den Aufgabenschwierigkeiten bei Aufgaben die das gleiche Merkmal abprüfen	Unterscheidet sich die Schwierigkeit von Aufgaben ungewollt voneinander, wird Varianz in Messergebnissen erzeugt, die als Messfehler zu bewerten ist. Eine Ursache hierfür könnte sein, dass sich die Instruktionsqualität von Aufgaben unterscheidet.
	Vorerfahrungen mit den Anforderungssituationen	Bestehen Vorerfahrungen hinsichtlich bestimmter Anforderungssituationen, kann die Schwierigkeit von Aufgaben ungewollter Weise verringert werden.
	Zufällig erzeugte Varianz durch Aufgaben	Bei der Bearbeitung von Aufgaben kann durch unerwartete Einflüsse Varianz in den Ergebnissen auftreten, z.B. durch einen plötzlichen Konzentrationsabfall oder durch andere systematische aber nicht identifizierte Einflüsse, wie die Testung verschiedener Personen zu unterschiedlichen Zeiten.
Messzeitpunkt	Konstanter Effekt durch inkonsistentes Verhalten bei allen getesteten Personen zwischen zwei oder mehreren Messzeitpunkten	Der Messzeitpunkt ist neben den Aufgaben eine der Hauptquellen für ungewollte Varianz in Messergebnissen. Typischerweise werden Prüfungen nur einmal an einem bestimmten Tag zu einer bestimmten Uhrzeit durchgeführt. Dies kann systematische Einflüsse erzeugen, die für alle getesteten Personen gleichermaßen wirken.
Bewerter ⁴	Konstanter Effekt durch unterschiedliche Stringenz bei Bewertern für alle getesteten Personen	Die Bewertungen unterschiedlicher Bewerter können sich unterscheiden, wenn eine Person stringenter, im Sinne von strenger oder genauer, korrigiert als eine andere Person, die einen liberaleren Maßstab bei der Korrektur anlegt.
Messfehler durch Interaktionen zwischen einer Person und Merkmalen eines Tests		
Interaktion zwischen Person und Aufgaben	Inkonsistenzen im Verhalten einer Person bei der Bearbeitung gleichwertiger Aufgaben, die dasselbe Merkmal abprüfen	Neben den beschriebenen Haupteffekten, die sich systematisch auf alle Personen beziehen, treten Interaktionen zwischen den genannten Facetten auf. Die Interaktion zwischen einzelnen Aufgaben und Personen kann zu ungewollter Varianz führen. Konstruktirrelevante Einfachheit kann entstehen, wenn die in Aufgaben abgebildeten Anforderungssituationen einen bestimmten Bekanntheitsgrad für eine getestete Person haben beziehungsweise wenn Vorerfahrungen bestehen. Bezüglich des Bekanntheitsgrads könnte es Unterschiede im Zugang zu ähnlichen Aufgabentypen geben oder eine Vertrautheit mit

³ Aus Gründen der besseren Lesbarkeit werden die Trennlinien zwischen Zeilen und Spalten in der Tabelle eingeblendet.

⁴ Soweit im Folgenden Berufs- Gruppen- und / oder Personenbezeichnungen in der maskulinen Form Verwendung finden, so ist dabei auch stets die jeweils weibliche Form mit angesprochen. Der Verfasser sieht zur Wahrung des Leseflusses bewusst von einer genderneutralen Ausdrucksweise ab.

2.1.2 Gütekriterien der pädagogischen Diagnostik im Kontext der Prüfungspraxis an Hochschulen

		bestimmten Aufgabentypen, mit denen das gleiche Merkmal erfasst werden kann. Vorerfahrungen können bezüglich der Aufgabeninhalte oder Kontexte bestehen, die die Bearbeitung möglicherweise einfacher machen. Konstruktirrelevante Schwierigkeit kann beispielsweise durch sprachliche Aspekte entstehen, oder durch unklare Anweisungen, die speziell für eine Person Auswirkungen haben. Andererseits können Aufgaben schwieriger werden für Personen, die nicht mit dem Aufgabentyp oder den darin enthaltenen Inhalten und Kontexten vertraut sind. In allen diesen Fällen tritt eine Interaktion zwischen Aufgabe und Person auf.
Interaktion zwischen Person und Prüfungsformat	Inkonsistenzen im Verhalten einer Person, die durch verschiedene Prüfungsformate hervorgerufen wird	Hier ist Varianz gemeint, die bei einer Person durch den Einsatz verschiedener Prüfungsformate auftritt. Beispielsweise könnte es sein, dass eine Person in einer mündlichen Prüfungssituation nicht in der Lage ist, die gestellten Aufgaben zu bewältigen obwohl sie in einem schriftlichen Test dazu in der Lage war. Auch Vorerfahrungen mit bestimmten Formaten spielen hier eine Rolle.
Interaktion zwischen Person und Bewerter	Inkonsistenzen eines Bewerter bei der Bewertung des Verhaltens einer Person	Ein Bewerter kann möglicherweise, bedingt durch konstruktferne Eigenschaften der betreffenden Personen, wie z.B. ein bestimmtes Schriftbild, Ausdrucksweise oder Herkunft, von seiner Bewertung bei einzelnen Personen abweichen.
Interaktion zwischen Bewerter und Messzeitpunkt	Konstanter Effekt für alle getesteten Personen durch unterschiedliche Stringenz des Bewerter bei verschiedenen Messzeitpunkten	Die Konsistenz eines Bewerter kann sich zwischen den Bewertungszeitpunkten unterscheiden. Werden mehrere Prüfungen hintereinander korrigiert, kann die Konsistenz beispielsweise durch nachlassende Konzentration beeinträchtigt werden. Außerdem können durch bereits korrigierte Antworten bestimmte Anker gesetzt werden, die die Bewertungsperspektive verzerren.
Interaktion zwischen Person und Messzeitpunkt	Inkonsistenzen im Verhalten einer Person zu verschiedenen Messzeitpunkten	Bei konstanter Ausprägung des konstruktrelevanten Personenmerkmals (das als zeitlich stabil angenommen wird) bei einer Testperson, kann das Abschneiden bei einem Test aufgrund von Schwankungen in der Leistungsfähigkeit einer Person variieren.
Interaktionseffekt zwischen Person, Bewerter, Aufgaben, Prüfungsformat, Messzeitpunkt und zufällig erzeugter Varianz	Effekt durch die einzigartige Kombination von Person, Bewerter, Aufgaben, Prüfungsformat, Messzeitpunkt und zufällig aufgetretener Varianz	Durch das Aufeinandertreffen von Personen auf Aufgaben, die in einem bestimmten Prüfungsformat umgesetzt sind und zu einem bestimmten Zeitpunkt bearbeitet und von einem bestimmten Bewerter bewertet werden und noch weitere zufällige Ereignisse, kann ebenfalls weitere ungewollte Varianz entstehen. Diese Art von Varianz ist im Vorfeld schwierig vermeidbar, jedoch ist eine nachgelagerte statistische Einschätzung des Effekts möglich.

Als Ursachen für Messfehler die konstruktferne Varianz in Messergebnissen erzeugen können, wurden die Prüfungsaufgaben, die Bewerter, die die Prüfungsantworten auswerten, und der Messzeitpunkt identifiziert. Diese Aspekte können einen systematischen Einfluss erzeugen, der auf alle Personen, die an einer Prüfung teilnehmen, gleichermaßen wirkt. Zudem können Interaktionseffekte zwischen einzelnen Personen und diesen Aspekten auftreten, die das Messergebnis bei diesen Personen beeinflussen. Bei der Erstellung und Auswertung von Prüfungen müssen die Leh-

renden versuchen, diese Fehlerquellen zu vermeiden, um die Reliabilität der Prüfungsergebnisse sicherzustellen. Nachfolgend wird dargelegt, durch welche Maßnahmen die Lehrenden Messfehler vermeiden können. Die Prüfungspraxis der Hochschullehrenden wird in dieser Arbeit dahingehend analysiert, ob die Lehrenden diese Maßnahmen umsetzen.

Eine Fehlerquelle bilden die Aufgaben. Als erste Ursache für Messfehler wurden bei den Aufgaben *unintendierte Unterschiede in Aufgabenschwierigkeiten* genannt. Um derartige ungewollte Schwierigkeitsunterschiede in den Aufgaben zu vermeiden, müssen die Lehrenden auf inhaltlicher Seite sicherstellen, dass in den Aufgaben ausschließlich konstruktrelevante Inhalte, Situationen und Anforderungen abgebildet sind. Auf formaler Seite sind bei den Aufgaben auf Gestaltungsmerkmale wie zum Beispiel die Klarheit der Instruktion, die Angabe eines Erwartungshorizonts und die Verwendung verständlicher Sprache zu achten (Bücker et al., 2015; Gronlund, 1974; National Assessment Governing Board, 2007). Auf Ebene eines Tests kann die Anzahl der Aufgaben, die sich auf ein Merkmal beziehen, erhöht werden, um Effekte einzelner Aufgaben auszugleichen (Ingenkamp & Lissmann, 2007).

Der zweite Aspekt bei Aufgaben, der zu konstruktferner Varianz führen kann, sind bestehende *Vorerfahrungen mit den Anforderungssituationen* seitens der Testpersonen. Sind Studierende beispielsweise hauptsächlich mit der Bearbeitung offener Aufgabentypen vertraut, kann durch eine Umstellung auf ein Multiple-Choice-Format, mit dem die gleichen Merkmale geprüft werden, die Aufgabenschwierigkeit ungewollt erhöht werden (Katz, Bennett, & Berger, 2000). Außerdem sind inhaltliche Vorerfahrungen mit Aufgabenkontexten zu berücksichtigen (haben bestimmte Personen eine Affinität zu einem Aufgabenkontext, kann dies die Bearbeitung erleichtern). Auf Ebene der *Prüfungsformate* kann ebenfalls konstruktferne Varianz entstehen. Sind Testpersonen an Paper-Pencil Formate gewöhnt, kann durch den Einsatz computerbasierter Testverfahren der Testwert verschlechtert werden (Bugbee, 1996). Hier gilt es die Testpersonen entsprechend auf die neuen Formate vorzubereiten. Diesbezüglich soll überprüft werden, ob auf eine gewisse Konstanz bezüglich der eingesetzten Prüfungs- und Aufgabenformate und Anforderungen geachtet wird.

Neben den Aufgaben wurden die *Bewerter* als eine Quelle für Messfehler beschrieben. Im Hinblick auf die Bewertung von Prüfungsantworten muss eine hohe Konsistenz angestrebt werden, um Messfehler zu vermeiden, die durch die Anwendung unterschiedlicher Bewertungsmaßstäbe durch Korrektoren auftreten können. Dazu können Bewertungsraster und Schulungen von Bewertern durchgeführt werden. Auf diese Aspekte wird in Abschnitt 2.1.3.4 vertieft eingegangen.

Der *Prüfungszeitpunkt* stellt eine weitere Quelle für konstruktferne Varianz dar. Typischerweise werden Prüfungen nur einmal an einem bestimmten Tag zu einer bestimmten Uhrzeit durchgeführt. Dadurch können Einflüsse entstehen, die für alle getesteten Personen gleichermaßen wir-

ken. Dieser Aspekt trägt erstmal zur Vergleichbarkeit der Prüfungsergebnisse bei, da alle Prüfungsteilnehmenden unter den gleichen Bedingungen die Prüfung ablegen. Dennoch sollte bei der Festlegung des Prüfungstermins beachtet werden, dass beispielsweise bei einer Durchführung der Prüfung an einem bestimmten Tag, an dem die Gruppe von Studierenden, die an der Prüfung teilnimmt, bereits mehrere Vorlesungen besuchte, das Prüfungsergebnis im Mittel (nicht zwingend bei allen Prüfungsteilnehmern) schlechter ausfallen könnte, als an einem anderen Tag am Vormittag, an dem zuvor keine anderen Verpflichtungen bestanden. Neben einem systematischen Einfluss des Prüfungszeitpunkts auf eine Gruppe von Studierenden, können Interaktionseffekte zwischen einer Person und dem Prüfungszeitpunkt auftreten (ebd.). Eine Möglichkeit, um auf diesen Einfluss zu reagieren, wäre, einen längeren Prüfungszeitraum anzugeben, innerhalb dessen sich die Prüfungsteilnehmer selbstständig entscheiden, an welchem Tag und zu welcher Uhrzeit sie die Prüfung ablegen. Diese Möglichkeit könnte beispielsweise mit computerbasierten Prüfungssystemen umgesetzt werden (Wannemacher, 2007). In dieser Arbeit wird jedoch auf den Messzeitpunkt nicht näher eingegangen, da die Möglichkeiten zu Veränderungen vermutlich den Handlungsspielraum einzelner Lehrender erstmals übersteigt und daher auch für das geplante Qualitätsentwicklungsprogramm nicht als realistische Zielsetzung erscheint.

Basierend auf diesen Ausführungen zu möglichen Ursachen für Messfehler, die zu einer Einschränkung der Reliabilität von Prüfungen führen können, werden in dieser Arbeit folgende Aspekte analysiert. Die Prüfungsaufgaben werden dahingehend untersucht, ob formale Gestaltungsaspekte wie die Klarheit der Instruktion, die Angabe von Erwartungshorizonten und die Verwendung verständlicher Sprache, umgesetzt werden. Bezüglich der eingesetzten Prüfungs- und Aufgabenformate soll überprüft werden, ob auf eine gewisse Konstanz geachtet wird. Hinsichtlich der Messfehler, die durch Bewerter auftreten können, soll überprüft werden, inwieweit die Lehrenden Maßnahmen zur Sicherung der Konsistenz bei der Bewertung von Prüfungsantworten einsetzen.

2.1.2.3 Objektivität

Unter dem Konzept der Objektivität werden Maßnahmen zusammengefasst, die dazu beitragen, Ursachen für Messfehler zu reduzieren, indem verschiedene Aspekte im Prüfungsprozess standardisiert werden (Bühner, 2011). Typischerweise wird zwischen drei Arten von Objektivität unterschieden (Bühner, 2011). Mit der *Durchführungsobjektivität* soll sichergestellt werden, dass alle Prüfungsteilnehmer die gleichen Informationen über den Ablauf der Prüfung, die Bearbeitungszeit, erlaubte Hilfsmittel und Hinweise zur Bearbeitung und Beantwortung der Aufgaben erhalten. Die Instruktionen müssen daher für alle Prüfungsteilnehmerinnen und –teilnehmer verständlich und einheitlich sein. Soziale Interaktionen zwischen Prüfungsleiter und einzelnen Studierenden

müssen minimiert werden, damit alle Prüfungsteilnehmenden die gleichen Informationen besitzen.

Die *Auswertungsobjektivität* ist erfüllt, wenn unterschiedliche Korrektoren bei der Klassifizierung von Prüfungsantworten zum gleichen Ergebnis kommen. Die Auswertungsobjektivität kann bei Aufgaben mit offenen Antwortformaten durch die Verwendung von Bewertungsrastern erhöht werden (Gorman & Rentsch, 2009). Ein Bewertungsraster enthält Bewertungskategorien (z.B. richtig-falsch), in denen definiert wird, welche Merkmale eine Aufgabe aufweisen muss, um einer Kategorie zugeordnet werden zu dürfen. Zusätzlich sollte zu jeder Bewertungskategorie ein Ankerbeispiel angegeben werden, um so die Interpretationsspielräume bei der Bewertung von Antworten zu reduzieren und eine Konsistenz zwischen Bewertern herzustellen (ebd.). Bei Wahlantwortformaten können Lösungsschablonen hilfreich sein, jedoch können auch dabei bei unsachgemäßer Handhabung, zum Beispiel durch falsches Anlegen der Schablone, Fehler entstehen (Bühner, 2011). Inwieweit diese Instrumente von den Lehrenden eingesetzt werden, soll in dieser Arbeit geklärt werden.

Die *Interpretationsobjektivität* bezieht sich auf die Frage, ob unterschiedliche Bewerter die gleichen Schlüsse aus dem Testergebnis ziehen und so beispielsweise für einzelne Aufgaben beziehungsweise Antworten die gleiche Anzahl an Punkten vergeben oder die gleichen Noten gebildet werden. Nach welchen Kriterien Punkte auf einzelne Aufgaben beziehungsweise Antwortelemente vergeben werden kann in einem Bewertungsraster angegeben werden. Für die Notenvergabe sollte ein Notenschlüssel festgelegt werden (Bühner, 2011; Sedlmeier & Renkewitz, 2007). Die Umsetzung dieser Aspekte durch Hochschullehrende soll ebenfalls in dieser Arbeit betrachtet werden.

Zwischenfazit

In diesem Abschnitt wurden Indikatoren erarbeitet, anhand derer die Qualität von Hochschulprüfungen eingeschätzt werden kann. Zu den Gütekriterien der Validität, Reliabilität und Objektivität wurden jeweils Aspekte und Kriterien erarbeitet, die bei der Erstellung, Durchführung und Auswertung von Prüfungen und bei der Interpretation von Prüfungsergebnissen beachtet werden müssen, um die Gütekriterien zu erfüllen und so die Qualität von Prüfungen zu sichern. Mit der Betrachtung der Gütekriterien konnten wesentliche Gestaltungsaspekte geklärt werden, die als Indikatoren für die Einschätzung der Prüfungsqualität verwendet werden können. Zur Sicherstellung der Validität sind sowohl auf Ebene der theoretischen Modellierung, als auch auf Ebene der Prüfung und der Prüfungsaufgaben verschiedene Maßnahmen vorzusehen. Bei der theoretischen Modellierung des Prüfungsgegenstands ist es wesentlich, dass Prüfungsinhalte und –anforderungen systematisch und nachvollziehbar identifiziert und beschrieben werden. Auf Ebene der Prüfung ist

ein Kriterium für die Validität, dass das vordefinierte Konstrukt repräsentativ durch Aufgaben abgedeckt wird. Auf Ebene der Aufgaben ist abzusichern, dass die theoretisch beschriebenen Prüfungsanforderungen übersetzungsgetreu in Aufgaben übersetzt werden können. Hinsichtlich der Reliabilität wurden Ursachen für Messfehler identifiziert und genauer beschrieben. Demnach können Messfehler durch mehrere Aspekte hervorgerufen werden. Bei der Erstellung und Auswertung von Prüfungen sind dabei vor allem Merkmale in den Aufgaben, die Konsistenz der Bewertung von Prüfungsantworten und der Messzeitpunkt zu beachten. Die Objektivität bezieht sich auf Maßnahmen zur Standardisierung des Prüfprozesses, damit alle Prüfungsteilnehmenden die gleichen Chancen bei der Bearbeitung der Prüfung vorfinden. Die Berücksichtigung und Umsetzung dieser Maßnahmen durch Hochschullehrende wird als Indikator für die erreichte Qualität von Prüfungen verwendet. Im nächsten Abschnitt werden die hier beschriebenen Aspekte und Maßnahmen weiter ausdifferenziert, um sie empirisch erfassbar zu machen. Dabei wird ein analytischer Rahmen entwickelt, anhand dessen die Prüfungspraxis von Hochschullehrenden empirisch analysiert sowie Stärken und Schwächen in bestehenden Prüfungen identifiziert werden können.

2.1.3 Vorgehensmodell zur Erstellung und Auswertung von Prüfungen

In diesem Abschnitt werden die wesentlichen Erkenntnisse aus den theoretisch-konzeptionellen Betrachtungen zur Kompetenzmessung (Abschnitt 2.1.1) und den Gütekriterien der pädagogischen Diagnostik (Abschnitt 2.1.2) anwendungsbezogen auf den Prozess zur Erstellung einer Hochschulprüfung übertragen. Die in den vorangegangenen Abschnitten beschriebenen Kriterien und Maßnahmen zur Kompetenzmessung und der Erreichung der Gütekriterien Validität, Reliabilität und Objektivität werden soweit konkretisiert, dass sie in der Prüfungspraxis von Hochschullehrenden direkt beobachtet werden können. Wie bereits beschrieben, werden bei den Analysen in dieser Arbeit sowohl die von den Lehrenden erstellten Produkte, wie Lehrziele, Prüfungen und Prüfungsaufgaben (materialbasierte Daten) als auch prozessbezogene Daten (Verbaldaten) herangezogen. Mit der Analyse der materialbasierten Daten soll festgestellt werden, welche Charakteristika Lehrziele und Prüfungen aufweisen und inwiefern zentrale Gestaltungskriterien, die zu einer hohen Prüfungsqualität beitragen, umgesetzt werden. Mit der Betrachtung der prozessbezogenen Daten können authentische Einblicke in die Prüfungspraxis von Hochschullehrenden bei der Erstellung und Auswertung von Prüfungen und die damit verbundenen Überlegungen gewonnen werden.

Um eine prozessnahe Analyse der Prüfungspraxis vornehmen zu können wird nachfolgend ein Vorgehensmodell entwickelt, das den Prozess der Erstellung und Auswertung von Prüfungen in einzelne Schritte differenziert. Die in den vorangegangenen Kapiteln erarbeiteten Kriterien für kom-

petenzorientiertes und an Gütekriterien ausgerichtetem Prüfen sowie die dabei zu beachtenden Maßnahmen werden diesen Schritten zugeordnet. Das entwickelte Vorgehensmodell ist damit als Synthese der erarbeiteten Erkenntnisse zum Prüfen von Kompetenzen und den dabei zu erfüllenden Gütekriterien zu betrachten. Die inhaltliche Gestaltung des geplanten Qualitätsentwicklungsprogramms ist ebenfalls entlang dieses Vorgehensmodells ausgerichtet. Im Folgenden werden die Schritte überblicksartig skizziert. Zu jedem der Schritte werden Leitfragen angegeben, die im Rahmen der empirischen Erhebungen beantwortet werden sollen.

- (1) **Festlegung des Prüfungsgegenstands:** Welches Wissen und welche Kompetenzen möchten die Lehrenden prüfen? Wie strukturieren und gewichten sie einzelne Wissens- und Könnensaspekte zueinander? Welche Referenzpunkte und Auswahlvorschriften legen sie bei der Identifikation von Wissens- und Könnensaspekten zugrunde? Werden Bearbeitungsniveaus bei den einzelnen Anforderungen unterschieden? Werden die Wissens- und Könnensaspekte durch Lehrziele beschrieben?
- (2) **Erstellung von Prüfungsaufgaben:** Erfolgt die Erstellung von Prüfungsaufgaben bezogen auf Lehrziele? Wie gut stimmen die Anforderungen in Lehrzielen und Aufgaben überein? Berücksichtigen die Lehrenden bestimmte Gestaltungsaspekte bei der Entwicklung von Aufgaben, damit (möglichst ausschließlich) die angestrebten Wissens- und Könnensaspekte übersetzungskonform abgebildet werden?
- (3) **Rückbindung der Aufgaben an den Prüfungsgegenstand:** Mit welchen Maßnahmen überprüfen die Lehrenden, ob die entwickelten Aufgaben tatsächlich die intendierten Wissens- und Könnensaspekte abprüfen oder auch anderweitig gelöst werden?
- (4) **Bewertung von Prüfungsantworten:** Setzen die Lehrenden Instrumente ein, um zu gewährleisten, dass in den Prüfungsantworten die relevanten Verhaltensaspekte bewertet und eine konsistente Bewertung möglich ist.
- (5) **Punkte- und Notenvergabe:** Vergeben die Lehrenden Punkte auf die relevanten Aspekte in den Prüfungsantworten? Ist die Notenvergabe für individuelle Studierende an konstruktrelevanten Wissens- und Könnensunterschiede ausgerichtet (Kriteriumsorientiert) oder am Abschneiden der Bezugsgruppe (soziale Norm) orientiert?
- (6) **Rückschlüsse aus den Prüfungsdaten:** Welche Informationen ziehen die Lehrenden aus den Prüfungsergebnissen? Welche statistischen Kennzahlen werden bestimmt?

Bevor auf die einzelnen Schritte eingegangen wird, gilt es grundlegend zu beachten, dass die Erstellung einer Prüfung im Normalfall nicht durch ein lineares Abarbeiten der einzelnen Schritte

erfolgt, sondern durch ein iteratives Vorgehen. Jeder Schritt kann Informationen und Erkenntnisse über vorangegangene und nachfolgende Schritte liefern und somit zum Ausgangspunkt für eine weitere Überarbeitung werden. Die Festlegung des Prüfungsgegenstands bildet den Ausgangspunkt für die weiteren Schritte (Wilson, 2005). Das Vorgehensmodell mit den sechs Schritten zur Erstellung einer Prüfung ist in Abbildung 3 schematisch abgebildet. In den folgenden Unterkapiteln wird auf die sechs Schritte jeweils vertiefend eingegangen, um handlungsnahe Indikatoren für eine empirische Analyse der Prüfungspraxis von Hochschullehrenden zu erarbeiten.

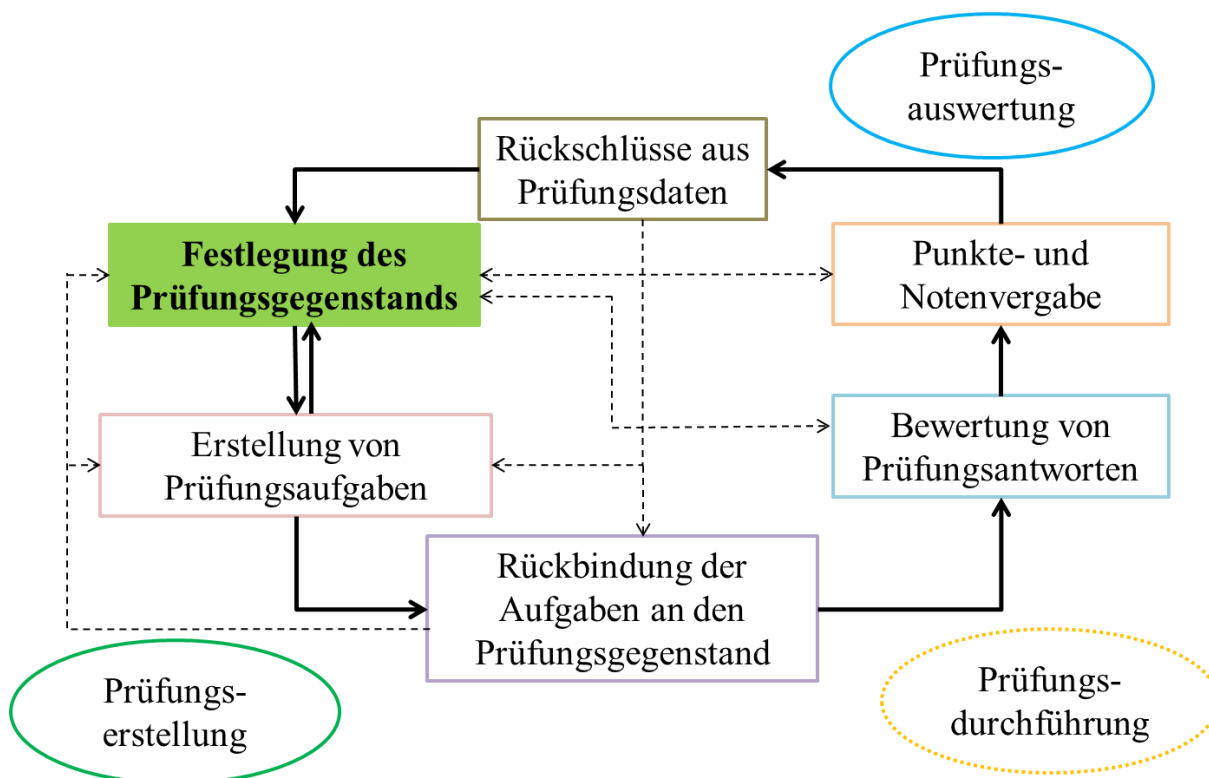


Abbildung 3: Vorgehensmodell für die Prüfungserstellung und -auswertung

2.1.3.1 Festlegung der Prüfungsgegenstandes

Im ersten Schritt der Prüfungserstellung legen die Lehrenden fest, welche Kompetenzen oder Wissensbestände bei Studierenden geprüft werden sollen. Durch die Festlegung des Prüfungsgegenstands wird ein inhaltlicher Rahmen gesetzt, anhand dessen entschieden werden kann, welche Wissens- und Könnensaspekte für einen Lernbereich relevant sind. Nachfolgend wird darauf eingegangen, welche Schritte bei der Festlegung des Prüfungsgegenstands von Prüfenden durchzuführen sind und wie diese Aspekte für die empirischen Analysen in dieser Arbeit verwendet werden.

Konzeptionelle Analyse des Prüfungsgegenstands

Die konzeptionelle Analyse des Stoffgebiets bildet den ersten Schritt bei der Festlegung des Prüfungsgegenstands. Die Lehrenden identifizieren im Rahmen der theoretischen Modellierung des

Prüfungsgegenstands die für eine Kompetenz relevanten Anforderungssituationen, zugehörige Kontexte, Inhalte und Verhaltensaspekte und beschreiben diese anhand von Lehrzielen. Als Referenzpunkte und Informationsquellen zur Bestimmung der genannten Aspekte auf Studiengangsebene kann dabei der Qualifikationsrahmen für Deutsche Hochschulabschlüsse (Kultusministerkonferenz, 2005) verwendet werden. Auf Ebene von Lehrveranstaltungen können die Lehrenden, die Lehrziele typischerweise aus der darüber liegenden Modulebene beziehungsweise der Studiengangsebene ableiten. Eine weitere Möglichkeit besteht darin, Anforderungen zu identifizieren, die Studierende in nachgelagerten Veranstaltungen oder in potenziellen Berufsbildern bewältigen müssen. Dazu kann beispielsweise eine empirisch-demoskopische Lehrzielforschung durchgeführt werden, durch die Aufgaben beziehungsweise Anforderungssituationen identifiziert werden, denen Lernende in einer Domäne, einem Berufsfeld oder einem Curriculum begegnen können (Klauer, 1974).

Neben diesen Ansätzen können Lehrzieltaxonomien für die Lehrenden eine Orientierungshilfe zur Strukturierung von Lehrzielen bieten. In der deutschen Hochschuldidaktik stellt die Lehrzieltaxonomie nach Bloom, Englhart, Furst, Hill und Krathwohl (1956) nach wie vor einen wichtigen Standard dar (Braun et al., 2014). Die darin beschriebenen kognitiven Niveaus (Erinnern, Verstehen, Anwenden, Analysieren, Synthese und Evaluation) werden häufig für die Formulierung von Lehrzielen auf Modulebene verwendet (ebd.).

Mit der Festlegung von Auswahlvorschriften und der Angabe eines Suchraums (z.B. Kompetenzstandards) legen die Lehrenden handlungsleitende Annahmen für die Auswahl von Lehrzielen fest, indem sie bestimmen, aus welchen Quellen und mit welcher Methode Lehrziele abgeleitet werden (Klauer & Leutner, 2012). Die Festlegung von Auswahlvorschriften und des Suchraums wird als Suchanleitung bezeichnet (ebd.). Durch die Explizierung einer Suchanleitung können die Lehrenden die Passung von Kompetenzen für einen Studiengang, eine Lehrveranstaltung, etc. bewerten, da nachvollzogen werden kann, welche Annahmen und Gründe für die Auswahl der einzelnen Komponenten, die eine Kompetenz näher bestimmen, zugrunde gelegt wurden.

Möchten die Lehrenden mit einer Prüfung feststellen, ob Studierende in der Lage sind reale berufliche Situationen zu bewältigen, müssen sie typische Anforderungssituationen und erfolgreiche Verhaltensweisen zur Bewältigung dieser Situationen in einem bestimmten Anwendungs- beziehungsweise Berufsfeld identifizieren und durch Lehrziele beschreiben. Außerdem müssen dabei möglicherweise variierende Kontextvariablen in den Anforderungssituationen berücksichtigt werden (z.B. kann sich bei einer Lehrkraft die kompetent unterrichten soll, u.a. die Kontextvariable der Klassenstärke verändern). Geht es den Lehrenden darum, festzustellen, ob Studierende die Voraussetzungen für kompetentes Handeln in diesen Situationen besitzen, zum Beispiel in Form

eines bestimmten Fachwissens, müssen die entsprechenden Wissenseinheiten identifiziert werden. Außerdem muss festgelegt werden, auf welche Art und Weise an welchen Aufgabensituationen dieses Wissen kognitiv verarbeitet werden muss und welche Kontextbezüge in den Aufgabensituationen angelegt sein müssen.

Die beschriebenen Schritte zur konzeptionellen Analyse des Prüfungsgegenstands werden für die empirischen Erhebungen in dieser Arbeit wie folgt aufgegriffen. Zunächst soll anhand der Verbaldaten ermittelt werden, welche Wissens- und Könnensaspekte Hochschullehrende als relevant wahrnehmen und welche Bezugspunkte und Auswahlvorschriften sie bei der Auswahl von Prüfungsinhalten und –anforderungen, Kontextvariablen und Verhaltensaspekten zugrunde legen. Hinsichtlich der Kompetenzorientierung von Studieninhalten ist insbesondere die Frage relevant, ob Lehrende reale und kontextualisierte Anforderungssituationen als Ausgangspunkt für die Festlegung von Prüfungsanforderungen und –inhalten wählen oder ob eine rein innerfachliche Auswahl von Prüfungsinhalten erfolgt (siehe dazu auch Abschnitt 2.1.1.2).

Zerlegung von Kompetenzen in Teilkompetenzen und Festlegung von Bearbeitungsniveaus

Nach Festlegung der Anforderungssituationen und den Verhaltensweisen zur Bewältigung der Situationen muss von den Lehrenden festgelegt werden, inwieweit die Anforderungen in Teilkompetenzen zerlegt und strukturiert und welche Bearbeitungsniveaus unterschieden werden können. Außerdem müssen sie, basierend auf einer Lerntheorie zur Kompetenzentwicklung im betreffenden Lernbereich, ein Lernentwicklungsmodell angelegen, um festzulegen, in welcher Reihenfolge oder Kombination Teilkompetenzen und Wissens Elemente vermittelt werden sollen (Klauer & Leutner, 2007).

Die ausgewiesenen Teilkompetenzen werden jeweils durch ein Aufgabenuniversum repräsentiert (Klauer & Leutner, 2012). Mit Lehrzielen werden Eigenschaften und Merkmale des Aufgabenuniversums definiert, die spezifizieren, welche Gestaltungsmerkmale eine Aufgabe aufweisen muss, um einem Aufgabenuniversum zugeordnet werden zu können (ebd.). Die Anforderungssituationen und das erwartete Verhalten müssen von den Lehrenden nachvollziehbar definiert werden, damit die Rückschlüsse, die aus dem Verhalten der Testpersonen gezogen werden, validierbar sind. Lehrziele sollen daher spezifizieren welche (1) Verhaltensweise an einem bestimmten (2) Inhalt, bei (3) welchen Personen (z.B. Studierende), in (4) welchen Situationen und auf (5) welchem Niveau beobachtet werden müssen, um auf ein dahinterliegendes Merkmal schließen zu können (Mager, 1962; Rost, 2004). Das zu messende Kriterium muss so beschrieben sein, dass Unklarheiten und Interpretationsmöglichkeiten ausgeschlossen sind (Mager, 1962). Wieberg (1983) unterscheidet zwischen zwei Verfahren zur Definition eines Aufgabenuniversums. Mit umgangssprachlich orientierten Verfahren wird versucht, Lehrziele, die auf einem allgemeinen

Niveau formuliert sind, begrifflich zu präzisieren. Außerdem werden Auswahlvorschriften angegeben, nach welchen Regeln eine Aufgabenstichprobe aus dem definierten Aufgabenuniversum gezogen wird (Ingenkamp & Lissmann, 2007). Auf formalsprachlich orientierte Verfahren wird in dieser Arbeit nicht näher eingegangen, da sie aufgrund ihrer Komplexität nur von Testexperten eingesetzt werden können (ebd.) und sich nicht für eine praktische Anwendung durch Lehrende eignen (Wieberg, 1983). Die folgenden Ausführungen beziehen sich daher auf die Variante der umgangssprachlichen Festlegung von Lehrzielen. Anhand eines Beispiels wird nachfolgend dargestellt, wie die Lehrziele empirisch analysiert werden. Das Beispiel stammt aus dem Qualifikationsrahmen für Deutsche Hochschulabschlüsse: [die Studierenden sind in der Lage;] *sich mit Fachvertretern und mit Laien über Informationen, Ideen, Probleme und Lösungen auf wissenschaftlichem Niveau auszutauschen* (Kultusministerkonferenz, 2005, S. 5).

In dieser Arbeit werden die folgenden Aspekte eines Lehrziels als Inhaltskomponenten bezeichnet: „Informationen“, „Ideen“, „Probleme“ und „Lösungen“. Die Inhaltskomponenten sind auf einem abstrakten Niveau formuliert und müssen fachspezifisch ausdifferenziert werden, indem für die Überbegriffe „Ideen“, „Probleme“ und „Lösungen“ jeweils entsprechende fachbezogene Begrifflichkeiten eingesetzt werden. Das Verb „austauschen“ definiert, was am Inhalt gekonnt werden soll und spezifiziert damit die Verhaltenskomponente in diesem Beispiel. Die Personengruppe sind Master-Studierende deutscher Hochschulen. Die Situationskomponenten sind beschrieben durch die Angabe der Personengruppen der Fachvertreter und Laien, mit denen sich die Studierenden über die genannten Aspekte austauschen können sollen. Das Niveau wird angegeben über die Festlegung, dass der Austausch über die enthaltenen Inhaltskomponenten auf wissenschaftlichem Niveau erfolgen soll.

Lehrziele müssen so beschrieben sein, dass sie in Aufgaben operationalisierbar sind. Die Operationalisierbarkeit ist dann gegeben, wenn „eine Kette ausübbarer Tätigkeiten (Meßvorschrift) angegeben ist, die eine Erfassung der Variablen zulässt“ (Herbig, 1976, S. 15). Die Beschreibung von Lehrzielen erfolgt auf unterschiedlichen Ebenen mit unterschiedlichem Auflösungsgrad. Unterschieden wird zwischen einem groben Auflösungsgrad auf einer übergeordneten gesellschaftlichen Ebene (z.B. allgemeine Erziehungsziele oder die Abdeckung eines anerkannten Wissenskanons der sich bspw. orientiert an wissenschaftlich generierten Fachsystematiken), einem mittlerer Auflösungsgrad zumeist auf institutioneller Ebene (z.B. Studiengangprofile) und einer feinen Auflösung auf Ebene von Lehrveranstaltungen (z.B. Modul- beziehungsweise Veranstaltungsbeschreibung) (Klauer & Leutner, 2012).

Der inhaltliche Rahmen im HQR beispielsweise ist bewusst breit gesetzt, damit den Universitäten Raum für eine eigene Profilbildung bleibt, wenn sie die Lehrziele des Qualifikationsrahmens auf

ihre Studiengänge anwenden (Kultusministerkonferenz, 2010). Die Lehrziele sind daher auf einem allgemeinen Niveau und domänenunabhängig formuliert. Für eine zielgerichtete Aufgabenerstellung fehlen noch Informationen darüber, an welchen konkreten Inhalten das Verhalten ausgeführt werden soll. Lehrziele müssen derart konkretisiert werden, dass die Übersetzung der Theorieelemente (Inhalte, Verhalten, Situationen, Niveau) in Anforderungssituationen nicht willkürlich erfolgt, sondern nachvollziehbar ist, um eine Einschätzung der Gültigkeit der vorgenommenen Übersetzungsprozesse ermöglichen. Dieser Anspruch ist nicht trivial, da die Übersetzung der Lehrziele in konkrete Erfassungssituationen typischerweise durch einen mehrschrittigen Übersetzungsvorgang erfolgt, bei dem Beeinträchtigungen der Wiedergabetreue auftreten können (Leuders, 2014). Der benötigte Auflösungsgrad beziehungsweise die Spezifität der Lehrziele können, abhängig von der Komplexität des festzulegenden Konstrukts, höher und damit spezifischer, oder niedriger und damit unspezifischer, sein. In eindeutigen Fällen, bei denen unter Fachleuten klar ist, was gemeint ist, kann eine kurze Definition ausreichen, in anderen Fällen müssen die Komponenten sehr ausführlich angegeben werden, bis deren Bedeutung unmissverständlich ist (Klauer, 1974). Gegebenenfalls sind weitere Situationsmerkmale, wie zum Beispiel die Verfügbarkeit von Hilfsmitteln anzugeben (Mager, 1962).

Hinsichtlich der Auflösungsgrade von Lehrzielen wird eine, in dieser Arbeit eigens entwickelte, Abstufung bezüglich der Spezifität von Lehrzielen verwendet, anhand derer Lehrziele von unspezifisch bis sehr spezifisch eingeteilt werden:

- **Lehrziel unspezifisch:** Unspezifische Lehrziele sind nach der hier gewählten Kategorisierung nicht aussagekräftig genug, da sie mehrdeutig bleiben oder mindestens eine der beschriebenen Komponenten (Inhalt oder Verhalten) nicht enthalten.
- **Lehrziel eher spezifisch:** Die als eher unspezifisch bewerteten Lehrziele sind auf Modulebene informativ, jedoch für die Aufgabenerstellung noch nicht zwingend aussagekräftig, da die Begrifflichkeiten nicht ausreichend konkret sind, um daraus Hinweise für die Aufgabengestaltung ablesen zu können.
- **Lehrziel spezifisch:** Bei spezifischen Lehrzielen werden die zu behandelnden Inhalte und ausführbaren Verhaltensweisen konkret benannt, anstatt übergeordnete Bezeichnungen zu verwenden.
- **Lehrziel sehr spezifisch:** Sehr spezifische Lehrziele grenzen die Reichweite der gewählten Begrifflichkeiten weiter ein, indem einschränkende Angaben zu Inhalten und Verhaltensweisen gegeben werden.

Als Ausgangspunkt für eine kriteriengeleitete Aufgabenerstellung eignen sich, je nach Komplexität des zu messenden Konstrukts, die als spezifisch und sehr spezifisch eingestuften Lehrziele.

Unspezifische Lehrziele sind nicht aussagekräftig genug für die Erstellung von Aufgaben und Bewertungsraster, da sie mehrdeutig bleiben oder entweder Inhalt oder Verhalten nicht beschreiben. Die als eher unspezifisch bewerteten Lehrziele sind auf Modulebene informativ, jedoch für die Aufgabenerstellung noch nicht zwingend aussagekräftig. Bei der Darstellung der Ergebnisse werden Lehrziele, die als *spezifisch* oder *sehr spezifisch* bewertet wurden mit der Kategorie „Lehrziele auf Prüfungsebene“ dargestellt, da sie sich aufgrund ihrer Spezifität für die Ableitung von Aufgaben und Bewertungsrastern eignen. Sowohl Inhalts- als auch Verhaltenskomponente werden dabei als mindestens spezifisch eingestuft. Als *eher spezifisch* bewertete Lehrziele sind auf Modulebene aussagekräftig und werden daher in den Ergebnissen als „Lehrziele auf Modulebene“ berichtet. Eine umfassende Operationalisierung der vier Abstufungen ist im Kodiermanual in Tabelle 35 angegeben.

Basierend auf den Ausführungen zur Festlegung von Teilkompetenzen und Bearbeitungsniveaus und der Formulierung von Lehrzielen soll in dieser Arbeit geklärt werden, inwieweit Lehrende übergeordnete Kompetenzen in Teilkompetenzen zerlegen, ob Bearbeitungsniveaus von den Lehrenden differenziert werden, und welche Annahmen die Lehrenden dabei zugrunde legen. Diese Aspekte werden aus den Verbaldaten gewonnen. Die Lehrziele werden hinsichtlich deren Spezifität analysiert, um einschätzen zu können, ob Inhalte und Verhaltensaspekte konkret und unmissverständlich formuliert sind, sodass sie in Aufgaben übersetzt werden können.

Überführung in eine Lehrzielmatrix

In einem nächsten Schritt sollten die Lehrenden die Lehrziele in eine Lehrzielmatrix überführen, um die Übersichtlichkeit über die Lehrziele und die Teilaspekte, aus denen sich die Lehrziele zusammensetzen, zu erhöhen (Klauer & Leutner, 2007). Auch die Festlegung, welche Anzahl an Aufgaben beziehungsweise welche Punktezahl zu jedem Lehrziel in der Prüfung angelegt werden, kann damit übersichtlich und nachvollziehbar angegeben werden. Dadurch können die Lehrenden eine Gewichtung der Lehrziele vorgenommen werden (Fricke, 1974). Die Lehrzielmatrix ist als ein Hilfsmittel zur Erstellung von Prüfungen anzusehen, dass sich in der Prüfungs- und Testerstellung über Jahre bewährt hat und sich gerade auch für den praktischen Einsatz durch Lehrende eignet (Wieberg, 1983).

Folgende Schritte sind dabei von den Lehrenden durchzuführen. In den Zeilen der Matrix werden die Inhalts- und/oder Situationskomponenten der Lehrziele aufgeführt, in den Spalten die Verhaltenskomponenten. Jede der so entstehenden Zellen der Matrix, in denen sich Inhalts/Situations- und Verhaltenskomponenten kreuzen, kennzeichnet ein Aufgabenuniversum. Zur Zusammenstellung der Aufgaben für eine Prüfung ziehen die Lehrenden, nach einer festzulegenden Auswahlvorschrift, eine zufällige Stichprobe an Aufgaben (Ingenkamp & Lissmann, 2007). In der Matrix

muss nicht jede Zelle besetzt sein, da es sein kann, dass nicht jeder Verhaltensaspekt an allen Inhalten ausgeführt werden muss.

Tabelle 2⁵: Beispiel für eine Lehrzielmatrix

Inhalts- und/oder Situationskomponenten	Verhaltenskomponenten		
	Verhaltensaspekt 1	Verhaltensaspekt 2	Verhaltensaspekt 3
Inhalt 1 und/oder Situation 1	Lehrziel 1 (Angabe der Anzahl an Aufgaben oder den zu vergebenden Punkten für Lehrziel 1)		
Inhalt 2 und/oder Situation 2		Lehrziel 2 (Angabe der Anzahl an Aufgaben oder den zu vergebenden Punkten für Lehrziel 2)	
Inhalt 3 und/oder Situation 3			Lehrziel 3 (Angabe der Anzahl an Aufgaben oder den zu vergebenden Punkten für Lehrziel 3)
Aufgaben die keiner Zelle zugeordnet werden können			

Die Lehrzielmatrix kann ergänzt werden durch eine Lehrzieltaxonomie, in der die Zeilen und Spalten schon vorgeben sind (z.B. Anderson & Krathwohl, 2001). Im Zuge der Kompetenzorientierung wäre es außerdem denkbar, eine dritte Dimension (neben Inhalten und Verhalten) zu ergänzen, in der die entsprechenden Kontexte angegeben werden. In dieser Arbeit soll geklärt werden, ob dieses Hilfsmittel unter Hochschullehrenden bekannt ist und entsprechend genutzt wird. In einer Lehrzielmatrix werden die Lehrziele getrennt in zwei Dimensionen, einer Inhaltsdimension und einer Verhaltensdimension. In einer dritten Dimension können bei Bedarf Situationen beziehungsweise Kontexte mitangegeben werden. Den Lehrenden wurden drei Varianten vorgestellt. In der einfachsten Variante (siehe Tabelle 2) wurde lediglich eine leere Matrix vorgegeben, die Überschriften enthielt, um aufzuzeigen, dass die Verhaltenskomponenten der Lehrziele in den Spalten, die Inhaltskomponenten in den Zeilen angetragen werden sollen. In der zweiten Vorlage waren im Stile einer Lehrzieltaxonomie die kognitiven Niveaus nach Bloom als Verhaltenskomponenten vorgegeben, um den Lehrenden eine Orientierung zur Differenzierung von Anforderungsniveaus anzubieten (siehe Tabelle 3). Die Inhaltskomponenten wurden von den Lehrenden anhand der Lehrziele selbstständig ergänzt.

⁵ Aus Gründen der besseren Lesbarkeit werden die Trennlinien zwischen Zeilen und Spalten in der Tabelle eingeblendet. Quelle: Klauer & Leutner (2012)

Tabelle 3: Auszug aus der Beispielmatrix mit kognitiven Anforderungsniveaus⁶

Inhaltskomponenten der Lehrziele	Kognitive Anforderungsniveaus					
	Stufe 1 Erinnern	Stufe 2 Verstehen	Stufe 3 Anwenden	Stufe 4 Analysieren	Stufe 5 Bewerten	Stufe 6 Entwickeln
Inhalt 1	Lehrziel 1		Lehrziel 5		Lehrziel 7	
Inhalt 2	Lehrziel 2	Lehrziel 4		Lehrziel 6		
Inhalt 3	Lehrziel 3					Lehrziel 8

Anmerkung. Nicht alle Spalten müssen mit Lehrzielen belegt sein, da es sein kann, dass an einem Inhalt nicht alle Verhaltenskomponenten ausgeführt werden müssen.

In einer erweiterten Form wurden zusätzlich zu den kognitiven Anforderungsniveau die in dieser Arbeit verwendeten Wissensarten (deklaratives, prozedurales, konzeptionelles und strategisches Wissen) in den Zeilen der Matrix eingesetzt, anstatt konkreter Inhalte. Die Lehrenden erhielten zudem eine Beschreibung der Bedeutung der Wissensarten. Auch diese Vorgabe war als Hilfestellung gedacht, um die Lehrenden bei der Differenzierung von Anforderungen zu unterstützen.

2.1.3.2 Erstellung von Prüfungsaufgaben

Ausgehend von der Festlegung des Prüfungsgegenstands entwickeln die Lehrenden im nächsten Schritt ihre Prüfungsaufgaben. Die Funktion von Prüfungsaufgaben ist es, Studierende in Situationen zu bringen, die sie unter Rückgriff auf ihre Kompetenzen beziehungsweise ihre Handlungspotenziale bewältigen müssen (Li et al., 2006). In Prüfungsaufgaben sollen die durch die Lehrziele beschriebenen Anforderungssituationen möglichst gut beziehungsweise übersetzungsgetreu abgebildet werden. Die folgenden Ausführungen beziehen sich auf Aufgabenformate, wie sie in schriftlichen Prüfungen umgesetzt werden können, da die empirischen Analysen dieser Arbeit sich auf dieses Prüfungsformat beziehen.

Prüfungsaufgaben bestehen aus zwei Komponenten, dem *Aufgabenstamm* und dem *Antwortformat* (Rost, 2004). Der Aufgabenstamm stellt ganz allgemein die Anforderungssituation dar, zum Beispiel in Form einer Frage, einer Aussage, eines Bildes, einer Geschichte, Zeichnung oder Rechenaufgabe und enthält die zur Bearbeitung benötigten Informationen. Die Anforderungssituationen müssen so gestaltet sein, dass Studierende zum Abrufen des relevanten Verhaltens angeregt und angeleitet werden (ebd.). Bei der Erstellung von Aufgaben ist von den Lehrenden sicherzustellen, dass sowohl die notwendigen kognitiven Aspekten als auch die relevanten Inhalte und Wissensbestände berücksichtigt werden, um die inhaltliche und kognitive Validität der Aufgaben zu ge-

⁶ Quelle: Bloom et al. (1956)

währleisten. Dabei müssen die Lehrenden besonders darauf achten, dass die Lösung der Aufgaben nicht durch konstruktferne Fähigkeiten möglich ist.

Das Antwortformat dient zur Registrierung des Testverhaltens (Rost, 2004). Das Antwortformat muss es den Studierenden ermöglichen, das erwartete Verhalten zu zeigen (z.B. in Form einer anzufertigenden Zeichnung, einer Berechnung oder Beschreibung eines Sachverhalts). Soll beispielsweise eine Antwort konstruiert werden, darf kein Wahlantwortformat gewählt werden, bei dem lediglich eine richtige Auswahl aus bereits vorgegeben Antwortmöglichkeiten getroffen werden muss. Grundlegend wird bei Tests zwischen offenen (oder freien) und geschlossenen (gebundenen) Antwortformaten unterschieden (National Assessment Governing Board, 2007).

In dieser Arbeit werden Prüfungsaufgaben auf Basis der bisherigen Ausführungen hinsichtlich mehrerer Merkmale untersucht. Die Aufgabenanforderungen werden entlang der Ausführungen in Abschnitt 2.1.1.3 hinsichtlich der in den Aufgaben abgebildeten Wissensarten, der Aufgabenoffenheit, der Nutzung von Kontextbezügen, sowie der sprachlogischen Komplexität der Aufgabenstellung analysiert, um die in den Aufgaben umgesetzte Kompetenzorientierung einzuschätzen. Weiterhin wird die Qualität von Aufgaben hinsichtlich des Lehrzielbezugs, als entscheidende Voraussetzung für die Validität, bewertet. Um die Passung zwischen einer Aufgabe und einem Lehrziel einzuschätzen, wird in dieser Arbeit die Übereinstimmung zwischen den Wissensarten in Lehrzielen und Prüfungsaufgaben abgeglichen. Außerdem werden aus dem Vorgehen der Lehrenden bei der Erstellung von Aufgaben weitere Hinweise auf eine Abgleichung von Aufgaben und Lehrzielen gewonnen. Betrachtet werden soll dabei, ob und inwieweit Lehrende Lehrziele bei Entwicklung von Aufgaben zugrunde legen. Neben inhaltlichen Gesichtspunkten, wie dem Lehrzielbezug und der Differenziertheit von Anforderungen in Aufgaben, müssen zudem auch formelle Aspekte in Aufgaben berücksichtigt werden, die bei Nichtbeachtung als Störvariablen wirken und die Messung verfälschen. Diesbezüglich werden Gestaltungsmerkmale herangezogen, die als wesentlich angesehen werden, um die Aufgabenqualität auf formaler Ebene sicher zu stellen (Bücker et al., 2015; Gronlund, 1974; National Assessment Governing Board, 2007). Die empirischen Analysen beziehen sich konkret auf die Angabe eines Erwartungshorizontes in Aufgaben, die Unterteilung von Handlungsaufforderungen in Teilaufgaben (anstelle einer Aneinanderreihung mehrerer Handlungsaufforderungen in einer Aufgabe) und bestehende Abhängigkeiten zwischen Aufgaben (sind Folgeaufgaben unabhängig voneinander bearbeitbar). Eine differenzierte Beschreibung dieser Aspekte wird in Tabelle 37 gegeben.

2.1.3.3 Rückbindung der Aufgaben an den Prüfungsgegenstand

Bei der Aufgabenentwicklung treffen die Lehrenden verschiedene Annahmen darüber, welches Wissen durch die Aufgabe angesprochen wird oder welche Kompetenzen zur Bearbeitung benö-

tigt werden. Diese Annahmen können fehlerhaft sein (Brown & Wilson, 2011). Im Nachgang an die Aufgabenerstellung muss von den Lehrenden daher überprüft werden, ob die getroffenen Annahmen über die Anforderungen einer Aufgabe zutreffend sind, um die inhaltliche, kognitive und externe Validität von Rückschlüssen, die auf Grundlage von Prüfungsergebnissen getroffen werden, zu sichern. Dazu existieren verschiedene Techniken, wie etwa das Durchführen einer rationalen (Li et al., 2006; Shavelson et al., 2003) und kognitiven (Ayala, Yin, Shavelson, & Vanides, 2002) Aufgabenanalyse. Als Bezugspunkt für derartige Analysen dienen die vorab definierten Lehrziele. Im Rahmen einer rationalen Aufgabenanalyse sollen die Lehrenden einschätzen, welche Prozesse bei der Aufgabenbearbeitung bei einem Studierenden ablaufen, der aufgrund seiner Personenmerkmale potenziell dazu in der Lage ist, die Aufgabe zu lösen und welches Wissen, welche Fähigkeiten oder Kompetenzen zum Tragen kommen (Li et al., 2006; Shavelson et al., 2003). Außerdem sollen die Lehrenden überprüfen, ob eine Aufgabe auch durch die Anwendung konstruktferner Merkmale lösbar ist, wodurch das Potenzial einer Aufgabe unterlaufen werden könnte (siehe Abschnitt 2.1.2.1.). Aspekte, die von den Lehrenden dazu herangezogen werden können, sind die Inhaltsaspekte in Lehrziel und Aufgabe (stimmen Inhaltsbereich in Lehrziel und Aufgabe überein), die abzurufenden Wissensarten (deklarativ, prozedural, konzeptionell, strategisch), die kognitiven Niveaus (z.B. Reproduktion vs. Verständnis oder Anwendung), die Aufgabenoffenheit, die Funktion von Kontextbezügen und die Komplexität der Aufgabe.

In der vorliegenden Arbeit wird der Frage nachgegangen, ob Lehrende die Anforderungen in Aufgaben mit den in Lehrzielen beschriebenen Anforderungen abgleichen, welche Kriterien für den Abgleich von Lehrzielen und Prüfungsaufgaben herangezogen werden und ob die dabei getroffenen Einschätzungen von einer Kollegin, oder einem Kollegen, überprüft werden. Letztlich wird betrachtet, ob die Lehrenden auf Ebene der Prüfung absichern, dass zu jedem Lehrziel anteilig die entsprechende Anzahl an Aufgaben beziehungsweise die Anzahl an Punkten in die Prüfung aufgenommen wurde (American Educational Research Association et al., 2014).

2.1.3.4 Bewertung von Prüfungsantworten

Bei der Korrektur können sowohl Einschränkungen im Hinblick auf die Validität von Prüfungsergebnissen entstehen, als auch bei der Reliabilität. Die Validität kann bei der Korrektur beeinträchtigt werden, wenn bei der Bewertung einer Antwort (beziehungsweise des Testverhaltens) nicht die für das zu messende Konstrukt relevanten Aspekte betrachtet werden. Bei der Korrektur müssen die Lehrenden sicherstellen, dass in den Antworten die Aspekte bewertet werden, die valide Rückschlüsse auf das Erreichen eines Lehrziels zulassen. Die Reliabilität kann durch eine inkonsistente Bewertung von Antworten beeinträchtigt werden, insofern beispielsweise zwei Personen die gleiche Antwort unterschiedlich bewerten (interindividuelle Unterschiede). Unterschiede kön-

nen auch bei einer korrigierenden Person (intraindividuelle Unterschiede) auftreten, bedingt durch den Zeitpunkt der Korrektur oder aufgrund anderer konstruktferner Merkmale (z.B. Schrift). Untersuchungen an Schulen in den Fächern Englisch, Mathematik und Geschichte, konnten hohe Beurteilungsunterschiede zwischen Lehrkräften, die die gleichen Schülerarbeiten bewerteten, feststellen. Die größten Diskrepanzen in der Bewertung ergaben sich entgegen der Erwartungen für das Fach Mathematik (einen Überblick über diese Ergebnisse geben Ingenkamp & Lissmann, 2007). Andere Studien hingegen zeigen, dass Beurteilungsfehler durch bestimmte Maßnahmen, wie die Entwicklung und Nutzung von Bewertungsrastern oder durch die Schulung von Korrektoren, deutlich verringert werden können (Gorman & Rentsch, 2009; Graham, Milanowski, & Miller, 2012). Im Folgenden werden Maßnahmen beschrieben, mit denen die Qualität der Korrektur von Prüfungen erhöht werden kann (siehe auch Kapitel 2.1.2.3).

Eine wesentliche Grundlage, um Beurteilungsfehler bei der Korrektur zu unterbinden, ist die Verwendung eines Bewertungsrasters, welches in der Regel parallel zur Erstellung der Prüfungsaufgaben entwickelt wird. Ein vollständiges Bewertungsraster enthält Bewertungskategorien, die die vorab definierten Kompetenzbeschreibungen abbilden (Wilson, 2005). Die Bewertungskategorien sind durch ein Regelwerk definiert, sodass eine eindeutige Zuordnung von Prüfungsantworten ermöglicht wird. Die Angabe von Ankerbeispielen, sowie Angaben zur Punktevergabe, tragen zu einer weiteren Erhöhung der Objektivität, Reliabilität und Validität der Korrektur und Auswertung bei (Wilson, 2008). Mit einem Bewertungsraster kann die Wahrscheinlichkeit erhöht werden, dass in den Prüfungsantworten die für das zu überprüfende Konstrukt relevanten Aspekte identifiziert und bewertet werden (Wilson, 2005). Außerdem kann durch die Verwendung von Bewertungsrastern die Konsistenz und damit die Reliabilität der Bewertungen erhöht werden (Gorman & Rentsch, 2009).

Ein Bewertungsraster beinhaltet typischerweise Bewertungskategorien, wie zum Beispiel richtig oder falsch und eine Beschreibung beziehungsweise Charakterisierung dieser Kategorien in Form von Regelwerken sowie Ankerbeispiele. Außerdem muss angegeben werden, welche Scores (beziehungsweise Leistungspunkte) für jede Kategorie vergeben werden. Die Festlegung von Bewertungskategorien ergibt sich aus den Lehrzielen und der Art der gestellten Anforderungen. Die Lehrenden müssen die Aspekte in den Antworten bewerten, die valide Rückschlüsse auf das Erreichen eines Lehrziels zulassen. Eine weitverbreitete Form ist die Unterscheidung zwischen „richtig beantwortet“ oder „falsch beantwortet“. Sollen Antworten bewertet werden, die teilweise gelöst wurden, können eine oder auch mehrere Kategorien (z.B. „teilweise richtig“) ergänzt werden. Die Bewertungskategorien können jedoch auch Kompetenzniveaus beziehungsweise Stufen abbilden (Ingenkamp & Lissmann, 2007). In Tabelle 4 ist ein Bewertungsraster exemplarisch dargestellt.

Tabelle 4: Exemplarisches Bewertungsraster

optional⁷ Code	Bewertungskategorien	Charakterisierung der Kategorien	Ankerbeispiele	optional Punktevergabe
C2	Richtig	Als richtig werden Antworten gewertet, die....		1 Punkte
C1	teilweise richtig	Als teilweise richtig werden Antworten gewertet, die....		0,5 Punkte
C0	Falsch	Als falsch werden Antworten gewertet, die....		0 Punkte

Wie in Tabelle 4 angeführt, müssen die gewählten Bewertungskategorien beschrieben und charakterisiert werden. Die Beschreibungen beinhalten Eigenschaften, Charakteristika, Indikatoren, Bewertungs- und Entscheidungsregeln, die eine zweifelsfreie Zuordnung unterschiedlicher, aber qualitativ gleichwertiger, Antworten ermöglichen. Grundsätzlich kann zwischen zwei Arten von Beschreibungen unterschieden werden: (1) eine Aufzählung aller möglichen richtigen Antworten oder (2) eine generische Beschreibung von Eigenschaften, die eine Antwort erfüllen muss, um einer entsprechenden Kategorie zugeordnet zu werden (Klauer, 1974). Zu jeder Bewertungskategorie werden Ankerbeispiele angegeben, um die Eindeutigkeit bei der Zuordnung zu erhöhen. Neben der Musterlösung, als Beispiel für eine eindeutig richtige Antwort, sollten Grenzfälle angeführt werden, die gerade noch einer Kategorie zugeordnet werden können (Ingenkamp & Lissmann, 2007). Die Ankerbeispiele können konstruiert oder aus bereits bearbeiteten Prüfungen entnommen werden.

Neben der Verwendung von Bewertungsrastern zeigte sich, dass durch eine Schulung von Lehrenden, die Prüfungen korrigieren, die Treffsicherheit und Konsistenz von Korrekturen verbessert werden kann. (Gorman & Rentsch, 2009). In den meisten Fällen sind mehrere Schulungstermine vorzusehen, um Unstimmigkeiten in den Bewertungen zu klären und die Lehrenden schrittweise an eine treffsichere Beurteilung heranzuführen (ebd.). Die Lehrenden sollten außerdem ein fundiertes Fachwissen hinsichtlich des Bewertungsgegenstands besitzen (ebd.).

In dieser Arbeit werden die Prüfungen dahingehend analysiert, ob für die Aufgaben vollständige Bewertungsraster existieren, die Ankerbeispiele, Bewertungskategorien und Regelwerke sowie

⁷ Das Beispiel enthält zusätzlich die Möglichkeit, den Bewertungskategorien einen Code zuzuweisen, der eine mögliche statistische Auswertung der Prüfungsergebnisse erleichtert. Diese Codes können bei der Korrektur auch anstelle einer ausführlichen Beschreibung der Bewertung verwendet werden. Die Anzahl der Punkte kann ebenfalls in das Bewertungsraster aufgenommen werden, um die Zuweisung von Punkten bei der Korrektur zu erleichtern. Der Code und die Punktevergabe sind unabhängig voneinander und müssen sich nicht entsprechen.

Angaben zur Punktevergabe enthalten. Die Lehrenden werden außerdem dazu befragt, ob Schulungen mit den Personen durchgeführt werden, die die Prüfungen korrigieren.

2.1.3.5 Punkte- und Notenvergabe

Mit Prüfungen werden den Studierenden bestimmte Leistungen und der Erwerb von Kompetenzen bescheinigt. Das erreichte Leistungsniveau wird über Noten ausgedrückt. Zur Bestimmung der Note wird klassischerweise über alle Prüfungsaufgaben hinweg ein Summenscore gebildet (Von Davier, 2012). Durch die Vergabe von Punkten zu einzelnen Aufgaben wird eine inhaltliche Gewichtung vorgenommen. Die Gewichtung kann entweder theoretisch oder empirisch erfolgen. Erfolgt die Vergabe von Punkten unreflektiert, zum Beispiel durch die Vergabe eines Punktes pro Häkchen, das bei der Korrektur gesetzt wird, können periphere Anforderungen ungewollt einen größeren Beitrag zur Bildung der Endnote leisten, als solche Anforderungen, die inhaltlich bedeutender sind, bei der Korrektur jedoch weniger Häkchen enthalten. Die Endnote lässt in dem Fall keine zuverlässige und valide Aussage über die zu messende Kompetenz zu.

Die Punktevergabe muss inhaltlich begründet und an der Struktur und den Anforderungsniveaus des zu messenden Konstrukts ausgerichtet werden. Theoretische Überlegungen zur Vergabe von Punkten zu einzelnen Aufgaben können Aspekte, wie die Relevanz der in der jeweiligen Aufgabe enthaltenen Anforderung oder die erwartete Schwierigkeit der Aufgabe, mit einbeziehen, und werden bereits bei der Definition der zu prüfenden Kompetenzen festgelegt. Bildet die Prüfung mehrere Kompetenzen ab, stellt sich zudem die Frage, inwieweit es sinnvoll beziehungsweise zulässig ist, eine Gesamtscore über verschiedene Kompetenzen zu bilden, um die Endnote zu bestimmen (Von Davier, 2012). Eine Alternative ist, sofern laut Prüfungsrecht der Hochschule zulässig, einzelne inhaltliche Bestandteile der Prüfung gesondert voneinander zu bewerten beziehungsweise zu bepunkten, um die jeweiligen Blöcke zueinander zu gewichten. Dies eröffnet unter anderem die Möglichkeit eröffnen, Prüfungen nur dann bestehen zu lassen, wenn bei jeder Teilkompetenz ein Mindestniveau erreicht wird.

Basierend auf den erreichten Punkten werden dem erzielten Testresultat Noten zugeordnet. Diese Zuordnung erfolgt unter Zuhilfenahme eines Vergleichsmaßstabs (Ingenkamp & Lissmann, 2007). Ein häufig verwendeter Maßstab zur Vergabe von Noten ist eine bezugsgruppenorientierte Norm. Dabei wird der Testrohwert einer einzelnen Person in Bezug gesetzt zu den Werten einer Bezugsgruppe (ebd.). Die erhaltene Note ist ein Maß für die Positionierung einer Person innerhalb der Bezugsgruppe. Eine Einschätzung, ob und in welchem Ausmaß eine Kompetenz ausgeprägt ist, kann dabei jedoch nicht getroffen werden, da sich die Notenskala nicht am Kriterium (der Kompetenz mit ihren Kompetenzniveaus) orientiert, sondern am Abschneiden der Bezugsgruppe (ebd.).

Mit einer kriteriumsorientierten Interpretation der Testrohwerte, werden Noten gemäß dem festgesetzten Kriterium vergeben (Klauer & Leutner, 2012). Dabei wird der Testrohwert (die Anzahl der erzielten Punkte) in Bezug zur Kompetenzdefinition interpretiert (ebd.). Noten können bei entsprechender Zuweisung zur Kompetenzdefinition demnach ausdrücken, welches Kompetenzniveau erreicht wurde. Der Notenschlüssel, also die Einteilung von Punktebereichen, denen eine Notenstufe zugeordnet ist (z.B. Note 1,0 bei einer erreichten Punktzahl von 92 Prozent bis 100 Prozent der Gesamtpunktzahl), muss an den Kompetenzniveaus ausgerichtet sein (Wilson, 2005). Festgesetzte Notenschlüssel können eine Orientierung liefern, sind aber nicht zwingend geeignet, um Kompetenzniveaus adäquat abzubilden (Ingenkamp & Lissmann, 2007).

Erfüllt die Prüfung die Funktion Standards zu sichern, liefert eine kriteriumsorientierte Interpretation von Testrohwerten die benötigten Informationen, da die Note widerspiegelt, inwieweit eine Person die in einem Studiengang vorgesehenen Kompetenzen und Kompetenzniveaus erreichen konnte. Werden die Noten entsprechend einer bezugsgruppenorientierten Rangfolge vergeben, drückt die Note aus, auf welcher Position ein Studierender im Vergleich zu anderen Studierenden steht. Eine derart gebildete Einzelnote ist von der Leistungsstärke der Bezugsgruppe beeinflusst. Die Note ist in dem Fall kein Maß für das erreichte Niveau hinsichtlich des vordefinierten Kriteriums (Klauer & Leutner, 2012).

Bezüglich der Punktevergabe gilt es in der vorliegenden Arbeit zu überprüfen, nach welchen Kriterien Hochschullehrende Punkte auf einzelne Aufgaben vergeben, ob eine Ausrichtung der Punktevergabe hinsichtlich der abzuprüfenden Wissens- und Könnensaspekte vorgenommen wird und wie konsistent die Vergabe von Punkten erfolgt. Hinsichtlich der Notenvergabe ist zunächst zu klären, welcher Vergleichsmaßstab angelegt wird. In einer kompetenzorientierten Prüfungspraxis muss der Bewertungsmaßstab am Kriterium ausgerichtet sein, da nur so festgestellt werden kann, inwieweit die Kompetenzen ausgeprägt sind (Klauer & Leutner, 2012).

2.1.3.6 Rückschlüsse aus den Prüfungsdaten

Mit einer Prüfung gewinnen die Lehrenden Daten, die über den Zweck der Notenbildung hinaus wertvolle Informationen liefern. Das Abschneiden von Studierenden bei einzelnen Aufgaben kann bezüglich mehrerer Dimensionen interpretiert werden. Zum einen können die Lehrenden Informationen zur Qualität beziehungsweise Bearbeitbarkeit der eingesetzten Aufgaben gewinnen. Zum anderen können sie die Passung der abgeprüften Anforderungen und damit das Leistungspotenzial der geprüften Personen und Personengruppen einschätzen. Bereits anhand einfach zu erstellender statistischer Kennwerte wie z.B. der Aufgabenschwierigkeit, Standardabweichungen bei einzelnen Aufgaben, den bei einer Aufgaben erzielten Minimal oder Maximalpunktzahlen, sowie der Trennschärfe, können vielfältige Rückschlüsse gezogen werden. Da bislang nicht bekannt ist, inwieweit

Lehrende entsprechende Betrachtungen vornehmen, soll in dieser Arbeit grundlegend geklärt werden, ob die Prüfungsdaten über die Notenvergabe hinaus auch hinsichtlich anderer Aspekte interpretiert werden. Nachfolgend werden dazu mögliche Überlegungen skizziert, die auf Grundlage von Prüfungsdaten und den genannten statistischen Kennzahlen angestellt werden könnten.

Wird bei einer Aufgabe beispielsweise von keinem Studierenden die maximal erreichbare Punktzahl erzielt, könnte auf Aufgabenebene möglicherweise ein Problem in der Aufgabenstellung dafür verantwortlich sein (z.B. das der Erwartungshorizont unklar angegeben). Auf Ebene der Anforderungen wäre es möglich, dass der abgeprüfte Aspekt, bei dem kein Studierender die volle Punktzahl erreichte, nicht vermittelt beziehungsweise von den Studierenden nicht gelernt wurde. Dabei könnte eine Rolle spielen, dass die Studierenden diesen nicht als relevant wahrgenommen haben. Hinsichtlich der Aufgabenschwierigkeiten sind solche Aufgaben näher zu betrachten, bei denen die theoretisch erwartete Schwierigkeit nicht mit der empirischen Schwierigkeit übereinstimmt. Sind bestimmte Aufgaben viel einfacher oder schwieriger als erwartet, können hierfür ebenfalls formale Merkmale in der Aufgabenstellung dafür verantwortlich sein. Auf inhaltlicher Ebene können unerwartete Aufgabenschwierigkeiten ein Hinweis darauf sein, dass das Anforderungsniveau zu hoch oder zu niedrig eingeschätzt wurde. Möglicherweise wurde das Anforderungsniveau für die Zielgruppe auch falsch angesetzt. Die Standardabweichung sagt aus, ob Testwerte eher nahe dem Mittelwert sind oder ob die Werte stark streuen. Besteht bei einer Aufgabe auch unter den Leistungsstärksten Studierenden eine große Streuung, könnte dies ein Hinweis sein, dass zur der Lösung der Aufgabe konstruktferne Kompetenzen notwendig sind. Die Trennschärfe einer Aufgabe gibt an, wie gut diese zwischen Personen mit einer hohen und einer niedrigen Merkmalsausprägung trennt. Aufgaben mit niedriger Trennschärfe können möglicherweise auch aus einem Test entfernt werden, da damit nur eingeschränkt aufgeklärt werden kann, welche Personen das fragliche Personenmerkmal besitzen und welche nicht (Lienert & Raatz, 1998).

Damit die Lehrenden entsprechende statistische Auswertungen vornehmen können, müssen die Prüfungsergebnisse aufbereitet werden, so dass für jeden Studierenden in einem Datensatz festgehalten wird, wie viele Punkte pro Aufgabe erreicht wurden (siehe Tabelle 5). In dieser Arbeit soll aufgeklärt werden, ob die Daten zu den Prüfungen durch die Lehrenden derart aufbereitet werden, dass sie statistisch auswertbar sind. Außerdem wird betrachtet, inwieweit die Lehrenden Rückschlüsse aus den Prüfungsdaten ziehen, um Informationen zum Leistungsvermögen der Studierenden oder Hinweise auf die Qualität der eingesetzten Prüfungsaufgaben zu erhalten.

Tabelle 5: Beispiel für einen Datensatz zur Erfassung von Prüfungsergebnissen

	Aufgabe 1	Aufgabe 2	Aufgabe 3	Aufgabe 4
Studierender 1	3 Punkte	2 Punkte	4 Punkte	1 Punkt
Studierender 2	1 Punkt	3 Punkte	5 Punkte	2 Punkte
Studierender 3	3 Punkte	1 Punkt	4 Punkte	3 Punkte
.				
.				
.				
Studierender 100	1 Punkt	3 Punkte	5 Punkte	2 Punkte

Zwischenfazit

Die Ausführungen in diesem Abschnitt dienen dazu ein Vorgehensmodell zu entwickeln, das den Prozess der Erstellung und Auswertung einer Prüfung abbildet und in sechs Schritte unterteilt. Dabei wurden die zuvor erarbeiteten Kriterien für kompetenzorientiertes und an Gütekriterien ausgerichtetes Prüfen diesen sechs Schritten zugeordnet und soweit konkretisiert, dass deren Umsetzung durch Hochschullehrende bei der Erstellung und Auswertungen von Prüfungen direkt beobachtbar ist. Dadurch konnten Indikatoren erarbeitet werden, anhand derer die Prüfungspraxis von Hochschullehrenden, sowohl auf Prozessebene durch die Auswertung von Verbaldaten, als auch auf Ebene der Materialien (Lehrziele, Prüfungen und Aufgaben, charakterisiert werden kann. In Tabelle 6 sind die Indikatoren zur Charakterisierung der Prüfungspraxis zusammenfassend dargestellt und den Schritten zur Prüfungserstellung zugeordnet. Außerdem wird angegeben, welche Datenquelle für die Analysen jeweils zugrunde gelegt werden.

Tabelle 6⁸: Datenquellen zugeordnet zu den sechs Schritten der Prüfungserstellung und -durchführung

Schritte	Teilschritte	Leitfragen als Indikatoren	Datenquelle
Festlegung des Prüfungsgegenstands	Auswahl der abzuprüfenden Inhalte und Kompetenzen	Welche Inhalte und Kompetenzen werden als wichtig erachtet(z.B. Wissensorientierung, Fokus auf prozedurale Fähigkeiten, Kompetenzorientierung oder Praxisorientierung)?	Bestehende und neu erstellte Lehrziele Verbaldaten
	Konzeptionelle Analyse des Prüfungsgegenstands	Welche Referenzpunkte und Unterlagen werden bei der Auswahl von Kompetenzen, Anforderungssituationen und Wissens- und Könnensaspekten zugrunde gelegt?	Verbaldaten
	Strukturierung und Festlegung von Niveaus	Werden Kompetenzniveaus und -strukturen differenziert?	Verbaldaten
	Formulierung von Lehrzielen	Welche Wissensarten sind in Lehrzielen abgebildet? Welche Spezifität haben Lehrziele?	Inhaltsanalyse der Lehrziele und Modulbeschreibungen.
	Übertragung in eine Lehrzielmatrix	Werden die Lehrziele in eine Lehrzielmatrix übertragen?	Verbaldaten und ggf. erstellte Matrizen
Erstellung von Prüfungsaufgaben	Ausgangspunkt für die Aufgabenerstellung	Werden Lehrziele zur Aufgabenentwicklung herangezogen bzw. zugrunde gelegt? Werden für bestimmte Lehrziele gezielt Aufgaben entwickelt? Woher stammen Aufgaben bzw. Ideen für Aufgaben.	Verbaldaten
	Passung zwischen Lehrzielen und Prüfungsaufgaben	Stimmen die Wissensarten, die in Lehrzielen und Aufgaben angelegt sind, überein?	Inhaltsanalyse von Lehrzielen und Prüfungsaufgaben
	Formale Qualitätsindikatoren von Aufgaben	Sprachliche Merkmale: Sind Handlungsaufforderungen als Aussage oder Frage formuliert? Gibt es doppelte Verneinungen im Aufgabenstamm? Wird eine klare eindeutige Sprache verwendet? Instruktionsqualität: Sind Erwartungshorizonte klar beschrieben	Inhaltsanalyse Prüfung
	Wie lange ist die Bearbeitungsdauer pro Aufgabe?		
	Welche Darstellungsformen werden verwendet?		
	Kompetenzorientierung in Aufgaben	Welche Wissensarten sind in Aufgaben angelegt?	
Welche Aufgabenoffenheiten werden verwendet?			
Wird die sprachlogische Komplexität in Aufgaben variiert?			
Werden Kontext- und Situationsbezüge in Aufgaben hergestellt?			

⁸ Aus Gründen der besseren Lesbarkeit werden die Trennlinien zwischen Zeilen und Spalten in der Tabelle einblendet.

Rückbindung der Anforderungen an den Prüfungsgegenstand	Rationale Aufgabenanalyse	Findet ein nachvollziehbarer (explizierter) Abgleich zwischen den in den Lehrzielen formulierten Anforderungen und den in Aufgaben abgebildeten Anforderungen statt? Wird dieser Abgleich von einer zweiten Person überprüft? Werden die Ergebnisse festgehalten?	Verbaldaten und neu erstellte Lehrziel-taxonomien mit Aufgabenzuordnung
	Kognitive Validierung von Aufgaben	Werden Laut-Denken Protokolle von Studierenden beim Lösen der Aufgaben angefertigt, um Denkprozesse zu erfassen, die beim Lösen der Aufgabe zum Tragen kommen.	
Bewertung von Prüfungsantworten	Bewertungsraster	Sind Ankerbeispiele zu den Prüfungsaufgaben vorhanden?	Auswertung in den Prüfungen
		Sind jeweils Regelwerke für jede Bewertungskategorie vorhanden?. Werden Charakteristika beschrieben, die eine Antwort aufweisen muss, um einer Bewertungskategorie zugeordnet zu werden?	
	Werden Angaben zur Zuweisung von Punkten zu den Antworten bzw. Teilaspekten einer Antwort gemacht?		
	Korrektorenschulung	Wird eine Einweisung bzw. Schulung der Korrektoren vorgenommen? Welche Informationen werden gegeben?	Verbaldaten
Punkte- und Notenvergabe	Punktevergabe	Nach welchen Gesichtspunkten werden Punkte vergeben, z.B. benötigte Bearbeitungsdauer, Schwierigkeit der Aufgabe, Relevanz der Anforderung. Was hat sich bewährt?	Verbaldaten
		Besteht ein Bewusstsein für die Gewichtung, die durch die Punktevergabe vorgenommen wird (evtl. Überbetonung peripherer Anforderungen)?	
	Notenvergabe	Welche Überlegungen spielen bei der Festlegung des Notenschlüssels eine Rolle (kriteriums- vs. normorientierte Messung)?	
Rückschlüsse aus den Prüfungsdaten	Aufbereitung von Daten	In welcher Form werden Daten aufbereitet?	Verbaldaten
	Nutzung von Daten für Rückschlüsse	Welche Kennwerte werden betrachtet (Itemschwierigkeit, Minima-Maxima, Trennschärfen, etc.)?	
		Welche Rückschlüsse werden von den Lehrenden aus Daten gezogen?	

2.1.4 Kurzzusammenfassung

In diesem Kapitel wurde der analytische Rahmen für die Arbeit entwickelt, mit dem Ziel Indikatoren zu erarbeiten, anhand derer die Prüfungspraxis von Hochschullehrenden im Hinblick auf die Umsetzung der geforderten Kompetenzorientierung und der erreichten Prüfungsqualität charakterisiert werden kann. Den Ausgangspunkt bildete eine begriffliche Klärung des Kompetenzbegriffs. Kompetenzen wurden definiert als *die Fähigkeit in realen Situationen erfolgreich zu handeln*. Außerdem wurde differenziert zwischen Kompetenztests (Observation von Verhalten in

realen Situationen), kompetenzorientierten Tests (Observation von Verhalten in realitätsnahen Situationen) und kompetenzorientierten Wissenstests (Erfassung von Voraussetzungen für kompetentes Handeln). Weiterhin konnte geklärt werden, dass mit dem erarbeiteten Kompetenzbegriff und der Unterscheidung der drei Testvarianten die Qualifikationsstufen des Qualifikationsrahmens für deutsche Hochschulabschlüsse vollständig abgedeckt werden. Daher erscheint eine Anwendung des eingeführten Kompetenzkonzepts für die Anwendung auf den Hochschulbereich als geeignet. Da schriftliche Prüfungen an Hochschulen erfahrungsgemäß die höchste Reichweite aufweisen, wurden Indikatoren erarbeitet, anhand derer die Umsetzung einer Kompetenzorientierung in schriftlichen Hochschulprüfungen eingeschätzt werden kann. Die Prüfungsaufgaben werden hinsichtlich der angelegten Wissensarten, ihrer Offenheit, der Funktion von Kontextbezügen und der sprachlogischen Komplexität, analysiert.

Entlang der Gütekriterien der pädagogischen Diagnostik wurden Kriterien und Maßnahmen identifiziert und beschrieben, deren Berücksichtigung wesentlich ist, um belastbare Prüfungsergebnisse zu erhalten, auf deren Grundlage gültige Rückschlüsse auf den Lern- und Leistungsstand von Studierenden gezogen werden können. Aufbauend auf den Ausführungen zu Gütekriterien und der Erfassung von Kompetenzen wurde ein Vorgehensmodell entwickelt, das den Prozess der Erstellung und Auswertung einer Prüfung in sechs Schritte strukturiert. Die erarbeiteten Indikatoren für Kompetenzorientierung und Prüfungsqualität wurde den sechs Schritten zugeordnet und soweit konkretisiert, dass sie empirisch erfassbar sind. In Tabelle 6 sind diese Indikatoren aufgeführt.

Das erarbeitete Vorgehensmodell erfüllt zwei Funktionen. Es dient einerseits als analytischer Rahmen für die empirischen Erhebungen. Entlang des Vorgehensmodells wird die bestehende Prüfungspraxis analysiert, um Schwachstellen und Probleme in der Prüfungspraxis von Hochschullehrenden zu identifizieren. Zum anderen dienen die Erkenntnisse zur bestehenden Praxis als Ausgangspunkt für die Gestaltung des geplanten Qualitätsentwicklungsprogramms, im Rahmen dessen die Prüfungspraxis von Hochschullehrenden entwickelt werden soll. Das Programm ist ebenfalls entlang der sechs Schritte des Vorgehensmodells angelegt. Durch die vorhergehende Analyse der Prüfungspraxis kann das Qualitätsprogramm auf die in der bestehenden Prüfungspraxis identifizierenden Herausforderungen und Probleme zugeschnitten werden. Um bereits existierende Forschungsergebnisse zur Prüfungspraxis von Hochschullehrenden zu berücksichtigen, wird im nächsten Abschnitt der aktuelle Forschungsstand zur Prüfungspraxis von Hochschullehrenden betrachtet, um Erkenntnisse zu den sechs Schritten zur Erstellung und Auswertung von Prüfungen zu erhalten.

2.2 Erkenntnisse zur Prüfungspraxis von Hochschullehrenden

In diesem Kapitel wird der nationale Forschungsstand zur Prüfungspraxis von Hochschullehrenden betrachtet. Das entwickelte analytische Rahmenmodell mit den sechs Schritten der Prüfungserstellung und –auswertung bildet dafür die Basis. Der Fokus soll auf Studien liegen, die sich auf der Handlungsebene von Lehrenden mit deren Prüfungspraxis auseinandersetzen oder die Qualität von Hochschulprüfungen untersuchen. Das Ziel bei der Betrachtung des Forschungsstands ist es, Erkenntnisse zur Umsetzung einer kompetenzorientierten Prüfungspraxis und der erreichten Prüfungsqualität zu erhalten. Zentrale Fragen dabei sind, ob es Lehrenden gelingt, eine Kompetenzorientierung in Prüfungen umzusetzen und dabei testtheoretische Gütekriterien einzuhalten. Außerdem wird betrachtet, welche Probleme und Defizite diesbezüglich in Prüfungen derzeit bestehen. Für den Klärungsfortschritt der vorliegenden Arbeit bilden derartige Erkenntnisse die Grundlage, um weiterführende Forschungsfragen sowie Hinweise für die Gestaltung des geplanten Qualitätsentwicklungsprogramms zu erarbeiten. Im Folgenden wird der Fokus zunächst auf Studien zur Prüfungspraxis an deutschen Hochschulen gelegt. Anschließend werden ausgewählte Ergebnisse einer Befragung zur Prüfungspraxis unter Lehrenden und Studierenden berichtet, die als Vorstudie zu dieser Arbeit an der Technischen Universität München durchgeführt wurde.

2.2.1 Forschungsergebnisse zur Prüfungspraxis an deutschen Hochschulen

Im deutschen Hochschulkontext findet kaum Forschung zur Prüfungspraxis von Hochschullehrenden statt (Dany, Szczyrba, & Wildt, 2008). Als Ausgangspunkt der Literaturrecherche wurde daher eine systematische Suche in drei einschlägigen Datenbanken (FIS-Bildung, Deutscher Bildungsserver, ERIC) durchgeführt. Das Ziel war es, Zeitschriftenartikel zu identifizieren, die sich empirisch mit der Prüfungspraxis und der Qualität von Prüfungen, die von Hochschullehrenden erstellt wurden, befassen. Es wurden sowohl deutsche als auch englischsprachige Schlagwörter für die Suche verwendet. Die deutschen Schlagwörter lauten: (1) Prüfungspraxis, (2) Prüfungserstellung, (3) Prüfungsanforderungen, (4) Prüfungsaufgabe, (5) Aufgabengestaltung, (6) kompetenzorientiertes Prüfen + Prüfungspraxis. Die englischen Schlagwörter lauten: (1) Assessment practice, (2) Exam practice, (3) Teacher-made tests, (3) Classroom assessment, (4) Classroom-level assessment, (5) Exam requirements und (6) Assessment literacy. Für jedes Schlagwort wurde jeweils eine eigene Suche in den Datenbanken vorgenommen. Wenn möglich wurden für die Suche Filter gesetzt, um die Treffer auf Zeitschriftenartikel und den Bereich Hochschule oder Higher Education einzugrenzen. Als Ergebnis dieser Suche konnten zunächst 41 Artikel identifiziert werden. Diese Artikel wurden nachfolgend gesichtet, um Studien auszusortieren, die sich

nicht auf die Prüfungspraxis und die Qualität von Prüfungen an deutschen Hochschulen bezogen. Letztlich wurden drei Artikel näher betrachtet, die den genannten Suchkriterien entsprachen. Bei einem der Artikel (Schulz et al., 2014, in dieser Arbeit bereits in Abschnitt 2.1.1.5 zitiert), in dem eine Typisierung von Prüfungsanforderungen anhand der Anforderungsdimensionen im HQR vorgenommen wird. In einer weiteren Studie berichtet Kordts-Freudinger (2013) vom Versuch, eine Skala zur Erfassung der Einstellungen von Lehrenden hinsichtlich des kompetenzorientierten Prüfens zu entwickeln. Stefanica (2013) gliedert die in Modulbeschreibungen definierten Anforderungen mit den Inhalten und Anforderungen der zugehörigen Prüfungen ab. Die wenigen Studien unterstreichen nochmals das Forschungsdefizit hinsichtlich der Prüfungspraxis von Hochschullehrenden (Dany et al., 2008).

Studie zur Typisierung von Prüfungsanforderungen

Schulz et al. (2014) gehen in ihrer Untersuchung unter anderem der Frage nach, welche Anforderungen in Prüfungen an der Technischen Universität München abgebildet sind und wie sich diese Anforderungen mit den Qualifikationszielen des HQR decken. Sie betrachten außerdem, welche Lernstrategien bei den Studierenden zur Prüfungsvorbereitung zum Tragen kommen und wie hoch die Prüfungsbelastung eingeschätzt wird. An dieser Stelle wird vorwiegend auf die vorgenommene Typisierung von Prüfungsanforderungen eingegangen. Als Datengrundlage dient eine Online-Umfrage unter Studierenden (N=4615) zur Prüfungspraxis an der Technischen Universität München. Die Studierenden schätzten die Prüfungsanforderungen der für sie wichtigsten Prüfung im vorangegangenen Semester ein, indem sie anhand einer fünfstufigen Skala angaben, zu welchen Anteilen die Bereiche „Wissen wiedergeben“, Sachverhalte in eigenen Worten erklären“, „Wissen auf Beispiele und Problemstellungen anwenden“ und „Sachverhalte bewerten“ in der Prüfung vorkamen. Diese vier Anforderungsbereiche werden als Operationalisierung der Kategorien des HQR „Wissen und Verstehen“ und „Können“ betrachtet. Mittels hierarchischer Clusteranalysen wurden die Prüfungsanforderungen typisiert. Es konnten dabei drei Cluster gebildet werden. Prüfungen in Cluster 1 weisen laut Einschätzung der Studierenden einen hohen Anwendungsbezug mit geringen Anteilen von Wissenswiedergabe auf (Verhältnis 75:25). Prüfungen dieser Art werden als „anwendungsorientiert“ bezeichnet. In Cluster zwei treten die Anforderungen *Wissen*, *Verstehen* und *Anwenden* zu etwa gleichen Anteilen auf. Prüfungen in diesem Cluster werden deshalb als Prüfungen mit *gemischten Anforderungen* bezeichnet. Die Anforderungen in Prüfungen von Cluster 3 werden als *wiedergabeorientiert* deklariert, da der Anteil der Wissenswiedergabe relativ hoch ist, bei geringen Anteilen von Anwendungsorientierung (das Verhältnis zwischen Wiedergabe- und Anwendungsorientierung liegt bei 75:25). Fast die Hälfte der Studierenden (49,2 %) gab an, dass in der für sie wichtigsten Prüfung des vorangegangenen Semesters gemischte Anforderungen enthalten waren. 37 Prozent der Studierenden werteten die Anforderung der

Prüfung als vorwiegend anwendungsorientiert. Am wenigsten begegneten den Studierenden wieder-gabeorientierte Prüfungen, mit einem Anteil von 13,8 Prozent.

Besonders der hohe Anteil an Prüfungen mit gemischten Anforderungen wird von den Autoren als positiv gewertet, da die verschiedenen Anforderungsdimensionen *Wissen, Verstehen* und *Anwenden* aufeinander aufbauen und die Anforderungsdimensionen des HQR angemessen abbilden. Im Anschluss untersuchen die Autoren, wie die Studierenden der verschiedenen Cluster die jeweiligen Prüfungen in Bezug auf andere Variablen, wie etwa Belastung, Aufwand und Schwierigkeit, wahrnehmen und wie der Erfolg der entsprechenden Formate in Bezug auf Durchfallquote und Noten zu bewerten ist. Für die anwendungsorientierten Prüfungen zeigte sich, dass diese von den Studierenden als belastender wahrgenommen werden. Die Studierenden berichten zudem von einem höheren Vorbereitungsaufwand. Außerdem schätzten die Studierenden dieses Prüfungstyps die Schwierigkeit höher ein als die Studierenden anderer Prüfungstypen. Entsprechend ist die Durchfallquote höher und auch die Prüfungsnoten fallen schlechter aus als bei den anderen Gruppen. Als mögliche Ursachen werden Aspekte der Aufgabengestaltung genannt (z.B. ob die Aufgaben systematisch auf Vorwissen aufbauen), eine möglicherweise fehlende Transparenz bezüglich der Prüfungsanforderungen und eine eventuell nicht ausreichende Vorbereitung auf anwendungsorientierte Aufgaben in der Lehrveranstaltung.

Die Ergebnisse der Studie liefern einen ersten Eindruck von Prüfungsanforderungen an der Technischen Universität München. Inwieweit sich die Ergebnisse auch an anderen Hochschulen beobachten lassen, müsste geklärt werden. Die Ergebnisse deuten jedoch darauf hin, dass es den Lehrenden in gewissem Maße gelingt, die im HQR geforderten Qualifikationsziele derart in Prüfungen umzusetzen, sodass auch Studierende diese als differenziert wahrnehmen. Die Fokussierung auf die Wahrnehmung der Studierenden ist gleichzeitig aber auch eine Einschränkung der Studie. Außerdem ist die gewählte Operationalisierung der Anforderungsbereiche des HQR, mit je einer Aussage pro Kategorie, wenig ausdifferenziert. Wie auch die Autoren anmerken, könnte eine systematische inhaltsanalytische Betrachtung von Hochschulprüfungen, mit einer differenzierten Operationalisierung der Anforderungsdimensionen des HQR, vertiefende Einblicke liefern.

Entwicklung eines Inventars zur Erfassung von Einstellungen von Hochschullehrenden hinsichtlich des kompetenzorientierten Prüfens

In der Studie von Kordts-Freudinger (2013) wird ein Instrument zur Erfassung der Einstellung von Hochschullehrenden hinsichtlich des kompetenzorientierten Prüfens vorgestellt. Der Fragebogen enthält 12 Aussagen zum Thema kompetenzorientiertes Prüfen. Die Lehrenden wurden gebeten, die Items zweimal zu bearbeiten: einmal mit der Aufforderung, die persönliche Zustim-

mung hinsichtlich der im Fragebogen enthaltenen Aussagen zu bewerten; das zweite Mal mit der Bitte die Relevanz der Aussagen für das Konstrukt kompetenzorientiertes Prüfen einzuschätzen. Durch das zweimalige Bearbeiten mit unterschiedlichen Aufforderungen konnten zwei Skalen gebildet werden. Die Skala im ersten Bearbeitungsschritt bildete die *Zustimmung zum kompetenzorientierten Prüfen* ab, die Skala im Rahmen der zweiten Bearbeitung die *Relevanz für kompetenzorientiertes Prüfen*. Die beiden Skalen umfassen 12 identische Items. 17 Lehrende süddeutscher Universitäten füllten den Fragebogen im Rahmen einer Weiterbildung zum kompetenzorientierten Prüfen aus. Die erzielten Reliabilitätswerte, die über die interne Konsistenz der Skalen Auskunft geben, fielen eher niedrig aus. Für die Skala „Zustimmung zu kompetenzorientiertem Prüfen“ betrug Cronbachs $\alpha = .4$, nach Bereinigung um drei Items, die von den Lehrenden als wenig relevant eingestuft wurden, ergab sich ein Wert von Cronbachs $\alpha = .45$. Für die Skala „Relevanz für kompetenzorientiertes Prüfen“ betrug Cronbachs- $\alpha = .35$. Zur Validierung der Skalen füllten in einer Online-Umfrage $N = 32$ Lehrende den Fragebogen zur Erfassung der Einstellung zum kompetenzorientierten Prüfen und eine deutsche Übersetzung des „Approaches to Teaching Inventory“ (ATI) aus, mit dem die Einstellungen der Lehrenden zur Lehre erfasst werden. Bei dieser Befragung konnte für die Skala „Zustimmung zu kompetenzorientiertem Prüfen“ ein Cronbachs- $\alpha = .062$ erzielt werden. Mit $r(31) = .7$, $p < .001$ ergab sich ein starker Zusammenhang zwischen den Skalen „Zustimmung zu kompetenzorientierten Prüfen“ und der Skala „Studierendenfokussierung“ des ATI. Zur Skala „Lehrendenzentrierung“ des ATI konnten keine signifikanten Zusammenhänge nachgewiesen werden. Dieses Ergebnis wird als Hinweis auf die Validität des Fragebogens gewertet. Als Fazit wird festgehalten, dass der Fragebogen einer Überarbeitung bedarf, um die Reliabilität zu erhöhen. Außerdem müssen die Items nochmals kritisch auf ihre inhaltliche Passung zur Erfassung des Konstrukts „Einstellungen zum kompetenzorientierten Prüfen“ geprüft werden. Der Artikel bietet einen Einblick zur Pilotierung eines Instruments. Aussagekräftige Erkenntnisse zur Einstellung von Hochschullehrenden hinsichtlich des kompetenzorientierten Prüfens sind nicht enthalten.

Passung zwischen inhaltlichen Anforderungen in Modulbeschreibungen und Prüfungen

In einer Studie zu den inhaltlichen Anforderungen in Modulbeschreibungen glich Stefanica (2013) die in den Lehrzielen von Modulbeschreibungen formulierten Anforderungen mit den Prüfungsanforderungen der entsprechenden Modulprüfungen ab. In der Studie wurden Modulbeschreibungen und zugehörige Prüfungen von Maschinenbaustudiengängen einer Universität und zwei Hochschulen für die Fächer Mathematik I und II analysiert. Die Übereinstimmung zwischen Lehrzielen und Prüfungsaufgaben wurde gesondert für die Inhalte und die Anforderungen, die in Aufgaben und Lehrzielen abgebildet sind, bestimmt. Die Ergebnisse werden in den nächsten zwei Absätzen dargestellt.

Inhaltlich konnten, insgesamt betrachtet, weitgehende Übereinstimmungen zwischen Lehrzielen und Prüfungsaufgaben beobachtet werden. Bei einigen Themenbereichen wurde beobachtet, dass in den Lehrzielen Inhalte genannt sind, die nicht durch Aufgaben abgedeckt werden. Umgekehrt wurden Prüfungsaufgaben zu Themenbereichen eingesetzt, die in den Lehrzielen nicht aufgeführt sind. An einem der untersuchten Standorte war ein Themenbereich mit nur einer Aufgabe in der Prüfung stark unterrepräsentiert. Ein anderer Themenbereich, der in den Lehrzielen nicht genannt wird, ist durch vier Aufgaben in der Prüfung vertreten. Bei anderen Themenbereichen werden Klausuraufgaben zu komplexen Zahlen gestellt, die jedoch in den Lehrzielen nicht vorgesehen sind. Zu anderen Inhaltsbereichen, die in den Lehrzielen benannt sind, wie Extremwerte, Kurvendiskussion, Richtungsableitungen von Flächen oder die Rotation und Divergenz von Vektorfeldern, existieren keine Aufgaben. An einer der beiden Hochschulen sind die Anforderungen in den Lehrzielen und den Klausuraufgaben (N=115) annähernd deckungsgleich. In der Klausur wurde lediglich ein mathematisches Verfahren verlangt, das in den Lehrzielen nicht genannt ist. Am anderen Hochschulstandort waren für vier thematische Einheiten, die in den Lehrzielen benannt wurden, keine Prüfungsaufgaben enthalten. Eine Aufgabe bezog sich auf einen Themenbereich, der nicht in den Lehrzielen abgebildet war.

Zur Analyse der Prüfungsaufgaben wurde ein Kategoriensystem entwickelt, das sieben Kategorien (Aufgabenkontext, Methode, Prozessbezogenheit, Anspruchsniveau, Abstraktionsgrad, Anzahl der Lösungsschritte, Bekanntheit und Lösungsform) zur Einschätzung der Anforderungsstruktur in den Aufgaben enthält. Die Kategorien und deren Operationalisierung sind im Artikel ausführlich beschrieben. An dieser Stelle wird auf die erhaltenen Ergebnisse eingegangen. Am Universitätsstandort wird in den Lehrzielen der Anspruch formuliert, dass die Studierenden die behandelten Methoden selbstständig, sicher und kreativ anwenden sollen. In den Aufgaben ist dieser Anspruch jedoch nur ansatzweise abgebildet. Insbesondere der kreative und kritische Umgang mit den Methoden sind stark unterrepräsentiert. An der Hochschule 1 wird der Anspruch formuliert, ein Verständnis mathematischen Modellierens und die Fähigkeit zur Lösung ingenieurwissenschaftlicher Problemstellungen auszubilden. Beide Ziele werden nur ansatzweise durch Prüfungsaufgaben abgedeckt. Gleiches gilt für die Hochschule 2 hinsichtlich des Ziels, die Studierenden zur Bearbeitung ingenieurmäßiger Problemstellungen zu befähigen. Eine bessere Übereinstimmung liegt für das Lehrziel vor, die Studierenden mit der Anwendung mathematischer Methoden vertraut zu machen. Insgesamt kann festgehalten werden, dass es für Lehrziele mit einem höheren Anspruchsniveau nicht ausreichend gelingt, entsprechende Aufgaben zu erstellen. Während inhaltlich eine weitgehend zufriedenstellende Übereinstimmung zwischen Lehrzielen und Prüfungsaufgaben erreicht wird, sind die Anforderungsniveaus in den Lehrzielen oftmals deutlich höher angesetzt, als die in Prüfungsaufgaben realisierten Anspruchsniveaus.

Zwischenfazit

Auf nationaler Ebene existieren nur wenige Erkenntnisse zur Prüfungspraxis von Hochschullehrenden. Anhand einer Datenbanksuche mit verschiedenen Schlagwörtern konnten drei empirische Zeitschriftenartikel zur Prüfungspraxis von Lehrenden an Hochschulen identifiziert werden. Kordts-Freudinger (2013) stellt einen Versuch vor, ein Instrumentarium zur Erfassung der Einstellungen von Lehrenden hinsichtlich des kompetenzorientierten Prüfens zu entwickeln. Aufgrund niedriger Reliabilitätswerte des Instruments können keine belastbaren Ergebnisse zu den festgestellten Einstellungen berichtet werden. Schulz et al. (2014) typisierten Prüfungsanforderungen in schriftlichen Prüfungen an einer Hochschule anhand der retrospektiven Einschätzungen von Studierenden. Ihre Ergebnisse deuten darauf hin, dass die Prüfungsanforderungen von den Studierenden als heterogen wahrgenommen werden und den Anforderungsdimensionen des HQR entsprechen. Stefanica (2013) analysierte Modulbeschreibungen und Prüfungsaufgaben, um die Übereinstimmung der Anforderungen in Lehrzielen und Aufgaben abzugleichen. Dabei konnte festgestellt werden, dass es gerade für Lehrziele mit höheren Anspruchsniveaus nicht umfassend gelingt, diese entsprechend in Prüfungsaufgaben zu übersetzen. Auch die inhaltliche Übereinstimmung zwischen Lehrzielen und Prüfungsaufgaben gelingt laut den Ergebnissen von Stefanica nicht für alle Themenbereiche. Aufgrund der festgestellten Diskrepanzen zwischen Lehrzielen und Aufgaben stellt sich die Frage, inwieweit die Lehrenden sich bei der Erstellung von Aufgaben an den Lehrzielen orientieren und diese nutzen, um die Prüfungsanforderungen mit den in den Lehrzielen formulierten Anforderungen abzugleichen.

Die Ergebnisse der Studien von Schulz et al. (2014) und Stefanica (2013) liefern einen ersten Eindruck von Prüfungsanforderungen und bestehenden Diskrepanzen zwischen Lehrzielen und Prüfungsaufgaben. Weitere Veröffentlichungen zur Prüfungspraxis an deutschen Hochschulen konnten nicht ausfindig gemacht werden. Zu diesem Themenbereich liegt demnach ein großes Forschungsdefizit vor. Diese Arbeit nimmt sich diesem Forschungsdefizit an, indem sowohl Materialien wie Lehrziele, Prüfungen und Prüfungsaufgaben analysiert werden, als auch Prozessdaten, die Aufschluss darüber geben, welche Überlegungen und Annahmen bei der Gestaltung von Prüfungen handlungsleitend sind. Aufbauend auf den dabei erzeugten Erkenntnissen soll zudem erforscht werden, wie Lehrende dazu angeleitet werden können die Qualität von Prüfungen zu entwickeln.

2.2.2 Erkenntnisse zur Prüfungspraxis an der Technischen Universität München

Aufgrund der eingeschränkten Befundlage zur Prüfungspraxis an deutschen Hochschulen (siehe vorheriger Abschnitt), wurde an der Technischen Universität München versucht, die Prüfungspraxis von Lehrenden und Studierenden mittels zwei parallel angelegter Online Umfragen zu explorieren. An dieser Stelle werden ausgewählte Ergebnisse berichtet, um für diese Arbeit eine Ausgangsbasis zur Konkretisierung von Forschungsfragen und die Gestaltung des geplanten Qualitätsentwicklungsprogramms zu erhalten. Teile der nachfolgend dargestellten Ergebnisse wurden in zwei Projektberichten veröffentlicht, die auf der Webpräsenz der Technischen Universität München abrufbar sind (Prenzel, Schindler, et al., 2012; Prenzel, Schulz, & Schindler, 2012). Das Vorgehen bei der Befragung ist in den Projektberichten ausführlich dargestellt. Für die im vorherigen Abschnitt thematisierte Untersuchung von Schulz et al. (2014) wurde ebenfalls auf Daten aus der Online Umfrage zurückgegriffen.

An der Befragung beteiligten sich insgesamt 412 Dozierende, darunter 146 Professorinnen und Professoren, 199 wissenschaftliche Mitarbeiterinnen und Mitarbeiter sowie 68 Lehrbeauftragte. An der parallel durchgeführten Studierendenbefragung nahmen 4615 Studierende teil. Die Befragung der Lehrenden bezog sich auf die Anzahl der zu betreuenden Prüfungen pro Semester, die verwendeten Prüfungsformate, die Anzahl der Prüfungsfälle sowie auf das Vorgehen bei der Erstellung von Prüfungen, den Prüfungsanforderungen und mögliche Unterstützungsbedarfe bei der Erstellung und Korrektur von Prüfungen. Die erhaltenen Ergebnisse ermöglichen erste Einblicke in die Prüfungspraxis und deuten auf mögliche Herausforderungen und Probleme von Lehrenden bei der Erstellung, Durchführung und Auswertung von Prüfungen hin.

Im Schnitt hielten die Lehrenden an der TUM 3,1 Prüfungen im Semester ab, wobei Professorinnen und Professoren mit im Schnitt 3,9 Prüfungen das höchste Prüfungsaufkommen zu bewältigen haben. Dies ist vermutlich durch unterschiedlich hohe Lehrdeputate bedingt, die bei Professorinnen und Professoren im Regelfall höher liegen, als bei wissenschaftlichen Mitarbeiterinnen und Mitarbeitern. Das Prüfungsaufkommen unterschied sich zwischen den Fakultäten, was auf grundsätzliche Unterschiede zwischen den Disziplinen hindeuten könnte. Die Lehrenden an der TUM hatten bei ihren Prüfungen mit im Schnitt 200 Studierenden pro Prüfung, relativ hohe Studierendenzahlen zu bewältigen. Die hohe Anzahl an Prüfungsfällen muss bei der Auswahl des Prüfungsformates und der Aufgabengestaltung entsprechend berücksichtigt werden, um die Korrigierbarkeit zu gewährleisten.

Die größte Reichweite erzielten schriftliche Prüfungsformate. 90,2 Prozent der Studierenden legten im betreffenden Semester schriftliche Prüfungen ab. Als Aufgabentypen kamen am häufigsten offene Antwortformate zum Einsatz. Multiple-Choice Aufgaben spielen nur an den Fakultäten für Medizin (96,7 % aller Prüfungen enthielten ausschließlich Multiple-Choice Aufgaben) und der Fakultät für Sport- und Gesundheitswissenschaften mit einem Anteil von 48,9 Prozent an Multiple Choice Prüfungen. Als wichtigste Prüfungsanforderung wurde die Anwendung von Wissen zur Lösung von Problemstellungen genannt. Lehrende orientierten sich bei der Prüfungserstellung typischerweise an stark vorstrukturierten Medien wie Foliensätzen, Vorlesungsskripten und Lehrbüchern. Es fällt auf, dass die Studierenden bei der Prüfungsvorbereitung überwiegend die von den Dozierenden zugrunde gelegten Materialien wie Vorlesungsskripte (21,5 %), Foliensätze (15,5 %), eigene Mitschriften (10,4 %) und, obwohl von Dozierenden kaum genannt, insbesondere Altklausuren (31,6 %) als am hilfreichsten empfanden und entsprechend häufig nutzten. Die in der Veranstaltung angegebene Begleitliteratur und selbst recherchierte Literatur nannten lediglich 2,4 Prozent und 1,7 Prozent aller befragten Studierenden als relevante Quellen für ihre Prüfungsvorbereitung. Die Studierenden richteten ihr Lernverhalten offensichtlich sehr eng an der vorgegebenen Lehrsystematik aus. Dieses Ergebnis verdeutlicht daher die Notwendigkeit, Prüfungen als Steuerungsinstrument für das Lernverhalten zu betrachten und entsprechend zu gestalten. Aufgrund der einseitigen Fokussierung des Lernens auf Medien, in denen die Inhalte stark vorstrukturiert aufbereitet und präsentiert werden, ist zu hinterfragen, ob eine derartige Lernstrategie geeignet ist, um sich auf Prüfungen mit hohen Anwendungsbezügen vorzubereiten. Auch die starke Nutzung von Altklausuren deutet auf Prüfungsanforderungen hin, die weniger die Lösung neuartiger Problemstellungen erfordern, sondern das Lösen bekannter Aufgabentypen, die mit gleichen oder ähnlichen Prozeduren lösbar sind, wie in zuvor eingeübten Aufgaben aus vorangegangenen Klausuren. Die Lösungen zu wissensorientierten Fragen sind hingegen häufig in Skripten, Foliensätzen oder Lehrbüchern zu finden (Wollersheim, März, & Schminder, 2011).

Der Zeitaufwand, den Lehrende für die Prüfungserstellung aufwandten, fiel eher gering aus. Beinahe die Hälfte der Befragten investierte 1 bis 5 Stunden in die Erstellung einer Prüfung. Bei diesem geringen zeitlichen Aufwand für die Prüfungserstellung ist es fraglich, ob es gelingt, relevante Gestaltungsaspekte bei der Erstellung einer Prüfung einzuhalten, die eine hohe Qualität gewährleisten. Etwa 70 Prozent der Lehrenden begannen erst zum Semesterende oder kurz vor dem Prüfungstermin mit der Erstellung der Prüfung. Dies könnte ein Hinweis darauf sein, dass die Erstellung von Prüfungen als eigener Schritt betrachtet wird. Inwieweit die Prüfungen systematisch auf die Lehrziele und die Lehrveranstaltung zugeschnitten werden, gilt es daher in der vorliegenden Arbeit zu klären. Dazu soll zum einen anhand der Verbaldaten geklärt werden, ob die

Lehrenden Lehrziele bei der Prüfungsgestaltung verwenden. Zum anderen wird überprüft, ob die in Lehrzielen und Prüfungsaufgaben abgebildeten Anforderungen übereinstimmen.

Ein großer Teil der Lehrenden lehnte die Möglichkeit einer externen Unterstützung bei der Prüfungserstellung ab. Unter den Professorinnen und Professoren war die Ablehnung am höchsten, während über ein Viertel der wissenschaftlichen Mitarbeiterinnen und Mitarbeiter eine entsprechende Unterstützung als wünschenswert betrachteten. Einen Leitfaden zur Prüfungserstellung wünschten sich hingegen deutlich mehr Lehrende. Dieses Ergebnis könnte so interpretiert werden, dass die Lehrenden grundsätzlich Verbesserungsbedarfe bezüglich ihrer Prüfungen wahrnehmen und einen Leitfaden als hilfreich empfinden, um Schwachstellen zu beseitigen. Die Bereitschaft, Dritten Einblick in die Prüfungspraxis zu geben, scheint bei vielen Lehrenden hingegen nicht besonders ausgeprägt.

Zwischenfazit

Basierend auf den dargestellten Umfrageergebnissen konnten Hinweise auf Probleme in der Prüfungspraxis von Hochschullehrenden identifiziert werden. Das Lernverhalten der Studierenden, mit einer starken Orientierung an Altklausuren und Medien, in denen Inhalte strukturiert aufbereitet sind, deutet auf Prüfungsanforderungen hin, die durch die Reproduktion von Wissen oder das Einüben von bestimmten Lösungsmustern bewältigt werden können. Für die vorliegende Arbeit erscheint es daher vielversprechend, in weiteren Analysen der Frage nachzugehen, inwieweit eine Kompetenzorientierung in Prüfungen umgesetzt wird. Inwieweit die Gütekriterien der pädagogischen Diagnostik berücksichtigt werden, konnte auf Basis der Umfragedaten nicht geklärt werden. Die Lehrenden begannen relativ kurz vor dem Prüfungstermin mit der Prüfungserstellung. Ob dabei eine an Lehrzielen ausgerichtete Aufgabenerstellung stattfindet, erscheint fraglich. Die Durchführung eines Qualitätsentwicklungsprogramms, um herauszufinden, wie Lehrende bei der Erstellung und Auswertung von Prüfungen vorgehen, stellt schon deshalb einen großen Mehrwert dar.

Das geplante Qualitätsprogramm sollte schriftliche Prüfungen mit hohen Studierendenzahlen fokussieren, da mit diesem Format die größte Reichweite erzielt werden kann. Aufgrund der hohen Zahl an Lehrenden, die einer externen Unterstützung eher kritisch gegenüber stehen, sollte das Programm auf einer freiwilligen Teilnahme beruhen und interessierte und aufgeschlossene Lehrende ansprechen. Außerdem ist ein realistischer zeitlicher Aufwand für ein solches Programm vorzusehen, damit die (geringen) zeitlichen Ressourcen, die den Lehrenden für die Erstellung einer Prüfung zur Verfügung stehen, angemessen zu berücksichtigen.

2.2.3 Kurzzusammenfassung

Ziel der Betrachtung des Forschungsstands war es einzuschätzen, inwieweit es Hochschullehrenden gelingt, Prüfungen kompetenzorientiert zu gestalten und dabei die Gütekriterien der pädagogischen Diagnostik einzuhalten. Außerdem sollte geklärt werden, welche Probleme, Defizite und Herausforderungen in der Prüfungspraxis von Hochschullehrenden bestehen. Diese Fragen konnten anhand des aktuellen Forschungsstands nicht umfassend geklärt werden. Vor dem Hintergrund der Ergebnisse der Umfrage zur Prüfungspraxis an der Technischen Universität München, wonach die Lehrenden ihre Prüfungen erst kurz vor dem Prüfungstermin erstellen, sollte der Frage nachgegangen werden, inwieweit die Prüfungen systematisch mit Lehrzielen und Lehre verschränkt sind und ob es gelingt zentrale Aspekte bei der Prüfungskonzeption einzuhalten, um deren Qualität zu sichern. Auf inhaltlicher Seite erscheint daher die Frage relevant, inwieweit es den Lehrenden gelingt, die in Lehrzielen formulierten Inhalte und Anforderungen übersetzungsgetreu in Prüfungsaufgaben abzubilden. Die Ergebnisse von Stefanica (2013) deuten darauf hin, dass es den Lehrenden, gerade bei Lehrzielen die weiterführende Anforderungen beschreiben, nicht gelingt, entsprechende Aufgaben zu entwickeln. Aufgrund des einseitigen Lernens von Studierenden bei der Prüfungsvorbereitung (das Lernen erfolgte vorrangig mit Vorlesungsskripten, Foliensätzen und Altklausuren, was eher für Prüfungsanforderungen spricht, die schwerpunktmäßig durch die Wiedergabe von Wissen oder dem Ausführen routinemäßiger Prozeduren liegt) muss aufgeklärt werden, welche Anforderungen in Prüfungen umgesetzt werden.

Für die Gestaltung von Professionalisierungsmaßnahmen konnte geklärt werden, dass sich entsprechende Programme auf schriftliche Prüfungen mit hohen Studierendenzahlen beziehen sollten, da damit die größte Reichweite erzielt wird. Auch sollte daher ein Fokus darauf gelegt werden Prüfungsformate und Aufgabentypen zu erarbeiten, die standardisierbar und korrigierbar sind, es aber dennoch ermöglichen, heterogene Anforderungen abzubilden. Außerdem ist das geringe Zeitbudget zu beachten, das den Lehrenden für die Erstellung und Auswertung zur Verfügung steht. Aufgrund der tendenziellen Ablehnung gegenüber externer Unterstützungsmaßnahmen sollte die Teilnahme an entsprechenden Programmen auf Freiwilligkeit beruhen.

Als Fazit kann festgehalten werden, dass die Erforschung der Prüfungspraxis noch ganz am Anfang steht. Die eingangs gestellten Fragen zur Qualität von Prüfungen unter Einhaltung wesentlicher Kriterien der Testtheorie und bestehenden Problemen und Defiziten in der Prüfungspraxis konnten nur in Ansätzen beantwortet werden. Problematisch erscheint jedoch die von Stefanica (2013) festgestellte Diskrepanz zwischen den Anforderungen in Lehrzielen und Prüfungsaufgaben. In der Dissertation soll daher grundlegend geklärt werden welche Probleme Lehrende in ihrer Prüfungspraxis sehen, inwiefern es gelingt, die in Lehrzielen formulierten Anforderungen in Prü-

fungsaufgaben zu übersetzen und ob die Lehrenden zentrale Gestaltungsaspekte bei der Erstellung und Auswertung von Prüfungen berücksichtigen, um deren Qualität zu sichern. Außerdem wird betrachtet, ob es mit dem geplanten Qualitätsentwicklungsprogramm gelingt, Lehrende dabei zu unterstützen, die Qualität von Prüfungen zu entwickeln.

2.3 Theoretische Grundlagen zur Entwicklung eines Qualitätsentwicklungsprogramms

Die empirischen Erhebungen, sowohl zur bereits bestehenden als auch zur veränderten Prüfungspraxis, wurden im Rahmen des geplanten Qualitätsentwicklungsprogramms durchgeführt. Dieses Vorgehen bietet auf wissenschaftlicher Seite den Vorteil, dass Veränderungen der Prüfungspraxis und deren Bedingungen bzw. Voraussetzungen in ihrer Komplexität prozessnah und damit authentisch betrachtet und analysiert werden können. Für die Teilnehmenden entsteht gleichzeitig ein praktischer Nutzen, da sie im Rahmen eines sorgfältig entwickelten Programms zusammen mit Experten aus der Bildungsforschung an der Entwicklung ihrer eigenen Prüfungen arbeiten können. Bezüglich der Gestaltung des Qualitätsentwicklungsprogramms konnte im vorangegangenen Abschnitt (2.2.2) geklärt werden, dass schriftliche Prüfungen mit hohen Studierendenzahlen fokussiert werden und die Teilnahme auf Freiwilligkeit beruhen sollte sowie, dass die geringen zeitlichen Ressourcen, die Lehrenden für die Prüfungserstellung zu Verfügung stehen, berücksichtigt werden müssen. Nachfolgend werden basierend auf erprobten Prinzipien effektiver Professionalisierung weitere Konzeptionskriterien für das Programm erarbeitet. Dazu werden in Abschnitt 2.3.1 zunächst Prinzipien vorgestellt, die sich wissenschaftlich bewährt haben, um das Handeln von Lehrpersonen effektiv und nachhaltig zu professionalisieren. Anschließend werden die inhaltlichen Zielsetzungen des geplanten Programms beschrieben, die ausgerichtet sind an den erarbeiteten Kriterien zur Kompetenzmessung und an den Gütekriterien der pädagogischen Diagnostik (Abschnitt 2.3.2). Daran anknüpfend wird dargelegt, wie die beschriebenen Prinzipien effektiver Professionalisierung in ein Fortbildungskonzept für Hochschullehrende, das den Ausgangspunkt für die geplante und längerfristige angelegte Qualitätsentwicklungsarbeit bildet, übersetzt wurden (Abschnitt 2.3.3). Mit der Fortbildung sollte im Rahmen von vier Arbeitstreffen zunächst die Qualität von Prüfungen über ein Semester entwickelt werden, indem Lehrende mit ihrer fachlichen Expertise in einem gemeinsamen Dialog mit Prüfungsexperten über ihre Prüfungspraxis eintreten. Außerdem sollte die im Rahmen der Fortbildung begonnene Arbeit an Prüfungen in einen eigenständigen und nachhaltigen Qualitätsentwicklungsprozess an den teilnehmenden Lehrstühlen überführt werden. Der Ablauf der Arbeitstreffen wird in Abschnitt 2.3.4 beschrieben.

2.3.1 Prinzipien effektiver Professionalisierung

Die Professionalisierung von Lehrpersonen wird als zentraler Hebel verstanden, um die Qualität institutioneller Bildung zu verbessern (Ostermeier, Prenzel, & Duit, 2010). In der Forschung zur Professionalisierung von Lehrenden konnten Prinzipien identifiziert und beschrieben werden, die zur Wirksamkeit entsprechender Maßnahmen beitragen. Da der Großteil der betrachteten Studien

im Schulbereich und mit Lehrkräften durchgeführt wurde, müssen die beschriebenen Prinzipien jeweils auf die besonderen Gegebenheiten der Hochschullehre, respektive das Prüfen an Hochschulen übertragen werden, um abzuschätzen, inwieweit die verschiedenen Aspekte auf den Hochschulbereich anwendbar sind.

Das Lernen von Lehrpersonen im Rahmen von Professionalisierungsmaßnahmen muss situiert und aktiv erfolgen (Borko, 2004), indem die Teilnehmenden eigenverantwortlich an Materialien, Gegenständen und Problemen der eigenen Praxis arbeiten. Fortbildungsinhalte müssen am Wissensstand und den Erfahrungen der Teilnehmenden ausgerichtet werden (Desimone, 2009). Dieses Vorgehen erscheint zielgruppenadäquat, insofern Hochschullehrende qua Profession einen hohen Grad an Autonomie gewohnt sind. Um die Nachhaltigkeit von Fortbildungsmaßnahmen zu steigern, ist die kollektive Teilnahme einer Organisationseinheit (z.B. Lehrstuhl) anzustreben (Garet, Porter, Desimone, Birman, & Yoon, 2001). Zudem fördern Diskussionen unter Kolleginnen und Kollegen das Verständnis neuer Konzepte und unterstützen die Identifikation und Lösung von Problemen (Borko, 2004). Lehrende können sich dabei gegenseitig eigene Materialien vorstellen, diese diskutieren und Feedback dazu erhalten (Desimone, 2009). Eine dauerhafte Kooperation von Lehrenden kann die Grundlage bilden für einen eigenständigen und dauerhaft angelegte Qualitätsentwicklungsprozess. Außerdem können so gemeinsame Standards entwickelt und etabliert werden (Ostermeier, 2004).

Professionalisierungsinhalte müssen an professionellen und wissenschaftlich fundierten Standards ausgerichtet sein (Ostermeier et al., 2010). Das in Abschnitt 2.1.3 eingeführte Modell zur Erstellung und Auswertung von Prüfungen bildet dafür einen Referenzrahmen. Externe Moderatorinnen und Moderatoren, die mit den entsprechenden Standards vertraut sind, versuchen den Fokus der Professionalisierung auf diese Standards zu lenken. Zudem können die Moderatoren Arbeitsprozesse durch theoretische Inputs strukturieren und entsprechend Feedback geben, um Probleme in bestehenden Materialien sowie bestehende Routinen der Teilnehmenden sichtbar zu machen. Die Moderatoren sollten flexibel auf die Probleme und das Vorwissen der Teilnehmenden eingehen und die Fortbildungsziele, den Input und Unterstützungsmaßnahmen in der Folge an die Bedürfnisse der Teilnehmer anpassen (Borko, 2004). Ein weiterer wichtiger Aspekt für die Nachhaltigkeit und den Praxistransfer von Fortbildungsinhalten ist das kollaborative Festlegen von Zielen und die Reflexion über Fortschritte hinsichtlich der vereinbarten Ziele (Opfer & Pedder, 2011). Bei Professionalisierungsmaßnahmen muss daher darauf geachtet werden, dass Ziele vereinbart werden und Raum für Reflexion von Arbeitsprozessen und die Zielerreichung besteht.

2.3.2 Ziele des Qualitätsentwicklungsprogramm

Das Ziel der geplanten Maßnahme war es, auf Ebene von Lehrstühlen den Anstoß zur Aufnahme einer kollektiv angelegten Qualitätsentwicklungsarbeit zu geben, um die Qualität von Prüfungen kontinuierlich zu verbessern. Zudem wurde angestrebt, dass die Teilnehmenden dabei sukzessive weitere Kolleginnen und Kollegen miteinbeziehen, um gemeinsame Qualitätsstandards für Prüfungen zu etablieren und Best-Practices auszutauschen und dadurch eine nachhaltige Wirkung der Maßnahme zu gewährleisten. Ein Fortbildungsangebot bildete den Ausgangspunkt für die angestrebte Qualitätsentwicklungsarbeit. Die Fortbildung war in Form einer Workshop-Reihe mit kollaborativen Erprobungs- und Reflexionsphasen angelegt. Lehrende arbeiteten zusammen mit externen Moderatoren über den Zeitraum eines Semester an einer Prüfung, die zum Ende des Semesters eingesetzt wurde. Nach Durchlaufen dieses Prozesses entlang des entwickelten Vorgehensmodells zur Prüfungserstellung und –auswertung (siehe Abbildung 3) sollten die Lehrenden dazu in der Lage sein:

- (1) Ziele für Lehrveranstaltungen und Module systematisch auszuwählen und zu modellieren,
- (2) die in den Lehrzielen beschriebenen Aspekte des Wissens und Könnens in Anforderungen in Form von Prüfungsaufgaben zu übersetzen,
- (3) die Prüfungsaufgaben auf ihre inhaltliche und kognitive Validität zu überprüfen,
- (4) geeignete Korrektur- und Auswertungsverfahren zu erstellen und anzuwenden,
- (5) Punkte- und Noten nach testtheoretischen Grundsätzen zu vergeben. sowie
- (6) mittels statistischer Verfahren Informationen aus den Prüfungsdaten gewinnen.

Diese sechs Schritte strukturierten das gemeinsame Arbeiten zwischen Hochschullehrenden als Fachexperten und den externen Moderatoren im Rahmen der Fortbildung und bauten auf den theoretischen Überlegungen auf. Die in der Fortbildung gemeinsam begonnene Arbeit an den Prüfungen sollte in einen eigenständigen und selbstverantwortlichen Qualitätsentwicklungsprozess an den Lehrstühlen überführt werden. In der Fortbildung wurden daher, neben dem Feedback zu bestehenden Prüfungen, konkrete Vorschläge zur Etablierung von Organisationsstrukturen erarbeitet, die eine Kollaboration von Hochschullehrenden fördern. Dieser Aspekt wird nicht in die Analysen der vorliegenden Arbeit miteinbezogen, wurde jedoch an dieser Stelle kurz angesprochen, damit die folgenden Darstellungen zum Ablauf des Programms nachvollziehbar sind. Die folgenden Ausführungen beziehen sich aber im Kern auf das Fortbildungskonzept, das in das Qualitätsentwicklungsprogramm eingebettet ist.

2.3.3 Ablauf der Fortbildung

Ausgehend von den Zielsetzungen und den Prinzipien erfolgreicher Professionalisierung wurde ein idealtypischer Fortbildungsablauf festgelegt. In einem Informationsgespräch wurden die Teilnehmenden zuerst über die Fortbildungsinhalte und das geplante Vorgehen aufgeklärt. Die Teilnehmenden formulierten ihre Zielsetzungen für die Fortbildung. Im Anschluss an die Informationsgespräche hinterlegten sie ihre Unterlagen und Materialien wie bestehende Prüfungen, Lehrziele, Modulbeschreibungen und Veranstaltungsmaterialien (z.B. Skripte und Foliensätze) auf einem gemeinsamen Server. Die anderen Teammitglieder, sowohl lehrstuhlintern als auch lehrstuhlübergreifend, erhielten ebenfalls Zugriff auf die Materialien.

Auf Grundlage der gewonnenen Erkenntnisse zu den Zielen der Teilnehmenden aus dem Informationsgespräch und einer Analyse von bestehenden Prüfungen und Lehrzielen wurde das erste Arbeitstreffen geplant. Die im Theorieteil beschriebenen sechs Schritte zur Prüfungserstellung wurden auf vier Arbeitstreffen verteilt. Die ersten beiden Arbeitstreffen fanden zu Beginn des Semesters statt, um Lehrziele zu definieren, bestehende Aufgaben auf ihre inhaltliche Passung zu überprüfen und die bereits existierenden Prüfungsaufgaben formal und inhaltlich zu überarbeiten. Das dritte Treffen fand nach Abhaltung der Prüfung und vor der Korrektur statt, um an Korrekturverfahren und Möglichkeiten zur Punktevergabe und Benotung zu arbeiten. Im vierten Arbeitstreffen wurden statistische Kennzahlen zu den Prüfungen besprochen, um daraus Rückschlüsse auf die Qualität der Aufgaben, die Lehrveranstaltung, die Lehrziele und das Abschneiden der Studierenden zu ziehen. Auf jedes Arbeitstreffen folgte jeweils eine Erarbeitungsphase, in der die Lehrenden die gemeinsam begonnene Arbeit eigenständig fortsetzten.

Zu Beginn jedes Arbeitstreffens erfolgte eine Reflexion über die Fortschritte, Erfahrungen und aufgetretene Probleme während der Arbeitsphasen. Nach dem vierten Arbeitstreffen wurden zwei weitere Reflexionstermine für das darauffolgende Semester geplant, um Erkenntnisse darüber zu gewinnen, welche Aspekte der gemeinsamen Arbeit von den Lehrenden übernommen und eigenständig fortgesetzt wurden (siehe Abbildung 4).

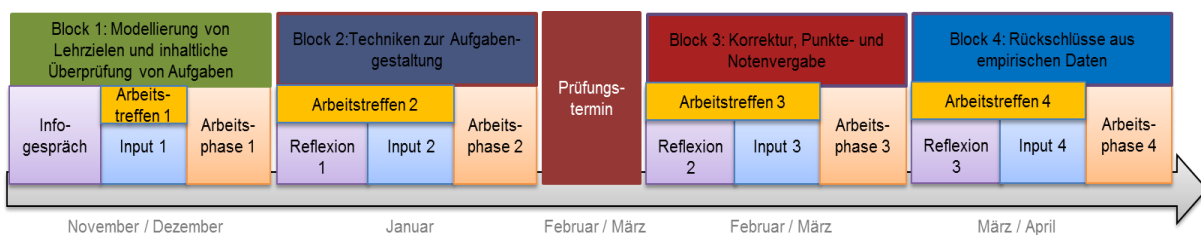


Abbildung 4: Ablauf der Fortbildung

2.3.4 Ablauf der Arbeitstreffen

Ziel der Arbeitstreffen war es, Lehrende und Moderatoren in einen problem- und materialbezogenen gegenseitigen Diskurs anhand bestehender Lehrziele und Prüfungen eintreten zu lassen. Der Ablauf der Arbeitstreffen war dabei einheitlich aufgebaut. Im ersten Teil der Treffen wurden die Lehrenden zu ihrem bisherigen Vorgehen hinsichtlich des behandelten Themenbereichs befragt. Dadurch konnte eine gemeinsame Ausgangsbasis etabliert werden, um bedarfsorientiert auf Probleme eingehen zu können. Ab dem zweiten Arbeitstreffen wurde außerdem ein Reflexionsteil zur vorangegangenen Arbeitsphase eingeführt, um festzustellen, welche Veränderungen in der Prüfungspraxis vorgenommen wurden, an welchen Stellen Probleme auftraten und ob Klärungsbedarf besteht. Anschließend folgte jeweils ein Input der Moderatoren zu einem der festgelegten Themenbereiche, der inhaltlich an wissenschaftlichen Standards ausgerichtet und gleichzeitig auf die identifizierten Probleme bezogen war. Das theoretisch vermittelte Wissen wurde dann auf die bestehenden Lehrziele, Prüfungen und Prüfungsaufgaben angewandt. Die Moderatoren wiesen dabei gezielt auf existierende Schwachstellen hin und schlugen exemplarisch Lösungsmöglichkeiten vor. Die Lehrenden wurden angeregt, sich in den Treffen gegenseitig Feedback zu geben und Lösungsvorschläge zu unterbreiten. Am Ende der Treffen wurden Ziele für die anschließenden Arbeitsphasen formuliert. Die Teilnehmenden erhielten zu jedem der Themenbereiche Handreichungen und Checklisten, in denen jeweils die entsprechenden testtheoretischen Grundlagen anwendungsbezogen dargelegt wurden.

2.3.5 Kurzzusammenfassung

In diesem Abschnitt wurde das Qualitätsentwicklungsprogramm „Herausforderung Prüfen“ vorgestellt. Das Programm dient als Rahmen für die empirischen Erhebungen der Arbeit. Ziel des Programms war es, dass Lehrende kollaborativ an ihrer Prüfungspraxis arbeiten und gemeinsame Qualitätsstandards etablieren, die sich an wissenschaftlich orientierten Kriterien orientieren. Dadurch sollte die Grundlage für eine eigenständige und nachhaltige Qualitätsentwicklung gelegt werden. Den Ausgangspunkt für die Qualitätsentwicklung bildet ein Fortbildungsprogramm, in dem Lehrende und Prüfungsexperten im Zeitraum eines Semesters an einer Prüfung arbeiten und diese entlang wissenschaftlicher Standards bewerten und entwickeln. Das in Abschnitt 2.1.3 eingeführte Vorgehensmodell zur Prüfungserstellung und -auswertung bildete hierbei den inhaltlichen Rahmen. Das Fortbildungskonzept wurde entlang von empirisch bestätigten Prinzipien effektiver Professionalisierung angelegt. Durch diesen Ansatz entstand auf wissenschaftlicher Seite der Vorteil, dass die Prüfungspraxis und deren Veränderung authentisch und prozessnah über einen längeren Zeitraum betrachtet werden konnte. Für die Lehrenden entstand durch die Fortbildung ein praktischer Mehrwert, da sich die Fortbildungsinhalte direkt auf die Entwicklung und

Überarbeitung einer Prüfung bezogen, die von den Lehrenden zu Ende des Semesters eingesetzt wurde.

3 Forschungsfragen

Aus den theoretischen Vorüberlegungen und der aufgezeigten Forschungslücke zur Prüfungspraxis von Hochschullehrenden lassen sich für diese Arbeit drei Forschungsfragen ableiten. Als analytischer Rahmen zur Beantwortung der Forschungsfragen dient das in Abschnitt 2.1.3 eingeführte Vorgehensmodell, das den Prozess zur Prüfungserstellung und –auswertung in sechs Schritte unterteilt. Die Forschungsfragen werden jeweils zu jedem der Schritte aus dem Vorgehensmodell betrachtet.

Forschungsfrage 1: Welche Ziele leiten Hochschullehrende aus den von ihnen wahrgenommenen Herausforderungen in ihrer existierenden Prüfungspraxis für die Teilnahme an einem Qualitätsentwicklungsprogramm ab?

Die Frage nach den Zielsetzungen von Lehrenden ist aus zweierlei Perspektiven relevant. Eine Grundvoraussetzung für das Wirksamwerden von Professionalisierungsmaßnahmen ist die Kohärenz zum Wissens- und Erfahrungsstand von Lehrenden und eine Fokussierung auf Probleme in deren täglicher Praxis (Desimone, 2009). Anhand der formulierten Zielsetzungen kann eingeschätzt werden, an welchen Stellen im Prozess der Prüfungserstellung und –auswertung die Lehrenden schwerpunktmäßig Probleme sehen. Dadurch können die Inhalte des Qualitätsentwicklungsprogramms an die Zielsetzungen der Lehrenden angepasst werden. Zum anderen kann die Problemeinschätzung der Lehrenden mit der Einschätzung der Prüfungspraxis anhand von standardisierten Bewertungskriterien durch externe Experten verglichen werden, um einen Abgleich zwischen Fremd- und Selbsteinschätzung zu erhalten.

Forschungsfrage 2: Inwieweit gelingt es den Lehrenden, Aspekte einer Kompetenzorientierung unter Einhaltung der Gütekriterien pädagogischer Diagnostik in ihrer Prüfungspraxis umzusetzen?

Wie im Theorieteil dargelegt, stellt die Vermittlung und Überprüfung von Kompetenzen mittlerweile einen zentralen Qualitätsanspruch an Studiengänge dar. Die entsprechenden Prüfungen müssen von Hochschullehrenden so gestaltet sein, dass die Kompetenzen valide (gültig) und reliabel (zuverlässig) erfasst werden können. Im Rahmen von Forschungsfrage 2 soll ermittelt werden, inwieweit es den Lehrenden gelingt, Prüfungen unter Einhaltung der Gütekriterien der pädagogischen Diagnostik kompetenzorientiert zu gestalten. Zur Beantwortung dieser Frage werden sowohl materialbasierte Daten (Lehrziele, Prüfungen und Prüfungsaufgaben) als auch Verbaldaten (basierend auf Audioaufzeichnungen der Arbeitstreffen) analysiert. Inwieweit Prüfungen kompetenzorientiert gestaltet werden, kann anhand des Vorgehens der Lehrenden bei der Festlegung des Prüfungsgegenstands, bei der Gestaltung von Aufgaben und der Notenvergabe ermittelt werden.

Bei der Festlegung des Prüfungsgegenstands müssen sich die Lehrenden an realen Anforderungssituationen und relevanten Kontexten einer (Lern) Domäne orientieren (siehe Abschnitt 2.1.1.1). Hier gilt es anhand der Verbaldaten zu klären, welche Suchanleitung - im Sinne von Referenzpunkten und Unterlagen - die Lehrenden bei der Festlegung von Lehrzielen verwenden. Eine kompetenzorientierte Suchanleitung liegt vor, wenn die Recherchen auf die Identifikation von realen Situationen abzielen (siehe Abschnitt 2.1.1.2). Auf Ebene von Aufgaben zeigt sich die Kompetenzorientierung in den angelegten Wissensarten, der Aufgabenoffenheit, der Funktion von Aufgabenkontexten und der sprachlogischen Komplexität von Aufgaben (siehe Abschnitt 2.1.1.3). Die Notenvergabe muss kriteriumsorientiert erfolgen, sofern über Noten Kompetenzniveaus ausgedrückt werden sollen (siehe Abschnitt 2.1.3.5).

Zur Einhaltung der Gütekriterien müssen die Lehrenden, bei allen sechs Schritten des Prüfungsmodells, die im Theorieteil erarbeiteten Maßnahmen und Kriterien berücksichtigen (siehe Abschnitt 2.1.3). Bei der Festlegung des Prüfungsgegenstands (Schritt 1 des Prüfungsmodells) ist relevant, ob die Auswahl von Prüfungsinhalten systematisch erfolgt und die zugrunde gelegten Auswahlvorschriften offen gelegt werden. Bei den Prüfungsaufgaben (Schritt 2 des Prüfungsmodells) ist der Lehrzielbezug zentral, also die Frage, ob die Aufgabenanforderungen den durch die Lehrziele beschriebenen Anforderungen entsprechen. Damit Aufgaben keine konstruktirrelevante Einfachheit oder Schwierigkeit aufweisen, müssen formale Gestaltungsmerkmale wie sprachliche Aspekte, die Instruktionsqualität und Aspekte in der Formatierung berücksichtigt werden. Die inhaltliche und kognitive Validität muss mit Maßnahmen wie einer rationalen und kognitiven Aufgabenanalyse abgesichert werden (Schritt 3 des Prüfungsmodells). Inwieweit Lehrende entsprechende Abgleiche durchführen soll geklärt werden. Hinsichtlich der Bewertung von Prüfungsantworten (Schritt 4 des Prüfungsmodells) wird überprüft, ob die Lehrenden Bewertungsraster mit Bewertungskategorien, Regelwerken und Ankerbeispielen verwenden und eine Schulung von Korrektoren durchführen. Die Konsistenz und Gültigkeit von Korrekturen kann durch diese Maßnahmen deutlich erhöht werden. Die Vergabe von Punkten zu Aufgaben und die Festlegung von Notenstufen (Schritt 5 des Prüfungsmodells) müssen theoriegeleitet und orientiert am zu messenden Konstrukt erfolgen. Letztlich wird ermittelt, inwieweit die Lehrenden die Prüfungsdaten nutzen, um Informationen und Rückschlüsse über die Prüfungs- und Aufgabenqualität und das Abschneiden von Studierenden und Studierendengruppen ziehen (Schritt 6 des Prüfungsmodells).

Forschungsfrage 3: Welche Prinzipien und Techniken zur Steigerung der Kompetenzorientierung und Einhaltung der Gütekriterien werden im Laufe des Qualitätsentwicklungsprogramms in die Praxis übernommen? Welche Veränderungen werden an den Prüfungen vorgenommen?

Mit Forschungsfrage 3 werden Veränderungen beschrieben, die Lehrende während des Qualitätsentwicklungsprogramms an ihren Prüfungen vornehmen. Das Erkenntnisinteresse richtet sich darauf, inwieweit die im Rahmen des Programms vorgestellten Maßnahmen und Techniken zur Gestaltung und Auswertung von Prüfungen von den Lehrenden unter den gegebenen Ressourcen als hilfreich zur Lösung von Problemen wahrgenommen und in die Prüfungspraxis übernommen werden. Zudem soll festgestellt werden, welche Grenzen bei der Umgestaltung der Prüfungspraxis auftreten und an welchen Stellen externe Ressourcen zur Unterstützung der Lehrenden zur Verfügung gestellt werden müssen.

4 Methoden

Im Methodenkapitel wird zunächst die Anlage der Untersuchung beschrieben. Das Ziel der Untersuchung ist eine fallbasierte Aufklärung der bestehenden Prüfungspraxis an verschiedenen Lehrstühlen aus der Fakultät für Maschinenwesen und der Fakultät für Mathematik der Technischen Universität München. Die empirischen Erhebungen finden im Rahmen eines Qualitätsentwicklungsprogramms statt. Nachfolgend werden das Untersuchungsdesign, die Stichprobe sowie die Datenquellen und das Vorgehen bei der Analyse und Auswertung der Daten dargestellt.

4.1 Untersuchungsdesign

Als Rahmen für die empirischen Erhebungen dient das beschriebene Qualitätsentwicklungsprogramm (siehe Abschnitt 2.3). Das Design der Studie sieht eine fallbasierte Analyse von Materialien und dem Vorgehen der Teilnehmenden bei der Prüfungserstellung und –auswertung zu mehreren Zeitpunkten vor. Zur Klärung von Forschungsfrage 1 wurden in einem Informationsgespräch von den Teilnehmenden des Qualitätsentwicklungsprogramms die Zielsetzungen erfragt, die sie im Rahmen des Programms erreichen möchten. Für die Beantwortung von Forschungsfrage 2 wurden zum einen materialbasierte Daten (Lehrziele, Prüfungen und Prüfungsaufgaben) analysiert, um Probleme und Herausforderungen in der bestehenden Prüfungspraxis zu identifizieren. Zum anderen wurden die Audioaufzeichnungen aus den Arbeitstreffen verwendet, um Erkenntnisse zum Vorgehen der Lehrenden bei der Prüfungserstellung- und Auswertung zu gewinnen. Die Datengrundlage zur Beantwortung von Forschungsfrage 3 bilden die neu erstellten Lehrziele und Prüfungen, die mit den gleichen Kategoriensystemen wie die bestehenden Materialien analysiert wurden, sowie die Erkenntnisse aus den Verbaldaten, die auf den Audioaufzeichnungen der Arbeitstreffen basieren (siehe Abbildung 5).

Untersuchungsdesign						
Bestehende Praxis		Veränderungen in der Prüfungspraxis				Veränderte Praxis
Existierende Materialien		Prozess der Prüfungserstellung				Neu erstellte Materialien
Lehrziele		Arbeits-treffen 1	Arbeits-treffen 2	Arbeits-treffen 3	Arbeits-treffen 4	Lehrziele
Prüfungen u. Aufgaben						Prüfungen u. Aufgaben
Analyse mit Kategoriensystem		Transkription der Audioaufzeichnung; Auswertung durch induktive Kategorienbildung				Analyse mit Kategoriensystem

Abbildung 5: Untersuchungsdesign

Gemeinsam mit den Teilnehmenden aus der Fakultät für Maschinenwesen wurden vier Arbeitstreffen durchgeführt. Das Vorgehen mit den Teilnehmenden aus der Mathematik wich davon ab,

insofern nur zwei Arbeitstreffen stattfanden, die inhaltlich andere Schwerpunkte beinhalteten. So waren in diesen Treffen keine Arbeitsphasen vorgesehen. Dies lag darin begründet, dass die Mathematiker bereits im Vorfeld mit hohem Aufwand alternative Prüfungskonzepte entwickelten. Die Treffen dienten zur Vorstellung der Konzepte und um weiterführende Entwicklungspotenziale und Möglichkeiten zur Verbreitung zu diskutieren. Daher liegen aus den Arbeitstreffen mit den Teilnehmenden aus der Mathematik keine vergleichbaren und auf das Vorgehensmodell (siehe Abschnitt 2.1.3) bezogene Informationen zu den Fällen aus dem Maschinenwesen vor. Zudem wurde nur von einem der Teilnehmenden aus der Mathematik während des Qualitätsentwicklungsprogramms eine neue Prüfung erstellt, bei den beiden anderen Fällen ist dagegen kein Vorher-Nachher-Vergleich von Prüfungen möglich. Aufgrund der unterschiedlichen Datenlage im Vergleich zu den Arbeitsgruppen aus der Fakultät für Maschinenwesen wird in dieser Arbeit auf eine Darstellung der Ergebnisse zu den Fällen aus der Mathematik verzichtet. Somit kann eine kohärente Darstellung der Ergebnisse gewährleistet werden. Stattdessen ist geplant, die Prüfungskonzepte aus der Mathematik in einem Format außerhalb der Dissertation, im Sinne eines Best-Practice Leitfadens umfassend darzustellen. Um eine Datengrundlage für die Beschreibung der Best-Practices zu schaffen, wurden die Prüfungen aus der Mathematik jedoch im Rahmen dieser Arbeit parallel zu den Prüfungen aus dem Maschinenwesen ausgewertet. Damit liegen belastbare Daten zu den Prüfungen vor.

4.2 Stichprobe und Fallbeschreibungen

In diesem Abschnitt wird zunächst auf das Vorgehen bei der Akquirierung von Teilnehmerinnen und Teilnehmern für das Programm eingegangen (Abschnitt 4.2.1). Anschließend erfolgt die Bildung der Fälle. Da die Prüfungen Gegenstand der Qualitätsentwicklung waren, bilden die betrachteten Prüfungen die Analyse- bzw. Falleinheiten. Es wird außerdem beschrieben, wie die Fälle für die Arbeit im Programm zu Teams gruppiert wurden. Bei der Beschreibung der Fälle wird darauf eingegangen, welche Position die teilnehmenden Personen an den jeweiligen Lehrstühlen ausüben, über welche Prüfungserfahrung sie verfügen und wie die Verantwortlichkeiten bei der Prüfungserstellung, und -auswertung verteilt sind. Außerdem werden Informationen zu den jeweiligen Rahmenbedingungen der Prüfungen gegeben (Abschnitt 4.2.2).

4.2.1 Akquirierung von Teilnehmerinnen und Teilnehmern

Die Möglichkeit zur Teilnahme am Programm „Herausforderung Prüfen“ wurde im Rahmen eines Treffens der Dekane der TU München bekannt gegeben. Die Dekane erhielten ein Informationsschreiben zur Weiterleitung an Kolleginnen und Kollegen an den Fakultäten. In diesem Schreiben waren die Ziele, Inhalte und der Ablauf des Programms beschrieben. Potentielle Interessenten

wendeten sich direkt an die Programmträger (Mitarbeiter des Lehrstuhls für Empirische Bildungsforschung und ProLehre), um weitere Fragen sowie die Bedingungen einer Teilnahme zu klären. Im ersten Jahr der Maßnahme wurde die Anzahl der Teilnehmer auf fünf Lehrstühle aus dem Maschinenwesen und drei Lehrstühle aus der Mathematik begrenzt. Es bestanden weitere Anfragen aus der Fakultät für Medizin, die aus Kapazitätsgründen für die erste Kohorte nicht berücksichtigt werden konnten.

4.2.2 Bildung der Fälle und Zusammensetzung der Teams

Die Analyse- beziehungsweise Falleinheiten bilden die Prüfungen, die im Rahmen des Programms bearbeitet wurden. Zur bestehenden Praxis wurden insgesamt acht Prüfungen betrachtet, wodurch acht Fälle gebildet wurden. Die betrachteten Fälle aus dem Maschinenwesen werden mit „Fall MW 1“ bis „Fall MW 5“ bezeichnet. Aus Gründen der Vollständigkeit werden die Fälle aus der Mathematik in der Fallbeschreibung zunächst mitaufgeführt, obwohl die Ergebnisse zu den Mathematikfällen in dieser Arbeit nicht berichtet werden. Die Fälle aus der Mathematik werden mit „Fall Mathe 1“ bis „Fall Mathe 3“ bezeichnet. Aufgrund der unterschiedlichen Konstellationen kann ein Fall aus nur einer Person bestehen oder aus mehreren Personen. So besteht der Fall MW 1 aus drei Personen eines Lehrstuhls, die gemeinsam an einer Prüfung arbeiten. Die Fälle MW 3 und MW 4 bestehen aus jeweils einer Person vom gleichen Lehrstuhl, die jedoch mit zwei verschiedenen Prüfungen am Programm teilnahmen. Die Zusammensetzung der Teams und die Bildung der Fälle ist in Abbildung 6 grafisch dargestellt.

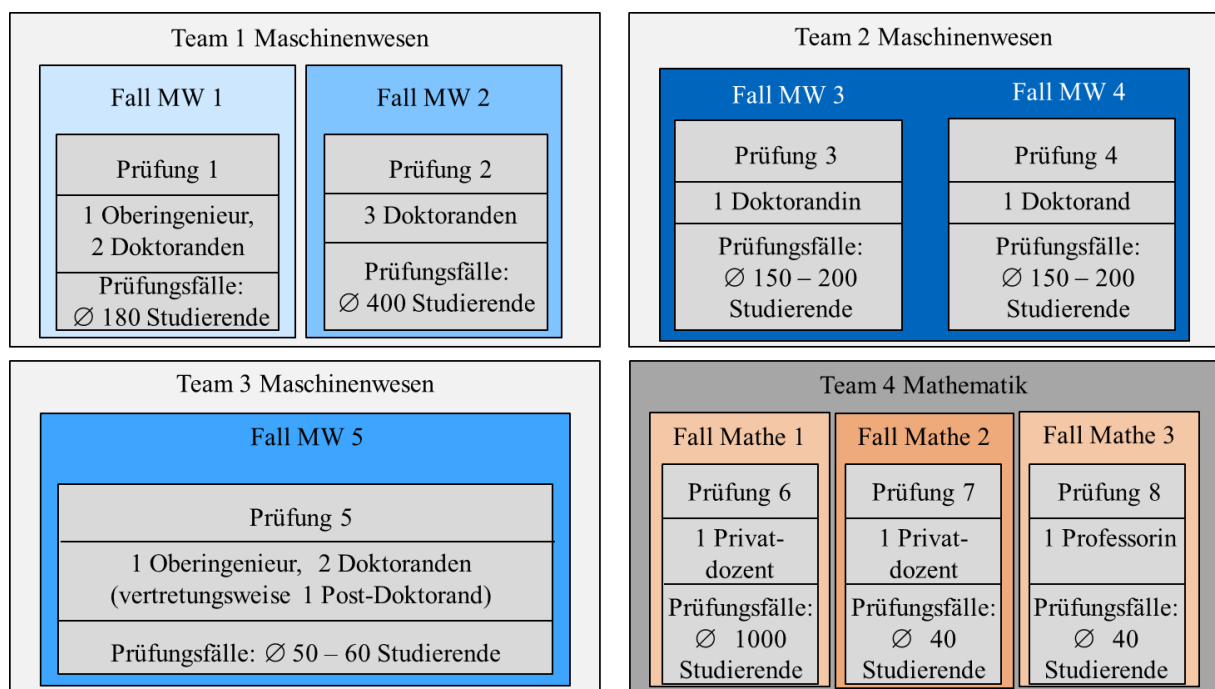


Abbildung 6: Zusammensetzung der Teams

Die Anzahl der Teilnehmenden pro Fall variierte zwischen einer und drei Personen. Insgesamt nahmen 14 Personen über die gesamte Dauer des Programms teil. Bei entsprechender inhaltlicher Nähe wurden zwei Fälle für die gemeinsamen Arbeitstreffen zu Teams zusammengeschlossen, um einen fallübergreifenden Austausch zu initiieren. Die Teilnehmenden wurden angeregt, den Austausch über die Arbeitstreffen hinaus auch in den Arbeitsphasen fortzuführen. Das so gebildete Team 1 Maschinenwesen setzt sich zusammen aus den zwei Fällen MW 1 und MW 2, die fallübergreifend kooperierten. Beide Fälle bestehen aus jeweils drei Personen, die zwei verschiedenen Lehrstühlen angehören. Team 2 Maschinenwesen setzt sich aus zwei Personen vom gleichen Lehrstuhl zusammen, die jedoch zwei Fälle bilden, da sie mit unterschiedlichen Prüfungen am Programm teilnahmen. Team 3 Maschinenwesen setzt sich aus drei Personen eines Lehrstuhls zusammen. Sie bilden einen Fall, da sie mit einer Prüfung am Programm teilnahmen. Das Team 4 besteht aus drei Teilnehmenden aus der Fakultät für Mathematik, von jeweils unterschiedlichen Lehrstühlen, die mit Fall Mathe 1 bis Fall Mathe 3 bezeichnet sind. Die Teilnehmerin des Falls Mathe 3 stieg zum zweiten Arbeitstreffen in das Programm ein. Auf die Fälle aus der Mathematik wird an dieser Stelle nicht ausführlicher eingegangen, da keine entsprechend detaillierten Informationen vorlagen und die Ergebnisse zu den Fällen aus der Mathematik in dieser Arbeit nicht berichtet werden.

Bei den Fällen aus dem Maschinenwesen waren sowohl akademische Räte mit langer Prüfungserfahrung, als auch Doktoranden mit geringer oder keiner Erfahrung mit Prüfungen vertreten. Die Fälle unterschieden sich außerdem dahingehend, ob die für die Prüfungserstellung beteiligten Personen auch in der Veranstaltung hauptverantwortlich dozierten. Dies beeinflusste, inwieweit von den Prüfungserstellern eigenständig Veränderungen an Lehrzielen und Prüfung vorgenommen werden konnten. In Tabelle 7 sind die Informationen zu den Fällen im Überblick dargestellt. Die Ausführungen beziehen sich auf den Stand zu Beginn der Zusammenarbeit im Programm im Oktober 2013.

Tabelle 7: Informationen zu den Fällen aus dem Maschinenwesen

Beschreibung der Fälle aus dem Maschinenwesen					
	Team 1		Team 2		Team 3
	MW 1	MW 2	MW 3	MW 4	MW 5
	Prüfung 1	Prüfung 2	Prüfung 3	Prüfung 4	Prüfung 5
personelle Zusammensetzung bei den Arbeitstreffen	Akademischer Oberrat, zwei Doktoranden	Drei Doktoranden	Eine Doktorandin	Ein Doktorand	Akademischer Oberrat, zwei Doktoranden
Weitere beteiligte Personen	Kollegen die Aufgaben erstellen und diese korrigieren	Kollegen die bei der Korrektur mitwirken	Kollegen die bei der Korrektur mitwirken	Kollegen die bei der Korrektur mitwirken	Kollegen die Aufgaben erstellen und diese korrigieren
Rolle bei Lehrveranstaltung	Prüfungserstellende sind <u>nicht</u> Dozierende	Prüfungserstellende sind <u>nicht</u> Dozierende	Prüfungserstellende sind nicht Dozierende	Prüfungsersteller ist <u>nicht</u> Dozierender sondern leitet die Übung	Akademischer Oberrat doziert in Veranstaltung, Doktoranden leiten Übungen
Gestaltungsspielräume	Lehrende können eigenständig Veränderungen an der Prüfung vornehmen	Veränderungen müssen vom Lehrstuhlleiter genehmigt werden	Veränderungen müssen vom Lehrstuhlleiter genehmigt werden.	Veränderungen müssen vom Lehrstuhlleiter genehmigt werden	Lehrende können eigenständig Veränderungen an der Prüfung vornehmen
Angaben zur Prüfung	Die Prüfung besteht aus einem wissensorientiertem Teil und einem Fallbeispiel. Durchschnittlich nehmen 180 Prüflinge aus verschiedenen Masterstudiengängen des Maschinenwesens an der Prüfung teil	Die Prüfung besteht aus einem wissensorientiertem Teil der zur Hälfte aus MC-Aufgaben besteht und einem Rechen-teil. Durchschnittlich nehmen 400 Bachelorstudierende des Maschinenwesens und einer anderen Fakultät an der Prüfung teil	Die Prüfung besteht aus einem Wissensteil und einem Teil mit Transferaufgaben, und soll im aktuellen Semester um einen Rechenteil erweitert werden. Durchschnittlich nehmen zwischen 150 und 200 Masterstudierende verschiedener Fakultäten an der Prüfung teil	Der erste Teil der Prüfung besteht aus einer großen, in sich geschlossenen Aufgabe, mit einem Schwerpunkt auf Berechnungen. Der zweite Teil enthält wissensorientierte Aufgaben. Durchschnittlich nehmen 150 bis 200 Studierende unterschiedlicher Studiengänge aus mehreren Fakultäten an der Prüfung teil	Die Prüfung ist nach den Inhaltsbereichen der Veranstaltung gegliedert. Sie beinhaltet mehrere Aufgabentypen (Berechnung, Diagramme, Schaubilder). Durchschnittlich nehmen 50 – 60 Studierende verschiedener Masterstudiengänge aus mehreren Fakultäten an der Prüfung teil

4.3 Datenquellen

Bei den Datenquellen wird unterschieden zwischen materialbasierten Daten und Verbaldaten. Bei den materialbasierten Daten handelt es sich um die von den Teilnehmenden formulierten Ziele für die Arbeit im Qualitätsentwicklungsprogramm und den vor Beginn des Programms bereits bestehenden Lehrzielen, Prüfungen und Prüfungsaufgaben. Die Verbaldaten bestehen aus den aufgezzeichneten Aussagen in den Arbeitstreffen. Beide Datenquellen werden nachfolgend ausführlich beschrieben.

4.3.1 Materialbasierte Daten

Die materialbasierten Daten zur Analyse der Prüfungspraxis vor Beginn der gemeinsamen Arbeit im Qualitätsentwicklungsprogramm bestanden aus den Prüfungen, die im Semester vor Beginn des Programms eingesetzt wurden, und den Lehrzielen, die in den Modulbeschreibungen hinterlegt waren. Die von den Teilnehmenden in den Informationsgesprächen formulierten Programmziele wurden von den Moderatoren verschriftlicht und für eine finale Abstimmung nochmals per E-Mail an die Teilnehmerinnen und Teilnehmer gesendet. Die schriftlichen Endfassungen der Programmziele sind daher den materialbasierten Daten zugeordnet. Insgesamt wurden 31 Programmziele von den Teilnehmenden definiert.

Die Prüfungsaufgaben konnten zum Teil in ihrer ursprünglichen Form nicht analysiert werden, da eine Aufgabe oftmals zwei oder mehrere Handlungsaufforderungen beinhaltete. Für die wissenschaftlichen Analysen wurden die Aufgaben daher segmentiert, damit jede der so gebildeten Aufgaben jeweils nur eine Aufforderung beinhaltete. Ein Beispiel dazu ist:

Originalversion der Aufgabe aus einer Prüfung:

Aufgabe 1: „Welcher Trend ist in den letzten fünf Jahren beim deutschen Energiemix auszumachen? Geben Sie drei mögliche Gründe für diese Entwicklung an.“

Segmentierte Form der Aufgabe:

Aufgabe 1: Welcher Trend ist in den letzten fünf Jahren beim deutschen Energiemix auszumachen?

Aufgabe 2: Geben Sie drei mögliche Gründe für diese Entwicklung an.

Begründung: In dieser Aufgabenstellung sind zwei Handlungsaufforderungen zusammengefasst. Im ersten Schritt soll der Trend hinsichtlich des deutschen Energiemix genannt werden. Im zweiten Schritt sollen mögliche Gründe für den ausgemachten Trend angegeben werden. Die Aufga-

benanforderungen sprechen gegebenenfalls unterschiedliche kognitive Niveaus und Wissensarten an. Außerdem scheint die Lösung von Aufgabe 2 abhängig von der Lösung von Aufgabe 1 zu sein. Diese Dimensionen werden bei der Analyse der Aufgaben erfasst, wodurch eine Unterscheidung in zwei Aufgaben notwendig wird.

Bei den Lehrzielen war ebenfalls eine Segmentierung erforderlich, da in einigen Fällen ein Lehrziel mehrere Inhalte oder Verhaltenskomponenten umfasste. Die Lehrziele in ihrer ursprünglichen Form kombinierten zum Teil Angaben mit unterschiedlicher begrifflicher Klarheit und verschiedenen inhaltlichen Reichweiten. Während manche Angaben relativ präzise auf kleine Teilausschnitte bezogen waren, blieben andere Angaben vage, indem sie einen Inhaltsbereich oder ein erwartetes Verhalten global beschrieben. Da die Lehrziele dahingehend analysiert werden, inwieweit sie für eine zielgerichtete Aufgabenerstellung tauglich sind, wurde eine Segmentierung vorgenommen. Die auf diese Weise gebildeten Lehrziele bestehen aus jeweils einer Inhalts- und einer Verhaltenskomponente. Ein Beispiel ist:

Originalversion des Lehrziels:

Lehrziel 1: Die Studierenden sind in der Lage, mathematische Verfahren zu nutzen, um ingenieurwissenschaftliche Problemstellungen mathematisch zu lösen. Außerdem können sie gewöhnliche Differentialgleichungen mit einer Veränderlichen berechnen.

Segmentierte Form des Lehrziels:

Lehrziel 1: Die Studierenden sind in der Lage mathematische Verfahren zu nutzen, um ingenieurwissenschaftliche Problemstellungen mathematisch zu modellieren.

Lehrziel 2: Die Studierenden können gewöhnliche Differentialgleichungen mit einer Veränderlichen lösen.

Begründung: Die Segmentierung ist für die wissenschaftlichen Analysen erforderlich, da zum einen unterschiedliche Inhaltsbereiche angesprochen werden, zum anderen sind verschiedene Verhaltenskomponenten beschrieben. Außerdem ist die Reichweite der Lehrziele unterschiedlich. Das neu segmentierte Lehrziel 1 grenzt weder mögliche mathematischen Verfahren ein, noch die ingenieurwissenschaftlichen Probleme, die es zu lösen gilt. Lehrziel 2 bezieht sich hingegen konkret auf das Lösen einer ganz bestimmten Art von Differentialgleichung. Die Reichweite dieses Lehrziels ist als geringer einzustufen als die von Lehrziel 1.

Die Segmentierungen wurden von einer Kodiererin beziehungsweise einem Kodierer durchgeführt und von einer zweiten Person überprüft. Bei Nichtübereinstimmung folgte eine Konsensvalidierung, so dass eine vollständige Übereinstimmungsquote erreicht wurde. Für die Einschätzung

der Prüfungspraxis, die vor Beginn des Programms existierte, standen somit 353 Prüfungsaufgaben und 37 Lehrziele zur Verfügung. Die neu erstellten Prüfungen bestehen insgesamt aus 319 Aufgaben, bei 35 neu erstellten beziehungsweise überarbeiteten Lehrzielen. In Tabelle 8 sind die Anzahl der Aufgaben und der Lehrziele für jeden Fall separat dargestellt. Die Fälle MW 5 und Mathe 1, Mathe 2 und Mathe 3 formulierten keine neuen Lehrziele. Die Fälle Mathe 2 und Mathe 3 erstellten zudem im Programmzeitraum keine neue Prüfung.

Tabelle 8: Anzahl der Prüfungsaufgaben und Lehrziele in den bereits existierenden und den neu erstellten Materialien

	Bereits existierende Materialien		Neu erstellte Materialien	
	Anzahl Prüfungs- aufgaben	Anzahl Lehrziele	Anzahl Prüfungs- aufgaben	Anzahl Lehrziele
MW 1	60	4	67	12
MW 2	63	4	64	10
MW 3	45	4	69	8
MW 4	58	4	62	5
MW 5	36	8	33	/
Mathe 1	27	4	24	/
Mathe 2	32	3	/	/
Mathe 3	32	6	/	/
Gesamt	353	37	319	35

4.3.2 Verbaldaten

Zusätzlich zu den materialbasierten Daten wurden Verbaldaten erhoben. Dazu wurden 14 Arbeitstreffen auf Tonband aufgezeichnet. Aufgrund terminlicher Schwierigkeiten variierte die Dauer der Treffen zwischen zwei und knapp vier Stunden. In Tabelle 9 ist die Länge der Arbeitstreffen für jeden Fall aufgeführt.

Tabelle 9: Dauer der Arbeitstreffen pro Fall in Stunden

	Treffen 1	Treffen 2	Treffen 3	Treffen 4	Gesamt
MW 1	02:05	02:01	03:09	03:33	10:48
MW 2	02:05	02:01	03:09	03:33	10:48
MW 3	02:02	02:15	02:03	02:12	08:32
MW 4	02:02	02:15	02:03	02:12	08:32
MW 5	02:16	02:10	02:18	02:05	08:49
Mathe 1	02:08	03:55	/	/	06:03
Mathe 2	02:08	03:55	/	/	06:03
Mathe 3	/	03:55	/	/	03:55
Gesamt	14:46	22:27	12:42	13:25	63:20

Der Ablauf der Arbeitstreffen wurde für die Teams aus dem Maschinenwesen standardisiert, um sicherzustellen, dass die wissenschaftlich relevanten Fragen an alle Teilnehmenden gestellt werden konnten. Der erste Teil der Arbeitstreffen bestand aus einem Fragenteil, der sich auf das jeweils behandelte Thema bezog. Das Thema des ersten Arbeitstreffens war die Festlegung des Prüfungsgegenstands. Die Teilnehmer beantworteten im ersten Teil des Treffens unter anderem Fragen dazu, wie sie bei der Auswahl von Prüfungsinhalten bislang vorgehen und welche Auswahlvorschriften sie dabei zugrunde legten. Im Anschluss an den Fragenteil folgte ein Input der Prüfungsexperten zum behandelten Thema. Die Inputphase wurde interaktiv gestaltet. Im dritten Teil wendeten die Teilnehmenden den Input auf die bestehenden Materialien (Lehrziele, Prüfungsaufgaben, Bewertungsraster, etc.) an, um diese gemeinsam mit den Prüfungsexperten zu überarbeiten. In den Überarbeitungsphasen wurden gezielt weitere Fragen eingebaut, um vertiefende Einblicke auf Prozessebene zu erhalten. Ab dem zweiten Arbeitstreffen wurde dem Fragenteil ein Reflexionsteil vorangestellt, in dem Fragen zu Problemen und Fortschritten in der vorangegangenen Arbeitsphase gestellt wurden. Aufgrund der unterschiedlichen Zielsetzungen der Teams und dem spontanen Aufgreifen von Fragen, die während der Arbeitstreffen aufkamen, unterschieden sich die inhaltlichen Ausdifferenzierungen innerhalb der Arbeitstreffen trotz des standardisierten Ablaufs voneinander. Dadurch ergaben sich Unterschiede hinsichtlich der Tiefe der Informationen, die für verschiedene Themenbereiche pro Team zur Verfügung stehen.

4.4 Datenauswertung

Für die Analyse der beschriebenen Datenquellen wurden unterschiedlichen Verfahren eingesetzt. Die Auswertung der materialbasierten Daten erfolgte durch eine Zuordnung von Lehrzielen, Prü-

fungen und Prüfungsaufgaben zu deduktiv gebildeten Hauptkategorien, die basierend auf den Ausführungen im Theorieteil zu bedeutsamen Qualitätsmerkmalen für Lehrziele, Prüfungen und Prüfungsaufgaben gebildet wurden (siehe Tabelle 6). Die Hauptkategorien werden weiter in Unterkategorien differenziert. Für jede der Unterkategorien sind in einem Kodierleitfaden Zuordnungsregeln definiert, durch die festgelegt ist, welche Eigenschaften eine Textstelle beziehungsweise Aufgabe aufweisen muss, um einer Kategorie zugeordnet werden zu dürfen (Mayring, 2001). Die Kategoriensysteme sind im Anhang in Kapitel A hinterlegt. Die Materialien (Lehrziele, Prüfungen und Prüfungsaufgaben) wurden von zwei Personen unabhängig voneinander kodiert. Die mit diesem Vorgehen erzeugten Daten stellen Auftretenshäufigkeiten von Kategorien dar, die in einem zweiten Analyseschritt quantitativ weiterverarbeitet werden (ebd.).

Die Verbaldaten bestehen aus den Aufzeichnungen der Arbeitstreffen (siehe Tabelle 9). Für die Auswertung wurden die Gespräche transkribiert. Als Analyseverfahren kam die Methode der induktiven Kategorienbildung zum Einsatz. Prinzipiell wäre auch eine deduktive Auswertung der Verbaldaten denkbar gewesen, indem aus den Prinzipien der pädagogischen Diagnostik und dem dokumentierten Vorgehen bei einer professionellen Testerstellung ein Soll-Stand für die Prüfungsentwicklung und -auswertung abgeleitet und durch Kategorien beschrieben worden wäre, um dann die Textpassagen den Kategorien zuzuordnen. Mit diesem Vorgehen könnte festgestellt werden, ob eine Kategorie erfüllt oder nicht erfüllt wird. Eine Beschreibung des tatsächlichen Ist-Stands der Prüfungspraxis und wie sich Prinzipien pädagogischer Diagnostik unter den gegebenen Rahmenbedingungen in der Prüfungspraxis niederschlagen und welche Strategien und Problemlösungen sich bewährten, kann dadurch jedoch nicht erreicht werden. Daher wurde für die Auswertung der Verbaldaten ein induktives Vorgehen gewählt, um den Ist-Stand der Prüfungspraxis zu charakterisieren. Das Vorgehen bei der induktiven Kategorienbildung ist geprägt durch ein systematisches Sichten, Ordnen und Verdichten von Textpassagen, in denen die Teilnehmenden ihr Vorgehen bei der Prüfungserstellung beschreiben. Aus den getroffenen Aussagen heraus werden Kategorien gebildet, die das Vorgehen bei der Prüfungserstellung charakterisieren. Der Prozess der induktiven Kategorienbildung wurde von einer Person durchgeführt. Um die interpretativen Prozesse und Annahmen bei der Bildung von Kategorien intersubjektiv nachvollziehbar darzulegen, wird für jede Kategorie auf die zugrunde gelegten Zitate verwiesen. An zwei Beispielen werden die interpretativen Prozesse verschriftlicht. Das Vorgehen bei der Datenanalyse wird in den beiden folgenden Unterkapiteln ausführlich beschrieben.

4.4.1 Auswertung der materialbasierten Daten

Die materialbasierten Daten wurden mit deduktiv gebildeten Kategoriensystemen kodiert. Es wird zwischen drei Analyseeinheiten unterschieden: (1) den formulierten Lehrzielen, (2) der Prüfung

als Ganzes und (3) den Einzelaufgaben. Zunächst wird das Vorgehen bei der Entwicklung der Kategoriensysteme beschrieben und ein Auszug dargestellt. Die vollständigen Kategoriensysteme sind im Anhang enthalten. Abschließend wird der durchlaufene Kodierprozess beschrieben. Vor Beginn der Kodierungen erfolgte eine Schulung der Kodierinnen und Kodierer, die das Kodiersystem an Probeklausuren testeten. Anschließend wurden die Lehrziele, Prüfungen und Prüfungsaufgaben kodiert. Nach dem ersten Kodierdurchlauf wurden die Übereinstimmungen zwischen den Kodierern bestimmt. Als Maß für die Übereinstimmung diente der Cohens-Kappa-Koeffizient κ (Wirtz & Caspar, 2002). Die Items, bei denen keine ausreichende Übereinstimmung vorlag, wurden nochmal überarbeitet, um damit in eine zweite Kodierrunde zu gehen. Die Kodierer wurden in die überarbeiteten Items eingeführt. Nach der zweiten Kodierrunde wurde die Übereinstimmung zwischen den Kodierern erneut bestimmt. Bei Items, die lediglich eine zufriedenstellende, jedoch keine gute Übereinstimmung, aufwiesen, erfolgte eine Konsensvalidierung, um eine möglichst hohe Datenqualität zu erreichen. Für die entsprechenden Items werden die κ -Werte von vor und nach der Konsensvalidierung berichtet.

4.4.1.1 Deduktive Entwicklung der Kategoriensysteme

In diesem Abschnitt werden das Vorgehen bei der Erstellung der eingesetzten Kategoriensysteme sowie deren inhaltliche Bestandteile beschrieben. Die vollständigen Kategoriensysteme sind im Anhang einsehbar. Die Kategoriensysteme wurden theoriegeleitet entwickelt. Es kann von Kategoriensystemen gesprochen werden, da ausschließlich nominale oder kategoriale Zuordnungen getroffen werden (Wirtz, 2004). Kategoriensysteme sind abzugrenzen von Rating- oder Einschätzskalen, anhand derer eine graduelle Abstufung von Merkmalsausprägungen erfasst werden kann zum Beispiel in der folgenden Einteilung: 1=trifft nicht zu, 2=trifft eher nicht zu, 3=trifft eher zu, 4=trifft zu (ebd.).

Zur Analyse der Datenquellen wurden drei verschiedene Kategoriensysteme eingesetzt für (1) die Lehrziele, (2) die Prüfung im Gesamten und (3) für die Einzelaufgaben. Ein Kategoriensystem besteht aus Hauptkategorien, die durch Unterkategorien weiter ausdifferenziert werden. Die Haupt- und Unterkategorien bilden auf formaler Ebene zentrale Gestaltungskriterien für Lehrziele, Prüfungen und Einzelaufgaben ab und sind aus den Ausführungen im Theorieteil abgeleitet (siehe Abschnitt 2.1.3 und Tabelle 6).

Im Kategoriensystem zur Analyse der Einzelaufgaben ist beispielsweise die Hauptkategorie „Qualität des Aufgabenstamms“ enthalten, die sich weiter untergliedert in die Unterkategorien *Aufgabenstellung klar ersichtlich*, *Angabe eines Erwartungshorizonts* und *Anzahl der Handlungsaufforderungen*. Die Unterkategorien wiederum enthalten zwei oder mehrere disjunkte Merkmalsausprägungen. Bei der Unterkategorie „Angabe eines Erwartungshorizonts“ muss unterschieden wer-

den, ob im Aufgabenstamm angegeben ist, welche Aspekte in der Antwort für eine vollständige Beantwortung erwartet werden. Die Merkmalsausprägungen dazu lauten „Trifft zu“ oder „Trifft nicht zu“. Die Unterkategorie „Wissensarten“ enthält die vier Merkmalsausprägungen „deklaratives Wissen, prozedurales Wissen, konzeptionelles Wissen und strategisches Wissen“. Zu jeder Merkmalsausprägung werden Kodierregeln definiert und Ankerbeispiele angegeben, um die Zuordnung von Lehrzielen, Einzelaufgaben und Merkmalen in der Prüfung den betreffenden Kategorien zuzuordnen. Bei Kategorien, bei denen im ersten Kodierdurchlauf Übereinstimmungswerte von $\kappa \leq 0,7$ erzielt wurden, erfolgte eine Spezifizierung der Kodierregeln, indem für manche Ankerbeispielen zusätzlich Begründungen angegeben wurden, die erklären, warum das Beispiel der betreffenden Kategorie zugeordnet wird. In Tabelle 10 ist der Aufbau der Kategoriensysteme dargestellt.

Tabelle 10: Aufbau und Darstellung der Kategoriensysteme⁹

Hauptkategorie		
Unterkategorie 1		
Ausprägung	Kodierregel	Ankerbeispiel
Ausprägung 1	Kodierregel zu Ausprägung 1	<i>Ankerbeispiel zu Ausprägung 1</i> <u>Begründung:</u> Begründung darüber, warum das gewählte Ankerbeispiel der jeweiligen Ausprägung zuzuordnen ist.
Ausprägung 2	Kodierregel zu Ausprägung 2	<i>Ankerbeispiel zu Ausprägung 2</i>
Unterkategorie 2		
Ausprägung	Kodierregel	Ankerbeispiel
Ausprägung 1	Kodierregel zu Ausprägung 1	<i>Ankerbeispiel zu Ausprägung 1</i>
Ausprägung 2	Kodierregel zu Ausprägung 2	<i>Ankerbeispiel zu Ausprägung 2</i>

Zur Analyse der Lehrziele wurden die inhaltliche Schwerpunktlegung sowie formale Gestaltungsmerkmale betrachtet. Zur Einteilung der in den Lehrzielen formulierten Anforderungen wird die Kategorie „Wissensarten“ verwendet, die die Ausprägungen (1) deklaratives Wissen, (2) prozedurales Wissen, (3) konzeptionelles Wissen und (4) strategisches Wissen umfasst. Außerdem wird betrachtet, ob in den Lehrzielen ein Kompetenzgrad angegeben wird, also ab welchem Grad des Könnens ein Lehrziel als erreicht gilt. Hinsichtlich der formalen Gestaltungsmerkmale wird überprüft, ob die Lehrziele jeweils eine Inhalts- und eine Verhaltenskomponente beinhalten und

⁹ Für eine bessere Lesbarkeit sind die Zwischenlinien in dieser Tabelle eingeblendet

wie präzise die dabei verwendeten Begriffe formuliert wurden (siehe Anhang Kapitel A.a und Abschnitt 2.1.3.1).

Auf Ebene der Prüfungen sind die Hauptkategorien „Instruktionsqualität“ und „Kenndaten zur Prüfung“ enthalten (siehe Anhang Kapitel A.b.). Die Hauptkategorie Instruktionsqualität enthält die Unterkategorie *Hinweise zur Bearbeitung der Aufgaben*. Darunter fallen Angaben, welche Aspekte bei der Bearbeitung der Aufgaben zu beachten sind, zum Beispiel die Angabe, dass stets Lösungswege anzugeben sind. In einer zweiten Kategorie wird überprüft, ob Hinweise zur Bewertung von Prüfungsantworten angegeben sind. Ein Beispiel hierfür ist, dass ein Ergebnis nur bewertet wird, wenn der Lösungsweg nachvollziehbar ist. Weitere Merkmale der Instruktionsqualität sind die Gliederung der Prüfung durch Überschriften, um die Orientierung und Navigation zu erleichtern sowie eine einheitliche Formatierung. Ablesbare Kenndaten zur Prüfung sind die Bearbeitungsdauer sowie allgemeine Hinweise, zum Beispiel zur Verwendung von Hilfsmitteln.

Das Kategoriensystem zur Analyse der Einzelaufgaben (siehe Anhang Kapitel A.a) gliedert sich in sechs Hauptkategorien. Zunächst werden die vorkommenden Aufgabentypen unterschieden. Unterkategorien dazu sind der Aufgabentyp (z.B. Multiple Choice oder offene Aufgabe), die verwendeten Handlungsaufforderungen (z.B. benennen Sie, berechnen Sie, definieren Sie, etc.), die Form der Antwort (z.B. Prosa, Formalsprache, Zeichnung, etc.) und die Verwendung von Illustrationen (z.B. Ablesen aus einer Grafik oder ein Diagramm zeichnen). Neben dem Aufgabentyp werden sprachliche Merkmale in den Aufgaben betrachtet. Es wird festgestellt, ob die Aufgaben in Frage- oder Aussageform formuliert sind, inwieweit die Satzstruktur der Aufgabenstellungen einfach gehalten ist und ob doppelte Verneinungen vorkommen, die zu Irritationen der Studierenden führen können. Die Instruktionsqualität auf Ebene von Aufgaben bildet die nächste Hauptkategorie und enthält die Unterkategorien *keine unnötigen Informationen im Aufgabenstamm*, *Vorstrukturierung des Antwortfelds* und *die Aufgabe enthält nur eine Handlungsaufforderung*. Daran anschließend wird überprüft, inwieweit die einzelnen Aufgaben unabhängig voneinander bearbeitbar sind beziehungsweise ob bei bestehenden Abhängigkeiten aufeinanderfolgender Aufgaben Zwischenergebnisse angegeben werden. Die Hauptkategorie „Inhaltliche Gestaltungsmerkmale“ enthält Unterkategorien zu den durch die Aufgaben angesprochenen Wissensarten, die Aufgabenoffenheit, die Funktion von Aufgabenkontexten und die Komplexität von Aufgaben. Bei den Einzelaufgaben werden zudem Angaben zur Korrektur, wie das Vorhandensein einer Musterlösung und eines Bewertungsschemas, sowie Hinweise zur Bepunktung erfasst.

4.4.1.2 Ablauf der Kodierungen

In diesem Abschnitt wird der Ablauf der Kodierungen beschrieben. Nach Fertigstellung der Erstversion der Kategoriensysteme erfolgten zunächst eine Schulung der Kodierinnen und Kodierer

und eine Pilotierung der Kategoriensysteme an Probematerialien. Mit den Erkenntnissen aus den Probekodierungen wurden die Kategoriensysteme überarbeitet, um aufgetretene Probleme zu beseitigen. Häufig handelte es sich in dieser Phase um strukturelle Aspekte in den Prüfungen, zum Beispiel wie entschieden wird, ob eine Teilaufgabe als Folgeaufgabe gewertet wird (als Folgeaufgaben werden Aufgaben gewertet, die auf vorangegangenen Aufgaben aufbauen). Inhaltliche Unklarheiten kamen zumeist erst nach der ersten Kodierrunde und nach Bestimmung der Inter-Kodierer-Übereinstimmung bei den Konsensvalidierungen zum Vorschein.

Nach dem ersten Kodierdurchlauf wurden Kappa-Werte berechnet, um die Übereinstimmung zwischen den Kodierenden zu bestimmen. Für einige Kategorien ergaben sich trotz hoher prozentualer Übereinstimmung niedrige Kappa-Werte, was durch Randverteilungen bei bestimmten Merkmalsausprägungen erklärt wird (Wirtz & Kutschmann, 2007). Kommt beispielsweise bei der Kategorie „Wissensart“ die Ausprägung „strategisches Wissen“ insgesamt zehn Mal vor, bei insgesamt 353 Aufgaben in den bestehenden Prüfungen, und unterscheiden sich die Einschätzungen der kodierenden Personen für das Auftreten dieser Merkmalsausprägung nur wenige Male, sinkt der Kappa-Wert relativ stark, auch wenn bei den restlichen 343 Aufgaben eine sehr hohe prozentuale Übereinstimmung vorliegt. Diese Eigenschaft des Kappa-Werts ist bei Merkmalen mit einer geringen Grundhäufigkeit, wie dies für die Ausprägung Strategiewissen erwartet werden kann, von Bedeutung, um das Auftreten solcher Merkmale möglichst treffsicher festzustellen. Die Betrachtung der prozentualen Übereinstimmung für die gesamte Kategorie *Wissensarten* könnte Nichtübereinstimmungen bei selten vorkommenden Ausprägungen kaschieren (Wirtz & Kutschmann, 2007).

Im nächsten Schritt erfolgte eine Überarbeitung der Kategorien, bei denen die Übereinstimmungswerte nicht ausreichend waren. Dabei wurden insbesondere die Kodierregeln präzisiert, Ankerbeispiele ergänzt und zum Teil Begründungen angegeben, die erläutern, warum ein bestimmtes Ankerbeispiel einer Kategorie zuzuordnen ist. Nach einer erneuten Schulung mit den überarbeiteten Kategoriensystemen wurden die Materialien rekodiert. Basierend auf einer erneuten Bestimmung der Inter-Kodierer Übereinstimmung wurde entschieden, von welchen Kategorien die Daten direkt für weitere Analysen übernommen werden, für welche Kategorien eine Konsensvalidierung vorgenommen wird und welche Kategorien aufgrund nicht zufriedenstellender Übereinstimmungen bei den unabhängigen Kodierungen aussortiert werden sollten. Die Güte der erzielten Übereinstimmungen wird eingeteilt in „fehlende oder schwache“, „mäßige“, „zufriedenstellende“, „gute“ und „sehr gute“ Übereinstimmung, wie sie in Tabelle 11 dargestellt ist. Diese Einteilung orientiert sich am Vorschlag von Wirtz und Kutschmann (2007), jedoch wird deren Kategorie „gute Übereinstimmung“, die sich im Wertebereich von $0,60 \leq \kappa \leq 0,74$ bewegt, noch-

mals differenziert in die Kategorien „zufriedenstellende Übereinstimmung“ ($0,60 \leq \kappa \leq 0,69$) und „gute Übereinstimmung“ ($0,70 \leq \kappa \leq 0,74$).

Kategorien bei denen bei der Rekodierung lediglich eine schwach ausgeprägte oder fehlende Übereinstimmung erzielt werden konnte, wurden für die weiteren Analysen aussortiert. Bei Kategorien mit mäßiger oder zufriedenstellender Übereinstimmung wurde die Qualität der Daten per Konsensvalidierung nochmals erhöht, um eine gute oder sehr gute Übereinstimmung zu erhalten. Bei Kategorien mit zufriedenstellender Übereinstimmung wurden lediglich die Merkmalsausprägungen konsensvalidiert, die sich als problematisch erwiesen. Dazu wurde zunächst betrachtet, bei welcher Ausprägung innerhalb einer Kategorie die Anzahl der Kodierungen am geringsten war (z.B. Ausprägung strategisches Wissen bei der Kategorie Wissensarten), um dann die Aufgaben zu identifizieren, bei denen die Kodierung nicht übereinstimmte. Bei Kategorien mit guter und sehr guter Übereinstimmung wurden die Daten für die weiteren Analysen unverändert übernommen.

Tabelle 11: Einteilung der Übereinstimmungsgüte von Kodierungen

Güte der Übereinstimmung	Maßnahme
$\kappa \geq 0,75$: sehr gute Übereinstimmung	Daten werden für weitere Analysen direkt übernommen.
$0,70 \leq \kappa \leq 0,74$: gute Übereinstimmung	Daten werden für weitere Analysen direkt übernommen.
$0,60 \leq \kappa \leq 0,69$: zufriedenstellende Übereinstimmung	Konsensvalidierungen problematischer Merkmalsausprägungen.
$0,40 \leq \kappa \leq 0,59$: mäßige Übereinstimmung	Konsensvalidierung bei allen Merkmalsausprägungen.
$\kappa \leq 0,39$: fehlende oder schwache Übereinstimmung	Kategorien werden aussortiert und nicht für weitere Auswertungen verwendet.

Um durchgängig eine möglichst hohe Datenqualität zu erreichen, erfolgte gemäß der Einteilung in Tabelle 11 die Durchführung einer Konsensvalidierung. Die Kodiererinnen und Kodierer betrachteten für jede Kategorie die Fälle, bei denen keine Übereinstimmung vorlag, um gemeinsam zu entscheiden, welcher Ausprägung der entsprechende Fall zuzuordnen ist. Gelang es auf Basis der zur Verfügung stehenden Informationen nicht, zu einem übereinstimmenden Urteil zu gelangen, wurden die Lehrziel- beziehungsweise Prüfungs- und Aufgabenersteller direkt konsultiert, um weitere Informationen zu erhalten. Im folgenden Unterkapitel werden die Übereinstimmungswerte für alle Kategorien berichtet.

4.4.1.3 Erzielte Übereinstimmungswerte

Als Maß für die Interrater Übereinstimmung wurde, wie schon beschrieben, Cohens-Kappa verwendet. Es werden für jede Kategorie, die bei der Rekodierung, also beim zweiten Kodierdurchlauf nach Überarbeitung der Kategoriensysteme, erzielten Kappa-Werte aufgeführt. Bei den Kategorien mit ursprünglich zufriedenstellenden oder mäßiger Übereinstimmung werden die Kappa-Werte von vor und nach der Konsensvalidierung berichtet.

Übereinstimmungswerte bei der Kodierung der Lehrziele

Tabelle 12: Erzielte Übereinstimmungswerte bei der Kodierung der Lehrziele

Kategorie	κ -Wert Kodierung	κ -Wert nach Konsensvalidierung
Kategorie 1: Wissensart	0,59	0,96
Kategorie 2: Angabe eines Kompetenzgrads	¹⁰	¹⁰
Kategorie 3: Lehrziel enthält Inhaltskomponente	¹¹	1
Kategorie 4: Spezifität der Inhaltskomponente	0,23 ¹²	0,72
Kategorie 5: Lehrziel enthält Verhaltenskomponente	0,61	0,87
Kategorie 6: Spezifität der Verhaltenskomponente	0,30	0,92

¹⁰ Für diese Kategorie wurde kein Kappa-Wert ausgegeben, da bei einem Kodierer nur eine Ausprägung auftritt und somit der Kappa Wert eine Konstante ist. Die prozentuale Übereinstimmung liegt bei 99,0 Prozent

¹¹ Für diese Kategorie wurde kein Kappa-Wert ausgegeben, da bei einem Kodierer nur eine Ausprägung auftritt und somit der Kappa Wert eine Konstante ist. Die prozentuale Übereinstimmung liegt bei 97,1 Prozent

¹² Die Kategorie wurde trotz niedrigem Kappa Wert bei der Erstkodierung in die Konsensvalidierung aufgenommen, da die Unterschiede in den Einschätzungen dadurch zustande kamen, dass der Zweitkodierer die Kodierung nicht anhand des aktuellsten Kategoriensystems vorgenommen hatte. Bei der Konsensvalidierung konnten daher alle Fälle mit Nichtübereinstimmung problemlos geklärt werden.

Übereinstimmungswerte bei der Kodierung der Prüfungen im Gesamten

Tabelle 13: Erzielte Übereinstimmungswerte bei der Kodierung der Prüfung im Gesamten

Kategorie	κ -Wert Kodierung	κ -Wert nach Konsensvalidierung
Kategorie 1: Hinweise zur Bearbeitung der fungsaufgaben	0,84	/
Kategorie 2: Hinweise zur Bewertung von Antworten	1	/
Kategorie 3: Prüfung enthält Überschriften	0,83	/
Kategorie 4: Angabe zur Bearbeitungsdauer	1	/
Kategorie 5: Allgemeine Hinweise zur Bearbeitung der Prüfung	0,63	1

Übereinstimmungswerte bei der Kodierung der Einzelaufgaben

Tabelle 14: Erzielte Übereinstimmungswerte bei der Kodierung der Einzelaufgaben

Kategorie	κ -Wert Kodierung	κ -Wert nach Konsensvalidierung
Kategorie 1: Aufgabentyp	0,83	/
Kategorie 2: Verwendete Handlungsaufforderung	0,84	/
Kategorie 3: Form der Antwort	0,85	/
Kategorie 4: Aufgabe enthält eine Illustration	0,63	0,91
Kategorie 5: Verwendung der Illustration	0,68	0,86
Kategorie 6: Aufgabenstamm in Frage- oder Auss- geform	0,96	/
Kategorie 7: Der Aufgabenstamm ist mit einfacher Satzstruktur formuliert	0,57	0,57
Kategorie 8: Der Aufgabenstamm enthält doppelte Verneinungen	¹³	/
Kategorie 9: Der Aufgabenstamm enthält keine nötigen Informationen	0,51	1
Kategorie 10: Erwartungshorizont klar definiert	0,50	0,99

¹³ Für diese Kategorie wurde kein Kappa Wert ausgegeben, da nur die Kategorie „Trifft nicht zu“ kodiert wurde. Die prozentuale Übereinstimmung zwischen den Kodierern liegt bei 100 Prozent.

Kategorie 11: Vorstrukturierung des Antwortfelds	0,62	0,99
Kategorie 12: Aufgabenstamm enthält nur eine Handlungsaufforderung	0,95	/
Kategorie 13: Anzahl der vergebenen Punkte	0,59	0,98
Kategorie 14: Die Aufgabe ist vom Ergebnis der vorangegangenen Aufgabe abhängig	0,72	0,86
Kategorie 15: Zwischenlösungen sind angegeben	0,72	0,86
Kategorie 16: Wissensarten in den Aufgaben	0,72	0,99
Kategorie 17: Aufgabenoffenheit	0,53	0,96
Kategorie 18: Kontext- und Situationsbezug	0,63	0,98
Kategorie 19: Komplexität von Aufgaben	0,51	0,99
Kategorie 20: Ankerbeispiele richtige Antworten	0,91	/
Kategorie 21: Ankerbeispiele falsche Antworten	0,85	/
Kategorie 22: Ankerbeispiele teilweise richtige Antworten	0,90	/
Kategorie 23: Regelwerk richtige Antworten	0,86	/
Kategorie 24: Regelwerk falsche Antworten	0,94	/
Kategorie 25: Regelwerk teilweise richtige Antworten	0,87	/
Kategorie 26: Punktevergabe richtige Antworten	0,87	/
Kategorie 27: Punktevergabe falsche Antworten	0,85	/
Kategorie 28: Punktevergabe teilweise richtige Antworten	0,83	/

4.4.2 Auswertung der Verbaldaten

Das Ziel bei der Auswertung der Verbaldaten war es, mittels induktiver Kategorienbildung das Vorgehen der Teilnehmenden bei der Prüfungserstellung und –auswertung zu analysieren und systematisch zu beschreiben. Der Prozess der induktiven Kategorienbildung ist geprägt durch ein systematisches Sichten, Ordnen und Verdichten von Text- beziehungsweise Gesprächspassagen (Flick, 2010). Ausgewählte Zitate aus den vorhandenen Materialien fungieren als Datengrundlage zur Bildung von Kategorien, die das untersuchte Phänomen charakterisieren. Bei einer Kategorienbildung anhand dieses Vorgehens kommen interpretative Prozesse der Person zum Tragen, die diese Auswertungen durchführt. Um das Zustandekommen der gebildeten Kategorien intersubjektiv überprüfbar zu machen, müssen die einer Kategorie zugrunde gelegten Zitate als Datengrund-

lage verfügbar sein. Nachfolgend wird beschrieben, wie das Prinzip der induktiven Kategorienbildung in der vorliegenden Arbeit umgesetzt wurde.

4.4.2.1 Vorstrukturierung der Textpassagen

In einem ersten Schritt erfolgte die Transkription der Audioaufzeichnungen der Arbeitstreffen. Für jeden der Fälle aus dem Maschinenwesen liegen jeweils vier Transkripte vor. Insgesamt beläuft sich die Anzahl der Transkripte auf 20 Stück. Der Umfang der Transkripte variiert zwischen 35 und 75 Seiten. Im Anschluss an die Transkription wurden die Texte segmentiert, um die Textpassagen zu strukturieren und auf theoretische bedeutsame Aspekte der Prüfungspraxis zu fokussieren. Die Aussagen der Teilnehmenden wurden dabei den sechs Schritten der Prüfungserstellung beziehungsweise deren Unterkategorien zugeordnet. Außerdem wurde unterschieden, ob sich eine Aussage zu einem bestimmten Themenbereich retrospektiv auf die bestehende Praxis bezog oder auf die veränderte Praxis. Als Grundlage für die Segmentierung der Textpassagen dienten zwei identische Kategoriensysteme, in denen die sechs Schritte der Prüfungserstellung und den zugehörigen Teilschritten abgebildet sind. In eines der Kategoriensysteme wurden Textpassagen eingeordnet, die sich auf die Prüfungspraxis vor Beginn des Qualitätsentwicklungsprogramms beziehen. Das zweite Kategoriensystem wurde für Textpassagen verwendet, die die veränderte Praxis beschreiben. Die Kategoriensysteme sind äquivalent zu den Ausführungen in Abschnitt 2.1.3 aufgebaut, in dem die sechs Schritte zur Prüfungserstellung ausführlich theoretisch beschrieben sind. Ein Auszug aus dem Kategoriensystem ist in Tabelle 15 dargestellt. Das vollständige Kategoriensystem ist im Anhang enthalten (Tabelle 38).

Eine Zweitkodierung der Segmentierung war nicht erforderlich, da die Auftretenshäufigkeiten der Kategorien nicht als Datenmaterial zur Beantwortung der Forschungsfragen herangezogen wurden. Die Segmentierung anhand des dargestellten Kategoriensystems diente lediglich der Vorstrukturierung des Textmaterials, um die nachfolgenden qualitativen Analysen zu erleichtern. Trotz des standardisierten Ablaufs der Arbeitstreffen waren die Informationen zu den einzelnen Themenbereichen häufig über mehrere Treffen verteilt. Dies lag zum einen daran, dass sich die inhaltliche Schwerpunktlegung innerhalb der Treffen an den Projektzielen der Teilnehmenden ausrichtete. Manche der Fragen zur Aufklärung der Prüfungspraxis konnten zudem nicht direkt im Rahmen des Frageteils am Anfang der Treffen beantwortet werden, da die Teilnehmenden dabei zum Teil nicht auf ihre impliziten Annahmen und Überzeugungen zu sprechen kamen. Diese Aspekte kamen häufig erst in den Arbeits- oder Reflexionsphasen zum Vorschein.

Tabelle 15: Auszug aus dem Kategoriensystem zur Segmentierung der Verbaldaten

Festlegung des Prüfungsgegenstand (bestehende Praxis)		
Kategorie 1: Prüfungsanforderungen		
Ausprägung	Kodierregel	Ankerbeispiel
Abzuprüfende Fähigkeiten und Kompetenzen	Aussagen darüber, was Studierende wissen oder können sollen, welche Fähigkeiten oder Kompetenzen geprüft werden sollen.	<i>In unserer Veranstaltung lernen die Studierenden die verschiedenen Konzepte der Unternehmensführung kennen.</i> <i>Wir wollen dass die Studierenden Zusammenhänge zwischen den behandelten Themenbereichen herstellen können.</i>
Kategorie 2: Konzeptionelle Analyse des Prüfungsgegenstandes		
Ausprägung	Kodierregel	Ankerbeispiel
Vorgehen und Bezugspunkte bei der Auswahl und Festlegung von Lehrzielen und Prüfungsinhalten	Aussagen darüber, wie die Lehrziele und Prüfungsinhalte ausgewählt werden, warum die angestrebten Kompetenzen und Fähigkeiten wichtig sind bzw. warum diese Kompetenzen und Fähigkeiten angestrebt werden. Wie wird entschieden was Studierende am Ende einer Lehrveranstaltung können sollen?	<i>Ich nehme das Skript der Veranstaltung und überlege mir dann welche Frage ich zu den Inhalten stellen könnte.</i>
Kategorie 3: Bildung und Strukturierung von Teilkompetenzen, Festlegung von Bearbeitungsniveaus		
Ausprägung	Kodierregel	Ankerbeispiel
Strukturierung von Kompetenzen und Inhalten	Aussagen dazu, wie Kompetenzen aufgebaut sind, ob gewisse Inhalte als Grundlage für andere Inhalte gesehen werden.	<i>Damit die Studierenden in der Lage sind den Prozess zu erläutern benötigen sie Wissen über die Bestandteile des Prozesses.</i>
Unterscheidung von Kompetenz- und inhaltlichen Niveaus	Aussagen dazu, dass Anforderungen auf bestimmten Niveaus ausgeführt werden können. Abstufungen in der Qualität der Bewältigung von Aufgaben	<i>Wir unterscheiden, ob Studierende die Begründung in Fachsprache formulieren und auf wissenschaftliche Theorien zurückgreifen oder ob lediglich Allerweltsaussagen kommen.</i>
Kategorie 4: Formulierung von Lehrzielen		
Ausprägung	Kodierregel	Ankerbeispiel
Vorgehen bei der Formulierung von Lehrzielen	Aussagen, zum Vorgehen bei der Formulierung von Lehrzielen beziehen und welche Kriterien beachtet werden	<i>Bei der Formulierung von Lehrzielen achten wir darauf, dass die verwendeten Begrifflichkeiten eindeutig sind.</i>
Kategorie 5: Lehrzielmatrix		
Ausprägung	Kodierregel	Ankerbeispiel
Nutzung einer Lehrzielmatrix	Aussagen darüber, ob das Konzept der Lehrzielmatrix genutzt wird und falls ja auf welche Weise	<i>Wir nutzen die Lehrzielmatrix, um darzustellen wie viele Punkte pro Lehrziel in der Prüfung vergeben werden sollen.</i>

4.4.2.2 Induktive Bildung von Kategorien

Im Zuge der induktiven Bildung von Kategorien wurden die Aussagen der Teilnehmenden im Anschluss an die Strukturierung der Textpassagen anhand des vorgestellten Kategoriensystems fallweise gesichtet und verdichtet, um daraus Kategorien abzuleiten, die das Vorgehen der Teilnehmenden hinsichtlich der verschiedenen Themenbereiche charakterisieren. Die dabei notwendigen interpretativen Prozesse und die Reichweite der angestellten Inferenzen können sich unterscheiden. Je eindeutiger die Teilnehmenden ihr Vorgehen, ihre Annahmen oder Überzeugungen artikulieren, desto geringer ist das Ausmaß der induktiven Prozesse zur Bildung einer entsprechenden Kategorie. Sind die Ausführungen zu einem Themenbereich diffuser oder vielschichtiger, müssen die entsprechenden Aussagen zueinander in Beziehung gesetzt werden, um deren Bedeutung zu verdichten. Die Reichweite der Inferenzen bei den interpretativen Prozessen, um eine Kategorie zu bilden, ist dadurch gegebenenfalls höher.

Das beschriebene Prinzip zur induktiven Bildung von Kategorien wird an einem Beispiel verdeutlicht. Die Teilnehmenden wurden in den Arbeitstreffen dazu befragt, welche Inhalte und Kompetenzen sie für ihre Veranstaltung beziehungsweise Prüfung als wichtig erachten (Item 1 im Kategoriensystem in Tabelle 15). Die Antworten wurden nun betrachtet und die getroffenen Aussagen zu Sinneinheiten verdichtet, um daraus eine (oder mehrere) Kategorie(n) abzuleiten, die charakteristisch für die als wichtig erachteten Inhalte und Kompetenzen ist. Der beschriebene Prozess wird an folgenden Zitaten, die von den Teilnehmenden von Fall 1 stammen, illustriert:

Interviewer: Was sollen die Studierenden am Ende der Veranstaltung können?

Teilnehmer 1: Ok, was ist das, was wir hier in unserer VL vermitteln wollen? Natürlich geht das auch in die Richtung von den Ingenieuren, weil hoffentlich haben wir irgendwo ein gesamtheitliches Bild, das stimmig ist und wir haben Professoren aus der Industrie und damit hat man natürlich auch die Anforderungen aus der Industrie. Oder wir haben ein Industrieprojekt, wo man sich austauscht, ja, was ist denn jetzt das Problem von den aktuellen Ingenieuren, was ist denn gut von den neuen Ansätzen und dann versucht man das natürlich in ein stimmiges Bild zu geben. Und ich sage jetzt mal ganz weit weg sind wir ja davon nicht, aber man muss ja ständig noch besser werden und ich sage jetzt mal das Resümee aus diesen Feedbacks, das kommt dann letztendlich auch wieder in die Vorlesung hinein. Wo man merkt, ok, was fehlt, es gibt sehr gute Spezialisten, aber wir wollen jetzt mit unserer, *mit dieser Vorlesung mehr den Allrounder, also die gesamtheitliche Betrachtung, fördern. Deswegen eben der Fokus auf zentrale Konzepte anstatt die Auslegung von kleinen Teilprozessen.* Sondern da geht es ein bisschen größer und mit dieser Vorlesung adressieren wir das.

Interpretative Prozesse zur induktiven Bildung von Kategorien: Die Aussagen zu den Veranstaltungsinhalten deuten auf zwei Schwerpunktlegungen hin. Der kursiv gekennzeichnete Abschnitt im Zitat benennt relativ konkret, dass der Fokus der Veranstaltung weniger auf Detailwissen liegt, sondern sich auf die Vermittlung eines ganzheitlichen Verständnisses zentraler Konzepte des Fachs konzentriert. Die restlichen Ausführungen deuten darauf hin, dass die Veranstaltung auf die berufliche Praxis vorbereiten soll, da darauf geachtet wird, Anforderungen und Ansätze aus der Industrie aufzunehmen. Zwei weitere Zitate verdichten diesen Eindruck:

Teilnehmer 1: [...] Ok, also für mich ist es jetzt erstmal wichtig, ist es das auch wirklich, was wir erzählen wollen. Also, ich meine, ja. Also ich weiß noch, vor allem früher bei Professor x [dem vorherigen Lehrstuhlleiter] war das genau was wir erzählen wollen. Jetzt weiß ich nicht, hat sich das bei Professor y [neuer Lehrstuhlleiter] geändert oder ist das weiterhin wirklich ein Inhalt der relevant ist [den Prozess xy darzulegen]? *Wir wollen jetzt, dass die Leute wissen, wie man das Produkt xy baut [...] anstatt, dass sie für einen Teilprozess eine bestimmte Rechnung durchführen können.*

Teilnehmer 2: Aber wir wollen auch etwas mehr als verstehen glaube ich. Wir wollen doch schon, es geht fast Richtung Anwendung, oder? Wenn ich da nämlich an die Prüfung denke, an den Teil B, da müssen die Studierenden das selber eigentlich teilweise schon anwenden. [...] *Und sie sollten das auch später im Beruf als Ingenieure anwenden können.*

Interpretative Prozesse zur induktiven Bildung von Kategorien fortgesetzt: In diesen Zitaten wird der Anspruch konkretisiert, mit der Veranstaltung auf die berufliche Praxis als Ingenieur vorzubereiten. Zum einen wird von Teilnehmer 1 gesagt, dass es wichtiger ist, zu wissen, wie man ein Produkt in seiner Gesamtheit herstellt, anstatt bestimmte Rechnungen für Teilprozesse zu berechnen. Damit wird auch nochmal der Anspruch an ein ganzheitliches Verständnis (anstatt Detailwissen) bekräftigt. Teilnehmer 2 unterstützt die Aussage hinsichtlich der Praxisorientierung, indem er konkret aussagt, dass die Studierenden in der Lage sein sollen, die Prüfungsinhalte später im Berufsleben anzuwenden. Auf Basis der Zitate und den angestellten interpretativen Prozessen werden nun zwei Kategorien gebildet, die ausdrücken, welche inhaltliche Schwerpunktlegung bei diesem Fall angestrebt werden. Die Ausrichtung der Veranstaltung ist an den Anforderungen der beruflichen Praxis orientiert. Der Name der Kategorie die dies ausdrückt lautet „Praxisorientierung“. Der Fokus der Veranstaltung wird durch die Kategorie „Ganzheitliches Verständnis“ charakterisiert.

Nachdem für Fall 1 die als relevant angesehenen Veranstaltungs- beziehungsweise Prüfungsinhalte durch die Kategorien „Praxisorientierung“ und „Ganzheitliches Verständnis“ charakterisiert wurde, erfolgte die Betrachtung der anderen Fälle zu diesem Themenbereich, um festzustellen,

welche Inhalte als relevant betrachtet werden. Dabei konnten weitere Kategorien gebildet werden, sofern die inhaltlichen Schwerpunktsetzungen durch die bereits gebildeten Kategorien nicht abgedeckt wurden. Im Ergebnisteil werden zu jedem Themenbereich zunächst die gebildeten Kategorien in einer Tabelle im Überblick dargestellt. In der Tabelle ist zudem gekennzeichnet, welche Kategorie auf welchen Fall zutrifft. Trifft eine Kategorie auf einen Fall zu, wird dies mit einem Häkchen (✓) gekennzeichnet. Bei aktiver Ablehnung einer Kategorie wird dies mit einem Minus Zeichen (-) dargestellt. Ein ○ zeigt an, dass die jeweilige Kategorie bei der Beschreibung des Vorgehens nicht aktiv als etabliertes Vorgehen beschrieben wurde, obwohl der Themenbereich, auf den sich die Kategorie bezieht, thematisiert wurde. Aufgrund der flexiblen inhaltlichen Gestaltung der Arbeitstreffen wurden zum Teil bei einigen Lehrstühlen bestimmte Themenbereiche nicht angesprochen. Dies wird durch die Angabe von „k.A.“ (keine Angabe) gekennzeichnet. Treffen auf einen Lehrstuhl innerhalb einer Überkategorie mehrere Unterkategorien zu, sind entsprechend mehrere Häkchen gesetzt. Ist eine Kategorie dominanter als die andere, ist das Häkchen in der dominanteren Kategorie in fett formatiert. In Fußnoten wird jeweils auf die Zitate verwiesen, auf deren Grundlage die Zuordnung der Lehrstühle zu den Kategorien vorgenommen wurde. Die Zitate sind im Anhang der Arbeit einsehbar. Dieses Prinzip ist in Tabelle 16 auszugsweise für die Fälle 1 und 2 dargestellt.

Aus der Tabelle wird ersichtlich, dass bei der Analyse der Aussagen von Fall 2 unter *Ausrichtung* die Kategorie *Wissensorientierung* und unter *Fokus* die Kategorie *Detailwissen* gebildet wurden. In den Aussagen der Teilnehmenden von Fall MW 1 werden eine Wissensorientierung sowie der Fokus auf Detailwissen aktiv abgelehnt. Dies ist in der Tabelle mit einem Minus-Zeichen ausgedrückt. Die Teilnehmenden von Fall 2 wiederum sagten aus, dass es nicht Ziel der Veranstaltung ist, Wissen auf eine Art zu vermitteln, sodass es in der Praxis angewendet werden kann. Dies ist mit einem Minus-Zeichen in der Kategorie *Praxisorientierung* gekennzeichnet. Neben *Detailwissen*, auf dem jedoch der Fokus liegt, soll auch ein ganzheitliches Verständnis gefördert werden. Deshalb wurde sowohl bei der Kategorie *Ganzeitliches Verständnis* als auch bei der Kategorie *Detailwissen* ein Häkchen gesetzt.

Tabelle 16: Beispiel für eine Tabelle zur Darstellung der induktiv gebildeten Kategorien

	Ausrichtung			Fokus	
	Praxis-orientierung	Wissens-orientierung	Mögliche weitere Kategorie	Ganzheitliches Verständnis	Detailwissen
MW 1	✓ ¹⁴	_ ¹⁵		✓ ²¹	_ ²¹
MW 2	_ ¹⁶	✓ ¹⁷		✓ ¹⁸	✓ ²³
MW 3					
MW 4					
MW 5					

Da ein stärkerer Fokus auf *Detailwissen* liegt, ist das dort gesetzte Häkchen fett markiert. Im weiteren Verlauf wurden nun die Aussagen der restlichen Fälle sukzessive ausgewertet und die Ergebnisse in der Tabelle festgehalten. Dabei konnten weitere Kategorien hinzukommen, sofern die inhaltlichen Zielsetzungen der anderen Fälle nicht durch die bereits bestehenden Kategorien abgedeckt sind.

Zusätzlich zur Darstellung der gebildeten Kategorien in der Tabelle wird die Bedeutung der Kategorien jeweils anhand des Vorgehens, beziehungsweise den Überzeugungen und Annahmen der Teilnehmenden, beschrieben, deren Aussagen der Bildung der Kategorien zugrunde gelegt wurden. Diese Beschreibungen vermitteln ein Verständnis über die Bedeutung der Kategorien, bilden jedoch die Tiefe der interpretativen Prozesse, die zur Bildung der Kategorien geführt haben, nicht vollständig ab. Damit das Zustandekommen der Kategorien nachvollziehbar ist, werden die interpretativen Prozesse und die dabei angestellten Überlegungen und Schlussfolgerungen zusätzlich zum oben angeführten Beispiel an zwei weiteren Beispielen exemplarisch illustriert (siehe Anhang Kapitel B; Interpretatives Vorgehen 1, S. 265). Für die Bildung der weiteren Kategorien wird auf eine Verschriftlichung der interpretativen Prozesse verzichtet.

¹⁴ Vgl. Zitat 1, S. 264 und Zitat 2, S. 264

¹⁵ Vgl. Zitat 4, S. 266

¹⁶ Vgl. Zitat 8, S. 267

¹⁷ Vgl. Zitat 7, S. 267

¹⁸ Vgl. Zitat 10, S. 268

5 Ergebnisse

Der Ergebnisteil gliedert sich in die sechs Schritte des Vorgehensmodells zur Erstellung und Auswertung von Prüfungen: (1) Festlegung des Prüfungsgegenstands, (2) Erstellung von Prüfungsaufgaben, (3) Rückbindung der Aufgaben an den Prüfungsgegenstand, (4) Bewertung von Prüfungsantworten, (5) Punkte- und Notenvergabe, (6) Rückschlüsse aus den Prüfungsdaten (siehe Abschnitt 2.1.3). Jeder der sechs Schritte bildet ein Kapitel im Ergebnisteil. Die einzelnen Kapitel zu den sechs Schritten sind wiederum entlang der drei Forschungsfragen untergliedert. Im Rahmen von Forschungsfrage 1 wird untersucht, welche Zielsetzungen die Lehrenden aus den Herausforderungen und Problemen, die sie in ihrer Prüfungspraxis wahrnehmen, für die Arbeit im Qualitätsentwicklungsprogramm ableiten. Forschungsfrage 2 bezieht sich darauf, inwieweit die Lehrenden in ihrer bestehenden Praxis vor Beginn des Qualitätsentwicklungsprogramms Prüfungen kompetenzorientiert gestalteten und dabei die Gütekriterien der pädagogischen Diagnostik beachteten. Als Datengrundlage zur bestehenden Praxis wurden die materialbasierten Daten (bestehende Lehrziele, Prüfungen und Prüfungsaufgaben aus dem vorangegangenen Semester) sowie die Verbaldaten aus den Arbeitstreffen verwendet. Die Ergebnisse zu den Forschungsfragen 1 und 2 dienen als Grundlage für die inhaltliche Schwerpunktlegung des Qualitätsentwicklungsprogramms. Bezüglich Forschungsfrage 3 wurde ermittelt, welche Veränderungen die Lehrenden während der Qualitätsentwicklungsarbeit an ihrer Prüfungspraxis vorgenommen haben und wie sich die Prüfungen dadurch veränderten. Dazu wurden die neu erstellten Lehrziele und Prüfungsaufgaben mit den gleichen Kategoriensystemen analysiert wie bereits die bestehenden Materialien. Die Ergebnisse zu den Forschungsfragen werden jeweils zunächst global über alle Fälle hinweg berichtet, gefolgt von Betrachtungen auf Fallebene, um Unterschiede zwischen den Fällen herauszuarbeiten.

5.1 Festlegung des Prüfungsgegenstandes

Die Festlegung des Prüfungsgegenstands ist der erste Schritt bei der Prüfungserstellung und bildet die Grundvoraussetzung für eine kriterienorientierte Aufgabenerstellung, Korrektur und Punkte- und Notenvergabe. Vor dem Hintergrund von Forschungsfrage 1 werden die Zielsetzungen der Lehrenden hinsichtlich der Festlegung des Prüfungsgegenstands berichtet. Zur Beantwortung von Forschungsfrage 2 wird betrachtet, inwieweit die Lehrenden in ihrer bestehenden Praxis den Prüfungsgegenstand kompetenzorientiert definieren und die dazu notwendigen Teilprozesse durchführen, um den Prüfungsgegenstand theoriegeleitet festzulegen. Im dritten Teilabschnitt wird bezogen auf Forschungsfrage 3 auf die während der Qualitätsentwicklungsarbeit vorgenommenen Änderungen am Vorgehen bei der Festlegung des Prüfungsgegenstands eingegangen.

5.1.1 Zielsetzungen der Hochschullehrenden

Insgesamt beziehen sich vier Programmziele, welche die Teilnehmenden vor dem Qualitätsentwicklungsprogramms formulierten, auf die Festlegung des Prüfungsgegenstands. Die Zielsetzungen hinsichtlich der abzu prüfenden Wissens- und Könnensaspekte sind sehr allgemein formuliert. Zwei Fälle (MW¹⁹ 3 und MW5) formulierten das Ziel, dass sie mit ihrer Prüfung mehr als reines Wissen abfragen möchten. Bei Fall MW 3 müssen in einer Prüfung zwei verschiedenen Studierendengruppen berücksichtigt werden, die unterschiedliche Eingangsvoraussetzungen mitbringen. Ziel der Teilnehmerin von Fall MW 3 ist es bei der Auswahl des Lehr- und Prüfungsstoff die Bedarfe und Interessen beider Gruppen zu berücksichtigen (siehe Tabelle 17). Aufgrund der allgemein gehaltenen Zielsetzungen konnten keine konkreten Hinweise auf das Vorgehen der Lehrenden bei der Festlegung des Prüfungsgegenstands, und ob dabei möglicherweise Probleme auftreten, gewonnen werden.

Tabelle 17: Programmziele der Lehrenden zur Festlegung des Prüfungsgegenstands

Fälle	Projektziele
MW 2	Mit der Prüfung mehr abfragen als nur auswendig gelerntes Wissen, da Beschwerden von Studierenden zur Sinnhaftigkeit der Prüfung vorliegen.
MW 3	Problemen mit Lehramtsstudierenden begegnen, indem die inhaltliche Relevanz für diese Gruppe gesteigert wird.
	Auf Lehramtsstudierende hinsichtlich des Prüfungsstoffs eingehen und dabei die Studierenden in den Ingenieurstudiengängen nicht vernachlässigen.
MW 5	Kompetenzen schaffen, statt nur Wissen vermitteln.

5.1.2 Erkenntnisse zur bestehenden Prüfungspraxis

Bezüglich Forschungsfrage 2 wird hinsichtlich des ersten Schrittes des Prüfungserstellungsprozesses betrachtet, inwieweit die Lehrenden in ihrer bestehenden Prüfungspraxis folgende Teilprozesse zur Festlegung des Prüfungsgegenstands durchführten: (1) Die konzeptionelle Analyse des Prüfungsgegenstands, (2) Zerlegung von Kompetenzen in Teilkompetenzen und Festlegung von Bearbeitungsniveaus, (3) Beschreibung der Anforderungen durch Lehrziele und deren (4) Übertragung in eine Lehrzielmatrix.

Bezüglich der konzeptionellen Analyse wird betrachtet, (1) welche Prüfungsanforderungen die Lehrenden fokussieren (z.B. Wissen vs. Kompetenzen) und (2) inwieweit sie bei der Auswahl und

¹⁹ MW = Maschinenwesen

Identifikation von Anforderungen verschiedene Informationsquellen benutzen (z.B. Lehrziele, Studiengangprofile, Kompetenzbeschreibungen aus dem Qualifikationsrahmen für deutsche Hochschulabschlüsse). Je umfassender sich Lehrende im Rahmen der konzeptionellen Analyse über den Prüfungsgegenstand informieren, desto wahrscheinlicher kann dieser valide und tragfähig definiert werden. Im nächsten Schritt müssen die Lehrenden die Anforderungen in Teilkompetenzen zerlegen und Bearbeitungsniveaus festlegen. Dadurch können zielgerichtet Aufgaben zu unterschiedlichen Kompetenzbereichen und Anforderungsniveaus erstellt werden. Damit kann der diagnostische Wert von Prüfungen erhöht werden. Für Lehrziele ist das entscheidende Kriterium, ob sie konkrete Gestaltungshinweise für die Aufgabenerstellung liefern, indem die Inhalte und Verhaltenskomponenten ausreichend präzise angegeben werden. Weiterhin wird der Frage nachgegangen, ob Lehrende eine Lehrzielmatrix benutzen, um Lehrziele zu strukturieren.

5.1.2.1 Konzeptionelle Analyse des Prüfungsgegenstands

Angestrebte Prüfungsinhalte und -anforderungen

Hinsichtlich der Prüfungsanforderungen ergab sich ein sehr heterogenes Bild zwischen den Fällen (siehe Tabelle 18). Die Lehrenden der Fälle MW 1 und MW 5 zielen laut eigener Aussagen darauf ab, praxisorientiertes Wissen mit einem Fokus auf ganzheitlichem Verständnis der betrachteten Inhalte und Konzepte zu vermitteln, um die Studierenden auf die Anforderungen in der beruflichen Praxis vorzubereiten. Die Lehrenden der Fälle MW 2 und MW 4 berichteten von einer wissensorientierten Ausrichtung mit einem Fokus auf einer detaillierten Wiedergabe von Begriffen, Theorien oder Definitionen. Die Lehrende von Fall MW 3 beschrieb zudem, dass mit ihrer Lehrveranstaltung die Grundlagen für das weitere Studium gelegt werden. Der Teilnehmende von Fall MW 4 beantwortete die Frage danach, was die Studierenden nach Durchlaufen der Veranstaltung können sollen, über eine Beschreibung der in den Prüfungsaufgaben abgebildeten Anforderungen.

Tabelle 18: Kategorien zur Beschreibung der angestrebten Prüfungsanforderungen²⁰

	Ausrichtung				Fokus	
	Praxis-orientiert	Wissens-orientiert	Grundlagen-orientiert	Aufgaben-orientiert	Ganzheitliches Verständnis	Detailwissen
MW 1	✓ ²¹	_ ²²	○	○	✓ ²¹	_ ²¹
MW 2	_ ²³	✓ ²⁴	○	○	✓ ²⁵	✓ ²³
MW 3	○	✓ ²⁶ _ ²⁶	✓ ²⁷	○	○	✓ ²⁷
MW 4	○	✓ ²⁸	○	✓ ²⁸	○	○
MW 5	✓ ²⁹	○	○	_ ³⁰	✓ ²⁹	○

Anmerkungen. ✓ = Trifft zu, - = aktive Ablehnung, ○ = wurde nicht genannt, k.A. = der Themenbereich, auf den sich die Kategorie bezieht, wurde nicht angesprochen. Sind einer Unterkategorie sowohl Häkchen als auch Minuszeichen zugeordnet, bedeutet dies, dass der Ist-Stand (fett gesetzt) zu einem Soll-Stand verändert werden soll.

Nachfolgend wird die Bedeutung der Kategorien anhand des Vorgehens und der Annahmen der Fälle beschrieben, deren Aussagen bei der Bildung der Kategorien zugrunde gelegt wurden.

- ***Praxisorientierung und ganzheitliches Verständnis³¹***

Praxisorientierung bezeichnet eine Ausrichtung der Lehrveranstaltungs- und Prüfungsinhalte an Anforderungssituationen, die den Studierenden in nachgelagerten Berufsfeldern begegnen können. Mit dem Fokus auf *ganzheitlichem Verständnis* möchten die Lehrenden erreichen, dass die Studierenden die behandelten Themenbereiche konzeptionell durchdringen. Diese Zielsetzung wurde von den Fällen MW 1 und MW 5 beschrieben. Bei Fall MW 1 liegt der Fokus der Veranstaltung laut Aussagen der Teilnehmenden auf tiefergreifenden Aspekten. Die tiefergreifenden Aspekte wurden charakterisiert als ein gesamtheitliches Verständnis von zentralen Konzepten des

²⁰ Die im Rahmen der induktiven Analysen gebildeten Kategorien sind in der Tabelle jeweils unter den beiden Überbegriffen Ausrichtung und Fokus, in den einzelnen Spalten dargestellt. Die Fußnoten in der Tabelle verweisen auf qualitative Exzerpte, die als Datengrundlage für die Bildung der Kategorien dienten (für eine genaue Beschreibung des Vorgehens bei der induktiven Kategorienbildung siehe Abschnitt 4.4.2.2).

²¹ Vgl. Zitat 1, S. 264 und Zitat 2, S. 264

²² Vgl. Zitat 4, S. 266

²³ Vgl. Zitat 8, S. 267

²⁴ Vgl. Zitat 7, S. 267

²⁵ Vgl. Zitat 10, S. 268

²⁶ Vgl. Zitat 12, S. 269, Zitat 13, S. 269, Zitat 14, S. 270 und Zitat 15, S. 270

²⁷ Vgl. Zitat 11, S. 269

²⁸ Vgl. Zitat 16, S. 270 und Zitat 17, S. 271

²⁹ Vgl. Zitat 21, S. 273 und Zitat 22, S. 274

³⁰ Vgl. Zitat 21, S. 273

³¹ Vgl. Interpretatives Vorgehen 1, S. 265.

Fachs und der Anwendung auf Praxisprobleme anstelle der spezialisierten Berechnung von Teilprozessen. Diese Zielsetzung war motiviert durch Anforderungen in der beruflichen Praxis, denen Ingenieure im Fachbereich, der durch den entsprechenden Lehrstuhl vertreten wird, begegnen können (vgl. Zitat 1 auf S. 264). Die Teilnehmenden von Fall MW 1 stimmten in ihrem Verständnis über die beschriebenen Prüfungsinhalte und –anforderungen überein, wie Zitat 2 (S. 264) verdeutlicht.

Die *Praxisorientierung*, wie sie hier beschrieben wird, entspricht somit dem im Theorieteil eingeführten Kompetenzbegriff, wonach Kompetenzen als die Befähigung einer Person zur Bewältigung von realen und für eine Domäne relevanten Anforderungssituationen betrachtet werden. Das angestrebte *ganzheitliche Verständnis* von Fachinhalten weist bei näherer Betrachtung der entsprechenden Zitate, die bei der Bildung der Kategorie zugrunde gelegt wurden, die Eigenschaften der Wissensart „konzeptionelles Wissen“ auf (siehe Abschnitt 2.1.1.5).

- *Wissensorientierung und Detailwissen*

Im Gegensatz zu Fall MW 1 liegt der Schwerpunkt bei Fall MW 2 auf der Vermittlung von Wissen über Fakten und Begriffe. Die Lehrenden von Fall MW 2 beschrieben, dass eine fundierte Kenntnis und Wissensbasis bezüglich grundlegenden Begriffen zu den behandelten Themenbereichen als Grundlage für die spätere berufliche Praxis dient (vgl. Zitat 8, S. 267). Es ist jedoch explizit nicht das Ziel, Praxiswissen, also Wissen, das zur Lösung von realen Problemstellungen eingesetzt werden kann, zu vermitteln (vgl. Zitat 8, S. 267). Da der Schwerpunkt in der Prüfung auf einer möglichst exakten Reproduktion der vermittelten Begrifflichkeiten und wissenschaftlichen Definitionen liegt (vgl. Zitat 7, S. 267), wird der Fokus der Prüfungsanforderungen mit der Kategorie *Detailwissen* charakterisiert. Diese Bezeichnung wird durch folgende Aussage der Lehrenden weiter illustriert, wonach die Lehrveranstaltung eine *abstrakte Vorlesung ist, mit wenig Transferwissen. Von daher geht es auch nicht anders als Wissen abzufragen. [...] es ist ja eine wissenschaftliche Veranstaltung und da muss es halt dann eben exakt sein* [das begriffliche Wissen]. Auch für Fall MW 2 kann ein gemeinsames Verständnis hinsichtlich der angestrebten Prüfungsinhalte und -anforderungen attestiert werden, wie die übereinstimmenden Aussagen der Teilnehmenden in Zitat 7 (S. 267) und Zitat 8 (S. 267) belegen. Die in der Veranstaltung vermittelten Inhalte werden zwar an einer Stelle als Voraussetzung für die berufliche Praxis betrachtet, da jedoch von den Lehrenden explizit ausgesagt wird, dass kein Praxiswissen vermittelt werden soll, wird eine kompetenzorientierte Ausrichtung zugunsten einer Fokussierung auf Detailwissen abgelehnt.

- ***Grundlagenorientierte Ausrichtung mit Fokus auf Detailwissen***

Eine grundlagenorientierte Ausrichtung bei den Prüfungsanforderungen würde bei Fall MW 3 festgestellt. Das Ziel der Veranstaltung ist laut Aussage der Teilnehmerin von Fall MW 3, Inhalte zu vermitteln, die als Grundlage in weiterführenden Lehrveranstaltungen vorausgesetzt werden. Die behandelten Inhalte werden zudem als Grundlage für Studienarbeiten am Lehrstuhl angesehen (vgl. Zitat 12, S. 269). Die Ausrichtung der angestrebten Prüfungsanforderungen wird deshalb als grundlagenorientiert bezeichnet, da Grundlagen für den weiteren Studienverlauf gelegt werden sollen. Die Prüfung weist laut Aussage der Teilnehmerin von Fall MW 3 zu hohe Anteile an Wissensreproduktion auf (vgl. Zitat 12, S. 269, Zitat 13, S. 269 und Zitat 15, S. 270). Die Anteile der Anwendungsbezüge sollen erhöht werden (vgl. Zitat 14, S. 270). Daher wird für Fall MW 3 in der Kategorie Wissensorientierung, die angestrebte Verschiebung von *stark wissenlastig* zu mehr *Anwendungsbezug* angezeigt. Der Fokus bei Fall MW 3 wird mit der Kategorie Detailwissen charakterisiert, da in der Veranstaltung die Inhalte, insbesondere grundlegende theoretische Modelle des betreffenden Fachgebiets, sehr detailliert besprochen und entsprechend von den Studierenden wiedergegeben werden sollen (vgl. Zitat 11, S. 269).

- ***Aufgabenorientierte Ausrichtung***

Bei einer aufgabenorientierten Ausrichtung erfolgt die Festlegung der Lehrziele ausgehend von den entwickelten Prüfungsaufgaben. Diese Kategorie wurde anhand der Aussagen des Teilnehmers von Fall MW 4 gebildet. Die Frage danach, was die Studierenden nach Durchlaufen der Veranstaltung können sollen, beantwortete der Teilnehmer durch eine Beschreibung der Aufgabenanforderungen, die in der Prüfung abgebildet sind (vgl. Zitat 16, S. 270 und Zitat 17, S. 271). Das Ziel der Lehrveranstaltung könnte in so einem Fall als die Fähigkeit, Prüfungsaufgaben zu bewältigen, ausgedrückt werden. Die Reichweite einer Kompetenz beschränkt sich daher auf die Lösung der Prüfungsaufgaben. Inwieweit die Prüfungsaufgaben stellvertretend für andere, weiterführende Anforderungssituationen stehen, konnte bei Fall MW 4 zunächst nicht festgestellt werden.

Die Ausrichtung und der Fokus bei Fall MW 5 wurde äquivalent zu Fall MW 1 als *praxisorientiert* mit einem Fokus auf *ganzheitlichem Verständnis* bezeichnet. Bei diesen beiden Fällen entsprechen die angestrebten Prüfungsinhalte und -anforderungen den Merkmalen des in Abschnitt 2.1.1.1 eingeführten Kompetenzbegriffs. Mit der wissensorientierten Ausrichtung bei Fall MW 2, mit einer Ablehnung von Praxisorientierung, wird eine Kompetenzorientierung indirekt abgelehnt. Die grundlagenorientierte Ausrichtung bei Fall MW 3 könnte möglicherweise als kompetenzorientiert bezeichnet werden, sofern systematisch Grundlagen für die Bewältigung weiterführender Anforderungen im Studienverlauf gelegt werden. Andernfalls werden lediglich innerfachliche

Grundlagen vermittelt. Ob die in der Ausrichtung der Zielsetzungen ausgedrückte Kompetenzorientierung im weiteren Verlauf der Prüfungserstellung und –auswertung konsequent in Prüfungsaufgaben umgesetzt wurde, gilt es zu klären.

Referenzpunkte und Auswahlvorschriften bei der Festlegung von Prüfungsanforderungen

Die Ergebnisse zu den Referenzpunkten und Quellen, die die Lehrenden bei der Bestimmung von Prüfungsinhalten und -anforderungen zugrunde legten, zeigen, dass bei allen Fällen ausschließlich die bestehenden Veranstaltungsunterlagen oder Aufgabenpools, die in Form von Altklausuren oder, wie im Falle einer Person aus MW 5, in einem Tabellenkalkulationsprogramm, angelegt sind (siehe Tabelle 19). Die Fälle unterschieden sich dahingehend, nach welchem Prinzip Inhalte aus den Materialien ausgewählt wurden. Die Lehrenden der Fälle MW 1 und MW 5 achteten darauf, die Veranstaltungskapitel repräsentativ durch Prüfungsaufgaben abzudecken. Bei Fall MW 2 wurde bisher eine Zufallsauswahl an Inhalten vorgenommen, ohne jedoch darauf zu achten, dass alle Themenbereiche repräsentativ abgedeckt sind. Die Bestimmung der Prüfungsanforderungen (im Sinne von Verhaltenskomponenten, mit denen festgelegt wird, was am Inhalt gekonnt werden soll) erfolgt entweder *orientiert am abzuprüfenden Inhalt* (welche Fragestellung bietet sich bei einem bestimmten Inhalt an) oder durch *Aufgabentypen*, zwischen denen die Lehrenden unterscheiden (z.B. Berechnungsaufgabe, Begründungsaufgabe, etc.). Die Teilnehmenden der Fälle MW 3 und MW 4 hatten sich diesbezüglich noch für kein bestimmtes Vorgehen entschieden.

Tabelle 19: Referenzpunkte und Auswahlvorschriften bei der Festlegung von Prüfungsinhalten und -anforderungen

	Referenzpunkte		Auswahl der Inhalte		Auswahl Verhaltenskomponenten	
	Vorlesungsunterlagen	Aufgabenpool	Repräsentative Auswahl	Zufallsauswahl	Inhaltsgeleitet	Geleitet durch Aufgabentypen
MW 1	✓ ³²	○	✓ ³²	○	✓ ³³	✓ ³³
MW 2	✓ ³⁴	○	- ³⁴	✓ ³⁴	○	✓ ³⁵
MW 3	○	✓ ³⁶	Nicht geklärt	Nicht geklärt	Nicht geklärt	Nicht geklärt
MW 4	✓ ³⁷	✓ ³⁷	Nicht geklärt	Nicht geklärt	Nicht geklärt	Nicht geklärt
MW 5	○	✓ ³⁸	✓ ³⁸	○	○	✓ ³⁹

Anmerkungen. ✓ = Trifft zu, - = aktive Ablehnung, ○ = wurde nicht genannt, k.A. = der Themenbereich, auf den sich die Kategorie bezieht, wurde nicht angesprochen. Sind einer Unterkategorie sowohl Häkchen als auch Minuszeichen zugeordnet, bedeutet dies, dass der Ist-Stand (fett gesetzt) zu einem Soll-Stand verändert werden soll.

Wie auch schon bei den Prüfungsanforderungen wird nachfolgend die Bedeutung der Kategorien anhand des Vorgehens und der Annahmen der Fälle beschrieben, deren Aussagen bei der Bildung der Kategorien zugrunde gelegt wurden.

- Repräsentative Auswahl der Inhalte

Die Teilnehmer des Lehrstuhls MW 1 orientierten sich bei der Auswahl der Prüfungsinhalte an den Kapiteln der Veranstaltung und den Veranstaltungsmaterialien (Vorlesungsskript und Foliensatz). Bei der Festlegung der Prüfungsinhalte wurde darauf geachtet, dass zu jedem Kapitel Prüfungsaufgaben entstehen. Die Anzahl der Prüfungsaufgaben pro Kapitel sollte anteilig den Umfang, den ein bestimmtes Kapitel in der Veranstaltung einnimmt, widerspiegeln. Als Maß wurde die Anzahl der Vorlesungsstunden verwendet, die für die Kapitel jeweils aufgewendet wurden (vgl. Zitat 3, S. 265). Innerhalb der Kapitel existierten bisher keine gesetzten Regeln bei der konkreten Auswahl von Inhalten. Beim Durchgehen der Veranstaltungsmaterialien zu den einzelnen Kapiteln, entschieden die Lehrenden spontan, zu welchen Inhalten sich Aufgaben erstellen lassen.

³² Vgl. Zitat 3, S. 265 und Zitat 4, S. 266

³³ Vgl. Zitat 4, S. 266

³⁴ Vgl. Zitat 5, S. 266

³⁵ Vgl. Zitat 6, S. 267

³⁶ Vgl. Zitat 12, S. 269

³⁷ Vgl. Zitat 18, S. 271

³⁸ Vgl. Zitat 23, S. 274

³⁹ Vgl. Zitat 24, S. 274, Zitat 25, S. 275 und Zitat 26, S. 275

- ***Zufallsauswahl der Inhalte***

Bei Fall MW 2 wurden bei der Auswahl von Prüfungsinhalten ebenfalls die Veranstaltungskapitel und -unterlagen zugrunde gelegt. Die Auswahllogik ist jedoch eine andere. Während am Lehrstuhl MW 1 darauf geachtet wurde, dass die Prüfungsinhalte anteilmäßig dem Umfang der Vorlesungskapitel entsprechen, achteten die Prüfungsersteller vom Lehrstuhl MW 2 darauf, eine Zufallsauswahl von Inhalten zu treffen, ohne dabei notwendigerweise eine Repräsentativität der Prüfungsinhalte zu den Vorlesungsinhalten zu erreichen. Dahinter steht die Annahme, dass die Studierenden alle Inhalte gleichermaßen lernen, um auf die Prüfung vorbereitet zu sein, solange sie nicht voraussagen können, welche Inhalte abgefragt werden. Eine Auswahl der Prüfungsinhalte in Relation zur aufgewendeten Zeit für einen bestimmten Inhalt in der Vorlesung wird mit dieser Auswahllogik als unvoreteilhaft angesehen, weil es den Studierenden helfen könnte die Prüfungsinhalte zu antizipieren und entsprechend selektiv zu lernen (vgl. Zitat 5, S. 266).

- ***Inhaltsgeleitete Festlegung von Verhaltenskomponenten***

Neben der Auswahl der Inhalte spielt es eine Rolle, wie die Verhaltenskomponenten, die beschreiben was die Studierenden am Inhalt können sollen, bestimmt wurden. Hierbei geht es um die Frage, ob beispielsweise die Fähigkeit Inhalte zu reproduzieren oder die Anwendung von Wissen, um ein Problem zu lösen, geprüft werden soll. Die Aussagen der Teilnehmenden von Fall MW 1 deuten auf mehrere implizite Auswahlvorschriften bei der Festlegung der Verhaltenskomponenten hin. Zum einen wurde differenziert zwischen wissensorientierten Anforderungen und Anforderungen die „tiefergreifend“ – die Bedeutung von tiefergreifend wurde zunächst nicht näher definiert - und damit auch schwieriger sind. Die Auswahl der Prüfungsinhalte erfolgte bisher im Zuge der Aufgabenerstellung und nicht als vorgelagerter Schritt. Beim Durchgehen der Veranstaltungsmaterialien bestimmten die Lehrenden die Inhalte, die in einer Prüfungsaufgabe abgefragt werden sollen. Dabei legten sie anhand der Charakteristika des Inhalts fest, welche Art von Fragestellung sich für den Inhalt eignet. Im Grunde wägten die Lehrenden ab, ob sich der Inhalt eher zur Wiedergabe von Wissen eignet, oder ob es möglich ist, eine weiterführende, tiefergreifende Frage zu stellen. Die Eigenschaften eines Inhaltsaspekts bestimmten also, welches Verhalten an diesem Inhalt ausgeführt werden soll (vgl. Zitat 4, S. 266). Das Vorgehen bei der Festlegung der Verhaltenskomponenten erfolgte demnach im Prozess der Aufgabenerstellung und wurde beeinflusst durch die Charakteristika des Inhalts. Die entsprechende Kategorie, die das Vorgehen bei der Festlegung von Verhaltenskomponenten kennzeichnet, wird demnach als „inhaltsgeleitet“ bezeichnet.

- *Aufgabentypengeleitete Festlegung von Verhaltenskomponenten*

Bei der aufgabentypengeleiteten Festlegung von Verhaltenskomponenten beeinflussen die Aufgabentypen, die den Lehrenden bekannt sind, welche Verhaltenskomponenten festgelegt werden. Die Lehrenden von Fall MW 5 versuchten bisher zu jedem Veranstaltungskapitel Wissensaufgaben, Rechenaufgaben, Aufgaben mit einer Grafik und Anwendungsaufgaben zu erstellen. Die Verhaltenskomponenten der Prüfung wurden bestimmt, indem die Lehrenden versuchten zu den Veranstaltungskapiteln jeweils die genannten Aufgabentypen zu entwickeln. Bei Fall MW 2 wurde die Festlegung der Verhaltenskomponenten ebenfalls über Aufgabentypen vorgenommen. Hierbei wurde jedoch nicht versucht, pro Veranstaltungskapitel verschiedene Aufgabentypen unterzubringen. Vielmehr unterschied man mit der Struktur der Prüfung zwischen zwei Typen von Aufgaben. Der erste Teil der Prüfung bestand aus der Abfrage von Wissen, der zweite Teil aus Rechenaufgaben. Die Struktur der Prüfung wurde angelehnt an die Aufteilung der Veranstaltung, bei der die Wissens Elemente in der Vorlesung und die Berechnungen in der begleitenden Übung behandelt werden (vgl. Zitat 6, S. 267). Die Lehrenden gingen zudem davon aus, mit den Berechnungsaufgaben auch Transferwissen abzufragen. Bei Fall MW 1 wurde die Festlegung der Verhaltenskomponenten ebenfalls vom Format der Prüfung, mit einem Wissens- und einem Anwendungsteil, beeinflusst (vgl. Zitat 2, S. 264). Dem Wissensteil wurden die oben genannten wissensorientierten Anforderungen zugeordnet, dem Anwendungsteil die als tiefergreifend bezeichneten Anforderungen. Beide Aspekte wurden über Operatoren, wie sie typischerweise in Aufgaben verwendet werden (nennen, beschreiben, berechnen), charakterisiert. Wissensorientierte Anforderungen werden mit dem Operator *nennen Sie* in Verbindung gebracht, wohingegen die relevanten tiefergreifenden Aspekte mit dem Operator *begründen Sie* assoziiert werden.

Bezüglich der Einhaltung von Qualitätsstandards der pädagogischen Diagnostik, kann festgehalten werden, dass die zielgerichtete Ableitung von Prüfungsinhalten und –anforderungen unter Berücksichtigung von bestimmten Auswahlvorschriften (z.B. Anlehnung an Studiengangprofil, Qualifikationsrahmen, Modullehrziele) bisher nicht konsequent umgesetzt wurde. Die Prüfungsinhalte wurden sehr operational festgelegt, sprich der Inhalt wurde gesichtet und gleichzeitig wurden Aufgaben dazu entwickelt. Die dabei zugrunde gelegten Auswahlvorschriften waren jedoch nicht explizit, jeder Lehrende verfolgt stattdessen seine eigene implizite Strategie. Es handelt sich hierbei anscheinend um eine tradierte Praxis zur Festlegung von Prüfungsanforderungen. Eine theoriegeleitete Festlegung des Prüfungsgegenstandes und Überlegungen im Sinne einer theoretischen Modellierung eines Konstrukts fanden nicht statt. Diese beiden Schritte sind jedoch als wesentliche Grundlage zu betrachten, an dem die restlichen Schritte der Prüfungserstellung (Erstellung von Aufgaben, und Überprüfung von deren Validität) und Auswertung (Korrektur, Punkte- und Notenvergabe, Erkenntnisse aus empirischen Daten).

Bezüglich einer Kompetenzorientierung kann zur konzeptionellen Analyse festgehalten werden, dass bisher keine Orientierung an realen Anforderungssituationen erfolgte, wie sie in beruflichen Kontexten oder im weiteren Studienverlauf auftreten können. Auch der Qualifikationsrahmen für deutsche Hochschulabschlüsse wurde nicht als Referenzrahmen genannt. Ob es mit dem beschriebenen Vorgehen der Teilnehmenden bei der konzeptionellen Analyse des Prüfungsgegenstands dennoch gelang, kompetenzorientierte Prüfungen zu gestalten, muss auf Ebene der Aufgaben geklärt werden. Bei der Festlegung des Prüfungsgegenstands sind jedenfalls keine Hinweise für eine systematische Umsetzung einer Kompetenzorientierung zu finden.

5.1.2.2 Zerlegung von Kompetenzen in Teilkompetenzen und Festlegung von Bearbeitungsniveaus

Durch eine Zerlegung von Kompetenzen in Teilkompetenzen und die Festlegung von Bearbeitungsniveaus lässt sich der diagnostische Wert von Prüfungen steigern. Wird beispielsweise eine übergeordnete Kompetenz nicht erreicht, kann bei einer entsprechenden Strukturierung und Zerlegung von Anforderungen festgestellt werden, bei welchen Teilkompetenzen Probleme bestehen und welche Stufe erreicht wurde. Mit der Festlegung von Niveaus können bei Studierenden unterschiedliche Kompetenzniveaus festgestellt werden, vorausgesetzt, dass für jedes Bearbeitungsniveau entsprechende Aufgaben (mit entsprechenden Aufgabenschwierigkeiten) existieren. Trotz der hohen Bedeutung dieses Schritts, führten die Lehrenden eine entsprechende Modellierung ihrer Prüfungsinhalte und -anforderungen nicht durch.

5.1.2.3 Formulierung von Lehrzielen

Lehrziele sind der zentrale Ausgangspunkt sowohl für die Erstellung von Aufgaben, die Zusammenstellung von Aufgaben zu einer Prüfung und die Entwicklung von Bewertungsrastern, als auch für die Punktevergabe und Notengebung. Wie sich in den Gesprächen zeigte, waren den Lehrenden die jeweiligen Lehrziele für ihre Veranstaltungen nicht bekannt. Bei allen Fällen wurden die Lehrziele von anderen Personen, die nicht bei der Prüfungskonzeption beteiligt waren, erstellt. Lediglich bei Fall MW 1 und MW 5 wussten die akademischen Oberräte darüber Bescheid, dass die Lehrziele existieren und zu einem nicht bekannten Zeitpunkt erstellt wurden (vgl. Zitat 27, S. 276). Dennoch konnten bestehenden Lehrziele zu den entsprechenden Veranstaltungen der Lehrenden im Zuge der Arbeit analysiert werden. Nachfolgend wird dargestellt, wie viele Lehrziele bei den einzelnen Fällen jeweils vorlagen, welche Wissensarten in den Lehrzielen angesprochen waren und ob die Lehrziele ausreichend spezifisch formuliert sind.

Betrachtet man zunächst die Anzahl der vorliegenden Lehrziele zeigt sich, dass bei vier von fünf Fällen jeweils vier Lehrziele formuliert wurden. Bei Fall 5 existierten gesonderte Lehrziele für die Lehrveranstaltung beziehungsweise das Modul und für die Prüfung (siehe Abbildung 7).

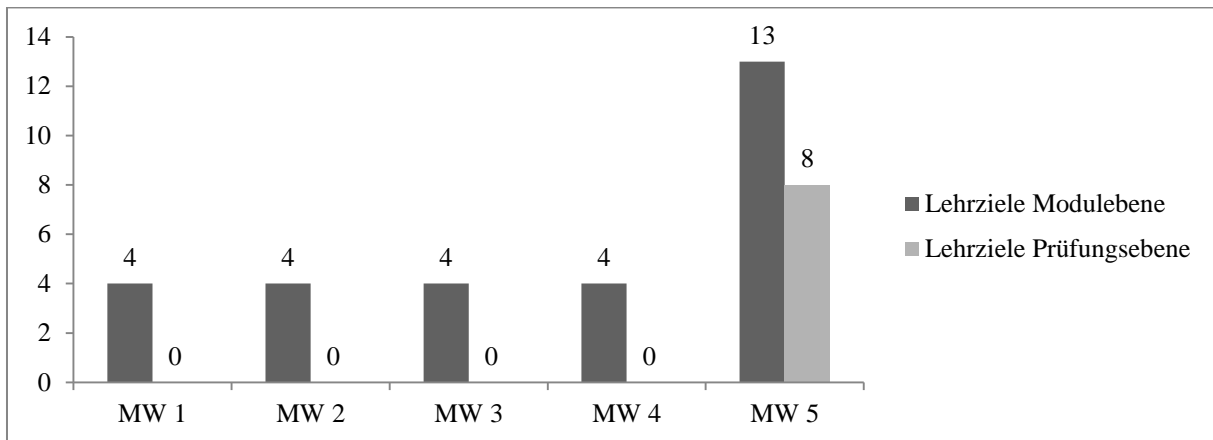


Abbildung 7: Anzahl der bereits existierenden Lehrziele pro Lehrstuhl ($n=37$)

Die in den Lehrzielen abgebildeten Anforderungen wurden über die Wissensarten, die in den Lehrzielen angesprochen sind, charakterisiert. Dabei wurde unterschieden zwischen deklarativem, prozeduralem, konzeptionellem und strategischem Wissen. In Abbildung 8 ist die Verteilung der Wissensarten in den Lehrzielen dargestellt. In die Analysen wurden von Lehrstuhl 5 die Lehrziele auf Prüfungsebene einbezogen. Die Analysen umfassen damit insgesamt 24 Lehrziele. Mit 66,7 Prozent wurde die Aneignung von konzeptionellem Wissen als häufigstes Lehrziel angegeben. Strategisches Wissen wurde hingegen in keinem der Lehrziele angesprochen. Etwa ein Fünftel der Lehrziele bezog sich auf prozedurales Wissen, 12,5 Prozent auf deklaratives Wissen. Auffallend sind die hohen Anteile an konzeptionellem Wissen und der geringe Anteil von deklarativem Wissen in den Lehrzielen.

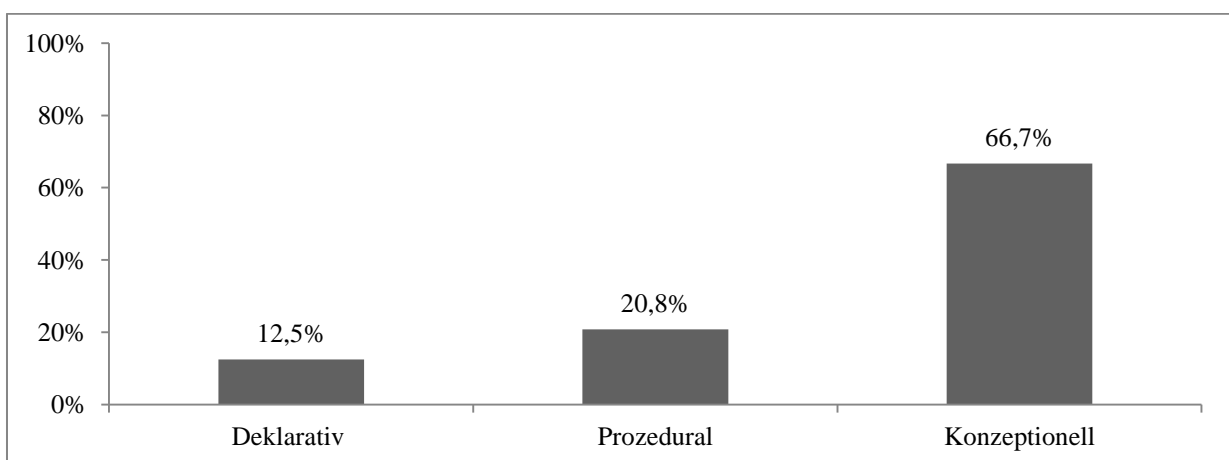


Abbildung 8: Verteilung der Wissensarten in den bereits existierenden Lehrzielen ($n=24$)

Betrachtet man die Verteilung auf Fallebene, so zeigen sich deskriptive Unterschiede in den Schwerpunktsetzungen (siehe Abbildung 9). Bei Fall MW 1 zielte die Hälfte der Lehrziele auf konzeptionelles Wissen ab. Deklaratives und prozedurales Wissen waren jeweils zu 25 Prozent vertreten. Diese Differenziertheit in den Wissensarten bei Fall MW 1 kann durchaus mit deren oben beschriebener Zielsetzung, praxisorientiertes Wissen vermitteln zu wollen, in Einklang stehen, wenn man annimmt, dass eine deklarative und prozedurale Wissensbasis die Grundlage für die Entwicklung von konzeptionellem Wissen bildet, welches wiederum verstärkt bei der Bewältigung von komplexen realen Anforderungssituationen zum Tragen kommt. Bei den restlichen Fällen fallen insbesondere die hohen Anteile an Lehrzielen ins Auge, die auf konzeptionelles Wissen abzielten. Insbesondere für die Fälle MW 2 und MW 3, bei denen sich zudem keines der Lehrziele auf deklaratives Wissen bezog, erscheint die Verteilung der Wissensarten widersprüchlich zu den oben beschriebenen Zielsetzungen, wonach beide Fälle detaillierte Wissens Elemente zu Begriffen, Fakten und Definitionen vermitteln wollten (siehe Tabelle 18). Ebenso fällt bei Fall MW 4 eine Diskrepanz zu den verbal geäußerten Zielsetzungen auf, da keines der Lehrziele auf prozedurales Wissen abzielt, obwohl die Zielsetzung sehr stark über Aufgaben beschrieben wurde, die auf die Analyse und Berechnung von Vorgängen beruhen.

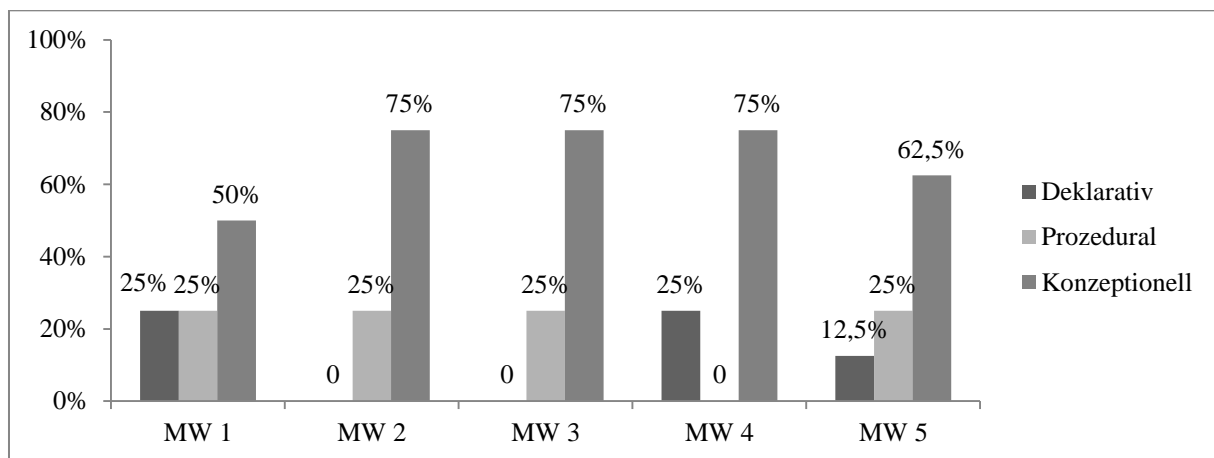


Abbildung 9: Verteilung der Wissensarten in Lehrzielen nach Fällen (n=24)

Mit der Betrachtung der Wissensarten in den Lehrzielen konnte gezeigt werden, dass in drei von fünf Fällen keine Kongruenz zwischen den verbal geäußerten Zielvorstellungen der Prüfungsersteller und den Lehrzielen bestand (Fälle MW 2, MW 3 und MW 4). Bei den Fällen MW 1 und MW 5 konnte auf dieser Ebene keine Diskrepanz zwischen den geäußerten Zielsetzungen für die Veranstaltung und den Lehrzielen beobachtet werden. Generell fällt der hohe Anteil an Lehrzielen auf, die sich auf konzeptionelles Wissen bezogen, gerade auch bei den Fällen, die verstärkt detailliertes deklaratives oder prozedurales Wissen vermitteln wollten. Hier bestand weiterer Klärungsbedarf, inwieweit die in den Lehrzielen der Veranstaltungen beschriebenen Wissens- und Könnensaspekte, entsprechend in gleichem Ausmaß in Prüfungsaufgaben umgesetzt wurden.

5.1 Festlegung des Prüfungsgegenstandes

Hinsichtlich der Spezifität der Lehrziele wurden nur 8,3 Prozent auf Prüfungsebene und damit als tauglich für die Gestaltung von Aufgaben und Bewertungsrastern eingestuft. Mit 54,2 Prozent der Lehrziele auf Modulebene bestand jedoch eine gute Ausgangsbasis für die Lehrenden, um die Lehrziele in einem weiteren Operationalisierungsschritt zu konkretisieren und so für die Aufgabengestaltung nutzbar zu machen. 37,5 Prozent der Lehrziele blieben unspezifisch und bedurften einer umfassenderen Überarbeitung (siehe Abbildung 10).

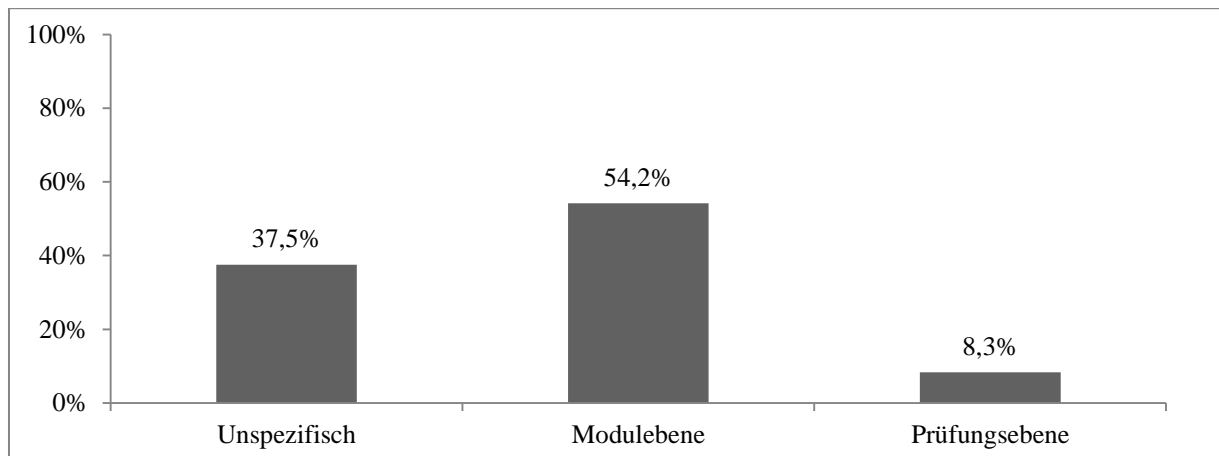


Abbildung 10: Spezifität der Lehrziele (n=24)

Bei fallweiser Betrachtung der Lehrziele zeigt sich, dass lediglich bei einem Fall (MW 5) zwei Lehrziele auf Prüfungsebene vorlagen. Bei Fall MW 2 hingegen waren alle vorhandenen Lehrziele unspezifisch. Bei den Fällen MW 3 und MW 4 waren drei von vier Lehrzielen auf Modulebene formuliert und boten damit eine gute Ausgangsbasis, um eine Spezifizierung auf Prüfungsebene zu erreichen (siehe Abbildung 11). Aufgrund der geringen Anzahl von Lehrzielen auf Fallebene sind die absoluten Werte angegeben, um die Verteilung hinsichtlich der Spezifität aufzuzeigen.

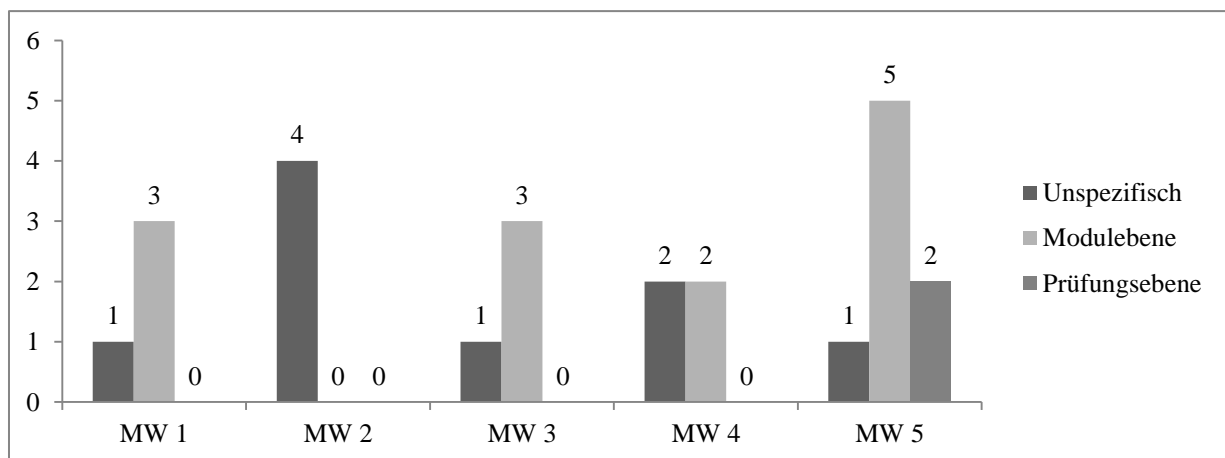


Abbildung 11: Spezifität der Lehrziele nach Lehrstühlen (n=24)

Die Ergebnisse verdeutlichen die Notwendigkeit, im Rahmen des Qualitätsentwicklungsprogramms mit den Lehrenden die bestehenden Lehrziele zu überarbeiten. Die Lehrziele mussten

spezifiziert werden, damit sie für die an der Prüfungserstellung beteiligten Kolleginnen und Kollegen einen gemeinsamen und verständlichen Bezugspunkt für die Prüfungskonzeption und Auswertung bilden konnten. Außerdem mussten die Lehrenden hinsichtlich der Bedeutung von Lehrzielen für den Prozess des Prüfens sensibilisiert werden.

5.1.2.4 Verwendung einer Lehrzielmatrix

Mit einer Lehrzielmatrix können Lehrenden Lehrziele strukturieren und übersichtlich darstellen. In der bestehenden Praxis wurde dieses Hilfsmittel jedoch nicht eingesetzt. Im Rahmen des Qualitätsentwicklungsprogramms wurde dieses Instrument daher vorgestellt.

5.1.3 Erkenntnisse zur veränderten Prüfungspraxis

In diesem Abschnitt werden die Ergebnisse zu Forschungsfrage 3 berichtet, die sich auf die Veränderungen beziehen, die von den Lehrenden während der gemeinsamen Qualitätsentwicklungsarbeit bei der Festlegung des Prüfungsgegenstands vorgenommen wurden. Der Fokus des Qualitätsentwicklungsprogramms lag auf den Problemen, die in der bestehenden Prüfungspraxis identifiziert werden konnten (siehe Abschnitt 5.1.2). Die zur Veranstaltung bestehenden Lehrziele, die jedoch von den Lehrenden in der bestehenden Praxis nicht genutzt wurden, dienten als Ausgangspunkt, um über die Festlegung des Prüfungsgegenstands zu diskutieren. Im ersten Arbeitstreffen wurde verstärkt darauf hingewirkt, dass die Teilnehmenden an ihren Lehrzielen arbeiten und diese als Ausgangspunkt für die Erstellung von Aufgaben verwenden. Die grundlegenden Zielsetzungen für die Lehrveranstaltungen, wie sie eingangs von den Lehrenden beschrieben wurden (siehe Tabelle 18), änderten sich dabei zunächst nicht. Stattdessen versuchten die Lehrenden ihre Lehrziele an diese Zielsetzungen anzupassen. Dabei wurde auch eine Strukturierung und Sequenzierung von Teilkompetenzen vorgenommen. Da sich die Gespräche in den Arbeitstreffen verstärkt auf die Lehrziele bezogen, gliedert sich dieser Abschnitt entlang der Veränderungen, die an den Lehrzielen vorgenommen wurden (und beginnt nicht wie im Falle der bestehenden Prüfungspraxis mit der konzeptionellen Analyse des Prüfungsgegenstands). Zunächst wird dargestellt, inwieweit sich die Anzahl der Lehrziele (Abschnitt 5.1.3.1) sowie deren Spezifität (Abschnitt 5.1.3.2) veränderten. Daran anschließend werden Veränderungen in den Wissensarten berichtet, die in den neu formulierten Lehrzielen angelegt sind (Abschnitt 5.1.3.3). Abschließend wird dargestellt, inwieweit es gelungen ist, die Lehrenden dazu anzuregen, eine Lehrzielmatrix zu verwenden (Abschnitt 5.1.3.4).

5.1.3.1 Anzahl der Lehrziele in der veränderten Praxis

Bei den Fällen MW 1 bis MW 4 hat sich die Anzahl der Lehrziele jeweils erhöht (siehe Abbildung 12). Die Lehrenden von Fall MW 5 arbeiteten auch in der veränderten Praxis nicht mit ihren Lehrzielen, so dass hier keine Veränderungen in deren Anzahl festzustellen waren. In Abbildung 12 ist der Unterschied zwischen der Anzahl, der vor Beginn der gemeinsamen Qualitätsentwicklungsarbeit bereits bestehenden Lehrziele (mit Messzeitpunkt 1 (MZP 1) bezeichnet), und der im Rahmen der Qualitätsentwicklungsarbeit neu erstellten Lehrziele (MZP 2) dargestellt.

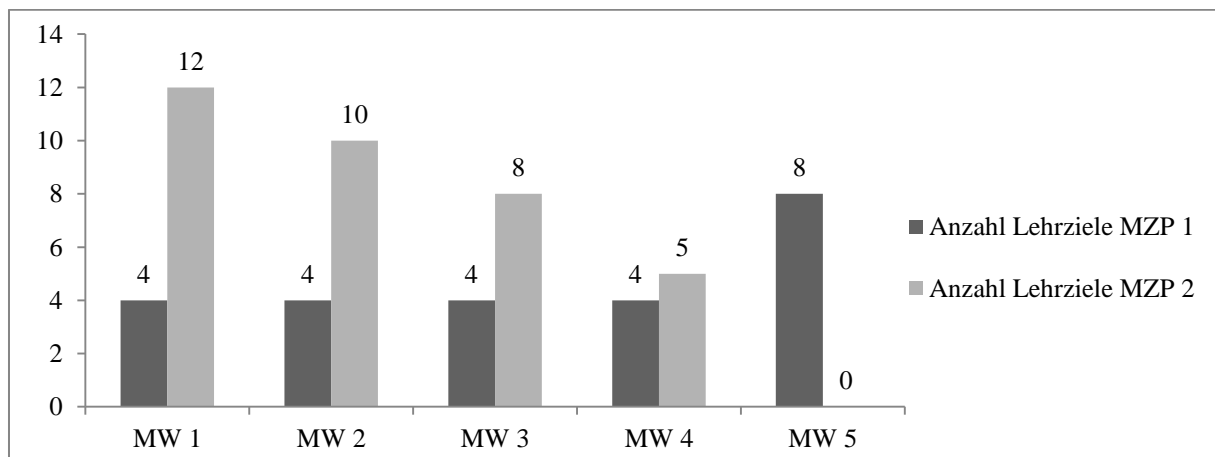


Abbildung 12: Anzahl der Lehrziele in der veränderten Praxis ($n=35$)

Am deutlichsten erhöhte sich die Anzahl der Lehrziele bei Fall MW 1. Die Gründe für die Erhöhung der Anzahl der Lehrziele sind bei allen Fällen ähnlich gelagert. Bei Fall MW 1 wurde beispielsweise ein Lehrziel ergänzt, da ein Themenbereich mit den bestehenden Lehrzielen nicht angesprochen wurde. Zwei der bereits bestehenden Lehrziele bezeichneten übergeordnete Kompetenzen, die jeweils in verschiedene Teilkompetenzen zerlegt und durch Lehrziele definiert wurden (siehe Zitat 28, S. 276). Durch die gemeinsame Diskussion der Lehrziele mit den Kolleginnen und Kollegen konnte bei Fall MW 1 eine Strukturierung der Prüfungsinhalte und -anforderungen erreicht werden, mit dem Ergebnis, dass mit den Lehrzielen nun systematisch zwischen Teilkompetenzen differenziert wird (siehe Zitat 29, S. 277). So konnte die beschriebene Zielsetzung nach Praxisorientierung und ganzheitlichem Verständnis (siehe Tabelle 18) von den Lehrenden durch die Formulierung von Lehrzielen konkretisiert und strukturiert werden. Als neue Referenzpunkte für die Auswahl der Prüfungsinhalte und -anforderungen werden nun die bestehenden Lehrziele verwendet, die durch gemeinsame Aushandlungsprozesse und Reflexion der Lehrveranstaltungsziele kooperativ festgelegt wurden.

Bei Fall MW 2 erfolgte ähnlich wie bei Fall MW 1 eine Zerlegung übergeordneter Themen- und Anforderungsbereiche in Teilbereiche, wodurch sich die Anzahl der Lehrziele erhöhte. Zu manchen Themenbereichen wurde in den bestehenden Lehrzielen lediglich ein Wiedergeben der In-

halte verlangt, während bei anderen Themen nur ein Anwenden in den Lehrzielen angesprochen wurde, obwohl für das Anwenden laut Aussage der Teilnehmenden eine Wissensbasis notwendig wäre. Zu jedem der Themenbereiche wurden daher die Anforderungen systematisch ergänzt, so dass zu jedem Bereich Anteile von Wissenswiedergabe, verstehensorientierte Anteile und anwendungsorientierte Anforderungen enthalten sind (vgl. Zitat 30, S. 277).

Die Teilnehmende von Fall MW 3 nahm die vier bestehenden Lehrziele als Ausgangspunkt und differenzierte diese weiter, indem die Inhaltskomponenten präzisiert wurden. Ein Lehrziel wurde neu hinzugefügt. Ein weiteres Lehrziel entstand bei der Übertragung in die Lehrzielmatrix, als in der Matrix sichtbar wurde, dass zu einem Themenbereich ein wesentlicher Verhaltensaspekt nicht in den Lehrzielen angesprochen wird. Eine Strukturierung und Sequenzierung von Anforderungen wurde im Prozess der Lehrzielüberarbeitung ebenfalls vorgenommen. Entlang des Aufbaus der Lehrveranstaltung wurden auch die Lehrziele sequenziert, wonach zuerst Grundlagen fokussiert werden, auf denen Anwendungs-, Bewertungs- und Transferaufgaben aufbauen. (Zitat 31, S. 279).

Der Teilnehmer von Fall MW 4 passte die bestehenden Lehrziele an die Anforderungen an, die in den Prüfungsaufgaben angelegt waren. Um die in den Aufgaben angelegten Anforderungen zu beschreiben, orientierte sich der Teilnehmer an den Begrifflichkeiten beziehungsweise den Verben aus der Lehrzieltaxonomie nach Bloom et al. (1956), die den Lehrenden als Hilfsmittel zur Verfügung gestellt worden war. Dabei ergänzte er ein Lehrziel zum Anforderungsbereich des Bewertens (siehe Zitat 32, S. 280).

Bei Fall MW 5 sollte die Prüfung mehrere Wochen früher stattfinden als bei den anderen Fällen, so dass im ersten Arbeitstreffen bereits über Fragen zur Aufgabenerstellung und Punktevergabe gesprochen wurde. Die Bedeutung von Lehrzielen und dem zielgerichteten Vorgehen bei der Erstellung und Auswertung von Prüfungen wurde zwar besprochen, die gemeinsame Arbeit an den Lehrzielen konnte während des ersten Arbeitstreffens aus Zeitgründen jedoch nicht aufgenommen werden. Die Teilnehmenden von Fall MW 5 erhielten lediglich die entsprechenden Handreichungen und Checklisten. Wie sich herausstellte wurden die Lehrziele auch in der veränderten Praxis nicht für die Aufgabenerstellung genutzt. Das bisherige Vorgehen, bei dem jede der an der Prüfungserstellung beteiligten Personen für ein Veranstaltungskapitel Aufgaben entwickelte, wurde beibehalten. Alle Aufgabenersteller⁴⁰ sollten jedoch nun ihre Aufgaben den Stufen der Lehrzielpyramide zuordnen (siehe Zitat 33, S. 280). So kann von einer systematischen Differenzierung von Anforderungen anhand der Stufen in der Lehrzielpyramide ausgegangen werden.

⁴⁰ Die maskuline Form ist hier zutreffend, da nur männliche Personen beteiligt waren.

Insgesamt betrachtet kann festgehalten werden, dass bei vier Fällen die Teilnehmenden ihre Lehrziele überarbeiteten und deren Anzahl dabei erhöhten. Die Erhöhung der Lehrziele war zum einen dadurch bedingt, dass Themenbereiche ergänzt wurden, die zuvor nicht in Lehrzielen berücksichtigt wurden. Zum anderen erfolgte mit der Arbeit an den Lehrzielen eine Strukturierung und Sequenzierung von Inhalten und Anforderungen durch die Zerlegung übergeordneter Kompetenzen in Teilkompetenzen. Dabei wurden sowohl inhaltliche Differenzierungen als auch Differenzierungen der Verhaltenskomponenten in den Lehrzielen vorgenommen. Auch bei Fall MW 5, die sich anstatt der Lehrziele auf eine Lehrzielpyramide nach Bloom et al. (1956) bezogen, wurde systematisch das Ziel verfolgt, unterschiedliche Anforderungen in Aufgaben abzubilden. Mit der Arbeit an und mit Lehrzielen konnten die Lehrenden dazu angeregt werden, ihre Inhalte und Anforderungen systematisch in Teilkompetenzen zu zerlegen und aufeinander abzustimmen. Dieser Prozess hatte in der bestehenden Praxis bisher nicht stattgefunden.

5.1.3.2 Spezifität der neu erstellten Lehrziele

Für die bestehende Prüfungspraxis wurde festgestellt, dass der Großteil der Lehrziele zu unspezifisch formuliert war. Wie in Abbildung 13 dargestellt, wurde der Anteil der Lehrziele auf Prüfungsebene leicht erhöht, bei gleichzeitiger Reduzierung der Lehrziele, die unspezifisch oder auf Modulebene formuliert waren. Laut der in dieser Arbeit eingeführten Einteilung in verschiedene Stufen der Spezifität, werden nur die Lehrziele, die auf Prüfungsebene eingestuft wurden, als eindeutig genug für die Erstellung von Aufgaben, Bewertungsrastern und der Vergabe von Punkten angesehen.

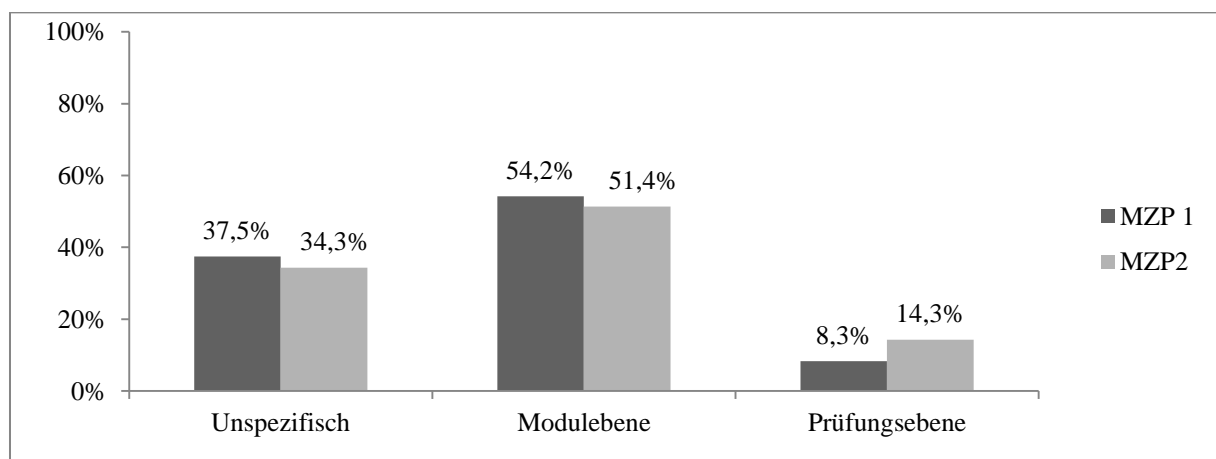


Abbildung 13: Spezifität der bestehenden (MZP 1) und der neu erstellten (MZP 2) Lehrziele ($n_{MZP1}=24$, $n_{MZP2}=35$)

Bei fallweiser Betrachtung zeigen sich zum Teil größere Unterschiede hinsichtlich der Spezifität von Lehrzielen (siehe Abbildung 14). Bei Fall MW 1, die am konsequentesten mit Lehrzielen arbeiteten, um ein gemeinsames Verständnis zwischen Kolleginnen und Kollegen bezüglich der

Bedeutung der Lehrziele zu schaffen, ist der größte Zuwachs an Lehrzielen auf Prüfungsebene zu verzeichnen. Bei den anderen Teams sind hingegen keine Lehrziele auf Prüfungsebene formuliert worden. Bei Fall MW 2 ist der größte Teil der Lehrziele nach der hier durchgeführten Analyse als mehrdeutig eingestuft worden. Dies lag daran, dass die Lehrenden die Inhaltskomponenten in den Lehrzielen bewusst sehr allgemein formulierten. Die Verhaltenskomponenten waren dabei jeweils entweder spezifisch oder eher unspezifisch. Dies wird an zwei der Lehrziele illustriert:

Lehrziel 1: Die Studierenden können Wissen und Terminologie in den relevanten Themenbereichen wiedergeben.

Lehrziel 2: Die Studierenden können Methoden und Konzepte aus den Themenbereichen anwenden.

Die Lehrenden hielten die Inhaltskomponenten bewusst allgemein. Die Kolleginnen und Kollegen, die zu den jeweiligen Veranstaltungskapiteln Aufgaben erstellten, wurden dazu angehalten, zu den allgemein gehaltenen Inhaltskomponenten jeweils die spezifischen Inhalte aus den Kapiteln einzusetzen. Mit diesem Vorgehen konnten die Prüfungsaufgaben ebenfalls systematisch an den Zielsetzungen der Veranstaltung ausgerichtet werden. Durch die Angabe der Verhaltenskomponenten wurde für die Aufgabenersteller transparent dargestellt welche Anforderungen in Aufgaben zu den Inhalten jeweils angelegt werden müssen. Obwohl die Lehrenden der Fälle MW 1 und MW 2 im gemeinsamen Arbeitstreffen den gleichen Input zur Formulierung von Lehrzielen erhielten, entschieden sie sich für ein deutlich unterschiedliches Vorgehen.

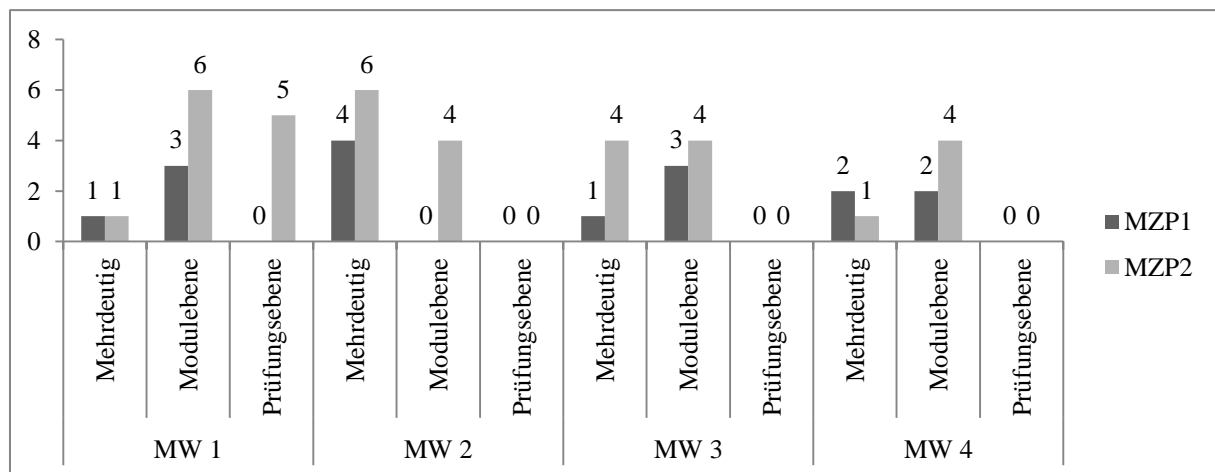


Abbildung 14: Veränderungen in der Spezifität der Lehrziele nach Fällen⁴¹ ($n_{\text{Lehrziele MZP 1}}=24$, $n_{\text{Lehrziele MZP 2}}=35$)

⁴¹ Auf Fallebene werden die absoluten Werte berichtet, da diese aufgrund der niedrigen Anzahl an Lehrzielen pro Fall besser interpretierbar sind als prozentuale Werte.

Bei Fall MW 3 erhöhten sich sowohl die Lehrziele auf Modulebene als auch die mehrdeutigen Lehrziele. Drei der Lehrziele wurden als mehrdeutig eingestuft, da die Inhaltskomponenten, ähnlich wie bei Fall MW 2, weiterhin relativ allgemein gehalten waren. Bei einem der mehrdeutigen Lehrziele war als Verhaltenskomponente das Verb *verstehen* angegeben, ohne dass weiter spezifiziert wurde, welche Aspekte in einer Aufgabe angelegt sein müssen, um darauf schließen zu können, dass ein Verständnis auf Seiten der Studierenden vorliegt (z.B. durch das Erklären eines Sachverhalts). Bei Fall MW 4 wurden die neu formulierten Lehrziele insgesamt betrachtet spezifischer formuliert, indem sich die Anzahl der Lehrziele auf Modulebene von zwei auf vier erhöhte.

Damit kann festgehalten werden, dass trotz des Inputs im Rahmen der Arbeitstreffen zur spezifischen Formulierung von Lehrzielen, nur bei Fall MW 1, mit knapp der Hälfte, ein größerer Teil der Lehrziele auf Prüfungsebene umgesetzt wurde. Trotz des Inputs im Rahmen der Arbeitstreffen zur Notwendigkeit, Lehrziele spezifisch zu formulieren, wurde dies nur von Fall MW 1 bei knapp der Hälfte der Lehrziele umgesetzt. Dies könnte bedeuten, dass für die Lehrenden eine genauere Spezifikation nicht nötig ist, um durch Lehrziele eine gemeinsame Ausgangsbasis für die Prüfungserstellung und -auswertung zu schaffen.

5.1.3.3 Wissensarten in den neu erstellten Lehrzielen

Zusätzlich zur Anzahl und Spezifität von Lehrzielen ist es relevant, welche Art von Wissen in den neu erstellten Lehrzielen angesprochen wurden. In der bestehenden Praxis stimmten bei den Fällen MW 2, MW 3 und MW 4 die in den Lehrzielen beschriebenen Anforderungen nicht mit den verbal geäußerten Zielsetzungen überein. Der Großteil der Lehrziele zielte auf konzeptionelles Wissen ab, während die Lehrenden die Veranstaltungsziele im Bereich der Vermittlung von deklarativem oder prozeduralem Wissen beschrieben. Die Betrachtung der neu erstellten Lehrziele zeigt insgesamt, dass der Anteil an konzeptionellem Wissen geringer ist, als in den bestehenden Lehrzielen. Dafür sind die Anteile von prozeduralem Wissen in den neu erstellten Lehrzielen höher als in den bestehenden Lehrzielen (siehe Abbildung 15). Die Anteile deklarativen Wissens befinden sich auf einem ähnlichen Niveau wie bei den bestehenden Lehrzielen.

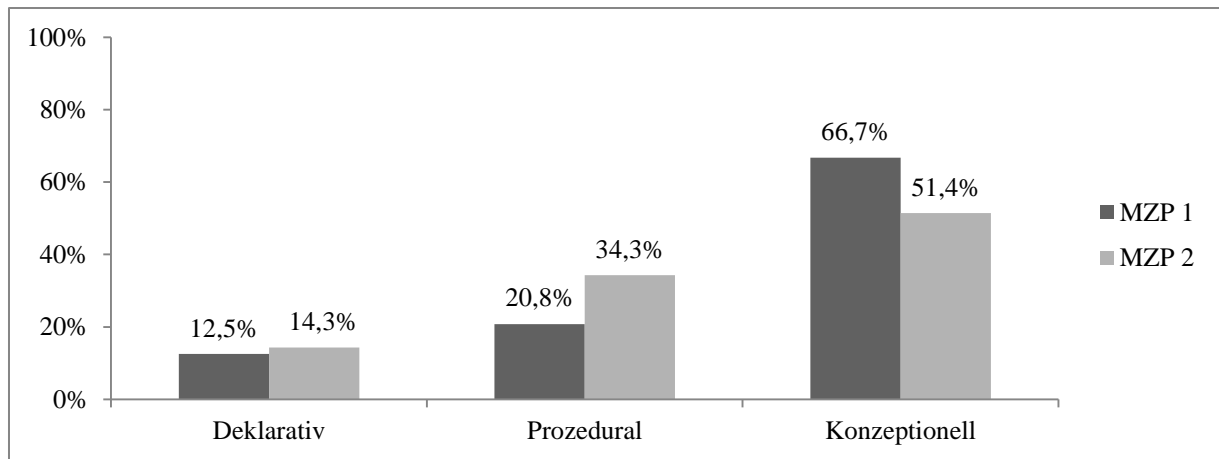


Abbildung 15: Vergleich der Wissensarten in bestehenden und neu erstellten Lehrzielen ($n_{\text{Lehrziele MZP 1}}=24$, $n_{\text{Lehrziele MZP 2}}=35$)

Auf Fallebene zeigt sich, dass bei drei Fällen (MW 2, MW 3 und MW 4) die Anteile der Lehrziele, die konzeptionelles Wissen fokussieren, reduziert wurden. Bei Fall MW 1 hingegen wurden die Anteile von konzeptionellem Wissen in den Lehrzielen um 25 Prozent erhöht, bei gleichzeitiger Verringerung der Lehrziele die prozedurales (-16,7 %) und deklaratives Wissen (-8,3 %) beinhalten (siehe Abbildung 16). Bei Fall MW 4 zielte keines der bestehenden Lehrziele auf prozedurales Wissen ab, obwohl die mündlich geäußerte Zielsetzung für die Lehrveranstaltung stark auf Analyse- und Berechnungsaufgaben abstellte. Unter den neu erstellten Lehrzielen sind Lehrziele zu prozeduralem Wissen mit einem Anteil von 40 Prozent entsprechend stark vertreten. Bei den Fällen MW 2 und MW 3 existierten keine Lehrziele zu deklarativem Wissen, obwohl die mündlich geäußerten Lehrveranstaltungsziele grundlagenorientiertes Detailwissen als Hauptziel ausgaben. Aufgrund dieser Zielsetzung sind die Anteile an deklarativem Wissen in den neu erstellten Lehrzielen sowohl bei Fall MW 2, mit einem Anteil von 10 Prozent, als auch bei Fall MW 3, mit einem Anteil von 12,5 Prozent, nach wie vor als gering anzusehen. Bei Fall MW 2 kann dies möglicherweise darüber erklärt werden, dass bei einem Lehrziel, dass auf die Wiedergabe von Wissen abzielte, mehrere Inhaltsbereiche zusammengefasst wurden. Bei beiden Fällen wurden die Anteile an prozeduralem Wissen in den Lehrzielen erhöht.

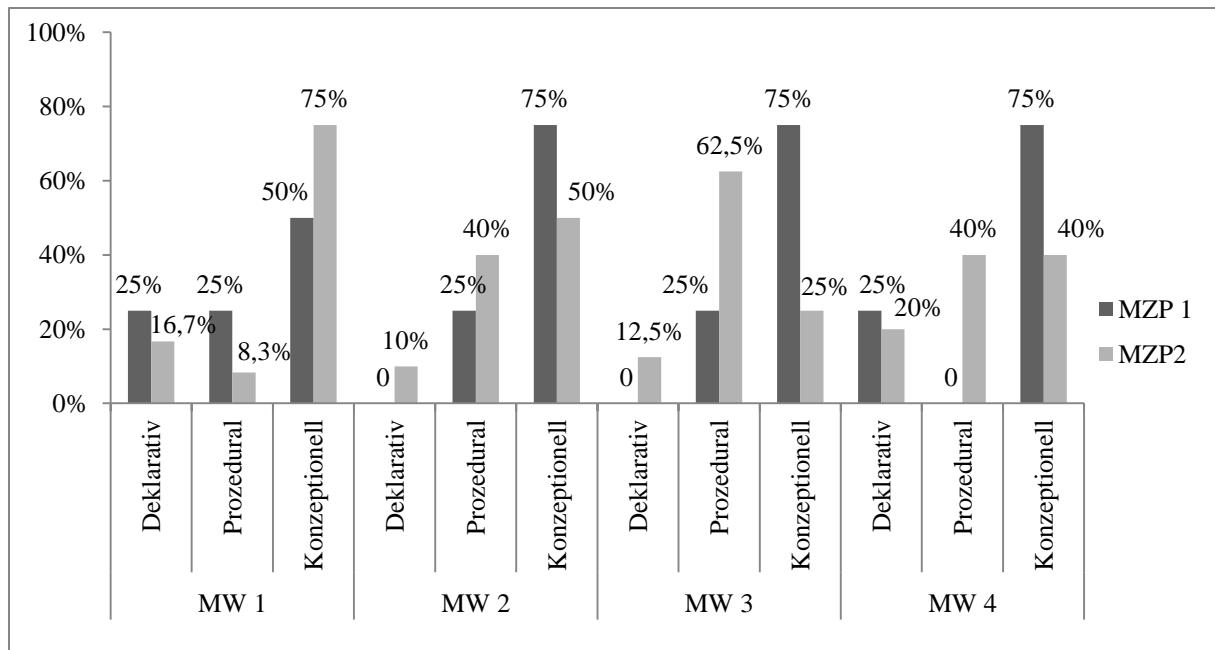


Abbildung 16: Unterschiede in den Wissensarten in bestehenden und neu erstellten Lehrzielen ($n_{MZP 1}=24$, $n_{Lehrziele MZP 2}=35$)

5.1.3.4 Verwendung einer Lehrzielmatrix

Nach der Überarbeitung der Lehrziele, sollten die Lehrenden diese in eine Lehrzielmatrix übertragen, um die Übersichtlichkeit von Lehrzielen und den darin beschriebenen Anforderungen zu erhöhen. In der bestehenden Praxis nutzten die Lehrenden dieses Hilfsmittel nicht für die Prüfungserstellung. Im Rahmen des Qualitätsentwicklungsprogramms setzten die Teilnehmenden der Fälle MW 1, MW 2, MW 3 und MW 4 eine Lehrzielmatrix ein, jedoch in unterschiedlichen Formen und Varianten. Bei Fall MW 1 wurden in den Zeilen der Matrix die Veranstaltungsinhalte eingesetzt. Als Verhaltenskomponenten wurden die kognitiven Anforderungsniveaus nach Bloom verwendet (siehe Tabelle 20). Zu jedem Inhalt wurde in der Matrix grafisch dargestellt, welche Niveaustufe diesbezüglich von den Studierenden erreicht werden soll. Zum Teil wurden in den Zeilen der Matrix komplette Lehrziele eingetragen, also mit Inhalts- und Verhaltenskomponente, bei anderen Inhalten wurde hingegen keine Verhaltenskomponente mit angegeben.

Tabelle 20: Nutzung der Lehrzielmatrix bei Fall MW 1

Inhaltskomponente der Lehrziele	Taxonomie-Stufen					
	Stufe 1 Erinnern	Stufe 2 Verstehen	Stufe 3 Anwenden	Stufe 4 Analysieren	Stufe 5 Bewerten	Stufe 6 Entwickeln
Bestandteile des PEP (Abkürzung für einen Prozess der behandelt wird)						
PEP Prozesse sowie Methoden						
Ziele samt Ergebnissen der Prozessphasen						
Entscheidungen zur Terminierung der Meilensteine treffen und begründen						
Fahrzeugkonzepte an Hand der Anforderungen bewerten						
Verbesserungspotenziale erkennen und Handlungsempfehlungen ableiten						
Simulationsmethoden erläutern und deren Einsatzbereiche beschreiben						
geeignete Simulationsmethoden in Abhängigkeit der Problemstellung auswählen						
Fahrzeugkomponenten und –systeme beschreiben						
nicht zuzuordnen						

Mit der schwarzen Linie wird visualisiert, zu welchen Inhaltsbereichen welche Anforderungsniveaus erreicht werden sollen. Für das grau hinterlegte Lehrziel *Verbesserungspotenziale erkennen und Handlungsempfehlungen ableiten* möchten die Lehrenden die Stufe *Entwickeln* erreichen. Die Lehrenden stellten dann in Diskussionen fest, dass sie nicht nur Aufgaben in den Randbereichen der Markierung benötigen (im genannten Beispiel wäre das die Stufe *Entwickeln*), sondern dass es sinnvoll ist, auch Aspekte abzuprüfen, die als Grundlage für höhere Stufen gelten. Folgende Aussagen verdeutlichen dies:

Akademischer Oberrat MW 1: *Wir haben die Tabelle [die Lehrzielmatrix] genommen und in die Felder eingetragen, wo wir die Leute [die Aufgaben erstellen] sehen, quasi das maximale Lehrziel. Und dann sollte versucht werden Aufgaben zu erstellen, die bis dahin gehen. Das ist natürlich eine multidimensionale Optimierung. Wir versuchen in Summe das Feld abzustecken. Wir müssen ja nicht außen [an der Markierung in der Matrix] entlang gehen, sondern brauchen natürlich auch Fragen bei den niedrigeren Lehrzielen. Dann haben wir versucht einen Stand zu verschaffen, und zu schauen, wie oft wir iterieren müssen, bis wir es geschafft haben.*

Doktorand 1: *Was wir uns dabei überlegt haben ist, wir haben den schönen Bereich mit den maximalen Stufen markiert. Dabei haben wir gemerkt, wenn wir nur am Rand Aufgaben haben [z.B.*

auf der Stufe entwickeln], *ist das die schwierigste Klausur, die wir stellen können. Wir haben gesagt, wenn ich was verstanden haben will, muss ich mich erst mal an die Begriffe erinnern haben. So decke ich alle unteren Stufen ab. Wenn ich etwas entwickeln will und nicht anwenden kann, kann ich es auch nicht entwickeln.*

Die gewählte Darstellungsform mit Abgrenzung der zu erreichenden Anforderungsniveaus diene den Lehrenden als Grundlage, um eine Modellierung der Prüfungsanforderungen vorzunehmen, indem Teilanforderungen ausgewiesen und sequenziert wurden. Für die Prüfung gaben die Lehrenden das Ziel vor, nicht nur für die höchste Entwicklungsstufe Aufgaben zu entwickeln, sondern auch für die Stufen darunter, um festzustellen zu können, welches Niveau die Studierenden erreichen. Die Notenvergabe kann ebenfalls entlang dieser Differenzierung ausgerichtet werden, indem die Punkte so verteilt werden, dass die Notenstufen die Anforderungsniveaus abbilden.

Bei Fall MW 2 wurden die Lehrziele auf einem allgemeineren inhaltlichen Niveau formuliert wie bei Fall MW 1 (siehe dazu auch Abschnitt 5.1.2.3). In den Lehrzielen wurden die einzelnen Inhalte nicht separat aufgeführt, da aufgrund der vielen unterschiedlichen Themenbereiche eine hohe Zahl an Lehrzielen benötigt würde. Den Lehrenden ging es vielmehr darum, das zu allen Themenbereichen, die in den verschiedenen Veranstaltungskapiteln angesprochen werden, *Wissen und Terminologien* behandelt werden, *Bezüge zwischen Themenbereichen hergestellt werden* und *Methoden und Konzepte angewendet* werden. Außerdem wurde differenziert und in der Lehrzielmatrix illustriert, welche kognitiven Anforderungsniveaus bei den jeweiligen Inhaltskomponenten fokussiert werden sollen (siehe Tabelle 21). Anschließend wurden den Zellen der Matrix die Veranstaltungskapitel zugeordnet, um ersichtlich zu machen, zu welchem Kapitel welche Art von Aufgaben erstellt werden muss.

Das Vorgehen bei Fall MW 3 ist vergleichbar zum Vorgehen von Fall MW 2. Bei Fall MW 3 sind die Inhaltskomponenten ebenfalls allgemeiner gefasst. Bei der Erstellung von Aufgaben werden jeweils die konkreten Inhalte aus den Veranstaltungskapiteln eingesetzt. Zur Differenzierung von Anforderungsniveaus wurden die Taxonomiestufen *erinnern*, *verstehen*, *analysieren* und *bewerten* verwendet. Der Teilnehmende von Fall MW 4 legte ebenfalls eine Lehrzielmatrix an, um die bestehenden Aufgaben den Lehrzielen zuzuordnen. Wie bereits in Abschnitt 5.1.3.3 dargelegt, passte er die Anforderungen in den Lehrzielen an die Prüfungsaufgaben an. Die Teilnehmenden von Fall MW 5 führten keine vorgelagerte Festlegung des Prüfungsgegenstands und verwendeten weder Lehrziele noch eine Lehrzielmatrix.

Tabelle 21: Auszug aus der Lehrzielmatrix von Fall MW 2

Inhaltskomponenten der Lehrziele	Kognitive Anforderungsniveaus				
	Kennen	wiedergeben	Verstehen	Anwenden/berechnen	bewerten
Wissen u. Terminologie in den relevanten Themenbereichen	Kapitel 1 a) ===== LZ 1 =====	Kapitel 2b		===== LZ 2 =====	
Zusammenhänge zwischen Themenbereichen	=====	LZ 3 =====	=====		
Methoden und Konzepte aus Themenbereichen				===== LZ 4 =====	

Insgesamt bleibt festzuhalten, dass die Lehrzielmatrix von allen Teilnehmenden der Fälle MW 1 bis MW 4 als hilfreiches Instrument zur Zerlegung, Strukturierung und Sequenzierung von Anforderungen und als Ausgangspunkt für die Erstellung von Aufgaben wahrgenommen und entsprechend verwendet wurde. Die Verwendung einer Lehrzielmatrix kann damit zu einer Erhöhung der inhaltlichen und kognitiven Validität von Prüfungen beitragen. In Abschnitt 5.3 wird berichtet, wie die Lehrzielmatrix genutzt wurde, um Prüfungsaufgaben mit Lehrzielen abzugleichen.

5.1.4 Kurzzusammenfassung

Hinsichtlich der Festlegung des Prüfungsgegenstands wurde bezüglich Forschungsfrage 1 berichtet, welche Zielsetzungen die Lehrenden diesbezüglich für die Arbeit im Qualitätsentwicklungsprogramm formulierten. Im Rahmen von Forschungsfrage 2 wurde ermittelt, inwieweit die Lehrenden in ihrer bestehenden Praxis Maßnahmen und Kriterien berücksichtigten, um den Prüfungsgegenstand kompetenzorientiert und gemäß der Gütekriterien der pädagogischen Diagnostik zu definieren. Bezogen auf Forschungsfrage 3 wurde dargelegt, welche Veränderungen die Lehrenden bei der Festlegung des Prüfungsgegenstands im Laufe des Qualitätsentwicklungsprogramms vornahmen. Die empirischen Befunde für die drei Forschungsfragen sind in Tabelle 22 überblicksartig dargestellt und werden später in Abschnitt 6.1 diskutiert.

Tabelle 22: Erkenntnisse zur Festlegung des Prüfungsgegenstands

Festlegung des Prüfungsgegenstands	FF 1: Zielsetzungen der Teilnehmenden	<ul style="list-style-type: none"> - Ziele zur Festlegung des Prüfungsgegenstands sind auf einem allgemeingültigen Niveau. - Konkrete Probleme wurden nicht angesprochen. - Aus den Zielsetzungen konnte nicht erschlossen werden, welche Schritte die Lehrenden zur Festlegung des Prüfungsgegenstands durchführen und wie sie dabei unterstützt werden können.
	FF 2: Bestehende Praxis	<ul style="list-style-type: none"> - Lehrziele wurden nicht für die Prüfungserstellung verwendet und waren den Lehrenden inhaltlich nicht bekannt. - Hohe Anteile an konzeptionellem Wissen in Lehrzielen, obwohl die verbal geäußerten Zielsetzungen bei drei Fällen verstärkt auf deklaratives und prozedurales Wissen abzielen. - Keine theoriegeleitete Festlegung von Prüfungsinhalten und -anforderungen. - Keine konzeptionelle Analyse, die eine Orientierung an Kompetenzen nahelegt. - Prüfungsanforderungen werden operational, während der Erstellung von Aufgaben, mitbestimmt. - Keine Strukturierung und Sequenzierung von Prüfungsanforderungen. - Der Großteil der bestehenden Lehrziele ist unspezifisch. - Keine Verwendung einer Lehrzielmatrix.
	FF 3: Veränderte Praxis	<ul style="list-style-type: none"> - Arbeit mit Lehrzielen und Erhöhung der Anzahl von Lehrzielen. - Reduzierung von Lehrzielen, die auf konzeptionelles Wissen abzielen. Realistischere Einschätzung tatsächlich umgesetzter Prüfungsanforderungen. - Festlegung von Prüfungsanforderungen erfolgt nicht mehr operational, während der Erstellung von Aufgaben, sondern in einem vorgelagerten Schritt. - Verwendung einer Lehrzielmatrix führte zu einer Strukturierung und Sequenzierung von Anforderungen. - Veranstaltungsunterlagen bilden nach wie vor den zentralen Bezugspunkt zur Festlegung von Lehrzielen. Eine Analyse weitere Informationsquellen (Kompetenzstandards, Lehrziele weiterführender Veranstaltungen, Arbeitsplatzanalysen oder Studien zu den betreffenden Lernbereichen) wurde nicht durchgeführt.

5.2 Erstellung von Prüfungsaufgaben

Wie im Theorieteil herausgearbeitet und dem Vorgehensmodell zur Erstellung und Auswertungen von Prüfungen entsprechend (siehe Abbildung 3), werden basierend auf der Festlegung des Prüfungsgegenstands im nächsten Schritt Prüfungsaufgaben erstellt. Wie auch in Abschnitt 5.1 gliedert sich der folgende Abschnitt entlang der drei Forschungsfragen. Zu Forschungsfrage 1 wird berichtet, welche Zielsetzungen die Lehrenden hinsichtlich der Erstellung von Prüfungsaufgaben formulierten (Abschnitt 5.2.1). Zu Forschungsfrage 2 wird beschrieben, wie die Lehrenden in der bestehenden Praxis bisher bei der Erstellung von Prüfungsaufgaben vorgehen, ob sie dabei Kriterien einhielten, um die Gütekriterien zu sichern und inwieweit die bestehenden Aufgaben Merkmale einer Kompetenzorientierung aufweisen (Abschnitt 5.2.2). Als Datengrundlage werden die bereits bestehenden Prüfungen beziehungsweise Prüfungsaufgaben, die im Semester vor Beginn des Qualitätsentwicklungsprogramms eingesetzt wurden, analysiert. Im dritten Abschnitt (5.1.3) werden bezogen auf Forschungsfrage 3, die Erkenntnisse zu Veränderungen, die während der Qualitätsentwicklung an den Aufgaben vorgenommen wurden, berichtet. Dazu werden die neu erstellten Prüfungsaufgaben analysiert.

5.2.1 Zielsetzungen der Hochschullehrenden

Hinsichtlich der Gestaltung von Aufgaben formulierten die Teilnehmenden insgesamt acht Programmziele. In Tabelle 23 sind die Zielsetzungen fallweise dargestellt. Alle Lehrenden äußerten das Ziel, neue Aufgabentypen kennenlernen zu wollen. Die Lehrenden sehen demnach einen Bedarf über die ihnen bekannten Aufgaben hinaus, weitere Aufgabenformate einzusetzen. Neben dem Wunsch, neue Aufgabentypen kennenzulernen, bestand bei Fall MW 4 anscheinend ein Problem darin, Prüfungen für Studierendengruppen mit unterschiedlichem Fachhintergrund zu erstellen und dabei Aufgabenbeschreibungen zu formulieren, mit denen die Studierenden unabhängig vom Studienfach zurechtkommen. Bei Fall MW 5 wurde das Spannungsverhältnis zwischen den zeitlichen Ressourcen für die Prüfungserstellung und der Aufgabenqualität, die in der zur Verfügung stehenden Zeit erreicht werden kann, thematisiert. Aufgrund solcher Effizienzüberlegungen wurden mit den Aufgaben möglicherweise nicht die tatsächlich angestrebten Kompetenzen abgeprüft. Diese Zielsetzung deutet darauf hin, dass zwar ein Bewusstsein hinsichtlich der Aufgabenvalidität besteht (prüfen die Aufgaben tatsächlich das was sie prüfen sollen?), diese jedoch bisher aufgrund fehlender zeitlicher Ressourcen nicht durchgehend erreicht werden konnte.

Tabelle 23: Projektziele zur Erstellung von Prüfungsaufgaben

Fälle	Projektziele
MW 1	Unsicherheit mit neuem Aufgabentyp reduzieren. Welche Art von Aufgaben ist optimal und nachhaltig einsetzbar?).
MW 2	Flexiblere Aufgabengestaltung / neue Aufgabenformate.
MW 3	Entwicklung von neuen Aufgaben und Aufgabentypen.
MW 4	Fokus auf großer Aufgabe: Aufgabe soll in sich geschlossen sein, Teilschritte müssen aufeinander passen, wie holt man alle ab in der ersten Aufgabe?
	Erarbeitung von Aufgabenbeschreibungen in der Prüfung, die alle Studiengänge abholt (besonders Lehramt).
	Generierung neuer, großer Aufgaben.
MW 5	Wie geht man mit dem Trade-off zwischen Qualität und Zeit um? Oftmals könnte man mit einer bestimmten Fragestellung / Aufgabentyp die Kompetenz bestmöglich abbilden, jedoch würde das Korrigieren der Aufgabe so viel Zeit in Anspruch nehmen, dass man Aufgaben verwendet, die die Kompetenz weniger gut abbilden
	Best Practice Beispiele für offene Fragen kennenlernen / entwickeln

5.2.2 Erkenntnisse zur bestehenden Prüfungspraxis

In diesem Abschnitt werden die Erkenntnisse zu den Prüfungsaufgaben berichtet, die im Semester vor Beginn der Qualitätsentwicklungsarbeit in den jeweiligen Prüfungen eingesetzt wurden. Bezüglich Forschungsfrage 2 werden die Aufgaben dahingehend bewertet, ob sie zentrale Qualitätskriterien der pädagogischen Diagnostik erfüllen und inwieweit sie Merkmale einer Kompetenzorientierung aufweisen (siehe Tabelle 6). Dazu werden sowohl formale als auch inhaltliche Aspekte in den Aufgaben betrachtet. Die formalen Aspekte (Abschnitt 5.2.2.1) sind insbesondere vor dem Hintergrund der Bearbeitbarkeit relevant und beeinflussen damit die Reliabilität. Auf inhaltlicher Seite wird zunächst die Passung zwischen Lehrzielen und Prüfungsaufgaben betrachtet, um festzustellen, inwieweit die Aufgaben, die in den Lehrzielen beschriebenen Anforderungen valide abbilden (Abschnitt 5.2.2.2). Bevor auf die Merkmale von Kompetenzorientierung eingegangen wird, werden die in den Prüfungen eingesetzten Aufgabentypen berichtet, um Charakteristika der eingesetzten Prüfungsaufgaben zu bestimmen (Abschnitt 5.2.2.3). Diese Informationen dienen, wie auch bei der Festlegung des Prüfungsgegenstands, als Grundlage für die Gestaltung und Schwerpunktsetzung der Inhalte des Qualitätsentwicklungsprogramms und vermitteln einen ersten Eindruck über die umgesetzte Anforderungsheterogenität der Prüfungen. In Abschnitt 5.2.2.4 wird dargelegt, inwieweit die im Theorieteil eingeführten Merkmale für Kompetenzorientierung (Wissensarten, Aufgabenoffenheit, Aufgabenkontexte und Aufgabenkomplexität) bisher in Aufgaben umgesetzt wurden.

5.2.2.1 Formale Qualitätsindikatoren von Prüfungsaufgaben

Prüfungsaufgaben müssen formal so gestaltet werden, dass aufgrund der Gestaltungsmerkmale keine konstruktferne Varianz erzeugt wird, durch die die Reliabilität einer Prüfung eingeschränkt werden könnte (siehe dazu auch die Abschnitte 2.1.2.2 und 2.1.3.2). Als wesentlich werden dabei sprachliche Merkmale (eindeutige Sprache, keine doppelten Verneinungen, keine unnötigen Informationen), die Angabe eines Erwartungshorizonts, die Unterteilung von Handlungsaufforderungen in Teilaufgaben (anstatt einer Aneinanderreihung mehrerer Handlungsaufforderungen in einer Aufgabe) sowie die Unabhängigkeit zwischen Aufgaben (sind Folgeaufgaben unabhängig voneinander bearbeitbar) betrachtet (für eine ausführliche Beschreibung dieser Aspekte siehe Tabelle 37).

Hinsichtlich sprachlicher Merkmalen konnten so gut wie keine Probleme in den Prüfungen festgestellt werden. Doppelte Verneinungen im Aufgabenstamm, die zu Irritationen führen könnten, gab es keine. Zu 81,8 Prozent waren die Handlungsaufforderungen anstatt als Frage als Aussage formuliert. Dies ist als positiv zu werten, da die Studierenden dadurch in der Regel konkretere Anweisungen erhalten, wie eine Aufgabe zu bearbeiten ist. In 99,3 Prozent der Fälle wurde die verwendete Sprache im Aufgabenstamm als eindeutig und klar eingeschätzt, was ebenfalls Irritationen unter den Studierenden minimieren sollte.

Hinsichtlich der Instruktionsqualität lag der Anteil an problematischen Aufgaben etwas höher als bei den sprachlichen Merkmalen. Bei 18,3 Prozent waren keine klaren Erwartungshorizonte angegeben. Die Studierenden haben bei diesen Aufgaben keine Anhaltspunkte, wann die Aufgabenstellung vollständig bearbeitet ist und welche Aspekte in einer Antwort keinesfalls fehlen sollten. Der Anteil bei den Aufgaben, deren Bearbeitung nur in Abhängigkeit des Ergebnisses der vorangegangenen Aufgaben möglich war, lag bei 28,2 Prozent. Im Falle dieser Aufgaben, die abhängig von der vorhergehenden Aufgabe waren, da sie nur mit dem dort zu erarbeitenden Ergebnis gelöst werden konnten ($n=74$), wurde wiederum in 94,6 Prozent keine Zwischenlösung angegeben. Hier besteht folglich Nachbesserungsbedarf, da ansonsten nicht aufgeklärt werden kann, ob die Studierenden die Folgeaufgabe aufgrund fehlender Kompetenzen oder lediglich aufgrund der fehlenden Zwischenlösung nicht bearbeiten konnten. Für die Studierenden sind Abhängigkeiten zwischen Aufgaben ein Nachteil, da sie Aufgaben möglicherweise auf Grund ihrer Fähigkeiten lösen könnten, jedoch eine Bearbeitung aufgrund der fehlenden Lösung einer vorangegangenen Aufgabe nicht bearbeitet werden kann. Bei knapp über der Hälfte der Fragen waren in einem Aufgabenstamm außerdem mehrere Handlungsaufforderungen zusammengefasst. Auch hier sollte eine Trennung vorgenommen werden, sodass pro Aufgabe eine Handlungsanweisung ausgewiesen wird. Dadurch erhöht sich die Übersichtlichkeit der Aufgaben. Außerdem können die Lehrenden die

Lösungshäufigkeiten bezüglich einzelner Handlungsaufforderungen differenzierter nachvollziehen. Möglicherweise wird außerdem die Korrektur erleichtert, insofern die Antworten getrennt nach Handlungsaufforderungen betrachtet werden können.

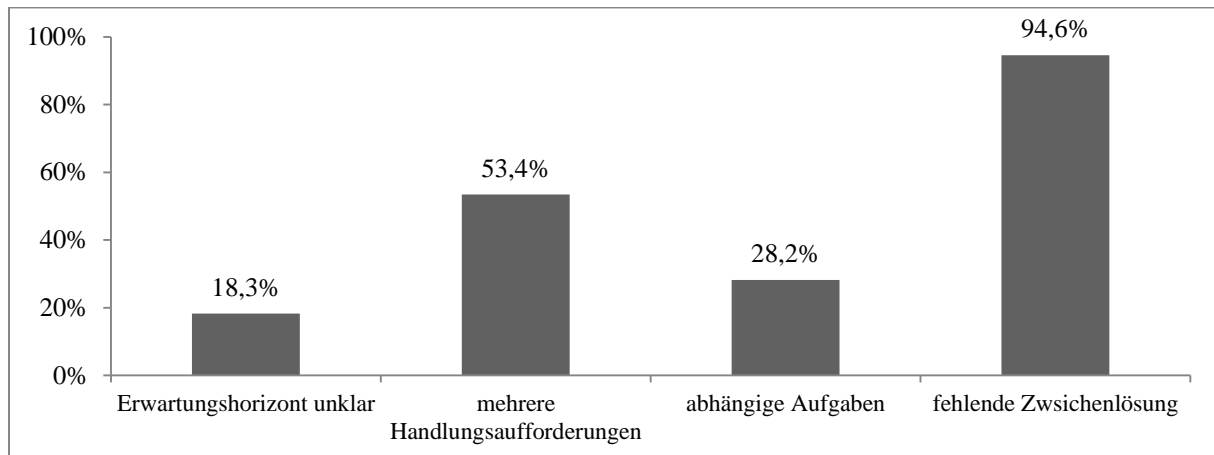


Abbildung 17: Formale Gestaltungsmerkmale von Aufgaben (Anzahl der Aufgaben mit fehlenden Zwischenlösungen bei $n=74$; $n_{\text{gesamt MZP I}}=262$)

Wie schon angemerkt, kann auf Basis dieser Ergebnisse nicht geschlossen werden, inwieweit die Studierenden tatsächlich Probleme bei der Bearbeitung aufgrund der Nichteinhaltung der dargestellten Qualitätsaspekte haben. Dieser Aspekt könnte beispielsweise durch eine kognitive Validierung, bei der die Studierenden ihre Denkprozesse beim Lösen der Aufgabe verbalisieren, beantwortet werden. Die Erkenntnisse liefern jedoch einen Eindruck der formalen Qualität der Prüfungen. Verbesserungspotenziale sind insbesondere in Bezug auf Abhängigkeiten zwischen Aufgaben zu sehen. Diesbezüglich besteht möglicherweise kein wirkliches Problembewusstsein unter den Lehrenden, da die Abhängigkeiten kaum durch die Angabe von Zwischenlösungen aufgelöst werden.

5.2.2.2 Passung zwischen Lehrzielen und Prüfungsaufgaben

Das zentrale inhaltliche Qualitätskriterium von Prüfungsaufgaben im Hinblick auf die Validität ist deren Passung zu den Lehrzielen. In den Analysen konnte eine deutliche Diskrepanz zwischen Lehrzielen und den in Prüfungen umgesetzten Anforderungen festgestellt werden. Bei den Lehrzielen liegt der Schwerpunkt auf konzeptionellem Wissen (66,7 %), wohingegen deklaratives Wissen kaum angesprochen wurde (12,5 %). In den Aufgaben hingegen konnte ein umgekehrter Fokus festgestellt werden. Über die Hälfte der Aufgaben (55,9 %) zielen auf deklaratives Wissen ab, während nur 17,6 Prozent konzeptionelles Wissen abprüfen. Hinsichtlich prozeduralem Wissen ist die Diskrepanz zwischen Lehrzielen (Anteil von 26,5 % der auf prozedurales Wissen abzielt) und Prüfungsaufgaben (20,8 % prozedurales Wissen), deutlich geringer (siehe Abbildung 18). Die hier dargestellten Unterschiede zwischen den Wissensarten in Lehrzielen und Aufgaben wurde von den Lehrenden während der Fortbildung bestätigt, als sie anhand einer rationalen Auf-

gabenanalyse ihre Aufgabenanforderungen auf die Lehrziele rückbezogen (siehe Abschnitt 5.3.3). Den Studierenden werden folglich mit dem Bestehen einer Prüfung Qualifikationen im Bereich konzeptionellen Wissens zugeschrieben, die jedoch nicht ausreichend geprüft wurden.

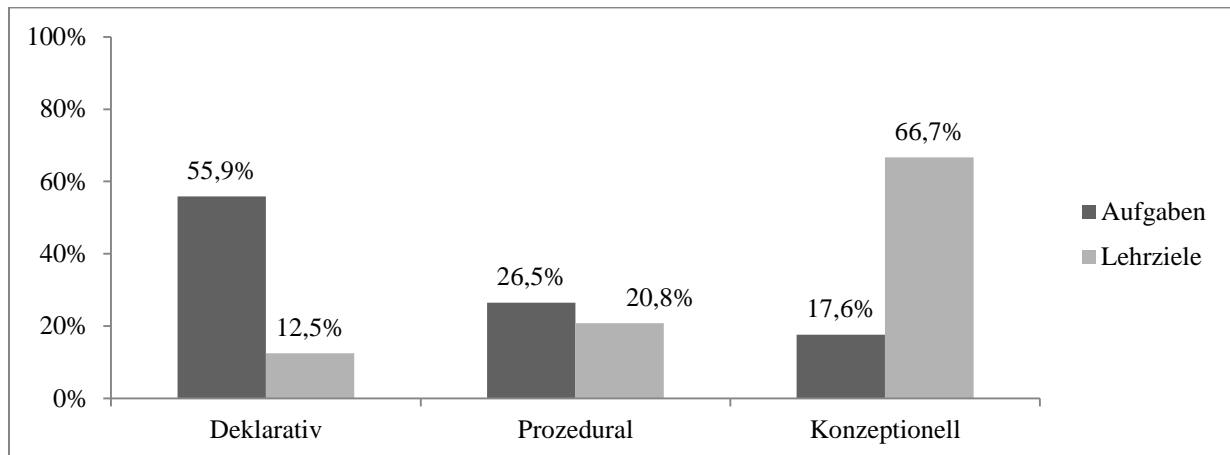


Abbildung 18: Verteilung der Wissensarten in Lehrzielen und Aufgaben ($n_{\text{Lehrziele}}=24$; $n_{\text{Aufgaben}}=262$)

Bei fallweiser Betrachtung ergibt sich ein ähnliches Bild. Die Anteile an deklarativem Wissen sind bei allen Fällen in den Aufgaben stärker vertreten, die Anteile an konzeptionellem Wissen hingegen in den Lehrzielen (siehe Abbildung 19). Bei Fall MW 1 ist die Diskrepanz zwischen Lehrzielen und Aufgaben hinsichtlich konzeptionellen Wissens relativ gering. Hier gelingt es am ehesten, auch die höherwertigen Zielsetzungen im Sinne konzeptionellen Wissens adäquat in Prüfungsaufgaben umzusetzen. Bei den Fällen MW 2 und MW 3 ist die Diskrepanz zu konzeptionellem Wissen besonders hoch (jeweils 75 % der Lehrziele bei 2.2 bzw. 7,5 % der Aufgaben). Demzufolge müssten die in den Lehrzielen beschriebenen Anforderungen nach unten korrigiert werden, indem die Anzahl an Lehrzielen mit deklarativem Wissen erhöht werden, bei gleichzeitiger Reduktion von Lehrzielen, die auf konzeptionelles Wissen abzielen. Bei Fall MW 5 hingegen wurden die verbal geäußerten Zielsetzungen (Praxisorientierung und ganzheitliches Verständnis) für die Veranstaltung als kongruent zu den Lehrzielen eingeschätzt. Hier scheint es also eher nicht ausreichend zu gelingen, die selbst gesetzten Ziele, gerade im Hinblick auf die anspruchsvolleren Aspekte zu konzeptionellem Wissen, entsprechend in Prüfungsaufgaben umzusetzen.

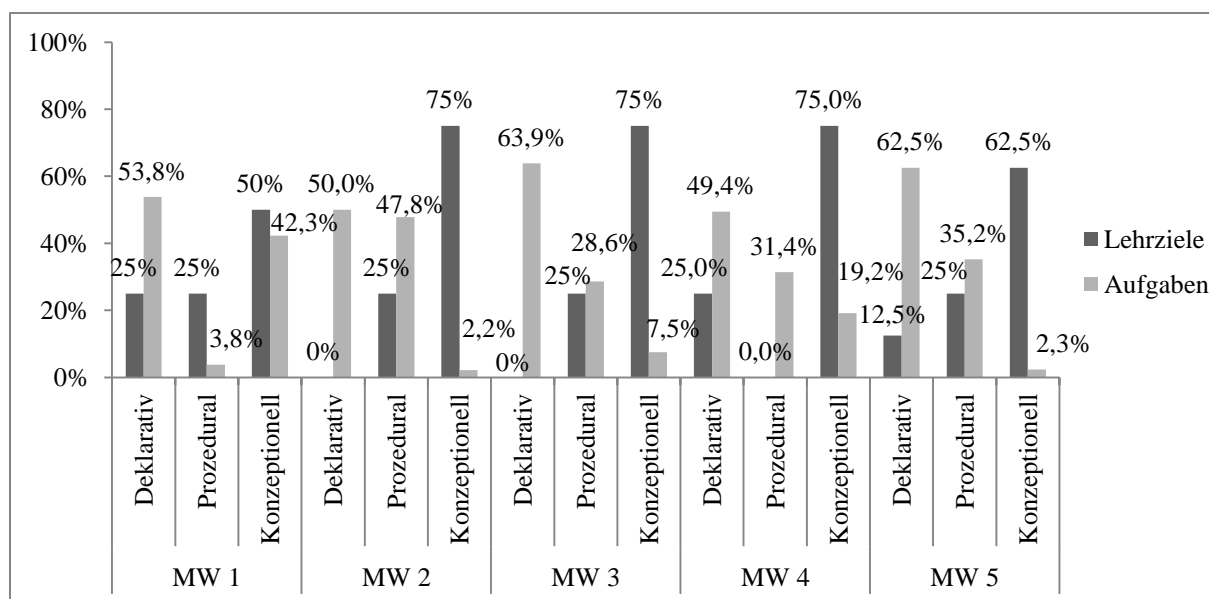


Abbildung 19: Prozentuale Verteilung der Wissensarten in Lehrzielen und Prüfungsaufgaben ($n_{\text{Lehrziele}}=24$; $n_{\text{Aufgaben}}=262$)

Insgesamt konnte eine deutliche Diskrepanz zwischen Lehrzielen und den in Prüfungen umgesetzten Anforderungen festgestellt werden. Die Validität der Aufgaben in Bezug auf die Lehrziele war damit bisher nicht ausreichend gegeben. Im Rahmen des Qualitätsentwicklungsprogramms wurden die Lehrenden auf diese Diskrepanz aufmerksam gemacht. Außerdem wurden Möglichkeiten erarbeitet, um eine Angleichung zwischen Lehrzielen und Aufgaben zu erreichen, die entweder darauf abzielten die in den Lehrzielen beschriebenen Anforderungen nach unten zu korrigieren (im Sinne einer Reduktion von konzeptionellem Wissen bei gleichzeitiger Erhöhung von Lehrzielen mit deklarativem Wissen) oder aber mehr Aufgaben einzusetzen, die auf konzeptionelles Wissen abzielen.

5.2.2.3 Differenzierung von Aufgabentypen

In diesem Abschnitt werden die Prüfungsaufgaben charakterisiert, indem Aufgabentypen hinsichtlich von Bearbeitungsumfängen, Antwortformaten, den zugelassenen Antwortarten, den eingesetzten Operatoren und der Verwendung verschiedener Darstellungsformen näher beschrieben werden. Dies diente der zielorientierten Ausrichtung des Qualitätsentwicklungsprogramms an den Charakteristika der bestehenden Prüfungen.

Bearbeitungsumfänge

Die fünf Prüfungen, die im Semester vor Aufnahme der Qualitätsentwicklungsarbeit eingesetzt wurden, enthielten insgesamt 252 Prüfungsaufgaben (nach der Segmentierung, die im Zuge der Kodierung vorgenommen wurde, siehe Abschnitt 4.4.1). Die Anzahl der Aufgaben unterscheidet sich zwischen den Fällen. Am meisten Aufgaben waren in den Prüfungen der Fälle MW 1 (60

Aufgaben) und MW 2 (63 Aufgaben) enthalten, am wenigsten bei Fall MW 5 mit 36 Aufgaben. Die Bearbeitungsdauer der Prüfungen lag entweder bei 90 (MW 1, MW 2, MW 4) oder 60 Minuten bei den Fällen MW 3 und MW 5 (siehe Abbildung 20). Die Bearbeitungsdauer pro Aufgabe fiel pro Fall unterschiedlich aus. Am wenigsten Zeit hatten die Studierenden bei Fall MW 3 mit im Schnitt 1,3 Minuten pro Aufgabe, gefolgt von MW 2 (1,4 Minuten) und MW 1 mit 1,5 Minuten pro Aufgabe. Geringfügig mehr Zeit für die Bearbeitung einer Aufgabe erhielten die Studierenden bei den Prüfungen der Fälle MW 4 (1,6 Minuten) und MW 5 (1,7 Minuten). Auch wenn eine Bewertung der Bearbeitungsdauer ohne Kenntnis der jeweiligen Inhalte schwierig ist, erscheint die Zeit, die zur Lösung einer Aufgabe im Schnitt zur Verfügung steht, als relativ kurz. Eine Strukturierung und Lösung einer komplexeren Problemstellung ist in 90 bis 120 Sekunden vermutlich eher nicht leistbar. Die richtige Lösung muss bei der Bearbeitung von Aufgaben mehr oder weniger vorgefertigt abrufbar sein, ohne dass größere Veränderungen an eingeübten Lösungsroutinen notwendig sind. Auf formaler Seite können sich die kleinteiligen Aufgaben hingegen positiv auswirken, da die Studierenden ihr Wissen und Können an mehreren kleinen Aufgaben zeigen können, anstatt an wenigen großen Aufgaben. Das Nichtbestehen einzelner Aufgaben ist bei kleineren Aufgaben weniger schwerwiegend als bei wenigen großen Aufgaben. Jedoch sollten Abhängigkeiten zwischen den Aufgaben unbedingt vermieden werden.

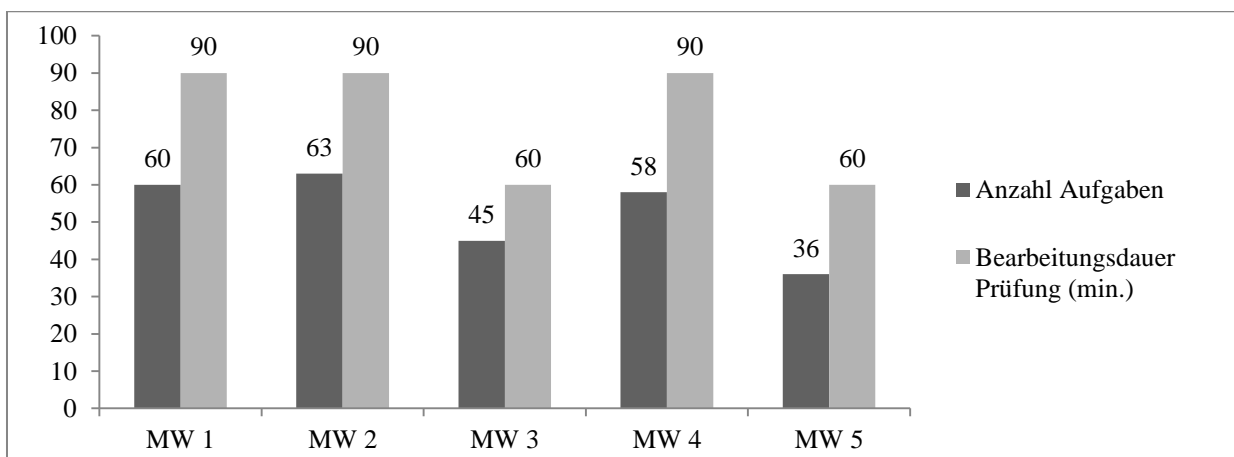


Abbildung 20: Anzahl der Aufgaben und Bearbeitungsdauer in den bestehenden Prüfungen ($n=262$)

Antwortformate

Die Lehrenden setzten in ihren Prüfungen zu 91,6 Prozent Aufgaben mit einem offenen Kurzantwortformat ein (siehe Abbildung 21; für eine genaue Beschreibung der Antwortformate siehe Tabelle 37). Die 5,7 Prozent der Aufgaben mit Wahlantworten waren in der Prüfung von Fall MW 2 enthalten, bei der von 63 Aufgaben 15 im Wahlantwortformat umgesetzt waren. Die anderen vier Prüfungen enthielten keine Wahlantwortaufgaben. Eine Variation von Antwortformaten fand demnach bisher so gut wie nicht statt.

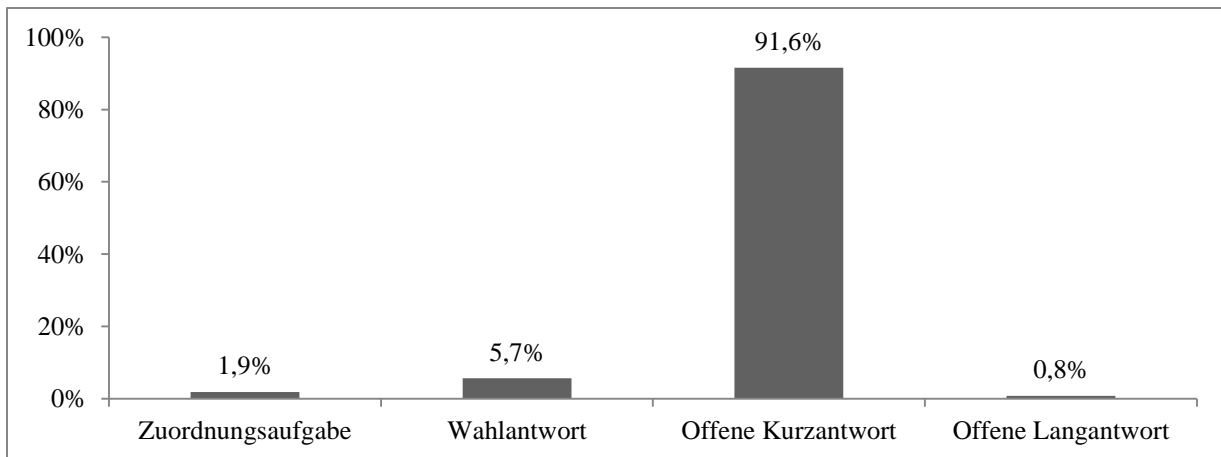


Abbildung 21: Eingesetzte Antwortformate in den bestehenden Prüfungsaufgaben ($n=262$)

Antwortarten

Erste Hinweise auf die Umsetzung verschiedener Anforderungen im Rahmen von offenen Kurzantworten liefert eine Betrachtung der Antwortarten, die von den Studierenden bei der Lösung der Aufgaben verwendet werden musste. Wie in Abbildung 22 dargestellt, erfolgten über zwei Drittel der Antworten in Prosa (geschriebenen Text) und knapp über 20 Prozent anhand von Formalsprache (z.B. Formeln oder mathematische Sätze). Das Anfertigen von Grafiken oder Zeichnungen fällt mit 8,3 Prozent der Fälle relativ gering aus. Die Möglichkeit, den Studierenden freizustellen, welche Antwortart sie zur Beantwortung der Aufgabenstellung wählen, wurde lediglich bei zwei Aufgaben umgesetzt (0,8 %). Demnach wurde auch bezüglich der zugelassenen Antwortarten, relativ wenig variiert.

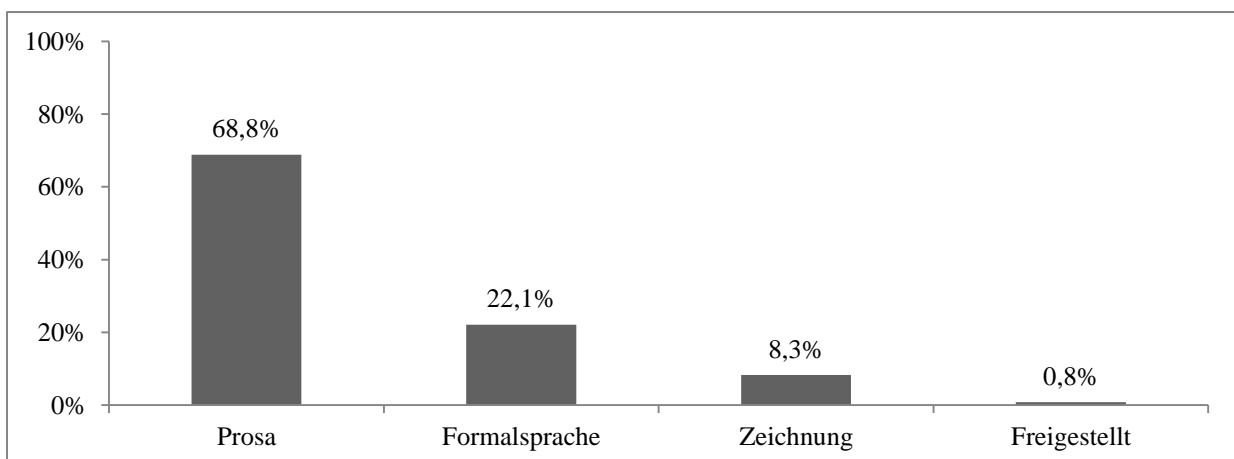


Abbildung 22: Art der Antwort in Kurzantwortformaten ($n=240$)

Verwendete Operatoren

Die Aufgaben wurden weiterhin anhand der darin verwendeten Operatoren analysiert. In 25 Prozent der Fälle war kein Operator in den Prüfungsaufgaben enthalten, da die Aufgabe als Frage formuliert war. Betrachtet man die restlichen Aufgaben, tritt mit 26,3 Prozent am häufigsten der Operator *nennen sie* (siehe Abbildung 23) auf. Dieser Operator wird in der Regel eingesetzt, um deklarative Wissensbestände abzufragen. Operatoren die verstärkt auf konzeptionelles Wissen abzielen kamen hingegen selten vor (Begründungen 3,3 %, Erklären 7,1 %, Nennen von Beispielen 2,1 %, Erklären anhand von Beispielen 1,3 %). 14,2 Prozent der Aufgaben enthielten andere als die genannten Operatoren, die jedoch bislang nicht weiter ausgewertet wurden.

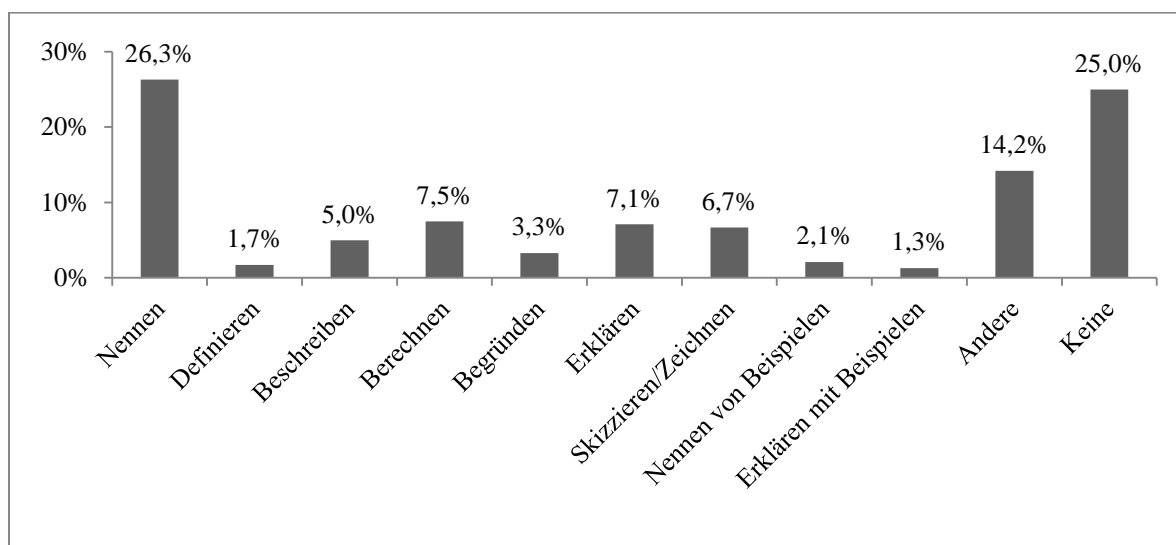


Abbildung 23: Verwendete Operatoren in den bestehenden Prüfungsaufgaben ($n=262$)

Verwendung von Illustrationen in Aufgaben

Bei Betrachtung der Antwortarten und den verwendeten Operatoren fällt der geringe Anteil an Aufgaben auf, bei denen als Antwort eine Zeichnung oder das Anfertigen einer Grafik erforderlich ist. Dementsprechend enthalten drei Viertel der Aufgaben keine Illustrationen. Bei 9,2 Prozent der Aufgaben sind Tabellen enthalten, in 8 Prozent der Fälle Diagramme. Bilder und Grafiken kommen kaum vor (Abbildung 24)

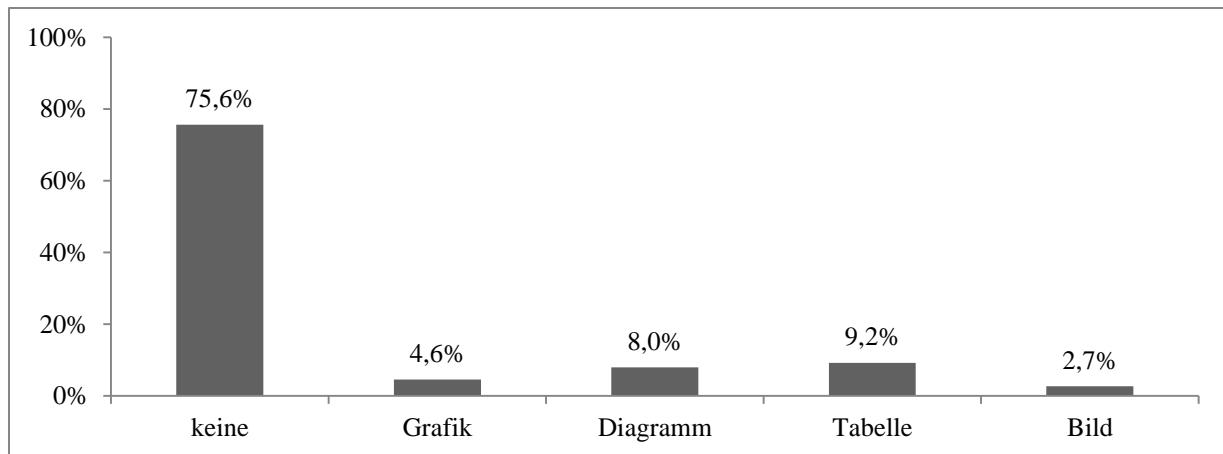


Abbildung 24: Anteil der bestehenden Aufgaben in denen Illustrationen verwendet werden ($n=262$)

Auf Grundlage der dargestellten Erkenntnisse wurden im Rahmen des Qualitätsentwicklungsprogramms Gestaltungsprinzipien und Techniken zur Erstellung offener Kurzantworten thematisiert, da dieses Format mit Abstand am häufigsten vorkommt. Damit die Lehrenden auch innerhalb dieses Formats heterogene Anforderungen umsetzen können, wurden Variations- und Gestaltungsmöglichkeiten für Aufgaben mit offenen Kurzantworten aufgezeigt. Außerdem wurde auf alternative Antwortformate eingegangen. Beispielsweise können Wahlantwortformate bei hohen Studierendenzahlen für die Lehrenden eine sinnvolle Alternative zur offenen Kurzantwort darstellen, da die Korrekturzeiten verringert werden können. Bei einer sorgfältigen Konstruktion ist es mit Wahlantworten für viele Anwendungsfälle ebenfalls möglich, komplexere Anforderungen abzufragen (Haladyna, Downing, & Rodriguez, 2002). Der Anteil an Aufgaben, die Illustrationen enthalten, könnte ebenfalls erhöht werden. Die Ergebnisse zur veränderten Praxis bezüglich dieser Aspekte werden in Abschnitt 5.2.3.4 berichtet.

5.2.2.4 Merkmale einer Kompetenzorientierung in bestehenden Aufgaben

In diesem Abschnitt werden die bestehenden Prüfungsaufgaben dahingehend analysiert, ob die folgenden Merkmale von Kompetenzorientierung umgesetzt wurden: (1) die Differenziertheit der Wissensarten in Aufgaben, (2) die Funktion von Kontextbezügen, (3) die Offenheit der in den Aufgaben abgebildeten Problemstellungen sowie (4) die sprachlogische Komplexität der Aufgaben im Sinne der Menge an zu verarbeitenden Informationen (eine Beschreibung der Indikatoren wird in Kapitel 2.1.1.5 gegeben). Bezüglich der Wissensarten wurde im Theorieteil dargelegt, dass bei der Bewältigung komplexer realer Anforderungssituationen insbesondere konzeptionelles und strategisches Wissen zum Tragen kommt. Demnach müssen in einer kompetenzorientierten Prüfung vorrangig konzeptionelles und strategisches Wissen abgefragt werden. Eine weitere Anforderung an kompetenzorientierte Aufgaben ist ein Aufgabenkontext, auf den das jeweilige Wissen bezogen werden und der bei der Bearbeitung einer Aufgabe berücksichtigt werden muss. Da-

her wird untersucht, ob in den Aufgaben Situationen beziehungsweise Kontexte beschrieben und darin enthaltene Informationen bei der Bearbeitung und Lösung der Aufgabe von den Studierenden berücksichtigt werden müssen. Weiterhin wird betrachtet, ob die Aufgabenoffenheiten variiert wurden. Dabei wird unterschieden zwischen *definierten und konvergenten* Aufgaben (klar definierter Anfangszugang, ein zugelassener Lösungsweg mit einer richtigen Lösung), *definierten und divergenten* Aufgaben (klar definierter Anfangszustand, mit mehreren zugelassenen Lösungen und Lösungswegen) und *undefinierte und divergente* Aufgaben (weder Ausgangszustand noch Zielzustand sind klar definiert). Flexibles Problemlösen kann insbesondere mit den beiden letztgenannten Aufgabentypen geprüft werden. Als letztes Merkmal kompetenzorientierter Aufgaben wird die sprachlogische Komplexität von Aufgaben betrachtet. Die Aufgaben wurden dazu in drei Komplexitätsstufen *einfach, weniger komplex* und *komplex* eingeteilt.

Wissensarten in den Prüfungsaufgaben

Bei 60 Prozent der Aufgaben wurde deklaratives Wissen abgefragt. Strategisches Wissen wurde in keiner Aufgabe umgesetzt, konzeptionelles Wissen kommt in 17 Prozent der Aufgaben vor und spielt damit eine untergeordnete Rolle. Betrachtet man die Verteilung der Wissensarten nach den vergebenen Punkten (anstatt der Häufigkeit der Aufgaben) ergibt sich das gleiche Bild (siehe Abbildung 25). Geht man davon aus, dass zur Bestimmung der Note ein Summenscore über alle Aufgaben gebildet wird, können die Ergebnisse so gelesen werden, dass bei der Bildung einer Endnote in den betrachteten Prüfungen bisher zu 55,9 Prozent das Abschneiden hinsichtlich deklarativer Wissensbestände einfluss, 26,5 Prozent der Punkte konnten über das Beherrschen prozeduraler Anforderungen erreicht werden, 17,6 Prozent der Punkte über konzeptionelles Wissen.

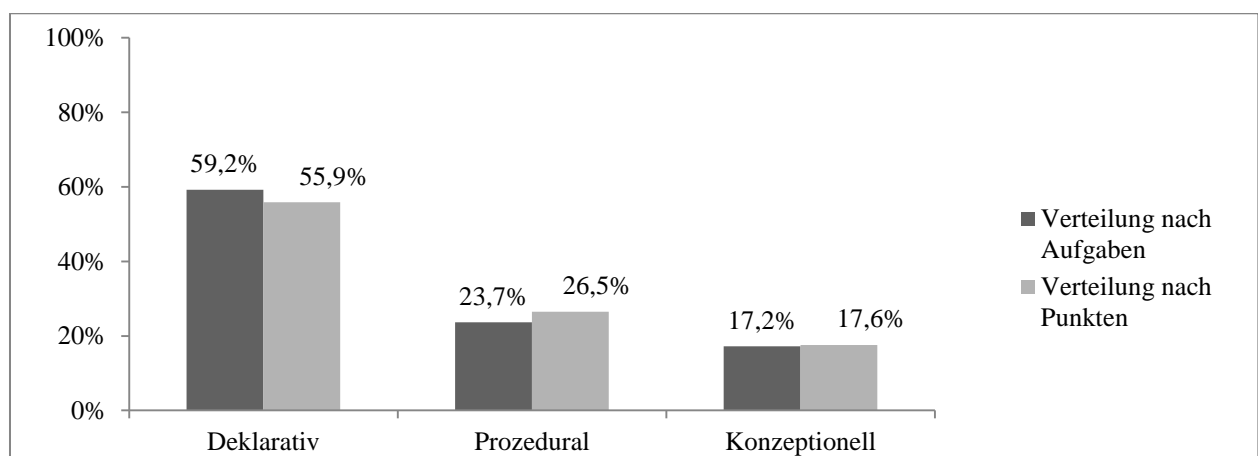


Abbildung 25: Verteilung der Wissensarten nach Aufgabenanzahl und nach der Anzahl der vergebenen Punkte ($n_{\text{Aufgaben}}=262$; $n_{\text{Punkte}}=518,5$)

Auf Fallebene zeigt sich, dass bei Fall MW 1 der Anteil an Punkten die für konzeptionelles Wissen vergeben wurden, mit 42,3 Prozent deutlich höher ausfiel, als bei den restlichen Fällen (siehe

Abbildung 26). Die verbal geäußerte Zielsetzung für die Lehrveranstaltung von Fall MW 1, die als praxisorientiert mit einem Fokus auf ganzheitlichem Verständnis charakterisiert wurde (siehe Abschnitt 5.1.2.1), schlug sich demnach in den Aufgaben nieder. Bei Fall MW 5 hingegen, die eine ähnliche Zielsetzung wie MW 1 äußerten, war der Anteil an Punkten die auf Aufgaben zu konzeptionellem Wissen vergeben wurden mit 2,3 Prozent sehr gering, was auf eine deutliche Diskrepanz zwischen angestrebten und realisierten Prüfungsanforderungen hindeutet. Der höchste Anteil an deklarativem Wissen ist bei Fall MW 3 zu finden. Die Einschätzung der Teilnehmerin von Fall MW 3 über eine zu hohe Wissenslastigkeit in der bestehenden Prüfung (siehe Zitat 12, S. 269) konnte durch die Analysen der Wissensarten in den Aufgaben bestätigt werden. Bei Fall MW 2 kamen Aufgaben zu deklarativem Wissen (50 %) und prozeduralem Wissen (47,8 %) etwa gleich häufig vor. Aufgaben zu konzeptionellem wurden mit 2,2 Prozent kaum eingesetzt. Bei Fall MW 4 liegt der Anteil von Aufgaben mit deklarativem Wissen bei 49,4 Prozent, 30 Prozent der Aufgaben fokussieren prozedurales Wissen, 19,2 Prozent konzeptionelles Wissen.

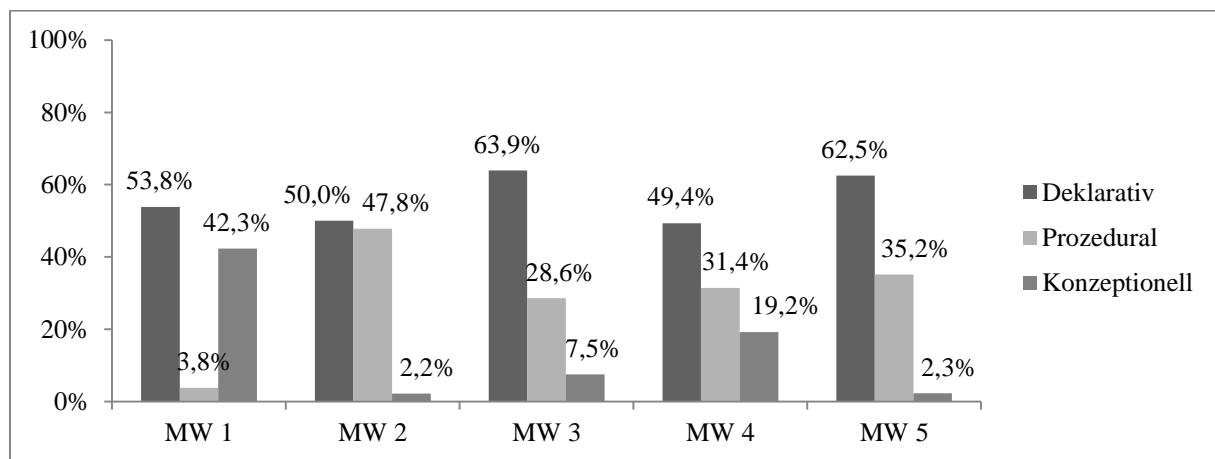


Abbildung 26: Fallweise Verteilung der Wissensarten anhand der vergebenen Punkte ($n_{\text{Punkte}}=518,5$)

Die Vermutung, die bereits nach Betrachtung der eingesetzten Operatoren in den Aufgaben aufgestellt wurde, dass die Prüfungsinhalte bisher stark auf deklaratives Wissen abzielen, kann nach Analyse der Wissensarten in den Prüfungsaufgaben bestätigt werden. Konzeptionelles Wissen wird in den bestehenden Prüfungen hingegen selten gefordert. Eine Aufgabe zu strategischem Wissen kam nicht vor. Im Hinblick auf eine kompetenzorientierte Gestaltung von Prüfungen, sollte daher eine Verschiebung der Schwerpunktsetzung hin zu mehr konzeptionellem und strategischem Wissen erfolgen.

Funktion von Aufgabenkontexten in den Prüfungsaufgaben

In knapp 60 Prozent der Aufgaben wurden keine Kontextbezüge hergestellt. Die Lösung der Aufgaben erfolgte rein innerfachlich. In 31,7 Prozent der Fälle wurde ein vorgeschalteter Kontext verwendet. Die tatsächliche Lösung der Aufgabe wurde dabei jedoch nicht beeinflusst. In diesem

Fall veränderten sich die Aufgabenanforderungen nicht, die Lösung kann ohne Beachtung der Kontextvariablen rein innerfachlich gelöst werden. Inhaltlich sind derartige Aufgabenkontexte somit wertlos. Hinsichtlich der Aufgabenbearbeitung durch die Studierenden muss jedoch beachtet werden, dass durch die Kontexte die Leselast in den Aufgaben erhöht wird, was als ungewollter Störfaktor wirken und damit die Reliabilität verringern kann. Als Teil der Problemstellung, wie es bei einer kompetenzorientierten Umsetzung von Aufgaben vorzusehen wäre, fungierte der Kontextbezug jedoch nur in 5,3 Prozent aller Aufgaben. Lediglich bei Fall MW 1 ist der Anteil der Aufgaben, in denen der Kontextbezug bei der Lösung der Aufgaben berücksichtigt werden muss, mit 16,7 Prozent etwas höher als bei den restlichen Fällen. In den Arbeitstreffen wurden daher Einsatz- und Gestaltungsmöglichkeiten von Kontextbezügen thematisiert. Vor dem Hintergrund einer kompetenzorientierten Prüfungsgestaltung wurde darauf hingewirkt, dass der Anteil an Kontextbezügen, die als Teil der Problemstellung fungieren, erhöht wird.

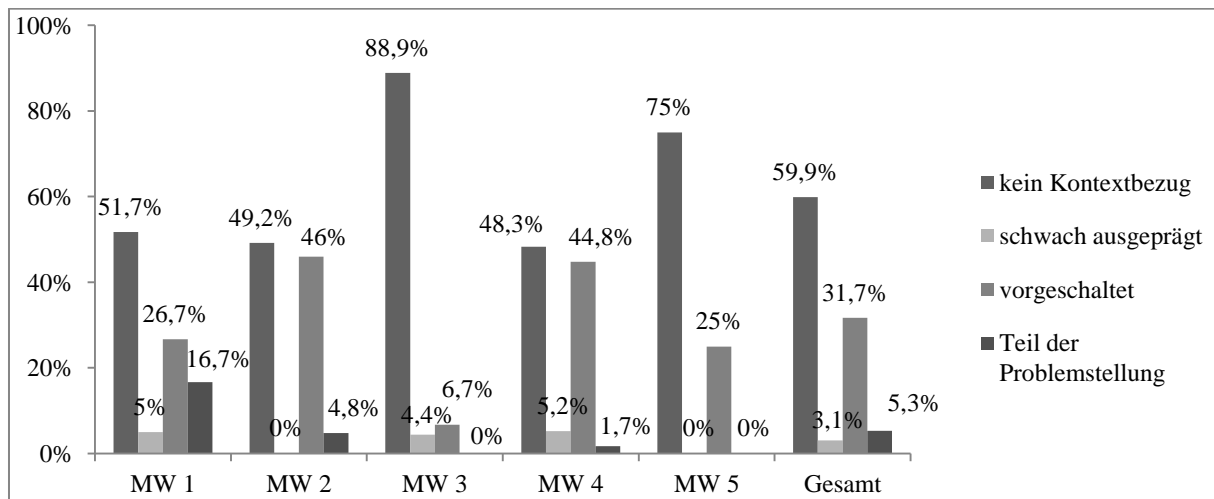


Abbildung 27: Funktion von Aufgabenkontexten in den bestehenden Prüfungsaufgaben ($n=262$)

Aufgabenoffenheiten

Wie die Analysen zeigen, wurde auch bei den Aufgabenoffenheiten bisher wenig variiert. 90,5 Prozent aller Aufgaben wurden als definiert und konvergent klassifiziert, 9,5 Prozent als definiert und divergent. undefinierte und divergente Aufgaben kamen nicht vor. Bei Fall MW 1 waren im Gegensatz zu den Prüfungen der anderen Fälle, etwa ein Drittel der Aufgaben definiert und divergent, wodurch den Studierenden bei zwar klarem Anfangszustand unterschiedliche Lösungswege oder Lösungen zur Beantwortung der Aufgaben eröffnet wurden (siehe Abbildung 28).

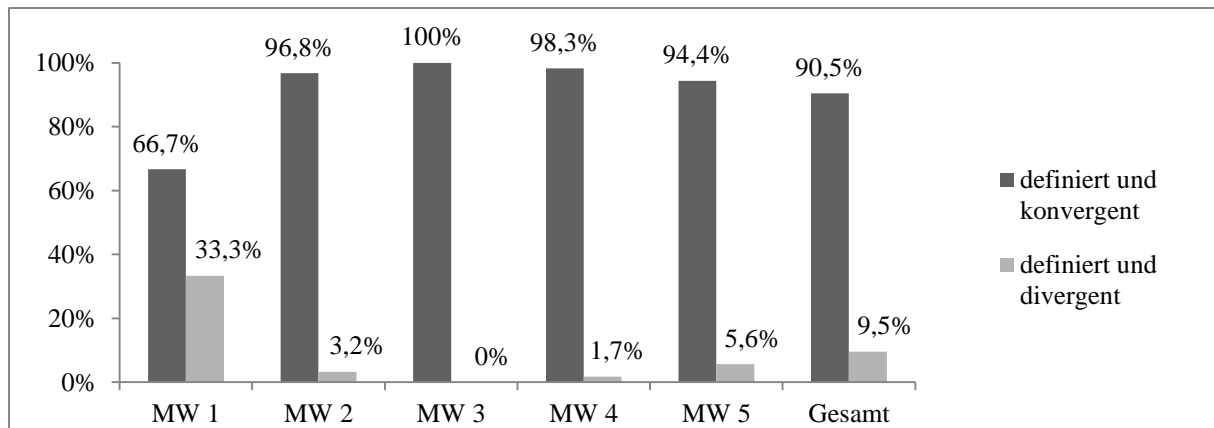


Abbildung 28: Offenheit der bestehenden Prüfungsaufgaben (n=262)

Aufgrund der geringen Anteile an offenen Aufgaben wurde im Rahmen des Qualitätsentwicklungsprogramms auf diesen Aspekt der Aufgabengestaltung hingewiesen. Außerdem wurden Möglichkeiten und Techniken aufgezeigt werden, um Aufgaben zu öffnen und weniger engzuführen. Eine Herausforderung dabei war es, Möglichkeiten zu finden, um die Aufgaben auf eine Art und Weise zu öffnen, ohne dass sich dadurch der Korrekturaufwand deutlich erhöht.

Sprachlogische Komplexität von Aufgaben

Wie bei den anderen Merkmalen einer kompetenzorientierten Aufgabengestaltung, wurde auch hinsichtlich der sprachlogischen Komplexität der Aufgaben bisher wenig variiert. Bei 84 Prozent aller Aufgaben wurde eine einfache Darstellung verwendet, bei elf Prozent eine wenig komplexe und bei fünf Prozent eine komplexe Darstellung. Bei fallweiser Betrachtung zeigt sich, dass die als weniger komplex charakterisierte Gestaltung von Aufgaben bei den Fällen MW 2 (25,4 %) und MW 4 (15,5 %) am häufigsten eingesetzt wurde. Die komplexe Darstellung kam bei Fall MW 1 mit 8,3 Prozent am häufigsten vor (siehe Abbildung 29).

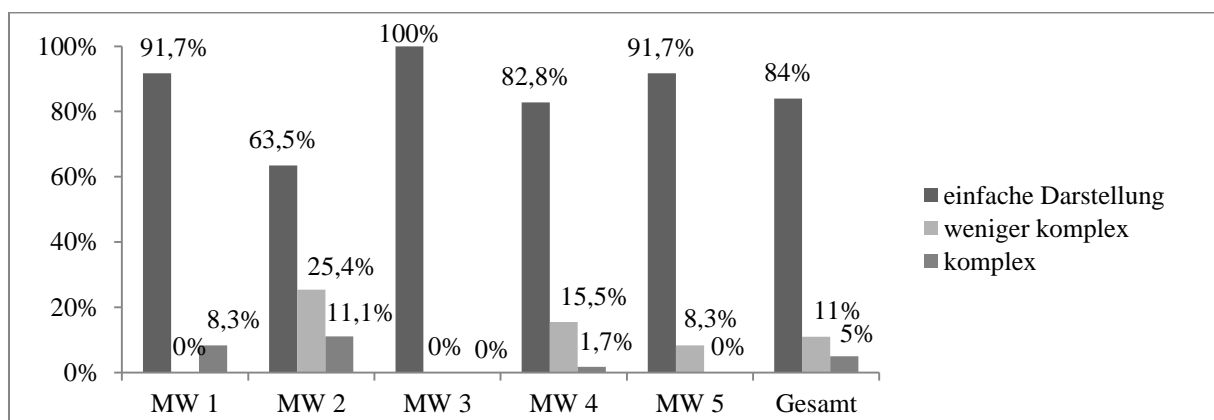


Abbildung 29: Sprachlogische Komplexität der Aufgaben (n=262)

Zusammenfassend betrachtet kann festgehalten werden, dass relevante Merkmale kompetenzorientierter Wissenstests, mit Ausnahme von Fall MW 1, in den bestehenden Prüfungen nicht oder

nur ansatzweise umgesetzt wurden. Insbesondere die starke Homogenität der Prüfungsanforderungen ist auffallend. Gerade anspruchsvollere Anforderungen im Sinne konzeptionellem oder strategischem Wissen werden kaum umgesetzt. Die Lehrenden müssten im Hinblick auf eine Kompetenzorientierung die Anteile an Aufgaben zu konzeptionellem (und je nach Zielsetzung der Veranstaltung auch zu strategischem) Wissen erhöhen. Eine weitere Differenzierung der Anforderungen sollte durch die Öffnung von Aufgaben oder der Nutzung von Aufgabenkontexten erreicht werden. Auch der Anteil an Aufgaben mit höherer sprachlogischer Komplexität war vor dem Hintergrund der im HQR vorgegebenen Kompetenzziele zum Finden und Nutzen von Informationen erstrebenswert. Im Rahmen des Qualitätsentwicklungsprogramms wurde daher verstärkt auf die genannten Aspekte eingegangen, indem Möglichkeiten aufgezeigt werden, um die Anforderungsheterogenität zu erhöhen. Dazugehörige Ergebnisse werden in Abschnitt 5.2.2.4 berichtet.

5.2.3 Erkenntnisse zur veränderten Prüfungspraxis

In diesem Abschnitt wird geklärt, welche Veränderungen die Lehrenden im Rahmen des Qualitätsentwicklungsprogramms an ihren Prüfungsaufgaben vornahmen. Dabei werden Aspekte fokussiert, die in der bestehenden Prüfungspraxis als problematisch eingestuft wurden. Ausgehend von der Erkenntnislage zur bestehenden Praxis wird zunächst beschrieben, ob die Lehrenden sich bei der Erstellung von Aufgaben an ihren Lehrzielen orientierten (Abschnitt 5.2.3.1). Anschließend wird hinsichtlich der formalen Aufgabenmerkmale berichtet, ob die Anzahl der voneinander abhängigen Aufgaben verringert werden konnte (Abschnitt 5.2.3.2). Bezüglich der Passung zwischen den Anforderungen in Lehrzielen wird dargelegt, ob es den Lehrenden gelungen ist, eine Angleichung zwischen Aufgaben und Lehrzielen zu erreichen (Abschnitt 5.2.3.3). Anschließend wird dargestellt, ob sich hinsichtlich der eingesetzten Aufgabentypen, den Antwortarten, den verwendeten Operatoren und der Verwendung von Illustrationen in Aufgaben Veränderungen ergaben (Abschnitt 5.2.3.4). In Bezug auf die Umsetzung einer Kompetenzorientierung werden die Veränderungen hinsichtlich der in Aufgaben umgesetzten Wissensarten, den Aufgabenkontexten und Aufgabenoffenheiten sowie der Komplexität von Aufgaben berichtet (Abschnitt 5.2.3.5).

5.2.3.1 Ausgangspunkt für die Erstellung von Aufgaben

Eine systematische Erstellung von Aufgaben muss an den Lehrzielen ausgerichtet sein, um die Validität der Aufgaben zu gewährleisten. In der bestehenden Praxis wurden die Aufgaben von den Lehrenden jedoch nicht ausgehend von den Lehrzielen erstellt, sondern während der Durchsicht von Veranstaltungsunterlagen oder durch Umarbeitung von bereits bestehenden Aufgaben. Nach Durchlaufen des Programms orientierten sich die Lehrenden der Fälle MW 1, MW 2 und MW 3 bei der Erstellung von Aufgaben stattdessen an ihren Lehrzielen. Damit schafften die Lehrenden

eine wesentliche Voraussetzung, um die inhaltliche und kognitive Validität der Aufgaben vor dem Hintergrund der Lehrziele zu sichern.

Bei Fall MW 1 berichtete ein Teilnehmer diesbezüglich, dass bislang bei einer nicht zielbezogenen Erstellung von Aufgaben in der Regel eher reproduktionslastige Anforderungen entstanden sind, da diese einfach in Aufgaben umzusetzen sind: *Also was vielleicht noch ganz vorteilhaft ist [bei der Nutzung von Lehrzielen bei der Aufgabenerstellung], was ich auch gemerkt habe. Diese typischen Aufgaben, die man als Ersteller von den Aufgaben leicht konstruieren kann sind diese typischen Erinnern Aufgaben. Das heißt ich gehe das Skript durch. Ah ja, das ist eine schöne Auflistung. Dann schreibe ich halt: Nennen Sie drei wesentliche... Und ich glaube dadurch, dass wir mit den Kollegen dieses Prinzip Skript und Lehrziele durchgegangen sind, wir deutlich weniger von diesen langweiligen Aufgaben bekommen haben. Das hat auf jeden Fall schon ganz gut funktioniert.*

Die Teilnehmende von Fall MW 3 achtete bei der Zusammenstellung der Prüfung darauf, dass zu jedem Lehrziel Aufgaben vorhanden sind. Zu den Lehrzielen, die mit den bestehenden Aufgaben nicht abgedeckt werden konnten, erstellte sie neue Aufgaben. Trotz der Orientierung an bestehenden Aufgaben waren die Lehrziele der maßgebliche Bezugspunkt bei der Erstellung von Aufgaben und der Prüfung, wie folgende Antwort, auf die Frage, welcher Aspekt für die Erstellung der Prüfung am hilfreichsten gewesen sei, verdeutlicht: *Auf jeden Fall die Lehrziele. Dass ich in diesem Sheet [der Lehrzielmatrix] schaue in welchem Bereich sind meine Lehrziele aufgestellt und in welchem Bereich brauche ich meine Aufgaben? Daran habe ich mich am meisten orientiert bei der Prüfungserstellung.*

Der Teilnehmende von Fall MW 4 erstellte die Prüfungsaufgaben, indem er Aufgaben aus Altklausuren übernahm und veränderte. Nach Erstellung der Aufgaben überprüfte der Lehrende, ob und inwieweit sich die Anforderungen in den Aufgaben mit den Lehrzielen decken (vgl. Zitat 35, S. 282). Da die Prüfungsaufgaben nach Einschätzung des Teilnehmers genau die Anforderungen enthielten, die mit der Veranstaltung angestrebt werden sollen (siehe Abschnitt 5.1.3.3), erfolgte eine Anpassung der Lehrziele an die bestehenden Aufgaben. Eine der Aufgabenerstellung vorgelegte theoretische Modellierung von Prüfungsinhalten und -anforderungen wurde daher nicht vorgenommen.

Die Teilnehmenden von Fall MW 5 arbeiteten als einziges Team nicht mit Lehrzielen. Bei der Aufgabenerstellung orientierten sie sich nach wie vor in erster Linie an den Veranstaltungsunterlagen oder an bereits bestehenden Aufgaben. Jedoch wurde bei der Erstellung von Aufgaben eine Lehrzielpyramide herangezogen, in der kognitive Niveaustufen unterschieden werden (siehe Abbildung 30).

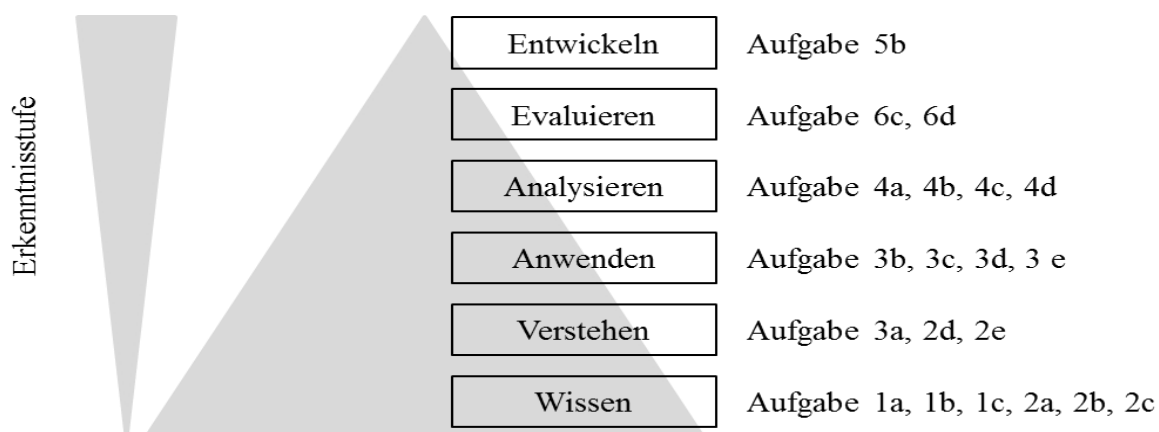


Abbildung 30: Lehrzielpyramide mit einer exemplarischen Zuordnung von Aufgaben

Mit dem Einsatz der Pyramide konnte erreicht werden, dass die Aufgabenanforderungen bei Fall MW 5 nicht mehr rein durch die Charakteristika der Inhalte oder anhand von geläufigen Fragetypen, zustande kamen, sondern sich auf die kognitiven Anforderungsniveaus der dargestellten Lehrzielpyramide beziehen. Insgesamt erstellten bei Fall MW 5 sechs Lehrende jeweils für ein Kapitel der Lehrveranstaltung Prüfungsaufgaben. Dabei wurden sie angewiesen, zu ihrem Kapitel Aufgaben auf unterschiedlichen Niveaustufen zu erstellen. Dadurch sollte eine systematische Variation von Anforderungen zu jedem Kapitel erreicht werden. Die Lehrenden, die Aufgaben erstellten, mussten bei der Abgabe ihrer Aufgaben mitangeben, auf welchen Stufen der Pyramide sie die Aufgaben verortet sehen. Dadurch konnten die Teilnehmenden dahingehend sensibilisiert werden, dass bei einer Aufgabenerstellung, die nicht von konkreten Zielvorgaben ausgeht, auch einmal unbemerkt Aufgabenanforderungen umgesetzt werden könnten, die nicht intendiert waren. Dies verdeutlicht die folgende Aussage des akademischen Oberrats bei Fall MW 5: *Ich hatte das erste Kapitel. Ich habe mir Aufgaben rausgesucht aus meiner Tabelle [einem bestehenden Aufgabenpool], die ich von früher noch habe, welche Aufgaben passen da rein, und habe die zusammengeschrieben und hab danach die Pyramide versucht auszufüllen beziehungsweise die Aufgaben zuzuordnen. Ich war dann nicht so ganz zufrieden ein bisschen einseitig ist unter erinnern und nur zwei Teilaufgaben bei entwickeln und bewerten steht. Ich habe meine Aufgaben dahingehend nochmal überprüft, ob ich es nicht irgendwie anders machen könnte und bin da aber auf keinen grünen Zweig gekommen.*

Aus dieser Aussage geht außerdem hervor, dass es den Lehrenden nicht unbedingt leicht fällt, Aufgaben zu entwickeln, die auf andere Anforderungen abzielen, wie die in schon einmal umgesetzten Aufgabentypen. Dies ist womöglich auch den begrenzten zeitlichen Ressourcen geschuldet. Dies legt zumindest eine Aussage eines anderen Teilnehmers nahe: *Im Prinzip habe ich mich daran orientiert, dass ich mehr Aufgaben aus unterschiedlichen Bereichen stellen möchte, ähm das ist aber natürlich begrenzt aufgrund des Pools an Aufgaben den wir auch schon haben. Ich hab dann eben versucht, Aufgaben zu nehmen, die auf unterschiedlichen Ebenen [der Pyramide]*

waren, sodass man nicht nur sehr einseitig das Repetieren von Studenten verlangt, sondern auch das selber Entwickeln. Ich glaube, ich habe auch eine Aufgabe mit drin, wo sie auch selber was entwickeln müssen. Ähm, genau von dem her habe ich eher versucht alles ein bisschen mit rein zu nehmen. Das hab ich zumindest für mich als Ziel gesetzt. Das hat dann auch ganz gut funktioniert. Wenn ich noch mehr Zeit, hätte würde ich noch viel schönere Aufgaben stellen können. Also das ist ja das was ich letztes Mal auch schon gesagt habe. Irgendwo ist halt ein begrenzender Faktor die Zeit, die wir aufwenden können.

5.2.3.2 Formale Gestaltungsmerkmale von Aufgaben

Bei den formalen Gestaltungsmerkmalen konnten hinsichtlich sprachlicher Merkmale weder in den bestehenden noch in die neu erstellten Aufgaben Schwachstellen oder Probleme festgestellt werden. Besonders auffallend war bei den bestehenden Aufgaben jedoch der Anteil an Aufgaben die voneinander abhängig waren (28,2 %). Bei fast keiner dieser Aufgaben wurde eine Zwischenlösung angegeben. Im Rahmen des Qualitätsentwicklungsprogramms wurde auf diesen Aspekt hingewiesen, jedoch konnten diesbezüglich keine Verbesserungen erkannt werden. Auch bei den Erwartungshorizonten konnten keine Veränderungen festgestellt werden. Lediglich die Trennung von Handlungsaufforderungen wurde geringfügig häufiger umgesetzt als bei den bestehenden Aufgaben (siehe Abbildung 31).

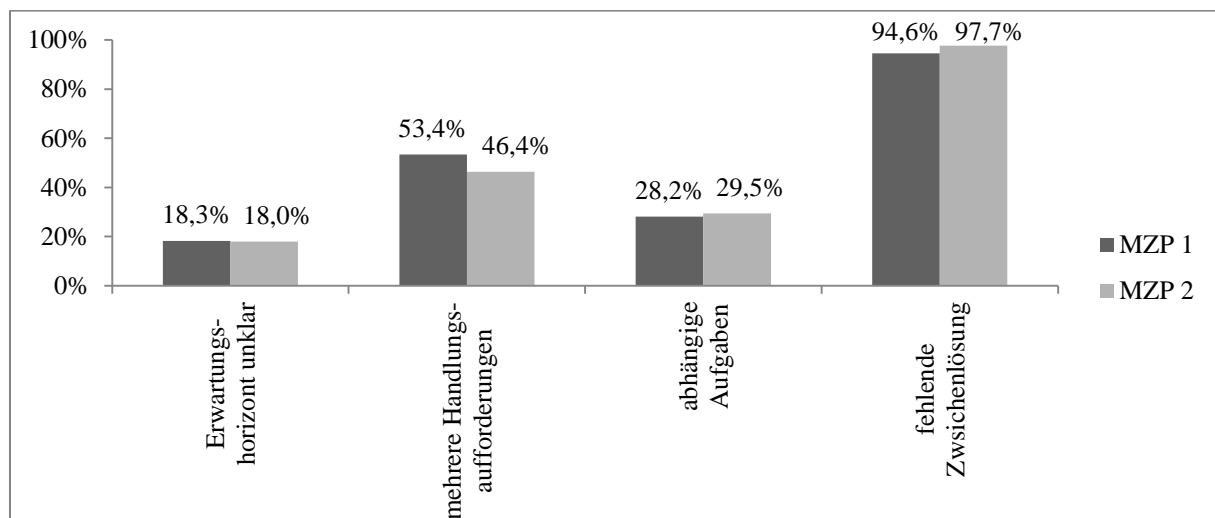


Abbildung 31: Formale Gestaltungsaspekte in den neu erstellten Prüfungen ($n=295$; $n_{\text{fehlende Zwischenlösung}}=87$)

5.2.3.3 Passung zwischen Lehrzielen und Aufgaben

In der bestehenden Prüfungspraxis wurde eine große Diskrepanz zwischen den Anforderungen, die in Lehrzielen beschrieben waren, und den in Prüfungsaufgaben umgesetzten Anforderungen festgestellt. In der veränderten Praxis konnte die Diskrepanz zwischen Lehrzielen und Aufgaben insgesamt betrachtet, sowohl bezüglich konzeptionellen Wissens, als auch beim deklarativen Wis-

sen, verringert werden (siehe Abbildung 32). Dennoch zielen nach wie vor anteilig deutlich mehr Lehrziele auf konzeptionelles Wissen ab (54 %) als Aufgaben (22 %). Beim deklarativen Wissen ist es umgekehrt mit anteilig 52 Prozent an Aufgaben bei nur 14,3 Prozent an Lehrzielen. Hinsichtlich prozeduralen Wissens besteht nach wie vor eine gute Passung. Die Angleichung beim konzeptionellen Wissen wurde vor allem durch eine Reduzierung der Lehrziele erreicht, die auf diese Wissensart abzielen. Die Anteile der Aufgaben mit konzeptionellem Wissen erhöhten sich ebenfalls, wie später noch gezeigt wird.

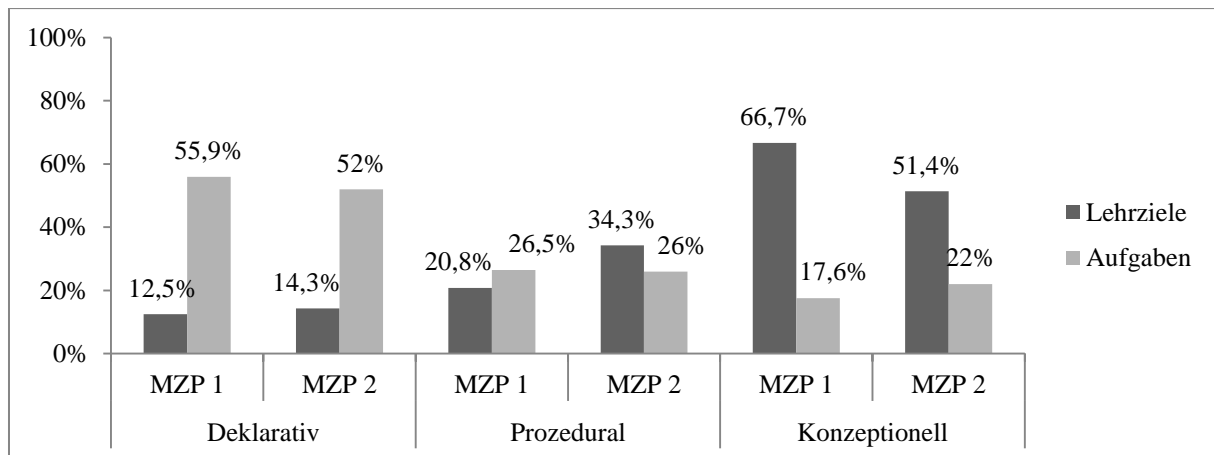


Abbildung 32: Passung zwischen Lehrzielen und Aufgaben in der bestehenden und in der veränderten Praxis ($n_{\text{Aufgaben MZP 1}} = 262$, $n_{\text{Aufgaben MZP 2}} = 295$, $n_{\text{Lehrziele MZP 1}} = 24$, $n_{\text{Lehrziele MZP 2}} = 35$)

Auf Fallebene zeigen sich Unterschiede im Ausmaß der Verringerung der Diskrepanz zwischen Lehrzielen und Aufgaben (siehe Abbildung 33). Bei Fall MW 1 verringerte sich die Diskrepanz zwischen Lehrzielen und Aufgaben hinsichtlich deklarativem und prozeduralem Wissen. Beim konzeptionellen Wissen erhöhte sich die Diskrepanz um 14 Prozent im Vergleich zur bestehenden Praxis. Die Lehrenden konnten zwar die Aufgaben zu konzeptionellem Wissen anteilig um 11,4 Prozent erhöhen, doch gleichzeitig wurden auch Lehrziele zu konzeptionellem Wissen um 25 Prozent erhöht. Bei Fall MW 2 konnte hinsichtlich aller Wissensarten eine Angleichung erzielt werden, ebenso wie bei Fall MW 4. Besonders auffallend ist die starke Reduzierung der Diskrepanz beim konzeptionellen Wissen zwischen Lehrzielen und Aufgaben bei Fall MW 3 (-60 % zwischen MZP 1 und MZP 2). Diese Annäherung konnte durch eine deutliche Reduzierung der Lehrziele, die auf konzeptionelles Wissen abzielen, (75 % in der bestehenden Praxis, 25 % in den veränderten Lehrzielen) bei gleichzeitiger Erhöhung der Aufgaben zu konzeptionellem Wissen (7,5 % der Aufgaben in der bestehenden Praxis, 17,5 % der Aufgaben in der veränderten Praxis), erreicht werden. Zusätzlich zur Angleichung beim konzeptionellen Wissen erhöhte sich jedoch bei Fall MW 3 die Diskrepanz hinsichtlich prozeduralen Wissens (+39 %). Diese Veränderung ist durch eine, im Vergleich zu den Aufgaben unverhältnismäßig starke Erhöhung von Lehrzielen, die prozedurales Wissen fokussieren, bedingt. Bei Fall MW 5 konnte beim konzeptionellen Wis-

sen (Differenz -10 %) eine leichte Annäherung verzeichnet werden, während sich die Differenz zwischen Lehrzielen und Aufgaben bei deklarativem (-12 %) und prozeduralem Wissen (16 %) leicht erhöhte (siehe Abbildung 33).

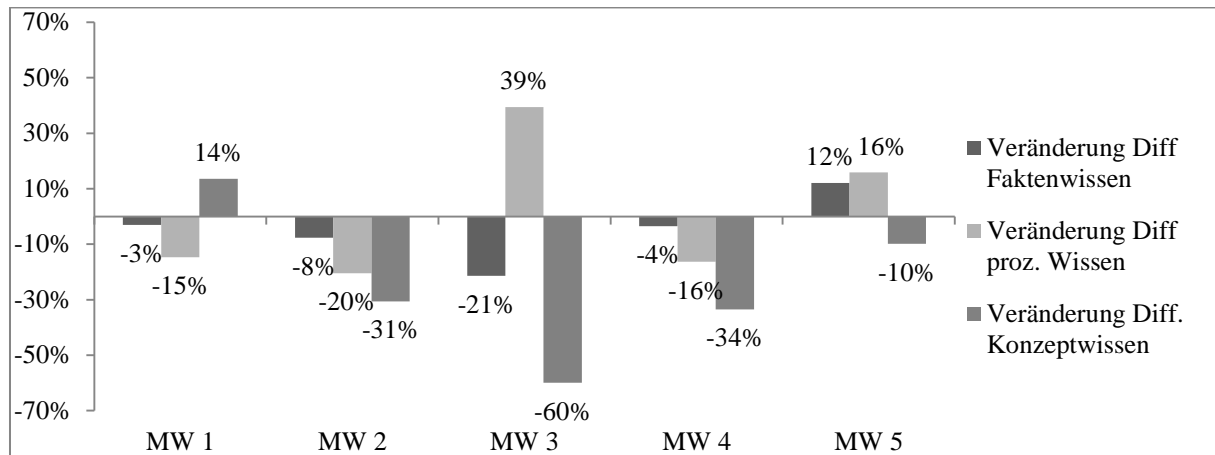


Abbildung 33: Unterschied zwischen den Differenzen in Aufgaben und Lehrzielen in der bestehenden und veränderten Praxis (bei einem positiven Wert vergrößerte sich die Differenz zwischen Lehrzielen und Aufgaben zwischen MZP 1 und MZP 2, bei einem negativen Wert wurde die Differenz verringert).

Im Rahmen des Qualitätsentwicklungsprogramms ist es den Lehrenden in der Tendenz gelungen, Aufgaben an Lehrziele anzupassen. Dies zeigte sich auch bei einer fallweisen Betrachtung der Differenzen zwischen den Wissensarten in Lehrzielen und Aufgaben, wonach beim Großteil der Wissensarten die Diskrepanz zwischen Lehrzielen und Aufgaben verringert wurde. Ausschlaggebend dafür war vor allem eine Reduzierung von Lehrzielen die auf konzeptionelles Wissen abzielen, zugunsten von Lehrzielen mit deklarativem und prozeduralem Wissen. Dadurch wurde erreicht, dass die tatsächlich umgesetzten Prüfungsanforderungen realistischer eingeschätzt und in den Lehrzielen entsprechend korrigiert wurden.

5.2.3.4 Differenzierung von Aufgabentypen

In den bestehenden Prüfungen kamen vorrangig klein geschnittene Aufgaben (kurze Bearbeitungsdauer mit im Schnitt 90 bis 120 Sekunden pro Aufgabe) mit einem offenen Kurzantwortformat vor. Dieser Aufgabentyp hatte sich anscheinend in der bisherigen Praxis der Lehrenden bewährt. Im Qualitätsentwicklungsprogramm wurde daher nicht versucht die Lehrenden grundsätzlich zum Einsatz anderer Aufgabentypen anzuleiten, sondern Möglichkeiten aufzuzeigen, wie mit dem bekannten Format unterschiedliche Anforderungen umgesetzt werden können. Auf Ebene der Aufgabentypen sind daher auch keine nennenswerten Unterschiede zu verzeichnen. Die Bearbeitungsumfänge blieben mit 90 – 120 Sekunden pro Aufgabe relativ kurz. In den neu erstellten Aufgaben wurde ebenfalls am häufigsten das offene Kurzantwortformat eingesetzt (86,1 % aller Aufgaben). Etwas erhöht hat sich dagegen der Anteil der offenen Langwortarten (von zwei Stück (0,8 %) in der bestehenden Praxis auf immerhin 14 Stück (4,7 %) in der veränderten Praxis.

Bei den Antwortarten dominieren nach wie vor Aufgaben, die in Form von Prosa zu beantworten sind (70,5 %). Der Anteil der Aufgaben, zu deren Beantwortung eine Zeichnung (z.B. Anfertigen eines Diagramms oder Graphen) verlangt wurde, blieb mit 8,5 Prozent auf dem gleichen (niedrigen) Niveau. Wie sich aus den Antwortarten schon vermuten lässt, hat sich auch der Einsatz von Illustrationen nicht nennenswert erhöht. In 27,8 Prozent der Fälle werden in den neu erstellten Aufgaben Illustrationen verwendet, in den bestehenden Prüfungen bei 24,4 Prozent aller Aufgaben. Hinsichtlich der eingesetzten Operatoren verringerte sich der Anteil an Aufgaben, in denen keine direkte Handlungsaufforderung angegeben wurde von 25 Prozent auf 10,8 Prozent. Dies ist als positiv zu bewerten, da bei Aufgaben mit einem Operator als Handlungsaufforderung für die Studierenden eindeutiger ersichtlich ist, wie die Aufgabe zu bearbeiten ist, als bei einer Frage. Auf inhaltlicher Seite konnte außerdem eine Reduzierung bei den Operatoren verzeichnen werden, die auf eine Reproduktion von Wissen hindeuten (Nennen, definieren und beschreiben kamen bei 33 % in den bestehenden Aufgaben vor, bei den neu erstellten Aufgaben betrug der Anteil nur noch 24,4 %).

5.2.3.5 Merkmale einer Kompetenzorientierung in veränderten Aufgaben

Eine Kompetenzorientierung war in den bestehenden Aufgaben kaum umgesetzt worden. So existierten anteilig wenige Aufgaben zu konzeptionellen Wissen (17,6 %). Aufgaben zu strategischem Wissen wurden gar nicht verwendet. Aufgabenkontexte, die als Teil der Problemstellung fungieren, kamen ebenfalls nur selten vor (5,3 %). Hauptsächlich wurden Aufgaben mit nur einer richtigen Lösung verwendet (91 %). Die sprachlogische Komplexität war bei den Analysen bei 84 Prozent als einfach bewertet worden. Nachfolgend wird dargelegt, inwieweit es den Lehrenden gelungen ist, die Anforderungen von einem starken Fokus auf deklarativem Wissen, hin zu konzeptionellem Wissen zu verschieben und die Aufgabenstellungen offener zu gestalten. Außerdem werden die Veränderungen hinsichtlich der Funktion von Aufgabenkontexten und der Komplexität von Aufgaben dargelegt.

Wissensarten

Über alle Fälle hinweg betrachtet hat sich die Verteilung der Wissensarten in den Aufgaben nur geringfügig verändert. Der größte Teil der Aufgaben bezieht sich nach wie vor auf deklaratives Wissen, lediglich der Anteil verringerte sich von 59,2 auf 52 Prozent. Leicht erhöht hat sich der Anteil der Aufgaben mit konzeptionellem Wissen, nämlich von 17,2 auf 22 Prozent. Die Anteile an Aufgaben mit prozeduralem Wissen änderten sich hingegen nicht nennenswert (siehe Abbildung 34).

5.2 Erstellung von Prüfungsaufgaben

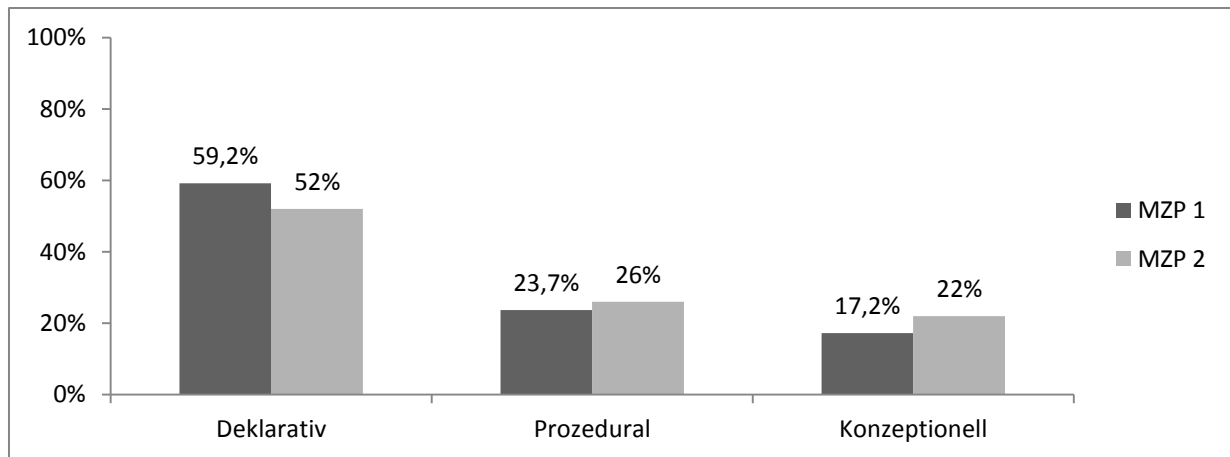


Abbildung 34: Verteilung der Wissensarten in den bestehenden und neu erstellten Aufgaben ($n_{MZP1}=262$; $n_{MZP2}=295$)

Auf Fallebene fielen die Unterschiede zwischen bestehenden neu erstellten Aufgaben zum Teil deutlicher aus (siehe Abbildung 35). Den Teilnehmenden von Fall MW 1 ist es gelungen, den bereits hohen Anteil von Aufgaben mit konzeptionellem Wissen um 14 Prozent auf 54 Prozent zu steigern.

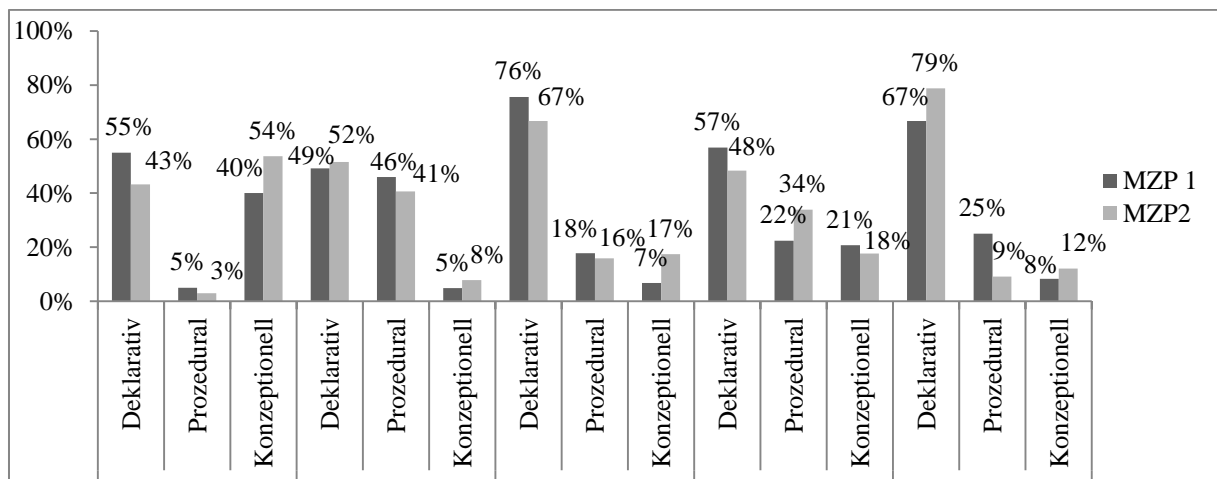


Abbildung 35: Wissensarten in den bestehenden und neu erstellten Prüfungsaufgaben pro Fall ($n_{MZP1}=262$; $n_{MZP2}=295$)

Damit unterschieden sich die Lehrenden von Fall MW 1 deutlich von den anderen Teams, die weit weniger Aufgaben zu konzeptionellem Wissen erstellten. Hier gelang es anscheinend Ideen für Aufgaben zu entwickeln, die über ein Wiedergeben von Wissen hinausgehen. Ein Aspekt, der hierbei zum Tragen kommt, ist, dass ein systematischer Abgleich zwischen den Lehrzielen und Aufgaben stattfand. Die Prüfungsaufgaben wurden dabei in mehreren Iterationen immer wieder überarbeitet, bis die gewünschten Anforderungen auch tatsächlich angesprochen wurden. Dabei wurde auch berücksichtigt, ob für eine Aufgabe die Lösung aus dem Skript erinnert werden oder eigenständig konstruiert werden muss (vgl. Zitat 34, S. 281). Die Lehrziele wurden bei Fall MW 1

zudem am differenziertesten beschrieben, indem übergeordnete Lehrziele nochmal in Teilziele zerlegt wurden. Diese genaue Beschreibung von Teilzielen wurde als sehr hilfreich wahrgenommen, um entsprechend unterschiedliche Anforderungen in der Prüfung abzubilden, wie diese Aussage des akademischen Oberrats verdeutlicht: *Hier haben wir es [ein Lehrziel] dann nochmal aufgeteilt damit wir es hier auch detaillierter beschreiben können. Das hat uns dann schon sehr gut geholfen bei der Arbeit hier und das glaube hilft uns auch bei der Prüfungserstellung, so dass wir genau sagen können, was wollen wir abfragen, da sehen wir da haben wir einen großen Fleck, den haben wir in der Prüfung nicht drin, wollen wir das und dann können wir gezielt unsere Prüfungsaufgaben stellen.*

Bei Fall MW 3 konnte die Teilnehmerin die Anteile an Aufgaben mit konzeptionellem Wissen gezielt erhöhten. Ihr Anliegen war es, zu jedem Themenbereich gezielt verstehensorientierte Aufgaben einzubauen, die auf konzeptionelles Wissen abzielen. Zur Erstellung entsprechender Aufgaben wendete sie die in der Fortbildung vermittelten Techniken zur Öffnung von Aufgaben an, wie folgende Aussage der Doktorandin von Fall MW 3 veranschaulicht: *Die Technik vom Begriff ausgehend habe ich öfter verwendet. Beispiele einfordern habe ich auch drinnen. Das mit den Grenzen erfragen habe ich eher nicht. Es war zielgerichtet diese Techniken zu verwenden, nicht per Zufall.* In Abbildung 36 ist in einem Profildiagramm dargestellt, wie die Wissensarten in der neu erstellten Prüfung verteilt sind. Der Teilnehmerin ist es durch die Anwendung der vorgestellten Techniken zur Aufgabenöffnung gelungen, zu allen Themenbereichen eine oder mehrere Aufgaben zu konzeptionellem Wissen zu erstellen (in der bestehenden Prüfung waren es dazu insgesamt nur drei Aufgaben).

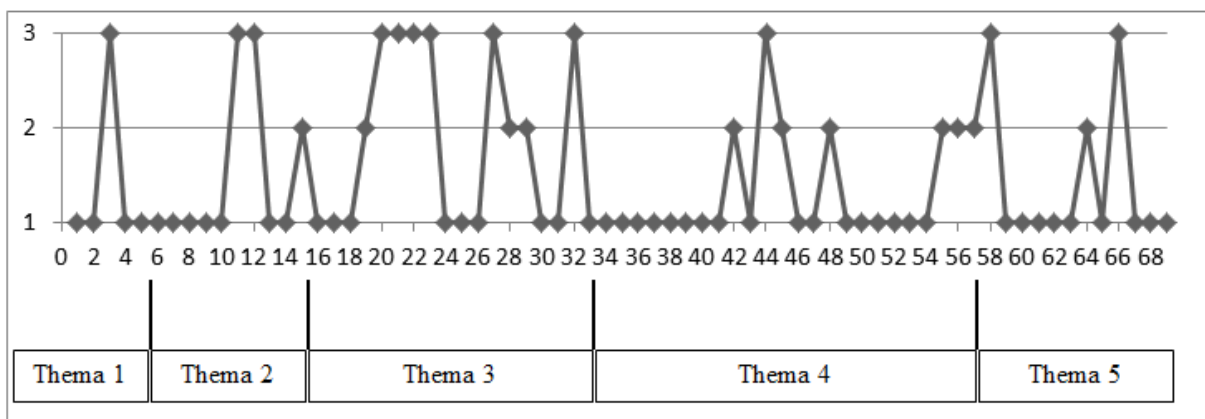


Abbildung 36: Verteilung der Wissensarten in der veränderten Prüfung von Fall MW 3; deklaratives Wissen = 1, prozedurales Wissen = 2, konzeptionelles Wissen = 3

Bei Fall MW 5 erhöhte sich der Anteil von Aufgaben zu konzeptionellem Wissen von acht auf zwölf Prozent. Damit fällt der Anteil an Aufgaben zu konzeptionellem Wissen trotz der Bemühungen die Aufgabenanforderungen an den kognitiven Anforderungsniveaus einer Lehrzielpyramide auszurichten, nach wie vor gering aus. Der Anteil an Aufgaben mit deklarativem Wissen

wurde in der neu erstellten Prüfung hingegen von 67 Prozent auf 79 Prozent erhöht. Einer der Teilnehmenden äußerte dazu, dass es ihm nicht gelungen sei, für die weiterführenden Stufen der Taxonomie, das heißt über die Stufe des Erinnerns hinaus, Aufgabenideen zu generieren (vgl. Zitat 38, S. 283).

Aufgabenoffenheiten

Die Öffnung von Aufgaben ist eine Stellschraube, um Aufgabenanforderungen auf konzeptionelles oder sogar strategisches Wissen auszurichten. In der bestehenden Praxis lag der Schwerpunkt mit anteilig 90,5 Prozent auf definierten und konvergenten Aufgaben. Die restlichen 9,5 Prozent entfielen auf definierte und divergente Aufgaben. Dieser Fokus änderte sich insgesamt nur geringfügig auf 85,5 Prozent bei definierten und konvergenten Aufgaben und 14,5 Prozent bei definierten und divergenten Aufgaben. Auf Fallebene sind die Entwicklungen wiederum differenzierter zu betrachten (siehe Abbildung 37). Der Anteil von definierten und divergenten Aufgaben konnte bei den Fällen MW 1 um 10 Prozent und bei Fall MW 3 sogar um 14,5 Prozent erhöht werden. Bei Fall MW 1 ist der Anteil an definierten und divergenten Aufgaben mit insgesamt 43,3 Prozent relativ hoch, insbesondere im Vergleich zu den anderen Fällen, die kaum Aufgaben einsetzten, in denen Lösungswege oder Antworten variabel sind.

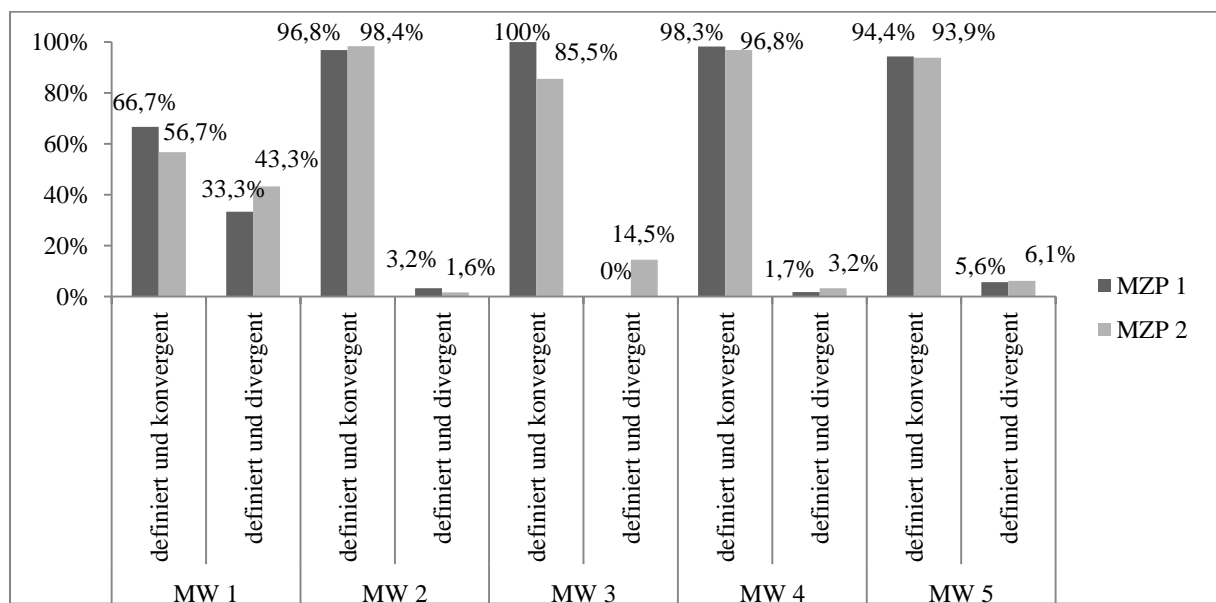


Abbildung 37: Aufgabenoffenheiten pro Fall in den bestehenden und der veränderten Praxis ($n_{MZIP1}=262$; $n_{MZIP2}=295$)

Aufgabenkontexte

Für eine Kompetenzorientierung sind insbesondere Kontexte relevant, die bei der Aufgabenbearbeitung zum Tragen kommen und für die Lösung der Aufgabe eine Rolle spielen. Als Teil der Problemstellung kam ein Aufgabenkontext in den bestehenden Prüfungen anteilig lediglich in 5,3

Prozent der Fälle vor. In den neu erstellten Prüfungen konnte dieser Anteil auf 12,5 Prozent gesteigert werden. Die größte Steigerung diesbezüglich wurde in der Prüfung von Fall MW 1 erzielt, die den Anteil von Aufgaben mit Kontexten, die als Teil der Problemstellung fungieren, von 16,7 Prozent auf 43,3 Prozent steigern konnten. Bei Fall MW 3 muss bei immerhin 7,2 Prozent der Aufgaben das Wissen auf einen Kontext angewendet werden (in der bestehenden Prüfung wurde dieses Element nie eingesetzt). Bei Fall MW 5 wurden in der neu erstellten keine Aufgaben mit einem Kontextbezug als Teil der Problemstellung eingesetzt (verglichen zu 5,3 % in der bestehenden Prüfung). Bei den Fällen MW 2 und MW 4 ergaben sich keine nennenswerten Veränderungen, insofern weiterhin kaum Aufgaben mit einem Kontextbezug als Teil der Problemstellung eingesetzt wurden.

Komplexität von Aufgaben

Als letztes Merkmal von Kompetenzorientierung wird die Komplexität von Aufgaben im Sinne der zu verarbeitenden Informationen betrachtet. Dieser Aspekt wurde weder in den bestehenden, noch in den veränderten Prüfungen systematisch genutzt. In der bestehenden Praxis wurden lediglich fünf Prozent der Aufgaben in den veränderten Prüfungen nur noch zwei Prozent als komplex eingestuft. Bei Fall MW 2 wurden die komplexen Aufgaben mit einer großen Menge an zu verarbeitenden Informationen komplett gestrichen, mit der Begründung, dass die Verarbeitung der gegebenen Informationen nicht für die abzu prüfenden Wissens- und Könnensaspekte relevant sei und daher eher ablenkend wirke. Sind Aufgabenkontexte und die gegebenen Informationen nicht konstruktrelevant, so sollte tatsächlich darauf verzichtet werden, da die erhöhte Leselast als Störvariable wirken kann. Die Komplexität wurde demnach von den Lehrenden nicht als hilfreiche Stellschraube zur Veränderung von Aufgabenanforderungen wahrgenommen und entsprechend selten genutzt.

Zusammenfassend kann festgehalten werden, dass in den neu erstellten Prüfungen die Merkmale von Kompetenzorientierung häufiger umgesetzt wurden, als in den bestehenden Prüfungen. Insgesamt zielten mehr Aufgaben auf konzeptionelles Wissen ab (+ 4,8 %), Aufgaben mit Kontextbezug wurden häufiger eingesetzt (+ 7,2 %), genauso wie Aufgaben mit mehreren möglichen Lösungswegen und Antworten (+ 5 %). Die Erhöhungen wurden hauptsächlich durch die Fälle MW 1 und MW 3 bedingt. Bei der neu erstellten Prüfung von Fall MW 1 sind alle Merkmale einer Kompetenzorientierung stark ausgeprägt, insofern über die Hälfte der Aufgaben auf konzeptionelles Wissen abzielt, und bei 40 Prozent der Aufgaben unterschiedliche Lösungswege und Antworten möglich sind. Ebenso mit etwa 40 Prozent werden in den Aufgaben Kontexte genutzt, die als Teil der Problemstellung fungieren. Bei Fall MW 3 wurden sehr gezielt Aufgaben zu konzeptionellem Wissen erstellt, um den starken Fokus auf deklarativem Wissen aufzubrechen. Bei beiden

Fällen war auffällig, dass die Aufgabenerstellung konsequent an den Lehrzielen ausgerichtet war. Bei Fall MW 1, die die größten Veränderungen in Richtung Kompetenzorientierung erzielen konnten, waren die Lehrziele am differenziertesten und spezifischsten ausgearbeitet worden. Außerdem wurde die Passung zwischen Aufgaben und Lehrzielen unter Kolleginnen und Kollegen ausführlich diskutiert und bei Bedarf überarbeitet. Neben der Orientierung an Lehrzielen betonte die Lehrende von Fall MW 3, dass die im Rahmen der Fortbildung vorgestellten Techniken zur Gestaltung von Aufgaben hilfreich waren, um unterschiedliche Anforderungen abzubilden. Eine konsequente Ausrichtung der Aufgabenerstellung an differenziert ausgearbeiteten Lehrzielen sowie der Einsatz von verschiedenen Techniken zur Öffnung von Aufgaben erscheinen daher als vielversprechende Instrumente, die Lehrende nutzen können, um die Anforderungsheterogenität in ihren Prüfungen zu erhöhen. Die Zuhilfenahme einer Lehrzielpyramide, in der verschiedenen kognitive Anforderungsniveaus abgebildet sind, trug bei Fall MW 5 hingegen nicht dazu bei, dass anteilig mehr Aufgaben erstellt werden konnten, die auf höheren Anforderungsniveaus liegen.

5.2.4 Kurzzusammenfassung

In diesem Abschnitt wurden die Erkenntnisse zur Erstellung von Prüfungsaufgaben berichtet. Im ersten Schritt wurden bezogen auf Forschungsfrage 1 die Zielsetzungen der Teilnehmer betrachtet. Anschließend wurden die Ergebnisse zu den vor Beginn der Qualitätsentwicklungsarbeit bestehenden Aufgaben berichtet (Forschungsfrage 2). Dabei stand im Vordergrund, ob die Aufgaben wesentliche Gestaltungsmerkmale zur Einhaltung der Gütekriterien erfüllen und inwieweit Merkmale einer Kompetenzorientierung umgesetzt sind. Ausgehend von der bestehenden Praxis erfolgte eine Beschreibung der Veränderungen, die die Lehrenden im Rahmen ihrer Qualitätsentwicklungsarbeit über ein Semester an den Prüfungen vorgenommen haben (Forschungsfrage 3). In Tabelle 24 sind die wesentlichen Ergebnisse zusammenfassend dargestellt.

Tabelle 24: Wesentliche Erkenntnisse zur Erstellung von Prüfungsaufgaben

Erstellung von Prüfungsaufgaben	FF 1: Zielsetzungen der Teilnehmenden	<ul style="list-style-type: none"> - Kennenlernen neuer Aufgabentypen als Hauptziel bei allen Fällen. - Ziel einer flexibleren Aufgabengestaltung, um unterschiedliche Anforderungen abzubilden, bei gleichzeitiger Wahrung der Korrigierbarkeit. - Handlungsanforderungen und Anweisungen innerhalb von Aufgaben präzisieren. - Zielsetzungen auf allgemeinem Niveau. Es wurden keine konkreten Probleme in Aufgaben oder bei der Aufgabengestaltung angesprochen.
	FF 2: Bestehende Praxis	<ul style="list-style-type: none"> - Erstellung von Aufgaben erfolgt implizit beim Sichten von Veranstaltungunterlagen - Keine Ausrichtung an Lehrzielen bei der Aufgabenerstellung - Auf formaler Ebene: wenige Probleme hinsichtlich sprachlicher Merkmale, jedoch bei 30 Prozent Abhängigkeiten zwischen Aufgaben. Nur bei einem Drittel der abhängigen Aufgaben werden Zwischenlösungen angegeben. - Große Diskrepanz zwischen den Anforderungen in Lehrzielen und Aufgaben. - Kaum Variation von Aufgabentypen, Antwortarten und Operatoren. Illustrationen in Aufgaben wurden kaum verwendet. Das dominante Format sind Aufgaben mit offenen Kurzantworten. - Merkmale einer Kompetenzorientierung wurden nur ansatzweise umgesetzt: Geringe Anforderungsheterogenität in Aufgaben hinsichtlich Wissensarten, Aufgabenoffenheiten, Aufgabenkontexten und der Aufgabenkomplexität. Der Schwerpunkt lag auf geschlossenen Aufgaben mit deklarativen und prozeduralen Wissen ohne Kontextbezüge.

- Erstellung von Aufgaben orientierte sich bei drei von fünf Fällen an den Lehrzielen. Bei Fall MW 4 wurden die Lehrziele an die Aufgaben angepasst. Bei Fall MW 5 erfolgte eine Zuordnung von Aufgaben zu den kognitiven Niveaustufen nach Bloom et al. (1956)
- Reduzierung von Abhängigkeiten zwischen Aufgaben
- Offenen Kurzantwort weiterhin das dominante Format
- Angleichung zwischen den Anforderungen in Lehrzielen und Aufgaben, insbesondere bei konzeptionellem Wissen.
- Veränderungen bei Wissensarten, Aufgabenoffenheiten und Aufgabenkontexten in Richtung Kompetenzorientierung nur bei fallweiser Betrachtung.
- Fall MW 1 konnte eine hohe Anforderungsheterogenität in der Prüfung im Sinne einer Kompetenzorientierung umsetzen.
- Vermittelte Techniken zur Öffnung von Aufgaben konnten von der Teilnehmenden von Fall MW 3 erfolgreich genutzt werden, um Aufgaben mit konzeptionellem Wissen umzusetzen.
- Bei Fall MW 5 Zuordnung von Aufgaben zu kognitiven Niveaustufen einer Lehrzieltaxonomie, jedoch keine Orientierung an Lehrzielen. Techniken zur Öffnung von Aufgaben waren zum Zeitpunkt der Aufgabenerstellung nicht bekannt. Anteile an Aufgaben mit konzeptionellem Wissen konnten nicht nennenswert erhöht werden obwohl Anteile in Lehrzielen gleich blieben.
- Umsetzung kompetenzorientierter Aufgaben im offenen Kurzantwortformat möglich. Differenzierte Lehrziele als Bezugspunkt bei Aufgabenerstellung und –überarbeitung erscheinen hilfreich, um heterogene Anforderungen hinsichtlich Wissensarten, Aufgabenoffenheiten und Aufgabenkontexten umzusetzen.
- Die vorgestellten Techniken zur Aufgabenöffnung konnten von den Lehrenden gezielt genutzt werden, um Aufgaben zu konzeptionellem Wissen zu erstellen.

5.3 Rückbindung der Aufgaben an den Prüfungsgegenstand

Nach der Erstellung von Aufgaben gilt es für die Lehrenden zu überprüfen, ob damit auch die relevanten Wissens- und Könnensaspekte geprüft werden, um eine inhaltliche und kognitive Validität von Aufgaben zu erreichen. In diesem Kapitel wird daher betrachtet, ob Lehrende ein Bewusstsein für den Aspekt der Validität von Aufgaben besitzen und bereits entsprechende Maßnahmen einsetzen, um diese sicherzustellen. Zunächst werden die von den Teilnehmenden formulierten Zielsetzungen zur Rückbindung von Aufgaben an den Prüfungsgegenstand betrachtet (Forschungsfrage 1). Anschließend wird darauf eingegangen, ob die Lehrenden in ihrer bestehenden Praxis vor Beginn der gemeinsamen Qualitätsentwicklungsarbeit Maßnahmen angewendet haben, um die Passung zwischen Aufgaben und Lehrzielen zu überprüfen (Forschungsfrage 2). Zur veränderten Praxis wird analysiert, ob es gelingt, den Einsatz solcher Maßnahmen anzuregen (Forschungsfrage 3). Die Erkenntnisse zu Forschungsfrage 2 und 3 wurden aus den Verbaldaten gewonnen.

5.3.1 Zielsetzungen der Hochschullehrenden

In Bezug auf die Rückbindung von Aufgaben an den Prüfungsgegenstand wurde nur ein Projektziel von den Teilnehmenden formuliert. Einer der Teilnehmenden von Fall MW 5 formulierte das Anliegen im Rahmen des Qualitätsentwicklungsprogramms Möglichkeiten kennenzulernen, um die Lehrinhalte und Lehrziele mit den Prüfungsfragen abzugleichen (siehe Tabelle 25). Hier besteht demnach ein Bewusstsein dafür, dass die in Aufgaben angelegten Anforderungen von den intendierten, beziehungsweise in Lehrzielen beschriebenen, Anforderungen abweichen könnten.

Tabelle 25: Projektziele zur Rückbindung von Aufgaben an den Prüfungsgegenstand

Fälle	Projektziele
MW 5	Abgleich von Lehrinhalten und Lehrzielen mit Prüfungsfragen und Aufgabenformaten: Bilden die bisher verwendeten Prüfungsformate und Aufgaben optimal die Inhalte der Veranstaltung und die zu erwerbenden Kompetenzen ab?

5.3.2 Erkenntnisse zur bestehenden Prüfungspraxis

Wie in den vorangegangenen Kapiteln bereits erarbeitet werden konnte, fand in der bestehenden Praxis keine Orientierung an Lehrzielen statt. Ohne einen expliziten inhaltlichen Referenzrahmen ist ein systematischer und zielorientierter Abgleich von Prüfungsaufgaben an vordefinierten Anforderungen wohl nicht möglich. Die Teilnehmenden von Fall MW 2 und MW 5 berichteten jedoch davon, dass die Prüfungen von Kolleginnen und Kollegen vor dem Prüfungstermin probe-

weise geschrieben wurden, um mögliche Schwierigkeiten hinsichtlich Aufgabenstellungen oder dem angesetzten Anforderungsniveau festzustellen.

Die Lehrenden von Fall MW 2 wendeten dabei ein mehrstufiges Verfahren an, um zu kontrollieren, ob die Prüfung eindeutig und von den Anforderungen her passend gestellt ist. Im ersten Schritt wurde die Prüfung dabei von wissenschaftlichen Mitarbeitern geschrieben, die idealerweise die Prüfung selber während ihres Studiums ablegen mussten. Dadurch sollen mögliche Schwierigkeiten wie unklare Handlungsaufforderungen in Aufgaben oder deren Nicht-Lösbarkeit (z.B. aufgrund fehlender oder falscher Angaben in einer Aufgabenstellung) aufgedeckt werden. Der „Eindruck“, den die Kollegen beim Durcharbeiten hinsichtlich der Aufgabenschwierigkeiten haben, dient als Indikator für das Anforderungsniveau. In einem weiteren Schritt überprüfte eine weitere Person formale Gestaltungsaspekte und schätzt die Aufgabenschwierigkeiten ab (vgl. Zitat 36, S. 282). Welcher Referenzrahmen zur Bewertung der Schwierigkeit verwendet wird, wurde nicht explizit genannt.

Bei Fall MW 5 stellte sich das Vorgehen ähnlich dar. Die Aufgaben wurden vor dem Prüfungstermin unter Prüfungsbedingungen von den Mitarbeiterinnen und Mitarbeitern bearbeitet, die hinterher auch die Korrektur durchführten. Basierend auf der probeweisen Bearbeitung wurden im Anschluss Fragen zur Punktevergabe, den Aufgabenstellungen und zur zeitlichen Einteilung, bewertet. Die Probearbeitung wurde bereits zweimal mit guten Erfahrungen durchgeführt (siehe Zitat 37, S. 282). Damit wurde bei den Fällen MW 2 und MW 5 im Vorfeld des Prüfungstermins eine Qualitätssicherungsmaßnahme durchgeführt, indem Kolleginnen und Kollegen die erstellte Prüfung unter Prüfungsbedingungen bearbeiteten. Bei den anderen drei Fällen wurden diesbezüglich keine Aussagen gemacht. Eine solche Qualitätssicherung ist aber für alle Fälle erstrebenswert, da mit diesem Vorgehen im Vorfeld der Prüfungsdurchführung Probleme aufgedeckt werden können. Der Fokus der Lehrenden lag dabei bislang aber hauptsächlich auf Aspekten der Bearbeitbarkeit und der geschätzten Schwierigkeit. Inwieweit eine kritische Auseinandersetzung darüber stattfindet, welche Inhalte und Kompetenzen mit den Aufgaben erfasst werden, wurde nicht beschrieben. Ohne explizierte Zielvorgaben, die in Lehrzielen festgehalten sind, fehlt dazu jedoch wahrscheinlich auch ein geeigneter Referenzpunkt. Im Rahmen des Qualitätsentwicklungsprogramms wurde daher auf das Konzept der inhaltlichen und kognitiven Validität von Aufgaben eingegangen. Außerdem wurden Möglichkeiten aufgezeigt, um Aufgaben auf ihre inhaltliche Passung zu überprüfen.

5.3.3 Erkenntnisse zur veränderten Prüfungspraxis

Da die Lehrenden in ihrer bestehenden Prüfungspraxis keine an Lehrzielen ausgerichtete Rückbindung von Aufgaben an den Prüfungsgegenstand vornahmen, wurde im Rahmen des Qualitätsentwicklungsprogramms angestrebt, die Lehrenden hinsichtlich Fragen zur inhaltlichen und kognitiven Validität von Aufgaben zu sensibilisieren. Zu diesem Zweck wurde im Qualitätsentwicklungsprogramm die Methode der rationalen Aufgabenanalyse vorgestellt. Bei drei von fünf Fällen (MW 1, MW 2 und MW 3) wurde im Anschluss an das entsprechende Arbeitstreffen eine auf Lehrziele bezogene rationale Aufgabenanalyse durchgeführt. Die Lehrenden konnten dadurch die in den Prüfungsaufgaben umgesetzten Anforderungen besser einschätzen, indem Anforderungsbereiche identifiziert wurden, zu denen keine Aufgaben vorlagen. Dabei stellte sich bei allen Fällen heraus, dass insbesondere bei weiterführenden Anforderungsbereichen, im Sinne von konzeptionellem Wissen, Aufgaben fehlten. Diese Einschätzung der Lehrenden deckt sich mit den Ergebnissen der Analysen, die in dieser Arbeit zur Passung zwischen Lehrzielen und Aufgaben durchgeführt wurden (siehe Abschnitt 5.2.2.2).

Das Vorgehen der Lehrenden bei der rationalen Aufgabenanalyse unterschied sich auf Fallebene aufgrund der unterschiedlichen Spezifität der Lehrziele und den verschiedenen Verwendungsformen der Lehrzielmatrix. Bei Fall MW 1 bildeten zwölf Lehrziele die Grundlage für die Rückbindung von Aufgaben an den Prüfungsgegenstand. Mit den Lehrzielen wurde bei Fall MW 1 besonders detailliert zwischen Teilanforderungen unterschieden, da es den Lehrenden nach eigener Aussage mit allgemeiner formulierten Lehrzielen nicht gelungen ist, die Anforderungen ausreichend differenziert zu beschreiben, um entsprechend Aufgaben zu entwickeln (vgl. Zitat 28, S. 264). Die Grundlage für die rationale Aufgabenanalyse bildete die bereits dargestellte Lehrzielmatrix (siehe Tabelle 20, S. 125), in die Inhalte und kognitive Anforderungsdimensionen der Lehrziele übertragen wurden. Die Lehrenden diskutierten, auf Basis des von den Moderatoren beschriebenen Vorgehens zur Durchführung einer zur rationalen Aufgabenanalyse, zu welchen Zellen der Matrix (und damit zu welchem Lehrziel) eine Aufgabe aufgrund der darin abgebildeten Anforderungen eingeordnet werden kann.

Folgendes Zitat des akademischen Oberrats illustriert das Vorgehen: *Vom Gefühl her waren wir relativ kritisch. Wo wir immer das Problem hatten, wir haben die Aufgaben oft weiter unten eingruppiert [hinsichtlich der kognitiven Anforderungsniveaus in der Lehrzielmatrix], weil wir gesagt haben: Stimmt, eigentlich haben wir das in der Vorlesung so erzählt. Würde ich irgendjemand die Aufgabe geben, wäre es vermutlich analysieren und bewerten, aber wenn man das in der Vorlesung sehr ähnlich durchgemacht hat, dann ist man einfach nicht mehr auf den höheren Anforderungsniveaus, sondern die Studierenden können sich erinnern, vielleicht haben sie es*

auch noch verstanden, aber mehr ist es halt nicht. [...] Wir waren dann trotzdem positiv überrascht, dass es halbwegs passt. Also dass das, was wir eigentlich vermitteln wollen und was wir prüfen, relativ nah beieinander liegt. Aber ein bisschen Potenzial haben wir uns noch nach oben gelassen, damit wir hier optimieren können. Ein paar Punkte [in den Lehrzielen] sind nicht bei den Aufgaben dabei. Bei anderen Punkten sieht man, die sind vom Schwerpunkt zu stark mit Aufgaben besetzt. Aber so vom großen und ganzen haben wir uns eigentlich ganz gut gefühlt.

Aus diesen Ausführungen wird ersichtlich, dass die Aufgabenanforderungen differenziert und kritisch vor dem Hintergrund der Lehrziele diskutiert wurden. Außerdem konnten die Lehrenden dadurch feststellen, dass einerseits zu manchen Lehrzielen keine Aufgaben in der Prüfung enthalten waren, dafür aber andererseits Aufgaben eingesetzt wurden, die keinem Lehrziel zugeordnet werden konnten. Aufgaben, die augenscheinlich auf weiterführende Anforderungen wie analysieren oder bewerten abzielen, wurden als reproduktiv eingestuft, sofern die Lösungen aus dem Skript erinnert werden konnten. Insgesamt wurde eine zufriedenstellende Passung zwischen den Prüfungsaufgaben und den Zielen der Veranstaltung festgestellt. Hierbei ist jedoch anzumerken, dass die beiden Doktoranden sich intensiv mit den Aufgaben, die sie von den Kolleginnen und Kollegen erhalten haben, auseinandersetzen und diese auf die in den Lehrzielen beschriebenen Anforderungen anpassten, wie folgende Aussage von Doktorand 2 aufzeigt: *Was ein bisschen ärgerlich war und auch viel Zeit gekostet hat, ist, dass man nicht alle Aufgaben direkt übernehmen konnte. Die wenigsten eigentlich ohne Überarbeitung.*

Die Einschätzung der Lehrenden von Fall MW 1 zur guten Passung zwischen Lehrzielen und Aufgaben deckt sich mit den Analysen der Aufgaben, die in Abschnitt 5.2.3.3 beschrieben wurden, wonach bei Fall MW 1, den Lehrzielen entsprechend, insbesondere weiterführende Anforderungen im Bereich konzeptionellen Wissens in Aufgaben umgesetzt wurden. Hier ist es durch eine differenzierte Beschreibung von Lehrzielen und der konsequenten Durchführung einer rationalen Aufgabenanalyse gelungen, die in Lehrzielen vorgesehene Anforderungsheterogenität in Aufgaben umzusetzen.

Die Lehrenden von Fall MW 2 stellten bei der Durchführung der rationalen Aufgabenanalyse fest, dass gerade zu den verstehensorientierten Anforderungen im Bereich konzeptionelles Wissen keine entsprechenden Prüfungsaufgaben vorlagen, wie aus folgender Aussage von Doktorand 1 hervorgeht: *Ein Ergebnis der rationalen Aufgabenanalyse war, dass wir den Zusammenhang zwischen den einzelnen Themenbereichen in der Prüfung gar nicht abfragen. Wir hatten es aber als Ergebnis so formuliert und haben nochmal darüber nachgedacht, ob es eigentlich in der Vorlesung vermittelt wird und das konnten wir auch nicht so richtig bejahen. Letztendlich hat es uns vor Augen geführt, das war eigentlich das Gravierendste, dass es nicht so richtig da ist. So haben*

wir die Aufgabe auch verstanden, dass wir unsere Lehrziele, also den Anspruch den wir an die Vorlesung haben, vergleichen mit dem was die Klausur abprüft.

Neben der Feststellung, dass Aufgaben fehlen, die sich auf Zusammenhänge zwischen den Themenbereichen beziehen, wurde in Bezug auf die Veranstaltung hinterfragt, ob dieser Aspekt in der Lehre genügend betont wird. Doktorand 1 stellt dazu folgendes fest: *Man muss die Verknüpfung der Themenbereiche auch in der Vorlesung noch verstärken. Also in der Vorlesung sind die Kapitel relativ stark getrennt. Man müsste dann, wenn man das an die Lehrziele anpassen will, so wie wir die gerne hätten, dann muss man auch die Vorlesung selbst anpassen. Ob wir die Lehrziele an die Aufgaben anpassen oder umgekehrt die Aufgaben an die Lehrziele, da sind wir jetzt in der Diskussion. Das müssen wir auch mit dem Lehrstuhlleiter abklären. Die Veranstaltung können wir nicht einfach selbstständig ändern. Das ist auch immer ein ziemlich großer Aufwand.*

Als Konsequenz auf die festgestellte Diskrepanz zwischen Lehrzielen und umgesetzten Prüfungsanforderungen wurde der Anteil an Lehrzielen, die auf konzeptionelles Wissen abzielen, reduziert. Die Anzahl der Aufgaben, in denen konzeptionelles Wissen abgeprüft wird, konnte hingegen nicht nennenswert erhöht. Für die kommenden Prüfungen ist jedoch geplant, mehr Aufgaben zu entwickeln, die sich auf Zusammenhänge zwischen Themenbereichen beziehen. Mit der Durchführung der rationalen Aufgabenanalyse erhielten die Lehrenden ein realistischeres Bild davon, welche Anforderungen tatsächlich in der Prüfung abgefragt und in der Veranstaltung gelehrt wurden. Die Ausrichtung der Prüfungsanforderungen an Lehrzielen wurde als hilfreich empfunden, um nicht willkürlich zu prüfen. Aus diesem Grund möchten die Lehrenden von Fall MW 2 die begonnene Arbeit mit Lehrzielmatrix und rationaler Aufgabenanalyse fortsetzen und diese Prinzipien auch an Kolleginnen und Kollegen weitertragen, damit auch die restlichen Veranstaltungen beziehungsweise Prüfungen am Lehrstuhl sukzessive überarbeitet werden. Als Fazit hält Doktorand 1 fest: *Ich denke dass ich den Professoren vorschlagen werde, dass wir zumindest diese Matrix für alle Vorlesungen aufstellen und ich glaube nicht, dass jeder der eine Veranstaltung übernimmt die neu aufstellen muss, aber es ist auf jeden Fall eine sehr gute Orientierung beim Erstellen der Prüfung, in welchen Bereichen soll ich mich mit meinen Fragen aufhalten weil das wird bei der Übergabe von einem Kollegen auf den anderen, da macht man das ein zwei Mal mit so eine Prüfung zu erstellen aber so Ziele was will ich abprüfen, welche Aufgaben soll ich stellen, sondern es heißt da ist das Skript such dir einfach ein paar relevante Fragen überleg dir was. Ohne wirklich so konkret was man prüfen will. Da wäre sowas eben für jedes Fach hilfreich.*

Die Lehrende von Fall MW 3 überprüfte ihre Aufgaben ebenfalls nach dem Prinzip der rationalen Aufgabenanalyse. Dabei stellte sie fest, dass die Anteile der Lehrziele, die auf konzeptionelles Wissen abzielen, deutlich höher ausfallen als der Anteil an entsprechenden Aufgaben: *Letztend-*

lich bin ich zu dem Schluss gekommen, dass ich bei dem Lehrziel keine Aufgaben habe, die sich auf die Anwendung beziehen und auf Grund dessen hab ich meine Lehrziele angepasst. Ich sehe, dass ich in vielen Feldern Lücken habe [in der Lehrzielmatrix] und dass in den Lehrzielen Anforderungen dargeboten werden, die nicht abgefragt werden. Ich wollte aber auf keinen Fall alle Lehrziele anpassen sonst befindet man sich in der ersten Spalte [auf dem untersten Anforderungsniveau der Lehrzielmatrix]. Ich möchte lieber probieren in den Spalten dazwischen mit Aufgaben zu füllen.

Durch die Durchführung der rationalen Aufgabenanalyse wurde die Lehrende von Fall MW 3 auf die Diskrepanz zwischen Lehrzielen und Aufgaben hinsichtlich weiterführender Anforderungen aufmerksam und konnte diese verringern. Das von der Lehrenden erzielte Resultat bei der rationalen Aufgabenanalyse deckt sich mit den in Abschnitt 5.2.3.3 berichteten Ergebnis einer Verringerung der Diskrepanz zwischen den Anforderungen in Lehrzielen und Aufgaben bei Fall MW 3, wonach sich die Anzahl der Lehrziele die auf konzeptionelles Wissen abzielen, verringerte, bei gleichzeitiger Erhöhung entsprechender Aufgaben.

Der Teilnehmende von Fall MW 4 führte nach eigener Aussage keine umfassende rationale Aufgabenanalyse nach den Vorschlägen der Moderatoren und der ausgegeben Anleitung durch. Allerdings glich er nach einem nicht näher beschriebenen Vorgehen die Anforderungen, die in Lehrzielen formuliert sind, mit den Aufgaben in den Prüfungsaufgaben ab. Dabei passte er die Anforderungen in Lehrzielen an die Prüfungsaufgaben an, wodurch er ebenfalls eine Annäherung zwischen Lehrzielen und Aufgaben erreichen konnte (siehe Abschnitt 5.2.3.1). Bei Fall MW 5 versuchte jeder der Teilnehmenden jeweils die eigens erstellten Aufgaben den kognitiven Anforderungsniveaus einer Lehrzielpyramide zuzuordnen. Eine an Lehrzielen ausgerichtete Überprüfung von Aufgaben fand nicht statt. Wie die Analysen der Wissensarten in den neu erstellten Aufgaben zeigen, konnte mit der Orientierung an der Lehrzielpyramide keine nennenswerte Angleichung zwischen Lehrzielen und Aufgaben erreicht werden. Der Anteil an Aufgaben zu konzeptionellem Wissen konnte geringfügig von acht Prozent in der bestehenden auf zwölf Prozent in der neu erstellten Prüfung erhöht werden. Diese Erkenntnisse verstärken den oben bereits dargestellten Eindruck (Abschnitt 5.2.3.1), dass bei einer Erstellung von Aufgaben, die nicht an Lehrzielen ausgerichtet ist, tendenziell Aufgaben entstehen, die auf die Reproduktion von deklarativem Wissen abzielen.

Abschließend kann festgehalten werden, dass es den Lehrenden mit dem Instrument der rationalen Aufgabenanalyse bei adäquater Durchführung gelingen kann, die in Prüfungsaufgaben umgesetzten Anforderungen realistisch einzuschätzen. Lehrende können dafür sensibilisiert werden, dass bei einer unreflektierten Erstellung von Aufgaben vorwiegend reproduktionslastige Aufgaben im

Bereich deklarativen Wissens entstehen. Bei Fall MW 2 wurden zudem Rückschlüsse auf die Passung der Lehrveranstaltung zu den Lehrzielen gezogen. Der Teilnehmende von Fall MW 4 führte keine umfassende rationale Analyse von Aufgaben durch. Jedoch passte er die Anforderungen in den Lehrzielen an die Anforderungen in den Aufgaben an. Bei Fall MW 5 erfolgte ebenfalls keine umfassende und an Lehrzielen ausgerichtete rationale Aufgabenanalyse. Hier konnte auch keine Verringerung der Diskrepanz zwischen Lehrzielen und Aufgaben festgestellt werden.

5.3.4 Kurzzusammenfassung

In diesem Abschnitt wurden Ergebnisse zur Rückbindung von Aufgaben an den Prüfungsgegenstand durch die Lehrenden berichtet. In Tabelle 26 sind die Ergebnisse zu Forschungsfrage 2 und 3 überblicksartig dargestellt. Auf die Programmziele (Forschungsfrage 1) wird in der Tabelle nicht gesondert eingegangen, da diesbezüglich nur ein Ziel genannt wurde, das wenig aussagekräftig war.

Tabelle 26: Wesentliche Erkenntnisse zur Rückbindung von Aufgaben an den Prüfungsgegenstand

Rückbindung von Aufgaben an den Prüfungsgegenstand	FF 2: Bestehende Praxis	<ul style="list-style-type: none"> - Bei zwei Fällen (MW 2 und MW 5) erfolgen probeweise Bearbeitungen von Prüfungen durch Kolleginnen und Kollegen, um mögliche Probleme aufzudecken. Im Mittelpunkt standen dabei Aspekte der Bearbeitbarkeit (z.B. vollständige und richtige Angaben) und weitere formale Gestaltungsmerkmale. Außerdem wurde dabei das Anspruchsniveau eingeschätzt. Ein explizierter Referenzrahmen zur Bewertung des Anspruchsniveaus scheint dagegen nicht zu existieren. - Von den Lehrenden der Fälle MW 1, MW 3 und MW 4 wurden keine Maßnahmen zur Rückbindung von Aufgaben an den Prüfungsgegenstand beschrieben.
--	-------------------------	---

	FF 3: Veränderte Praxis	<ul style="list-style-type: none">- Die Lehrenden der Fälle MW 1, MW 2 und MW 3 führten eine an Lehrzielen ausgerichtete rationale Aufgabenanalyse durch.- Die Lehrenden stellten dabei jeweils Diskrepanzen zwischen Lehrzielen und Prüfungsaufgaben fest. Diese Einschätzungen decken sich mit den Ergebnissen der Analysen in dieser Arbeit zur Passung zwischen Lehrzielen und Aufgaben.- Dies führte zu einer realistischeren Einschätzung der umgesetzten Prüfungsanforderungen. In der Konsequenz erfolgte eine Reduzierung von Lehrzielen zu weiterführenden Anforderungen im Sinne von konzeptionellem Wissen zusammen mit der Erstellung von Aufgaben in diesen Bereichen.- Dadurch wurde eine Angleichung zwischen Lehrzielen und Aufgaben erreicht.- Bei Fall MW 5 erfolgte keine rationale Aufgabenanalyse. Die Diskrepanz zwischen Lehrzielen und Aufgaben verringerte sich bei Fall MW 5 nicht bedeutend.- Fazit: Die rationale Aufgabenanalyse ist von den Lehrenden umsetzbar und ein effektives Instrument, um Lehrende hinsichtlich der Aufgabenvalidität zu sensibilisieren. Die Durchführung der rationalen Aufgabenanalyse gab den Lehrenden Anlass, sowohl Aufgaben als auch Lehrziele zu überarbeiten.
--	-------------------------	---

5.4 Bewertung von Prüfungsantworten

Nachdem die Prüfung durchgeführt wurde, müssen im Rahmen der Korrektur die Antworten der Studierenden bewertet werden. Bei der Bewertung von Prüfungsantworten können sowohl Einschränkungen im Hinblick auf die Validität von Prüfungsergebnissen entstehen, als auch bei der Reliabilität. Hinsichtlich Forschungsfrage 1 wird betrachtet, welche Zielsetzungen die Lehrenden bezüglich der Bewertung von Prüfungsantworten formulierten (Abschnitt 5.4.1). Im Rahmen von Forschungsfrage 2 wird beschrieben, ob die Lehrenden in ihrer bestehenden Praxis Maßnahmen getroffen haben, um die Validität und Reliabilität bei der Korrektur beziehungsweise der Bewertung von Prüfungsantworten zu sichern (Abschnitt 5.4.2). Dazu wurden die bestehenden Prüfungen dahingehend analysiert, ob für die Korrektur Bewertungsraster bereit gestellt wurden, die Bewertungskategorien, Regelwerke, Angaben zur Punkteverteilung und Ankerbeispiele (bzw. Musterlösungen) enthalten. Auf Basis der Verbaldaten wurde zudem ermittelt, ob die Lehrenden Schulungstermine durchführen, um die Kolleginnen und Kollegen auf die Korrektur vorzubereiten und um ein gemeinsames Verständnis über die Bedeutung der Bewertungskategorien zu schaffen. Bezüglich Forschungsfrage 3 wird in Abschnitt 5.4.3 dargelegt, ob die Lehrenden während der Qualitätsentwicklungsarbeit Veränderungen bei der Korrektur und Bewertung von Prüfungsantworten vornahmen.

5.4.1 Zielsetzungen der Hochschullehrenden

Insgesamt formulierten die Lehrenden acht Programmziele hinsichtlich der Bewertung von Prüfungsantworten. Dieser Themenbereich ist damit, gemessen an der Anzahl der diesbezüglich formulierten Programmziele, zusammen mit der Erstellung von Aufgaben, das zentrale Thema für die Lehrenden beim Prüfen. In Tabelle 27 sind die Projektziele fallweise aufgeführt. Ein wesentlicher Aspekt, der von den Fällen MW 1 und MW 5 gleichermaßen genannt wurde, bezieht sich auf den Umgang mit Prüfungsantworten, die zwar sinngemäß richtig sind, jedoch die Musterlösung nicht genau widerspiegeln, beziehungsweise, die einer weiteren Begründung bedürften, um eindeutig als richtig gelöst zu gelten. Diese Zielsetzung benennt relativ konkret das Problem zum Umgang mit frei konstruierten Antworten und der treffsicheren Identifikation konstruktrelevanter Merkmale. In einer Zielsetzung von Fall MW 5 wird eine mögliche Lösung angedeutet, indem hier das Anliegen formuliert wird, eigens ausgearbeitete Lösungsvorschläge um abweichende, jedoch richtige, Antworten von Studierenden zu erweitern.

Die Teilnehmenden von Fall MW 2 möchten die Korrektur für Studierende nachvollziehbarer machen, um Beschwerden zu vermeiden. Etwas anders gelagert ist die Zielsetzung bei Fall MW 4, bei der es nicht um eine qualitative Verbesserung der Korrektur geht sondern sehr allgemein der

Wunsch nach einer Ökonomisierung geäußert wird. Ähnlich ist das Anliegen der Lehrenden von Fall MW 5 zu interpretieren, die gerne diskutieren möchten, an welchen Stellen sie den Korrekturaufwand verringern könnten.

Tabelle 27: Projektziele zur Bewertung von Prüfungsantworten

Fälle	Projektziele
MW 1	Unsicherheit bei der Korrektur bzw. Bewertung von Antworten zum Fallbeispiel reduzieren (wegen der Offenheit der Fragestellung können unterschiedliche Lösungen produziert werden, es gibt oftmals kein eindeutig <i>richtig</i> oder eindeutig <i>falsch</i>).
	Bisher entwickelte Lösungen zur Bewertung von Aufgaben daraufhin überprüfen, ob eine objektive und nachvollziehbare Bewertung gewährleistet ist.
MW 2	Beschwerden von Studierenden reduzieren, indem die Korrektur von Aufgaben für Studierende nachvollziehbarer gemacht wird.
MW 4	Schneller zu korrigierende Prüfungen.
MW 5	Abklären, ob wir uns mehr Aufwand für Prüfungen machen als (insb. rechtlich) notwendig (Beispiel Doppelkorrekturen).
	Umgang mit Antworten die zwar nicht wortgetreu wiedergegeben werden (so wie im Skript beschrieben), jedoch (sinngemäß) richtig sind.
	Umgang mit ungewöhnlichen Antworten, die als Spezialfall richtig sein könnten, die aber nicht begründet werden.
	Erweiterung von ausgearbeiteten Lösungsvorschlägen, um studentische Antworten, die auch als richtig gewertet werden können.

Hinsichtlich einer Erhöhung der Konsistenz von Korrekturen wurden keine Ziele formuliert (die Konsistenz ist eine Voraussetzung für die Reliabilität der Ergebnisse). Hier gilt es zu klären, ob die Lehrenden entweder bereits Maßnahmen verwenden, die zum Erreichen einer hohen Konsistenz bei Korrekturen beitragen, oder ob die Sicherstellung einer konsistenten Korrektur nicht als relevant wahrgenommen und entsprechend nicht berücksichtigt wird.

5.4.2 Erkenntnisse zur bestehenden Prüfungspraxis

Zur bestehenden Prüfungspraxis wird betrachtet, ob bei der Korrektur von Prüfungen bisher Ankerbeispiele, Regelwerke zur Charakterisierung von Bewertungskategorien und eine Vorgabe zur Verteilung von Punkten verwendet wurden. Wie im Theorieteil erörtert, tragen diese Aspekte dazu bei, die Validität und Reliabilität der Korrektur zu gewährleisten (siehe Abschnitt 2.1.3.4). Bei den Analysen der bestehenden Prüfungen und den darin enthaltenen Musterlösungen zeigte sich, dass bei 95 Prozent der Aufgaben ein Ankerbeispiel (bzw. eine Musterlösung) angegeben wurde. Bei immerhin 79,4 Prozent aller Aufgaben wurden Angaben zur Vergabe von Punkten gemacht. Regelwerke zur Definition und Beschreibung der Bewertungskategorien waren bei nur

3,4 Prozent aller Aufgaben angegeben (siehe Abbildung 38). Bei 1,9 Prozent der Aufgaben (5 Aufgaben) wurden keine Angaben zur Bewertung gemacht.

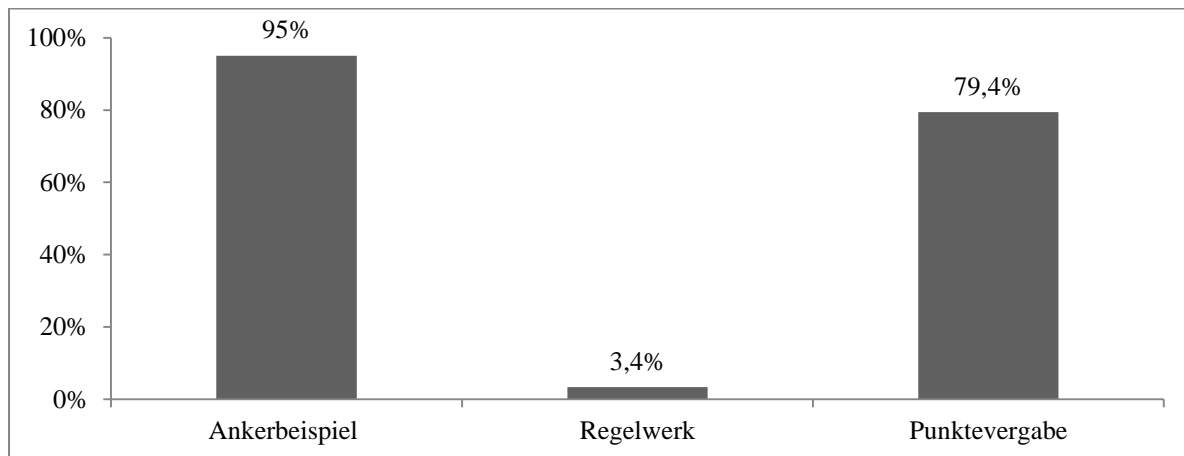


Abbildung 38: Angabe von Ankerbeispielen, Regelwerken und Punkteverteilung bei den bestehenden Prüfungsaufgaben (n=257)

Zu fast allen Aufgaben stehen den Lehrenden damit als Referenzpunkt für die Bewertung von Prüfungsantworten Ankerbeispiele zur Verfügung. Möglicherweise sind aus den Angaben zur Vergabe von Punkten weitere Hinweise ablesbar, welche Aspekte in einer Aufgabe vorhanden sein müssen. Wie aus den Programmzielen in Bezug auf die Bewertung von Prüfungsantworten hervorgeht, reicht den Lehrenden eine Orientierung an Ankerbeispielen und Punktevergabe nicht immer aus, um eindeutig entscheiden zu können, ob eine Aufgabe richtig beantwortet wurde. Die Erstellung von Regelwerken, um Bewertungskategorien zu definieren, wurde daher im Rahmen des Qualitätsentwicklungsprogramms thematisiert.

Neben den Bewertungskategorien kann die Qualität bei der Bewertung von Prüfungsantworten durch weitere Maßnahmen, wie der Schulung von Korrektoren, gesichert werden. Bei Betrachtung der Verbaldaten stellte sich heraus, dass die Lehrenden weitere Maßnahmen zur Sicherung der Qualität der Korrektur einsetzten (siehe Tabelle 28). Bei allen Fällen waren aufgrund der hohen Anzahl an Prüfungsfällen mehrere Personen bei der Korrektur eingebunden. Dabei war ein Korrektor jeweils für eine Aufgabe zuständig, die sie oder er bei allen Prüfungsfällen bewertete. Dadurch sollte die Konsistenz der Korrektur erhöht werden. Bei den Fällen MW 2, MW 3 und MW 4 berichteten die Teilnehmenden von einer Schulung der Korrektoren, bei der die Aufgaben mit den zugehörigen Musterlösungen vorgestellt wurden. Bei den Fällen MW 1 und MW 5 wurden diesbezüglich keine Angaben gemacht. Bei Antworten, die nicht eindeutig zugeordnet werden konnten, wendeten sich die Korrektoren bei den Fällen MW 1, MW 3 und MW 4 zur Klärung an den Ersteller der Aufgaben. Bei Fall MW 2 einigte man sich darauf, dass die die Korrektoren nach der Schulung zu den Musterlösungen, eigenständig Maßnahmen treffen sollen, um eine konsistente Korrektur zu gewährleisten. Bei Fall MW 5 wurden zusätzlich zur Musterlösung sowohl Ant-

worten von Kolleginnen und Kollegen aus den Probearbeitungen, als auch Antworten von Studierenden, die zwar nicht der Musterlösung entsprechen, jedoch trotzdem als richtig gewertet werden können, als Ankerbeispiele mitaufgenommen. Dadurch existierten hier bereits vor dem Programm zusätzlich zur Musterlösung weitere Referenzpunkte zur Bewertung von Prüfungsantworten.

Tabelle 28: Maßnahmen bei der Bewertung und Korrektur von Prüfungsantworten

	Mehrere Korrektoren	Aufgabenweise Korrektur	Schulung von Korrektoren	Rückfragen bei unklaren Antworten	Aufnahme von unerwarteten Lösungen als Ankerbeispiel	Aufnahme von Lösungen aus Probearbeitung
MW 1	✓ ⁴²	✓ ⁴²	○	✓ ⁴²	○	○
MW 2	✓ ⁴³	✓ ⁴⁴	✓ ⁴⁴	- ⁴⁴	○	○
MW 3	✓ ⁴⁵	✓ ⁴⁵	✓ ⁴⁵	✓ ⁴⁵	○	○
MW 4	✓ ⁴⁶	✓ ⁴⁶	✓ ⁴⁶	✓ ⁴⁶	○	○
MW 5	✓ ⁴⁷	✓ ⁴⁷	- ⁴⁸	○	✓ ⁴⁹	✓ ⁵⁰

Anmerkungen. ✓ = Trifft zu, - = aktive Ablehnung, ○ = wurde nicht genannt, k.A. = der Themenbereich, auf den sich die Kategorie bezieht, wurde nicht angesprochen. Sind einer Unterkategorie sowohl Häkchen als auch Minuszeichen zugeordnet, bedeutet dies, dass der Ist-Stand (fett gesetzt) zu einem Soll-Stand verändert werden soll.

5.4.3 Erkenntnisse zur veränderten Prüfungspraxis

Obwohl in der bestehenden Praxis bereits zentrale Maßnahmen zur Sicherung der Qualität bei der Bewertung von Prüfungsantworten durchgeführt wurden, bestand laut den Programmzielen für die Lehrenden das Problem in der Bewertung nicht zwischen sinngemäß richtigen Antworten, die aber von der Musterlösung abweichen, und fachlich substantielleren Antworten, differenziert werden konnte. Die Verwendung von Regelwerken, mit denen beschrieben wird, welche Merkmale und Eigenschaften Antworten aufweisen müssen, um einer Bewertungskategorie zugeordnet werden zu können, könnte dabei hilfreich sein. In der bestehenden Praxis wurde dieses Hilfsmittel

⁴² Vgl. Zitat 39, S. 283

⁴³ Vgl. Zitat 41, S. 284

⁴⁴ Vgl. Zitat 42, S. 284

⁴⁵ Vgl. Zitat 44, S. 284

⁴⁶ Vgl. Zitat 46, S. 285

⁴⁷ Vgl. Zitat 50, S. 286

⁴⁸ Vgl. Zitat 51, S. 286

⁴⁹ Vgl. Zitat 50, S. 286

⁵⁰ Vgl. Zitat 49, S. 286

jedoch kaum verwendet. In den Arbeitstreffen gingen die Moderatoren daher auf die Erstellung von Regelwerken ein, um aufzuzeigen, dass damit ein klares Kriterium zur Bewertung von Aufgaben geschaffen werden kann. Wie in Abbildung 39 dargestellt, wurden auch in der veränderten Praxis bei nur 6,1 Prozent (3,4 % in der bestehenden Praxis) der Aufgaben Regelwerke angefertigt. Der Input der Moderatoren zur Erstellung von Regelwerken wurde von den Lehrenden nur eingeschränkt angenommen. Für den Großteil der Aufgaben sahen die Lehrenden in den Regelwerken keinen Mehrwert zu ihren bereits durchgeführten Maßnahmen und bewerteten den Aufwand zur Erstellung deshalb als zu aufwändig. Insbesondere die Möglichkeit zur Rückfrage bei unklaren Antworten wurde von den Lehrenden als ebenso effektive und dabei weniger aufwändige Maßnahme genannt (vgl. Zitat 51, S. 286).

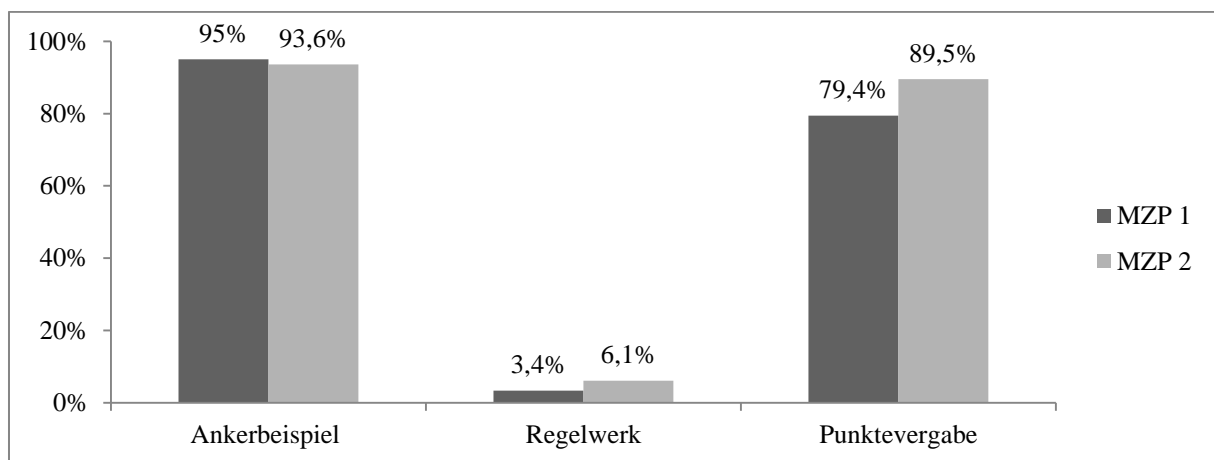


Abbildung 39: Angabe von Ankerbeispielen, Regelwerken und Punkteverteilung bei den bestehenden und neu erstellten Prüfungsaufgabe ($n_{MZP1}=262$; $n_{MZP2}=295$)

Bei Fall MW 2 wurde jedoch eine Lösung entwickelt, um Antworten, die zwar sinngemäß richtig sind, jedoch nicht die entsprechende Fachsprache enthalten, als falsch bewerten zu können. Auf dem Deckblatt der Prüfung wurde vermerkt, dass nur Antworten mit richtiger Fachsprache und -terminologie als richtig gewertet werden (vgl. Zitat 43, S. 284). Die Studierenden wurden zu Beginn der Prüfung auf diese Regelung aufmerksam gemacht und mussten auf dem Deckblatt per Unterschrift bestätigen diese Information erhalten zu haben. Den Korrektoren konnte damit ebenfalls ein klares Bewertungskriterium für die Bewertung von Prüfungsantworten vorgegeben werden. Weitere Veränderungen in der Vorgehensweise bei Bewertung und Korrektur von Prüfungsantworten wurden von den Lehrenden nicht berichtet.

5.4.4 Kurzzusammenfassung

In diesem Kapitel wurden die Erkenntnisse zur Bewertung und Korrektur von Prüfungsantworten berichtet. Dabei wurde betrachtet, ob die Lehrenden Maßnahmen, die sich in der Forschung als wirksam zur Verbesserung der Korrektur herausstellten, in ihrer Praxis umsetzten. Dazu gehören

die Schulung von Korrektoren sowie die Verwendung von Bewertungsrastern, die Bewertungskategorien, Regelwerke, Ankerbeispiele und Angaben zur Vergabe von Punkten enthalten. In Tabelle 29 sind die wesentlichen Erkenntnisse überblicksartig dargestellt.

Tabelle 29: Wesentliche Erkenntnisse zur Bewertung von Prüfungsantworten

Bewertung von Prüfungsantworten	FF 1: Zielsetzungen der Teilnehmenden	<ul style="list-style-type: none"> - Der Schwerpunkt in den Zielsetzungen betrifft die Frage, wie ein Bewertungskriterium erstellt werden kann, um zwischen sinngemäß richtigen Antworten, die sich jedoch fachlich auf unterschiedlichen Niveaus befinden, unterscheiden zu können. - Ein weiteres Anliegen war es, die Bewertung für Studierende nachvollziehbarer zu gestalten, um Beschwerden bei der Prüfungseinsicht zu vermeiden. - Ziel Möglichkeiten kennen lernen, um die Korrekturzeit zu verringern
	FF 2: Bestehende Praxis	<ul style="list-style-type: none"> - Ankerbeispiele im Sinne einer Musterlösung wurden bei fast allen Aufgaben angegeben. - Anweisungen zur Vergabe von Punkten auf einzelne Aspekte in Antworten wurden bei 80 Prozent der Aufgaben angegeben. - Regelwerke um Bewertungskategorien zu beschreiben wurden nicht verwendet. - Ein Korrektor bewertete jeweils nur eine Aufgabe für alle Prüfungsfälle. Dadurch möchten die Lehrenden Unterschiede in der Bewertung zwischen Korrektoren vermeiden. - Bei drei von fünf Fällen fand eine Schulung von Korrektoren statt. - Korrektoren konnten Rückfragen bei unklaren Antworten Rückfragen an Aufgabenersteller stellen. - Bei Fall MW 5, Erweiterung der Musterlösung um weitere Antworten die von der Musterlösung abweichen, jedoch sinngemäß richtig sind.
	FF 3: Veränderte Praxis	<ul style="list-style-type: none"> - Beibehaltung der Maßnahmen zur Qualitätssicherung der Korrektur aus der bestehenden Praxis - Bei fast allen Aufgaben wurden Ankerbeispiele angegeben - Angaben zur Vergabe von Punkten wurden bei 90 Prozent aller Aufgaben gemacht (80 % in der bestehenden Praxis). - Weiterhin kaum Verwendung von Regelwerken, da der Aufwand für die Erstellung im Vergleich zum Nutzen, als zu hoch eingestuft wurde.

5.5 Punkte- und Notenvergabe

Die Entscheidung über die Vergabe von Punkten zu einzelnen Aufgaben wird über die Lehrziele bestimmt und erfolgt spätestens nach der Bewertung von Prüfungsantworten. Mit der Punktevergabe können die Lehrenden inhaltliche Schwerpunktsetzungen vornehmen. Erfolgt die Vergabe von Punkten unreflektiert, so ist es möglich, dass die Endnote nicht valide ausdrückt, inwieweit die Lehrziele erreicht wurden. Ebenso wie die Punktevergabe muss sich die Festlegung von Notenstufen an den Lehrzielen orientieren, damit die Noten Aufschluss über das Erreichen der in den Lehrzielen beschriebenen Anforderungen geben (siehe Abschnitt 2.1.3.5). Zur Vergabe von Punkten und Noten wurden keine Programmziele von den Lehrenden formuliert. Möglicherweise wurden diesbezüglich keine Probleme wahrgenommen. Das in der Praxis etablierte Vorgehen bei der Punkte- und Notenvergabe wurde während des Qualitätsentwicklungsprogramms auch nicht verändert. Daher wird die Gliederungslogik mit einer Unterteilung in Programmziele, bestehende Praxis und veränderte Praxis in diesem Kapitel nicht verwendet. Die Ergebnisse werden stattdessen zum Vorgehen der Lehrenden bei der Vergabe von Punkten (Abschnitt 5.5.1) und der Festlegung von Noten (Abschnitt 5.5.2) berichtet. Die Ergebnisse basieren auf der Analyse der Verbaldaten.

5.5.1 Vergabe von Punkten

Die Vergabe von Punkten zu einzelnen Aufgaben erfolgt bei allen Lehrenden über die geschätzte Bearbeitungsdauer, die ein Studierender benötigt, um eine Aufgabe zu lösen. Bei Fall MW 5 werden zwei Punkte pro Minute Bearbeitungszeit vergeben (vgl. Zitat 65, S. 291). Auf eine Aufgabe, die in zwei Minuten vollständig bearbeitbar ist, werden also vier Punkte vergeben. Bei den Fällen MW 2, MW 3 und MW 4 wird pro Minute Bearbeitungszeit ein Punkt vergeben (vgl. Zitat 54, S. 288, Zitat 63, S. 290 und Zitat 64, S. 291). Bei Fall MW 1 wurde kein genauer Zeitwert genannt, jedoch ist auch die geschätzte Bearbeitungszeit maßgeblich für die Punktevergabe (vgl. Zitat 52, S. 287).

Zu den Fällen MW 2 und MW 5 konnte ermittelt werden, wie die Bearbeitungsdauer, die ein Studierender für die Bearbeitung einer Aufgabe aufwenden kann, geschätzt wird. Bei Fall MW 2 erfolgt zunächst eine Schätzung der Bearbeitungsdauer durch die Aufgabenersteller (vgl. Zitat 54, S. 288). Darüber hinaus wird überprüft, wie viele bepunktete Einheiten in einer Antwort enthalten sind (z.B. bei der Aufzählung von vier Stichpunkten, können vier Einheiten bepunktet werden). Werden auf einen Aufgabenblock (der mehrere Handlungsaufforderungen enthalten kann), basierend auf der geschätzten Bearbeitungsdauer, insgesamt zehn Punkte vergeben, sollten in den Antworten zur Lösung der Aufgaben ebenfalls zehn bepunktete Elemente vorhanden sein (vgl.

Zitat 54, S. 288). Die Lehrenden achteten in der Folge darauf, dass die Anzahl der bepunktbaren Einheiten mit der geschätzten Bearbeitungsdauer in Minuten übereinstimmt.

Bei Fall MW 5 wird die Prüfung vor dem Prüfungstermin probeweise von Kolleginnen und Kollegen bearbeitet (siehe dazu ausführlicher Abschnitt 5.3.2). Die dabei dokumentierten Bearbeitungsdauern werden bei der Punktevergabe zugrunde gelegt (vgl. Zitat 65, S. 291). Bei Berechnungsaufgaben, bei denen ein Aufgabenkontext angegeben ist, wurden zudem der Leseaufwand und die geschätzte Dauer, um sich in die Problemstellung einzudenken, mitberücksichtigt. Selbst wenn die Lösung der Aufgabe verhältnismäßig einfach oder in wenigen Schritten erfolgt, wurden entsprechend des geschätzten Zeitaufwands für das Lesen der Anweisungen mehr Punkte vergeben (vgl. Zitat 65, S. 291).

Im Rahmen des Qualitätsentwicklungsprogramms wurde versucht, die Lehrenden für mögliche ungewollte Auswirkungen, die durch die Vergabe von Punkten nach der Bearbeitungsdauer entstehen können, zu sensibilisieren. Beispielsweise wäre es möglich, dass für arbeitsintensive Aufgaben (z.B. Berechnungsaufgaben mit mehreren Zwischenschritten oder hohem Leseaufwand), die jedoch relativ einfach sind, mehr Punkte vergeben werden, wie auf eine anspruchsvollere Aufgabe, deren Bearbeitung jedoch weniger Zeit in Anspruch nimmt. Wie sich in den Diskussionen zeigte, begründeten die Lehrenden von Fall MW 1 und MW 2 die Punktevergabe nach Zeit mit dem Argument, dass anspruchsvollere Aufgaben auch eine längere Bearbeitungsdauer erfordern würden (vgl. Zitat 52, S. 287 und Zitat 55, S. 288). Ein weiteres Argument für die Bepunktung nach Zeit wird darin gesehen, dass die Studierenden dadurch einen Anhaltspunkt haben, wie viel Zeit sie für einzelne Aufgaben aufwenden können (vgl. Zitat 57, S. 289).

Obwohl die Lehrenden an ihrer Vorgehensweise mit der zeitgebundenen Punktevergabe auch für die neu erstellte Prüfung festhielten, konnte durch die Hinweise der Moderatoren während der Arbeitstreffen möglicherweise ein Problembewusstsein im Hinblick auf die mit der Punktevergabe vorgenommene Gewichtung erzeugt werden. Die Teilnehmenden von Fall MW 2 beispielsweise identifizierten im Nachgang zwei Aufgaben, bei denen aufgrund des geschätzten Zeitaufwands eine ähnlich hohe Punktzahl vergeben wurde, obwohl sich das Anspruchsniveau der Aufgaben unterschieden hat (vgl. Zitat 59, S. 289).

Die Vergabe von Punkten nach Zeit scheint jedoch ein bewährtes und etabliertes Vorgehen zu sein, um zu bestimmen, wie viele Punkte auf eine Aufgabe vergeben werden sollen. Damit ist ein klares Kriterium gesetzt. Inwieweit es mit einer Zeitskala gelingt, anspruchsvollere oder relevantere Anforderungen von weniger anspruchsvollen und relevanten Anforderungen zu unterscheiden, müsste in weiteren Untersuchungen geklärt werden. Die Lehrenden gingen davon aus, dass zur Lösung schwierigerer Aufgaben in der Regel mehr Zeit aufgewendet werden muss, als für das

Lösen einfacher Aufgaben. Damit wird angenommen, dass die Zeitskala mit der Aufgabenschwierigkeit korreliert. Gleichzeitig wurden die schwierigeren Anforderungen von den Lehrenden auch als relevanter wahrgenommen, als weniger schwierige Anforderungen. Ist diese Annahme nicht zutreffend, können jedoch periphere Anforderungen ein stärkeres Gewicht bei der Notenbildung haben, als die für ein Lehrziel tatsächlich relevanten Anforderungen. Hier wären weitere Analysen notwendig, um aufzuklären, ob die Annahmen hinter der Punktevergabe nach Zeit bei den eingesetzten Aufgaben auch zutreffend sind. Erste Hinweise diesbezüglich werden in Abschnitt 5.6 berichtet.

5.5.2 Notenvergabe

Die Lehrenden der Fälle MW 1, MW 2 und MW 5 beschrieben ein zweistufiges Verfahren zur Bestimmung von Notenstufen. Dabei wurde zunächst die Mindestpunktzahl für das Bestehen der Prüfung festgelegt, worauf die Notenstufe 4,0 vergeben wird. Die Grenze für das Bestehen liegt etwa bei 50 Prozent der maximal erreichbaren Punkte. Die weiteren Notenstufen erfolgen in den Abstufungen 4,0; 3,7; 3,3; 3,0; 2,7; 2,3;...; 1,0. Inklusiv der Note 1,0, als bestmöglichem Ergebnis, und der 4,0, als Grenze für das Bestehen, wird die Notenskala damit in 10 Notenstufen unterteilt. Die Punkte werden linear auf die einzelnen Notenstufen verteilt (vgl. Zitat 53, S. 287, Zitat 66, S. 291). Die Spannweite einer Notenstufe beträgt jeweils fünf Prozent der zu erreichenden Gesamtpunktzahl (siehe Tabelle 30). Bei den Fällen MW 3 und MW 4 befand man sich bezüglich der Skalierung von Noten noch in der Klärungsphase.

Tabelle 30: Exemplarische Notenverteilung bei einer Gesamtpunktzahl von 100

Notenstufen	1.0	1.3	1.7	2.0	2.3	2.7	3.0	3.3	3.7	4.0
Zu erreichende Punktzahl in Prozent	95 %	90%	85%	80%	75%	70%	65%	60%	55%	50%
Zu erreichende Punktzahl in absoluten Zahlen	95	90	85	80	75	70	65	60	55	50
Punktendifferenz zwischen zwei Notenstufen	5	5	5	5	5	5	5	5	5	5

Im Anschluss an die Generierung der Noten auf Basis des geschilderten Vorgehens betrachteten die Lehrenden im nächsten Schritt die erzielte Notenverteilung. Eine Normalverteilung der Prüfungsergebnisse wurde dabei bei allen Fällen als erstrebenswert betrachtet (vgl. Zitat 53, S. 287 und Zitat 67, S. 291). Neben der Notenverteilung achteten die Lehrenden darauf, dass der Notenschnitt in jeder Prüfung (über Semester hinweg) möglichst konstant ausfällt (vgl. Zitat 53, S. 287, Zitat 61, S. 290 und Zitat 66, S. 291). Um eine Normalverteilung und vergleichbare Notendurchschnitte zu erreichen, wurde an mehreren Stellschrauben gedreht. Zum einen können die Lehren-

den die maximal erreichbare Punktzahl flexibel anpassen, insofern mehr Aufgaben in die Prüfung aufgenommen wurden, als später in die Bewertung eingehen (sogenannte Überhangsaufgaben). Dadurch besteht die Möglichkeit, den Notenschnitt zu beeinflussen, indem entweder häufig gelöste Aufgaben (bei einem zu guten Schnitt) oder selten gelöste Aufgaben (bei einem zu schlechten Schnitt) nicht in die Bewertung eingehen. Bei Fall MW 2 wird die höchste von einem Studierenden erreichte Punktzahl als Maximalpunktzahl für die Prüfung festgelegt (vgl. Zitat 60, S. 290). Der oder die beste Studierende setzt damit die obere Punktegrenze fest. Die Bildung der Noten für die restlichen Studierenden hängt damit von dieser Maximalpunktzahl ab. Die Notenvergabe erfolgt demnach nach einer sozialen Bezugsnorm.

Bei Fall MW 5 wurde außerdem am unteren Ende der Punkteskala variiert, indem bei Bedarf die Bestehensgrenze nach unten versetzt wurde (vgl. Zitat 67, S. 291). Eine dritte Möglichkeit, die angewendet wurde, um den Notenschnitt und die Notenverteilung zu beeinflussen, bestand darin, die Spannweite einzelner Notenstufen zu verändern. Indem die Punktegrenze zum Erreichen einer bestimmten Notenstufe nach unten versetzt wird, können Studierende, die sich vorher knapp unterhalb dieser Grenze befanden, eine Notenstufe nach oben rücken (vgl. Zitat 53, S. 287 Zitat 62, S. 290, Zitat 67, S. 291).

Durch Veränderung der Spannweiten von Notenstufen wurde zudem versucht, die Anzahl an Studierenden, die sich dicht an der Stufe zur nächstbesseren Note befindet, möglichst gering zu halten, um Beschwerden von Studierenden bei der Prüfungseinsicht zu reduzieren (vgl. Zitat 67, S. 291). Dazu betrachteten die Lehrenden zunächst, ob sich mehrere Studierende an einer bestimmten Notengrenze befinden. Diese Notengrenze wurde dann entsprechend auf der Punkteskala nach unten geschoben, damit diese Gruppe in die nächsthöhere Notenstufe nachrückte. Durch diese Verschiebung können möglicherweise andere Studierende näher an die Notengrenze heranrücken. Bei diesen Fällen wurde zudem versucht, in den Prüfungen Möglichkeiten zu finden, um die benötigten Punkte noch zu vergeben (vgl. Zitat 53, S. 287 Zitat 62, S. 290, Zitat 67, S. 291). Von diesen Verschiebungen der Notengrenzen profitieren die Studierenden. Jedoch müsste geklärt werden, ob durch das Herabsetzen der Bestehensgrenzen, Studierende, die aus inhaltlicher Perspektive die Anforderungen nicht zufriedenstellend erreicht haben, die Prüfung dennoch bestehen.

Mit den von den Lehrenden durchgeführten Maßnahmen bei der Notenbildung erfolgt die Notenvergabe vor allem bezugsgruppenorientiert. Bei der Angleichung an eine Normalverteilung bildet das Leistungsvermögen der aktuellen Studierendengruppe den Bezugsmaßstab für die Bildung von Einzelnoten. Eine individuelle Note sagt in dem Fall aus, in welcher Rangfolge ein Studierender im Vergleich zur Gruppe der Studierenden steht, die die Prüfung ebenfalls abgelegt haben. Ob und inwieweit die Lehrziele von den Studierenden erreicht wurden, kann anhand von derart

gebildeten Noten nicht ausgesagt werden (theoretisch wäre es sogar möglich, dass die oder der beste Studierende das inhaltliche Mindestniveau nicht erreichen konnte, aber dennoch mit einer 1,0 abschneidet). Wie im Theorieteil dargelegt, ist bei einer kompetenzorientierten Prüfungspraxis eine kriteriumsorientierte Notenvergabe vorzusehen (siehe Abschnitt 2.1.3.5), damit die Noten Aufschluss über den erreichten Kompetenzstand geben. Dieser Aspekt wurde während der Arbeitstreffen durch die Moderatoren angesprochen. Eine Veränderung in der Praxis konnte jedoch nicht angestoßen werden. Das von den Lehrenden etablierte Vorgehen scheint sich in der Praxis bewährt zu haben, so dass auf Seiten der Lehrenden kein Änderungsbedarf gesehen wurde.

5.5.3 Kurzzusammenfassung

In diesem Kapitel wurden die Erkenntnisse zur Punkte- und Notenvergabe berichtet. Da die Lehrenden zur Punkte- und Notenvergabe keine Zielsetzungen formulierten und auch ihre Praxis während des Programms nicht veränderten, gliederte sich das Kapitel nicht in Programmziele, bestehende Praxis und veränderte Praxis. Stattdessen wurde gesondert auf das Vorgehen bei der Punktevergabe und die Festlegung einer Notenskala eingegangen. In Tabelle 31 werden die wesentlichen Erkenntnisse überblicksartig dargestellt.

Tabelle 31: Wesentliche Erkenntnisse zur Punkte- und Notenvergabe

Punkte- und Notenvergabe	FF 2: Punktevergabe	<ul style="list-style-type: none"> - Die Punktevergabe richtet sich nach der für eine Aufgabe geschätzten Bearbeitungsdauer. - Die Bearbeitungsdauer wird entweder von den Aufgabenerstellern geschätzt oder orientiert sich am Zeitaufwand den Kolleginnen und Kollegen bei der Probearbeitung der Prüfung für eine Aufgabe benötigten. - Laut Einschätzung der Lehrenden korreliert die Bearbeitungsdauer mit der Aufgabenschwierigkeit. - Bei der Punktevergabe werden Lehrzielen oder Aspekte wie die erwartete oder gemessene Aufgabenschwierigkeit oder der Relevanz von Anforderungen nicht berücksichtigt. - Die zeitbasierte Vergabe von Punkten wurde beibehalten, trotz Hinweisen, dass dadurch möglicherweise einfache jedoch zeitaufwändige Aufgaben genauso stark gewichtet werden, wie anspruchsvollere Anforderungen.
--------------------------	---------------------	--

	FF 2: Notenvergabe	<ul style="list-style-type: none">- Anwendung eines zweistufigen Verfahrens bei der Notenvergabe:<ul style="list-style-type: none">(1) Die Bildung von Noten erfolgt zunächst anhand einer Notenskala mit zehn Notestufen (zwischen 4,0 als Bestehensgrenze und 1,0 als Bestnote). Der Punkteabstand zwischen zwei Notestufen wird linear festgelegt. Die Spannweite einer Notestufe beträgt 5 Prozent der Gesamtpunktzahl.(2) Anschließend wird eine Angleichung der Notenverteilung an eine Normalverteilung vorgenommen. Außerdem soll der Notenschnitt Kohortenübergreifend konstant ausfallen Um dies zu erreichen werden:<ul style="list-style-type: none">a. Veränderungen an der maximal erreichbaren Punktzahl vorgenommen (über Überhangsaufgaben),b. Bestehensgrenzen herabgesetztc. die Spannweite bestimmter Notestufen verändert.- Notestufen sind nicht an Lehrzielen orientiert. Daher keine kriteriumsorientierte Notenvergabe.- Notenverteilung ist ausgerichtet an sozialer Bezugsnorm. Die Noten drücken aus, wie ein Studierender im Vergleich zu einer Bezugsgruppe abgeschnitten hat, jedoch nicht welches Kompetenzniveau erreicht wurde.
--	--------------------	--

5.6 Rückschlüsse aus den Prüfungsdaten

Mit einer Prüfung werden Daten gewonnen, die - über den Zweck der Notenbildung hinaus - wertvolle Informationen liefern. Zu Forschungsfrage 1 wird berichtet, welche Programmziele diesbezüglich von den Lehrenden formuliert wurden. Hinsichtlich Forschungsfrage 2 wird zunächst geklärt, ob Lehrenden in der bestehenden Praxis die Prüfungsergebnisse in Datensätzen so aufbereiteten, dass statistische Analysen durchgeführt werden können. Anschließend wird dargestellt, welche statistischen Kennzahlen die Lehrenden zu ihren Prüfungen bisher betrachteten und welche Rückschlüsse sie daraus zogen (Abschnitt 5.6.2). In Abschnitt 5.6.3 werden die Erkenntnisse zur veränderten Praxis beschrieben (Forschungsfrage 3). Zur Beantwortung von Forschungsfrage 2 und 3 wird auf die Verbaldaten zurückgegriffen. Außerdem wurden die Datensätze zu den neu erstellten Prüfungen gesichtet.

5.6.1 Zielsetzungen der Hochschullehrenden

Bezüglich der statistischen Analyse von Prüfungsergebnissen wurden zwei Projektziele von den Lehrenden der Fälle MW 1 und MW 5 definiert, in denen relativ allgemein der Wunsch nach einer Betrachtung von Aufgabenschwierigkeiten geäußert wird (siehe Tabelle 32). Konkrete Fragen, wie man bei entsprechenden Analysen vorgehen könnte oder Probleme die in bereits durchgeführten Analysen auftraten, wurden nicht genannt. Die Lehrenden der anderen Fälle gaben diesbezüglich keine Ziele an.

Tabelle 32: Programmziele zur Gewinnung von Rückschlüssen aus den Prüfungsdaten

Fälle	Projektziele
MW 1	Analysieren, welche Prüfungsaufgaben gut bzw. schlecht beantwortet werden.
MW 5	Statistische Analyse der bisherigen Aufgaben (z.B. wurde eine Aufgabe in einem Jahr gut und im anderen Jahr schlecht bearbeitet, Betrachtung von Aufgabenschwierigkeiten).

5.6.2 Erkenntnisse zur bestehenden Prüfungspraxis

Für die bestehende Praxis wird geklärt, ob zum einen die Prüfungsergebnisse in einem Datensatz bisher so aufbereitet wurden, dass statistische Analysen durchgeführt werden können (Abschnitt 5.6.2.1). Zum anderen wird der Frage nachgegangen, ob, beziehungsweise welche, statistischen Kennzahlen die Lehrenden bestimmt haben und welche Rückschlüsse sie daraus bisher gezogen haben (Abschnitt 5.6.2.2).

5.6.2.1 Aufbereitung von Datensätzen

Bei allen Fällen wurden für die neu erstellten Prüfungen Datensätze erstellt. In den Datensätzen wurde für jeden Studierenden separat erfasst, wie viele Punkte er oder sie pro Aufgabe erreichten. Die von den Lehrenden erstellten Datensätze unterschieden sich in der Granularität, mit der die Daten erfasst wurden. Bei Fall MW 5 lagen die Daten auf Ebene von Aufgabenblöcken vor, die zum Teil bis zu sechs Teilaufgaben umfassten. Bei einer zusammengefassten Darstellung der Daten tritt jedoch das Problem auf, dass bei den Auswertungen nicht differenziert werden kann, auf welche Aspekte innerhalb der Aufgabenblöcke Punkte vergeben wurden. Es kann somit beispielsweise nicht aufgeklärt werden, welche Anforderungen den Studierenden, beziehungsweise bestimmten Studierendengruppen, leichter oder schwerer gefallen sind. Bei den Fällen MW 1, MW 2, MW 3 und MW 4 wurden die Punkte auf Ebene von Teilaufgaben angegeben. Hierdurch ist eine differenziertere Auswertung der Daten möglich. Wie in Abschnitt 5.2.2.1 beschrieben, waren jedoch bei 50 Prozent der Aufgaben mehrere Handlungsaufforderungen, die möglicherweise unterschiedliche Anforderungen und Anforderungsniveaus ansprechen (z.B. ein Aspekt zielt auf deklaratives Wissen ab, der zweite Teil der Aufgabe auf konzeptionelles Wissen), in einer Aufgabenstellung zusammengefasst. Wird bei einer derartigen Aufgabe nicht die volle Punktzahl erreicht, kann mit den Daten nicht differenziert werden, welcher Aspekt in den Aufgaben nicht bearbeitet werden konnte. In Tabelle 33 ist exemplarisch ein Auszug aus dem Datensatz von Fall MW 5 dargestellt.

Tabelle 33: Exemplarischer Auszug aus dem Datensatz von Fall MW 5

Aufgaben Nr.	1	2	3	4	5	6	7	8	9	10	11	Erreichte Punkte
Erreichbare Punktzahl	11	13	6	6	21	7	22	11	18	6	12	133
Student 1	7	1	6	3	9	3	19	4	7	3	7	69
Student 2	7	12	6	6	18	7	21	11	16,5	6	5	115,5
Student 3	7	8	5	6	13	5,5	18	5	14	5	12	98,5

5.6.2.2 Rückschlüsse aus den Prüfungsdaten

Die Lehrenden nutzten die Datensätze vorwiegend, um die Notenverteilung der Klausur zu bestimmen. Die Datensätze wurden bei allen Fällen in einem Tabellenkalkulationsprogramm angelegt und so konfiguriert, dass pro Student die Gesamtpunktzahl und Note, sowie über alle Studenten hinweg, die Notenverteilung und der Notendurchschnitt automatisch erzeugt wurden. Wie bereits bei den Ausführungen zur Notenvergabe dargelegt (Abschnitt 5.5.2) wird eine Normalverteilung der Noten als ideal angesehen.

Zur Betrachtung weiterer statistischer Kennwerte liegen hauptsächlich Einzelaussagen vor. Eine systematische Bestimmung und Betrachtung von statistischen Kennwerten im Sinne einer etablierten Praxis wurde in der bestehenden Praxis nicht durchgeführt, zumindest konnten in den Gesprächen keine Hinweise darauf gefunden werden. Bei Fall MW 1 wurde zusätzlich zur Notenverteilung überprüft, ob alle Aufgaben bearbeitet wurden. Lag eine Normalverteilung vor und waren gleichzeitig keine Aufgaben in der Prüfung enthalten, die von keinem Studierenden bearbeitet wurden, wurde die Qualität der Prüfung als passend angenommen (vgl. Zitat 68, S. 292).

Die Lehrenden von Fall MW 2 nutzten den Datensatz zusätzlich auch, um Studierende zu identifizieren, die nur einen Punkt von der nächstbesseren Notenstufe entfernt liegen (siehe Kapitel 5.5.2). Der Datensatz ist dabei so konfiguriert, dass automatisch Aufgaben gekennzeichnet werden, bei denen ein Studierender keine Punkte (rot), nicht alle Punkte (weiß) oder alle Punkte erzielte (grün). Wenn ein Studierender auf eine Aufgabe mehr Punkte erhält als eigentlich erreichbar sind, wird dies mit gelb angezeigt. Damit wird Fehlern bei der Eingabe vorgebeugt.

Die Lehrende von Fall MW 3 nutzte diese automatische Anzeige ebenfalls, um Aufgaben zu identifizieren, bei denen mehr Punkte als die maximal mögliche Punktzahl vergeben wurden. Inwieweit hier weitere Informationen aus den Daten gewonnen wurden, konnte in den Gesprächen nicht ermittelt werden. Bei Fall MW 4 wurden die Aufgabenschwierigkeiten bestimmt - inwieweit diese jedoch interpretiert werden, konnte in den Gesprächen nicht ermittelt werden. Bei Fall MW 5 wurden neben der Betrachtung der Notenverteilung ebenfalls die Lösungshäufigkeiten betrachtet. Aufgaben, die von einem Teil der Studierenden fast vollständig und vom anderen Teil gar nicht gelöst wurden, dienten als Indikator für Abhängigkeiten zwischen Aufgaben (vgl. Zitat 77, S. 295).

Wie aus den Beschreibungen zum Vorgehen bei den einzelnen Fällen hervorgeht, nutzten die Lehrenden ihre Datensätze bisher vorwiegend zur Betrachtung von Notenverteilungen. Weitere Informationen, die aus den Daten gewonnen wurden, beziehen sich auf die bei Aufgaben erzielte Punktzahl, ob bestimmte Aufgaben gar nicht bearbeitet wurden und inwieweit bei Aufgaben mehr Punkte eingetragen wurden, als die erreichbare Gesamtpunktzahl zulässt. Bei Fall MW 2 wurden zudem Studierende identifiziert, die nur knapp unterhalb der nächstbesseren Notenstufe liegen. Bei den Fällen MW 4 und MW 5 wurden zusätzlich Aufgabenschwierigkeiten bestimmt. Die Lehrenden machten jedoch keine Angaben darüber, in welcher Weise die Aufgabenschwierigkeiten betrachtet und interpretiert wurden.

5.6.3 Erkenntnisse zur veränderten Prüfungspraxis

Im Rahmen des vierten Arbeitstreffens der Fortbildung wurden die Datensätze zu den neu erstellten Prüfungen besprochen. Die Lehrenden wurden dazu befragt, welche statistischen Kennwerte sie bislang in ihrer bestehenden Praxis gebildet hatten und welche Rückschlüsse sie daraus gezogen haben (siehe oben, Abschnitt 5.6.2). Im Vorfeld des vierten Treffens wurden die Lehrenden gebeten, folgende Fragen anhand der Prüfungsdaten zu beantworten:

- Gab es Aufgaben beziehungsweise Teilaufgaben, die unerwartet gut oder schlecht bearbeitet wurden?
- Welche Gründe könnte es haben, wenn eine Aufgabe für die Studierenden unerwartet einfach oder schwierig war?
- Bei welchen Aufgaben konnte keiner der Studierenden die Maximalpunktzahl erreichen? Was könnten Gründe hierfür sein?

Mit den Fragen wurde versucht, erste Impulse für eine umfassendere Nutzung der Daten zu geben. Die Moderatoren analysierten die Daten ebenfalls auf diese Fragen hin. Die Ergebnisse der Analysen wurden in den Arbeitstreffen besprochen. Die Diskussionen zu den statistischen Kennwerten waren sehr umfassend und bezogen sich hauptsächlich auf Detailfragen zu einzelnen Aufgabenstellungen, die ohne Kenntnis der jeweiligen Aufgaben, beziehungsweise dem dahinterstehenden fachlichen Hintergrund, für einen externen Leser wenig aussagekräftig sind. Nachfolgend werden daher auszugsweise solche Aspekte berichtet, die für die Lehrenden bei Betrachtung der statistischen Kennwerte unerwartet waren und Anlass zu Diskussionen über mögliche Veränderungen gaben.

Bei Fall MW 1 wurden auf eine Aufgabe 24 Punkte (etwa ein Fünftel der Gesamtpunktzahl) vergeben. Die Aufgabe beinhaltete umfangreiche Kontextinformationen, die bei der Bearbeitung berücksichtigt werden mussten, um hinsichtlich acht verschiedener Aspekte Entscheidungen zu treffen. Diese Entscheidungen mussten jeweils mit zwei Argumenten begründet werden. Es wurde je ein Punkt pro richtiger Entscheidung und für jede Begründung gegeben. Bei der Aufgabe waren die Ausgangsbedingungen definiert, jedoch waren mehrere unterschiedliche Lösungswege und Lösungen möglich. Die Aufgabe zielte auf konzeptionelles Wissen ab. Die Lehrenden gingen davon aus, dass mit der Aufgabe zwischen Studierenden differenziert werden kann, die dieses konzeptionelle Wissen besitzen und solchen, die es nicht besitzen. Die Moderatoren wiesen die Lehrenden darauf hin, dass bei der Aufgabe eine Lösungsquote von 84 Prozent erzielt wurde, womit sie sehr einfach war und auch von den Studierenden gelöst werden konnte, die insgesamt in

der Prüfung eher schlecht abschnitten. Die Lehrenden zogen daraus mehrere Schlüsse. Zum einen wurde bemerkt, dass bei der Bewertung der Antworten ein klares Bewertungskriterium fehlte, um bei der Korrektur zwischen elaboriert ausgeführten Antworten, in denen die nötigen fachlichen Aspekte enthalten waren, um auf das Vorhandensein eines konzeptionelles Verständnis zu schließen, und solchen Antworten, die zwar tendenziell in die richtige Richtung gingen, jedoch wesentliche Aspekte nicht aufwiesen, zu differenzieren. Dadurch konnten auch die Studierenden eine hohe Punktzahl erreichen, deren Antworten wenig substantiell waren (vgl. Zitat 69, S. 292). Anhand dieses Beispiels konnten die Moderatoren aufzeigen, dass es ohne klare Bewertungskriterien nicht notwendigerweise gelingt, gute von weniger guten Antworten zu trennen. Die Bedeutung von Regelwerken, in denen festgelegt wird, welche Merkmale eine Aufgabe aufweisen muss, um einer Bewertungskategorie zugeordnet werden zu können, konnte an diesen Beispielen noch einmal verdeutlicht werden. Die Lehrenden diskutierten außerdem, ob es aufgrund der hohen Lösungsquote weiterhin Sinn macht, die Aufgabe relativ stark mit Punkten zu gewichten. Die hohe Punktzahl kam insbesondere aufgrund des geschätzten Zeitaufwands zustande (für das Einlesen in die Kontextbedingungen und die Formulierung von je zwei Begründungen pro zu treffender Entscheidung), den die Lehrenden für die Lösung der Aufgabe veranschlagten (vgl. Zitat 70, S. 292). Durch Betrachtung der Aufgabenschwierigkeiten konnten die Teilnehmenden von Fall MW 1 dahingehend sensibilisiert werden, dass mittels einer Punktevergabe über die geschätzte Bearbeitungsdauer einfache Aufgaben ein sehr hohes Gewicht erhalten können. Diese Erkenntnis wurde außerdem an einem weiteren Aufgabenbeispiel diskutiert (vgl. Zitat 71, S. 293).

Bei den Prüfungen der Fälle MW 4 und MW 5 konnten bestimmte Aufgaben, die von den Lehrenden als relativ einfach eingestuft wurden, da sie auf die Wiedergabe von deklarativem Wissen abzielen, von den Studierenden jedoch nur selten gelöst werden. Bei beiden Fällen stellte sich heraus, dass die Aufgaben auf ein relativ spezialisiertes Wissen abzielten. Bei Fall MW 5 sollte ein sehr umfangreiches Diagramm detailliert beschriftet werden. Da das fragliche Diagramm vollständig im Skript abgebildet und die Aufgabe daher durch Erinnern der Beschriftung lösbar war, schätzte der Aufgabenersteller die Aufgabe als eher einfach ein. Keiner der Studierenden erreichte die volle Punktzahl. Die genaue Lösungsquote konnte nicht bestimmt werden, da die Daten nicht auf Teilaufgabenebene erfasst worden waren. Der Lehrende kam zu dem Schluss, dass eine detaillierte Wiedergabe der Diagrammbestandteile als Veranstaltungsziel weniger relevant ist, als ein konzeptionelles Verständnis der Graphen innerhalb des Diagramms. Für die nächste Prüfung wollte er die Aufgabe entsprechend verändern (vgl. Zitat 78, S. 295).

Bei der Aufgabe von Fall MW 4 sollte zu einer Tabelle und den darin aufgeführten Werten die zugehörige DIN Norm genannt werden, aus der die Werte stammen. Diese Norm wurde zwar in der Veranstaltung öfters angesprochen, jedoch existieren zu diesem Bereich mehrere Hundert

Normen. Die Lösungshäufigkeit bei der Aufgabe lag bei 3 Prozent (vgl. Zitat 75, S. 294). In beiden Fällen kamen die Lehrenden zu dem Schluss, dass Anforderungen, die auf ein zu spezialisiertes Wissen abzielen, von den Studierenden nicht entsprechend gelernt werden und auch für die Lehrziele in diesem Spezialisierungsgrad nicht relevant sind.

Durch die detaillierte Betrachtung von statistischen Kennwerten zu einzelnen Aufgaben konnten die Lehrenden hinsichtlich mehrerer Aspekte sensibilisiert werden. Anhand von unerwartet ausgefallenen Aufgabenschwierigkeiten begannen die Lehrenden, den angenommenen Zusammenhang zwischen Zeitaufwand und Schwierigkeit (je zeitaufwändiger eine Aufgabe desto schwieriger) zu diskutieren und hinterfragten die bislang bewährte Methode, die bei einer Aufgabe erreichbare Punktzahl anhand der geschätzten Bearbeitungsdauer zu vergeben. Nachdem bei einer Aufgabe in der Prüfung von Fall MW 1 die Aufgabenschwierigkeit unerwartet niedrig ausfiel, da ein klares Bewertungskriterium fehlte, um zwischen substantiellen und weniger reichhaltigen Antworten zu unterscheiden, wurde außerdem diskutiert, ob für bestimmte Aufgaben Regelwerke erarbeitet werden sollen. Die Regelwerke beschreiben welche Merkmale eine Aufgabe aufweisen muss, um einer Kategorie zugeordnet werden zu können. Wie in Kapitel 5.4.3 dargelegt, empfanden die Lehrenden das Anfertigen von Regelwerken zunächst als zu aufwendig im Vergleich zum erwarteten Mehrwert. Bei den Fällen MW 4 und MW 5 existierten Aufgaben, die für die Studierenden unerwartet schwierig waren. Die Lehrenden zogen daraus den Schluss, dass das in diesen Aufgaben abgefragte Wissen zu spezialisiert beziehungsweise detailliert und vor dem Hintergrund der Lehrziele letztlich auch gar nicht relevant war.

Zusammenfassend kann an dieser Stelle festgehalten werden, dass die Lehrenden durch die Betrachtung statistischer Kennzahlen (insbesondere den Aufgabenschwierigkeiten) Probleme und Schwachstellen in Bezug auf die gestellten Anforderungen, die Korrektur und die Punkte- und Notenvergabe ziehen konnten. Die Betrachtung dieser relativ einfach zu erzeugenden Kennzahlen ist demnach ein effektives Mittel, um Lehrende für vorhandene Probleme in den Prüfungen zu sensibilisieren, die andernfalls nicht wahrgenommen würden. Bei den Fällen MW 3 und MW 4 bezog sich die Diskussion statistischer Kennwerte auf einzelne Aufgabenstellungen, die für den Leser ohne Kenntnis der Aufgaben nicht aufschlussreich sind.

5.6.4 Kurzzusammenfassung

In diesem Kapitel wurde dargestellt, inwieweit die Lehrenden die mit den Prüfungen erzeugten Daten in Datensätzen aufbereiten und ob statistische Kennzahlen bestimmt und interpretiert wurden. Die Lehrenden formulierten zwei eher allgemein formulierte Zielsetzungen zu diesem Themenbereich aus denen keine weiterführenden Erkenntnisse zu möglichen Problemen oder Unter-

stützungsbedarfen abgeleitet werden konnten. In Tabelle 34 werden die wesentlichen Erkenntnisse zur bestehenden und veränderten Praxis überblicksartig dargestellt.

Tabelle 34: Wesentliche Erkenntnisse zu Rückschlüssen aus der Empirie

Rückschlüsse aus den Prüfungsdaten	FF 2: Bestehende Praxis	<ul style="list-style-type: none"> - Die Lehrenden legen Datensätze zu den Prüfungen an, die sich für statistische Auswertungen eignen. - Daten werden mit unterschiedlicher Granularität erfasst. Bei Erfassung der Daten auf Ebene von Aufgabenblöcken kann das Abschneiden bei einzelnen Anforderungen nicht differenziert ausgewertet werden. Anzustreben wäre eine Erfassung der Daten zu einzelnen Handlungsaufforderungen. - Die Daten werden in erster Linie genutzt, um die Notenverteilung abzubilden. Außerdem wird betrachtet, wie viele Studierende sich knapp unter einer Notenstufe bzw. der Grenze zum Bestehen befinden. - Auf Ebene von Aufgaben wurde zudem betrachtet, ob bestimmte Aufgaben gar nicht bearbeitet wurden oder mehr Punkte als möglich vergeben wurden (durch falsches Eintragen). Bei Fall MW 5 wurden Aufgaben gefiltert, die von einem Teil der Studierenden sehr gut, von anderen gar nicht bearbeitet wurden, da dies möglicherweise auf Abhängigkeiten zwischen den Aufgaben hindeutet. - Zu weiterführenden Interpretationen von Aufgabenschwierigkeiten oder weiterer Kennwerte wie Trennschärfen von Aufgaben wurden keine Angaben gemacht.
------------------------------------	-------------------------	--

- Eine fragengeleitete Betrachtung von statistischen Kennwerten stieß auf großes Interesse seitens der Lehrenden.
- Die Betrachtung von Aufgabenschwierigkeiten führte zu Aha-Momenten hinsichtlich mehrerer Themenbereiche:
 - o Punktevergabe nach der Bearbeitungsdauer wurde für unerwartet einfach ausgefallene Aufgaben hinterfragt.
 - o Als Ursache für unerwartet einfach ausgefallenen Aufgaben wurden von den Lehrenden unter anderem fehlende Bewertungskriterien bei der Korrektur genannt, wodurch nicht zwischen substantiellen Antworten und weniger reichhaltigen Antworten unterschieden werden konnte.
 - o Bei unerwartet schwierig ausgefallenen Aufgaben stellten die Lehrenden fest, dass die abgefragte Inhalte vor dem Hintergrund der Veranstaltungsziele nicht relevant waren.
- Fazit: Die fragengeleitete Betrachtung von Prüfungsdaten ist eine konstruktive Maßnahme, um Lehrenden Probleme und Schwachstellen in Prüfungen aufzuzeigen.

6 Diskussion

Prüfungen steuern das Lernen von Studierenden und tragen dadurch entscheidend zur Qualität der während eines Studiums erzielten Qualifikationen bei (Cilliers et al., 2010). Hochschulen entscheiden basierend auf Prüfungsergebnissen über den Studienfortschritt von Studierenden und die Verleihung von Abschlüssen (Secolsky & Denison, 2012). Die Lehrenden erhalten mit den Prüfungsergebnissen Rückmeldung über das Leistungsvermögen von Studierenden und können ihre Lehrveranstaltungen entsprechend anpassen (Shaeiwitz, 1996). Prüfungen können daher als wichtiger Hebel zur Verbesserung der Lehrqualität an Hochschulen verstanden werden. Damit mit Prüfungen der Lern- und Leistungsstand von Studierenden gültig (valide) und zuverlässig (reliabel) festgestellt werden kann, müssen sie von den Lehrenden so gestaltet werden, dass die Gütekriterien der pädagogischer Diagnostik – Validität, Reliabilität und Objektivität – eingehalten werden (Popham, 2009). Eine weitere Herausforderung stellt die im Rahmen der Bologna Reform geforderte Überprüfung von fachlichen und überfachlichen Kompetenzen dar, mit der sich Hochschulen und Lehrende konfrontiert sehen (Schaper, 2012). Die Gestaltung kompetenzorientierter und qualitativ hochwertiger Prüfungen verlangt eine entsprechende professionelle Kompetenz bei den Lehrenden (Popham, 2009). Bislang gibt es jedoch kaum empirisch gesichertes Wissen zur Prüfungspraxis von Hochschullehrenden und nur wenig Unternehmungen, Lehrende durch gezielte Qualitätsentwicklungsmaßnahmen an die geforderten Kriterien heranzuführen (siehe Abschnitt 2.2).

Das Anliegen der Dissertation war es deshalb, einen ersten Beitrag zu leisten, diese Wissenslücke zu schließen. Dazu wurde eine fallbasierte Untersuchung der Prüfungspraxis von Hochschullehrenden mit einer Stichprobe von N=14 Personen an sieben verschiedenen Lehrstühlen aus dem Maschinenwesen und der Mathematik der Technischen Universität München durchgeführt. Das fallbasierte Vorgehen, basierend auf einer kleiner Stichprobe, schien vor dem Hintergrund der Besonderheiten akademischer Lehre, die geprägt ist durch eine hohe Autonomie der Lehrenden, die ein spezialisiertes Wissen auf einem hohen fachlichen Niveau vermitteln (Braun et al., 2014), zielführend zu sein. Die empirischen Erhebungen wurden im Rahmen eines Qualitätsentwicklungsprogramms, das mehrere Arbeitstreffen umfasste, durchgeführt. Die Arbeitstreffen mit den Teilnehmenden aus dem Maschinenwesen⁵¹ unterschieden sich in der Anzahl sowie der inhaltlichen Schwerpunktsetzung von der Zusammenarbeit mit Teilnehmenden aus der Mathematik (siehe dazu ausführlicher Abschnitt 4.2.2). In dieser Arbeit wurden deshalb die Ergebnisse zu den Fällen aus der Mathematik nicht dargestellt, da aufgrund der Unterschiede in der Durchführung,

⁵¹ bezeichnet als Fall MW 1, MW 2, MW 3, MW 4 und MW 5

keine vergleichbaren Informationen zu den Fällen aus dem Maschinenwesen gewonnen werden konnten. Die Prüfungskonzepte aus der Mathematik werden stattdessen in einem Format außerhalb der Dissertation im Sinne einer Beschreibung von Best-Practices dargestellt. Die Forschungsfragen dieser Arbeit wurden zu den Fällen aus dem Maschinenwesen beantwortet, da mit den Teilnehmenden alle sechs Schritte des Prüfungsmodells angesprochen werden konnten und dementsprechend eine systematische Datenbasis vorliegt. In die Analysen der Fälle aus dem Maschinenwesen gingen sowohl materialbasierte Daten (bestehende und neu erstellte Lehrziele, Prüfungen und Prüfungsaufgaben) sowie Verbaldaten, die auf Audioaufzeichnungen aus den Arbeitstreffen basieren, ein.

In einem ersten Schritt wurde eine Ausgangsbasis zur bestehenden Prüfungspraxis der am Programm teilnehmenden Lehrenden ermittelt. Dazu wurde zunächst betrachtet, welche Ziele die Lehrenden auf Grundlage der von ihnen wahrgenommenen Herausforderungen in ihrer existierenden Prüfungspraxis für die Teilnahme am Qualitätsentwicklungsprogramm ableiteten (Forschungsfrage 1). Weiterhin wurde untersucht, inwieweit die Prüfungen, die von den Lehrenden im vorangegangenen Semester einsetzten, kompetenzorientiert angelegt waren und ob es den Lehrenden gelungen ist, diese Prüfungen so zu gestalten, dass die entsprechenden Fähigkeiten und Kompetenzen von Studierenden valide und reliabel erfasst werden konnten (Forschungsfrage 2). Die Ergebnisse zu diesen Forschungsfragen beschreiben deshalb vor allem Probleme aber auch Stärken in den bestehenden Prüfungen der Lehrenden und bildeten somit die Grundlage für die Gestaltung des Qualitätsentwicklungsprogramms, in dessen Rahmen intendiert wurde, die Lehrenden dazu anzuregen und anzuleiten, die Qualität ihrer Prüfung über ein Semester zu entwickeln. Die von den Lehrenden dabei vorgenommenen Veränderungen an ihrer Prüfungspraxis inklusive dazugehöriger Lehrziele, Prüfungen und Prüfungsaufgaben wurden in einem dritten Analyseschritt untersucht (Forschungsfrage 3).

Als inhaltlicher Bezugspunkt für die die empirischen Untersuchungen wurde ein analytischer Rahmen entwickelt, der auf der aktuellen Literatur zur Kompetenzmessung und den Gütekriterien der pädagogischen Diagnostik basiert. Aus der Synthese beider Forschungsstränge wurde ein Modell entwickelt, das den Prozess zur Erstellung und Auswertung von Prüfungen in sechs Schritte unterteilt und dabei jeweils Kriterien und Maßnahmen beschreibt, um Prüfungen kompetenzorientiert und unter Einhaltung der Gütekriterien der pädagogischen Diagnostik zu gestalten: (1) Festlegung des Prüfungsgegenstands, (2) Erstellung von Prüfungsaufgaben, (3) Rückbindung der Aufgaben an den Prüfungsgegenstand, (4) Bewertung von Prüfungsantworten, (5) Punkte- und Notenvergabe und (6) Rückschlüsse aus den Prüfungsdaten (siehe auch Abbildung 3). Die Berücksichtigung dieser sechs Schritte spielt bei der Erstellung und Auswertung von Prüfungen eine entscheidende Rolle, um die Prüfungsqualität zu sichern. Für das Qualitätsentwick-

lungsprogramm stellte dieses Modell deshalb den Arbeitsrahmen dar. Dadurch konnte mit den Lehrenden prozessbezogen an den - aus theoretischer Sicht notwendigen - Schritten bei der Erstellung und Auswertung von Prüfungen gearbeitet werden.

Bezüglich Teilschritt 1, der Festlegung des Prüfungsgegenstands, gilt es festzulegen, welche Wissens- und Könnensaspekte mit einer Prüfung abgerufen werden sollen. Durch eine theoriegeleitete Vorgehensweise wird erarbeitet und in Lehrzielen expliziert, welche Inhalte und Anforderungen in Aufgaben angelegt werden müssen (Brown & Wilson, 2011; Klauer & Leutner, 2012). Die Lehrziele bilden damit einen Ausgangs- und Referenzpunkt für die Erstellung von Aufgaben, die Entwicklung von Bewertungsrastern und die Punkte- und Notenvergabe (Pellegrino et al., 2001; Shavelson et al., 2003; Wilson, 2005).

Im zweiten Schritt werden die in den Lehrzielen beschriebenen Anforderungen in konkrete Aufgaben übersetzt. Mit den Aufgaben wird das Handlungspotenzial von Studierenden erfasst (Rost, 2004). Bei der Aufgabenerstellung ist darauf zu achten, dass die in den Lehrzielen beschriebenen Anforderungen valide abgebildet werden (Rost, 2004; Wilson, 2008). Außerdem müssen Gestaltungsaspekte in den Aufgaben vermieden werden, die den Studierenden eine Bearbeitung ungewollt erschweren und dadurch die Reliabilität einschränken können (Shavelson & Webb, 1991).

Nach Erstellung der Aufgaben muss im dritten Teilschritt überprüft werden, ob die Aufgaben zum einen bearbeitbar sind und zum anderen die Inhalte und Anforderungen abbilden, die in den Lehrzielen beschrieben sind. Mit einer Rückbindung der Aufgaben an den Prüfungsgegenstand (Schritt 3) kann so die Validität und Reliabilität der Prüfung erhöht werden, indem abgesichert wird, dass die Aufgaben tatsächlich die intendierten Inhalte und Anforderungen ansprechen (Ayala et al., 2002).

Im vierten Schritt werden die Prüfungsantworten der Studierenden bewertet und einer Bewertungskategorie (z.B. richtig oder falsch) zugeordnet. Zur Sicherung der Validität und Reliabilität bei der Bewertung von Prüfungsantworten müssen Maßnahmen getroffen werden, wie die Verwendung von Bewertungsrastern, die Bewertungskategorien, Regelwerke, Ankerbeispiele und Anweisungen zur Vergabe von Punkten enthalten. Schließlich muss diesbezüglich eine entsprechende Schulung von Korrektoren erfolgen (Graham et al., 2012).

Im fünften Schritt wird festgelegt, wie viele Punkte jeweils auf eine Aufgabe vergeben werden soll und nach welchem Kriterium Noten gebildet werden. Mit der Punktevergabe können Aufgaben zueinander gewichtet werden. Diese Gewichtung muss sich an den Lehrzielen orientieren, damit die Prüfungsergebnisse Aufschluss über das Erreichen der Lehrziele geben können. Die Festlegung der Notenstufen muss ebenfalls an den Lehrzielen ausgerichtet sein (kriteriumsorien-

tiert), damit die Note ausdrückt, zu welchem Grad das Lehrziel beziehungsweise die durch die Lehrziele beschriebenen Kompetenzen, erreicht wurden (Klauer & Leutner, 2012). Beim Anlegen einer sozialen Bezugsnorm drücken die Noten eine Rangfolge zwischen Studierenden aus. Die Erreichung der Lehrziele wird bei einer bezugsgruppenorientierten Notenvergabe nicht wiederspiegelt (ebd.).

Aus den mit der Prüfung erhaltenen Daten können die Lehrenden im sechsten Schritt, über den Zweck der Notenbildung hinaus, umfassende Informationen zu den Prüfungsaufgaben und der inhaltlichen Passung der Prüfungsanforderungen erhalten. Fallen bestimmte Aufgaben beispielsweise unerwartet einfach oder schwierig aus, kann dies ein Hinweis darauf sein, dass mit dieser Aufgabe die fraglichen Anforderungen nicht valide erfasst werden (Wilson, 2005). Zur Bestimmung von entsprechenden statistischen Kennwerten müssen die Prüfungsdaten so aufbereitet werden, dass in einem Datensatz jeweils für jeden einzelnen Studierenden erfasst wird, wie viele Punkte er oder sie pro Aufgabe erreichte.

Im Folgenden Abschnitt werden die zentralen Ergebnisse der Arbeit zusammengefasst und vor dem Hintergrund der theoretischen Annahmen und der Forschungsfragen diskutiert. Die Diskussion gliedert sich entlang der drei Forschungsfragen, die dieser Arbeit zugrunde liegen. Zu jeder Forschungsfrage werden jeweils die Befunde zu den sechs Teilschritten diskutiert und aufeinander bezogen, um herauszustellen, bezüglich welcher Schritte Probleme in der Prüfungspraxis vorlagen und an welchen Stellen Veränderungen von den Lehrenden vorgenommen wurden

6.1 Diskussion zentraler Befunde

6.1.1 Diskussion der Befunde zu den Zielsetzungen der Teilnehmenden für das Qualitätsentwicklungsprogramm

Bei Betrachtung der von den Lehrenden formulierten Programmziele stellte sich heraus, dass die Lehrenden vor allem bezüglich der Teilschritte Veränderungen in ihrer Prüfungspraxis anstreben, bei denen Beschwerden und Probleme von Seiten der Studierenden erwartet wurden. Genannt wurden hier in erster Linie Beschwerden von Studierenden bei der Prüfungseinsicht (vgl. Zitat 40, S. 283, Zitat 43, S. 284, Zitat 45, S. 285, Zitat 47, S. 285). Entsprechend fallen die Zielsetzungen zur Bewertung von Prüfungsantworten aus (siehe Abbildung 40). Der Schwerpunkt lag dabei vor allem darauf, die Bewertung von Prüfungsantworten (Teilschritt 4) nachvollziehbar und objektiv zu gestalten, um die Beschwerden von Studierenden zu reduzieren. Die Lehrenden wollten diesbezüglich Möglichkeiten kennen lernen, um Antworten von Studierenden differenziert bewerten zu können, die zwar sinngemäß richtig sind, jedoch keine zutreffende Fachsprache und Termino-

logie beinhalten. Ein anderes Problem, das zu Beschwerden von Studierenden führen kann, sind unklare Aufgabenstellungen, welche dem Teilschritt 2 des Rahmenmodells, der Erstellung von Prüfungsaufgaben, zugeordnet werden können. Hierzu wurde das Ziel formuliert, Aufgabenanweisungen klar und eindeutig zu formulieren, sodass alle Studierenden abgeholt werden. Die genannten Zielsetzungen, zu denen die Lehrenden von Beschwerden auf Seiten der Studierenden berichteten, wurden relativ konkret formuliert, während in den restlichen Zielen eher unkonkret der Wunsch nach möglichen inhaltlichen Veränderungen angesprochen wurde. Johannes und Seidel (2010) berichten ebenfalls davon, dass Teilnehmende an einer hochschuldidaktischen Fortbildungsmaßnahme vorwiegend unspezifische Angaben hinsichtlich der Zielsetzungen formulierten, ohne dabei konkrete Handlungsbezüge zu nennen. Sie führen dies darauf zurück, dass sich die Lehrenden hinsichtlich bestimmter lehrbezogener Themen weder auf existierende Handlungsmuster, noch auf ein expliziertes theoriebasiertes Wissen beziehen können, um diesbezüglich Zielsetzungen oder Probleme zu beschreiben (Johannes & Seidel, 2010).

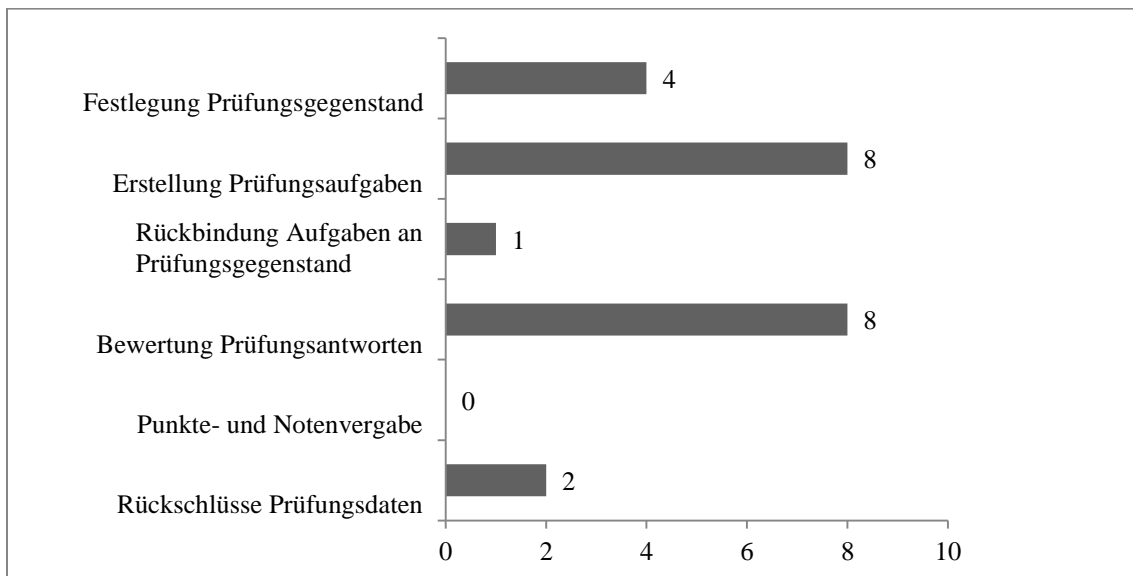


Abbildung 40: Verteilung der Programmziele auf die sechs Schritte des Prüfungsmodells

Die Erkenntnisse zu Forschungsfrage 1 und damit den Programmzielen deuten demnach darauf hin, dass die Lehrenden ein Problembewusstsein vor allem hinsichtlich der Aspekte im Prozess der Prüfungserstellung und -auswertung entwickelten, bei denen durch Beschwerden von Studierenden ein Veränderungsdruck entsteht, dem sie mit der Durchführung von gezielten Maßnahmen begegnen wollen. Zu diesen erwarteten Problemen, die vorrangig im Bereich der Bewertung von Prüfungsantworten gesehen wurden, formulierten die Lehrenden spezifische Ziele. Die restlichen Zielsetzungen waren dagegen eher allgemein formuliert, mit einem generellen Wunsch nach Verbesserungen. Eine Begründung der Zielsetzungen unter Bezugnahme auf Konzepte zur Kompetenzorientierung oder den Gütekriterien der pädagogischen Diagnostik wurde nicht vorgenommen. Diese Aspekte scheinen in der Prüfungspraxis der Lehrenden bislang eine untergeordnete

Rolle zu spielen. Ein theoriebasiertes und systematisches Vorgehen bei der Erstellung und Auswertung von Prüfungen scheint daher eher nicht zu existieren. Studien zur Lehrerprofessionalität legen nahe, dass Lehrerhandeln in weiten Teilen auf praktischem Wissen und Können basiert, in Kontexten gebunden und auf spezifische Problemstellungen fokussiert ist (Baumert & Kunter, 2006). Um Lehrende dazu anzuregen, ihr Handeln an expliziten Standards auszurichten, muss an deren impliziten Überzeugungen und Annahmen gearbeitet werden, indem ihnen daraus resultierende Probleme aufgezeigt werden (Kember, 1997). Solche Transformationsprozesse zur Veränderungen von lehrbezogenen Überzeugungen vollziehen sich in der Regel jedoch nur über längere Zeiträume (ebd.).

6.1.2 Diskussion der Befunde zur bestehenden Prüfungspraxis

Neben den Erkenntnissen darüber, was sich die Lehrenden vom Qualitätsentwicklungsprogramm erwarten, war es für die inhaltliche Ausrichtung des Programms des Weiteren relevant zu erfahren, zu welchen Teilschritten in der bisherigen Praxis Probleme bestanden. Die Ergebnisse zur bestehenden Prüfungspraxis beruhen auf einer standardisierten Analyse von Lehrzielen und Prüfungen, die im Semester vor Beginn des Qualitätsentwicklungsprogramms von den Lehrenden eingesetzt wurden, sowie auf der Auswertung von Verbaldaten. Insgesamt betrachtet zeigte sich, dass bei allen Fällen insbesondere die Aspekte, die von den Lehrenden in den Programmzielen verstärkt angesprochen wurden, bereits zufriedenstellend umgesetzt wurden. Die Lehrenden nahmen in der bestehenden Praxis bereits relativ umfangreiche Maßnahmen im Hinblick auf die Bewertung von Prüfungsantworten (Teilschritt 4) vor. Auch die Instruktionklarheit in den Aufgaben (im Rahmen von Teilschritt 2), etwa durch die Verwendung einer eindeutigen Sprache oder der Angabe von klaren Erwartungshorizonten, befanden sich auf einem hohen Niveau. Damit verstärkt sich der Eindruck aus der Betrachtung der Programmziele, dass die von den Lehrenden erwarteten Probleme und Beschwerden durch Studierende ausschlaggebend dafür sind, an welchen Stellen sie im Prozess der Prüfungserstellung und -auswertung verstärkt Aufwand investieren.

Bezüglich der anderen Teilschritte wurden, mit Ausnahme von Fall MW 1, kaum Maßnahmen bezüglich der Umsetzung einer Kompetenzorientierung und zur Sicherung der Gütekriterien der pädagogischen Diagnostik, umgesetzt. Bereits der erste Schritt bei der Prüfungserstellung, die Festlegung des Prüfungsgegenstands, wurde in der bestehenden Praxis bei keinem der Fälle durchgeführt. Die Prüfungsinhalte und -anforderungen wurden stattdessen von den Lehrenden während der Aufgabenerstellung implizit bestimmt, indem sie während der Sichtung von Veranstaltungsmaterialien Ideen für Aufgaben entwickelten oder bereits bestehende Aufgaben abwandeln. Die Aufgabenanforderungen entstanden damit ohne konkrete Bezugnahme auf ein klar definiertes Kriterium. Zu den Veranstaltungen existierten zwar Lehrziele, jedoch waren sie den

Lehrenden nicht bekannt und wurden dementsprechend nicht bei der Prüfungserstellung genutzt. Dieser Befund deckt sich zwar einerseits mit Untersuchungen aus dem Schul- sowie aus dem Hochschulbereich, wonach Lehrende so gut wie nie Lehrziele explizieren (Johannes & Seidel, 2012; Seidel, 2003); andererseits sind diese Ergebnisse vor dem Hintergrund der Modularisierung von Studiengängen und den damit verbundenen Bemühungen, eine an Lehrzielen ausgerichtete Lehr- und Prüfungspraxis an Hochschulen zu etablieren⁵², dennoch unerwartet. Zwar vermelden die Hochschulen, dass die formale Adaption der Lehrzielformulierungen so gut wie abgeschlossen sei (Wissenschaftsrat, 2012). Dennoch sollte vor dem Hintergrund der hier dargestellten Erkenntnisse der Frage weiter nachgegangen werden, inwieweit sich die angestrebte Lehrzielorientierung tatsächlich auf der Handlungsebene von Hochschullehrenden niederschlägt. Befunde aus der Organisationsforschung legen nahe, dass solche Reformvorgaben von den betreffenden Institutionen in der Regel formal adaptiert werden, um den Anforderungen einer externen Rechenschaftslegung gerecht zu werden (Coburn, 2004). Tatsächliche Veränderung im Bewusstsein und Handeln der beteiligten Akteure vollziehen sich dadurch in der Regel nicht eigenständig, sondern müssen durch unterstützende Maßnahmen, wie zum Beispiel durch Fort- und Weiterbildungen, initiiert, angeleitet und unterstützt werden (ebd.).

Ohne eine theoriegeleitete Festlegung des Prüfungsgegenstands fehlte den Lehrenden damit bislang eine wesentliche Grundlage, um im weiteren Prozess der Prüfungserstellung und Auswertung die Validität der Prüfung zu gewährleisten, da kein inhaltlicher Bezugspunkt für die Erstellung von Aufgaben, das Anlegen von Bewertungsrastern und die Punkte- und Notenvergabe gegeben war. In der Folge unterschieden sich die in den Lehrzielen beschriebenen Anforderungen zum Teil erheblich von den Veranstaltungszielen, die von den Lehrenden beschrieben wurden. Beispielsweise stimmten bei den Fällen MW 2 und MW 3 die hohen Anteile an konzeptionellem Wissen in den Lehrzielen nicht mit der von den Lehrenden beschriebenen wissensorientierten Ausrichtung mit einem Fokus auf Detailwissen überein. Die Lehrziele waren zudem auf einem zu abstrakten Niveau formuliert, um noch konkrete Hinweise auf die Erstellung von Aufgaben oder die Erstellung von Bewertungsrastern geben zu können. Im Rahmen des Qualitätsentwicklungsprogramms sollte den Lehrenden daher die Notwendigkeit zur Festlegung des Prüfungsgegenstands und zur Formulierung von Lehrzielen verdeutlicht werden.

Das Fehlen eines inhaltlichen Bezugspunkts in Form von Lehrzielen machte sich auch auf Ebene der Aufgaben (Teilschritt 2) bemerkbar. Die Lehrenden konnten zwar darlegen, zu welchen Veranstaltungskapiteln Aufgaben entwickelt wurden, welche inhaltlichen Anforderungen in den Auf-

⁵² In der Hochschuldidaktik wird in diesem Zusammenhang von dem Prinzip des *Constructive Alignment* gesprochen (Braun et al., 2014)

gaben abgebildet waren, konnte jedoch nicht gezielt erklärt werden. Die Analysen in dieser Arbeit zeigten, dass sich bei den Fällen MW 2, MW 3, MW 4 und MW 5 die in den Lehrzielen beschriebenen Anforderungen deutlich von den in den Prüfungsaufgaben umgesetzten Anforderungen unterschieden. Besonders hoch war hierbei die Diskrepanz im Bereich konzeptionellen Wissens. Dies zeigte sich darin, dass kaum Aufgaben zu konzeptionellem Wissen in den Prüfungen vorkamen, wohingegen bei 75 Prozent der Lehrziele konzeptionelles Wissen angesprochen wurde. Den Studierenden wurden somit bisher Fähigkeiten im Bereich konzeptionellen Wissens zertifiziert, die in dieser Form nicht geprüft wurden. Beim deklarativen Wissen lag der Schwerpunkt umgekehrt. Die entsprechenden Prüfungen waren somit in diesem Sinne nicht lehrzielvalide. Die Beobachtung deuten vielmehr darauf hin, dass mit dem bisherigen Vorgehen bei der Aufgabenerstellung verstärkt Aufgaben zu deklarativem Wissen entstanden und die Lehrenden andererseits Probleme zu haben schienen, weiterführende Anforderungen im Sinne von konzeptionellem Wissen in Aufgaben abzubilden. Diese Ergebnisse decken sich soweit mit den Befunden von Stefanica (2013), die in Abschnitt 2.2.1 beschrieben wurden. Sie stellte bei einer Untersuchung von Lehrzielen und Aufgaben ebenfalls fest, dass die in Lehrzielen beschriebenen Anforderungsniveaus oftmals deutlich höher angesetzt sind, als die in Prüfungsaufgaben realisierten Anspruchsniveaus. Dieses Ergebnis verstärkt den oben geschilderten Eindruck, dass die Bemühungen, die Lehrpraxis im Zuge der Modularisierung an Lehrzielen auszurichten, auf der Handlungsebene von den Lehrenden bisher nicht entsprechend umgesetzt wurde - zumindest wenn es um Prüfungen geht. Aus dieser empirischen Befundlage lässt sich nochmals die Relevanz von Maßnahmen zur Qualitätsentwicklungsarbeit ableiten, um Lehrende bei der Umsetzung von extern herangetragenen Anforderungen (lehrzielorientierte Lehr- und Prüfungspraxis) zu unterstützen (Coburn, 2004).

Bei Fall MW 1 zeigte sich hierzu ein leicht anderes Bild. Die Anzahl der Lehrziele zu konzeptionellem Wissen war in der bestehenden Praxis mit 50 Prozent geringer als bei den anderen Fällen. Auf der anderen Seite gelang es den Lehrenden bei Fall MW 1 deutlich mehr Aufgaben zu entwickeln, die auf konzeptionelles Wissen abzielen, als den Lehrenden der anderen Fälle (siehe Abbildung 25). Die Lehrenden von Fall MW 1 erklärten diesbezüglich, dass mit der Veranstaltung praxisorientiertes Wissen mit einem ganzheitlichen Verständnis vermittelt werden soll (siehe Abschnitt 5.1.1). Diese Zielsetzung konnte, gemessen am hohen Anteil von Aufgaben zu konzeptionellem Wissen, in der bisherigen Prüfung schon sehr gut umgesetzt werden, obwohl sich das Vorgehen bei der Aufgabenerstellung nicht grundlegend von den anderen Fällen unterschied. Möglicherweise bezogen sich die Lehrenden bei der Aufgabenerstellung implizit auf diese Vorgabe bezüglich Praxisorientierung und ganzheitlichem Verständnis und achteten darauf die Prüfungsaufgaben verständnisorientiert anzulegen.

Relevante Merkmale einer Kompetenzorientierung waren in den Aufgaben dagegen grundsätzlich nicht oder nur ansatzweise angelegt. Insbesondere die starke Homogenität der Prüfungsanforderungen war hierbei auffällig. Gerade anspruchsvollere Anforderungen im Bereich konzeptionellen Wissens wurden, mit Ausnahme der Prüfung von Fall MW 1, kaum umgesetzt (strategisches Wissen kam in keiner Prüfung vor). Kontextbezüge, die als Teil der Problemstellung fungieren, wurden lediglich bei 5,3 Prozent der Aufgaben verwendet. Auch hier ist der Anteil bei Fall MW 1 mit 16,7 Prozent höher als bei den anderen Fällen. Die Aufgabenoffenheiten wurden ebenfalls kaum variiert. 90,5 Prozent der Aufgaben waren definiert und konvergent, und hatten damit einen klar definierten Ausgangszustand mit nur einem richtigen Lösungsweg. Bei Fall MW 1 wurden immerhin bei 33,3 Prozent der Aufgaben divergente Lösungsmöglichkeiten eröffnet, bei denen die Studierenden die Problemstellung eigenständig strukturieren mussten. Die sprachlogische Komplexität war bei 84 Prozent der Aufgaben als einfach eingestuft worden. Das Finden, Filtern und Nutzen von fachbezogenem Wissen zur Lösung von Problemstellungen, wie es im Qualifikationsrahmen für deutsche Hochschulabschlüsse vorgesehen ist (Kultusministerkonferenz, 2005), wurde daher bei den betrachteten Fällen kaum geprüft. Im Qualitätsentwicklungsprogramm wurden daher die Merkmale von Kompetenzorientierung vermittelt und Möglichkeiten zu deren Umsetzung in Aufgaben aufgezeigt. Gemessen an der beschriebenen Aufgabenkultur, bleiben die in den Prüfungen umgesetzten Anforderungen hinter den Zielsetzungen, die mit akademischer Lehre in Verbindung gebracht werden, deutlich zurück. Demnach sollte eine verständnisorientierte Verarbeitung von Lerninhalten auf einem hohen kognitiven Niveau und deren breite Anwendbarkeit auf vielfältigen Situationen im Vordergrund stehen (Braun et al., 2014).

Stefanica (2013) kam zu ähnlichen Ergebnissen. Sie untersuchte von Lehrenden erstellte Prüfungsaufgaben zur höheren Mathematik an drei Hochschulstandorten. Sie konnte bei Aufgaben zur Mathematik - und trotz des Einsatzes eines auf die Mathematik bezogenen Kategoriensystems - ebenfalls eine homogene Verteilung von Anforderungen in den betrachteten Aufgaben feststellen. Der Schwerpunkt der Aufgaben lag jedoch auf prozeduralem Wissen (anstatt deklarativem Wissen, wie in dieser Arbeit), insofern der Großteil der Aufgaben von den Studierenden die Anwendung von mehrschrittigen mathematischen Prozeduren zum Umgang mit Variablen, Termen, Gleichungen und Funktionen, oder die Durchführung von mehrschrittigen Modellierungen erforderte (über 80 %⁵³). Aufgaben, in denen Begründungen oder ein mehrstufiges oder komplexes Argumentieren und die Bewertung verschiedener Argumente verlangt waren, wurden eher selten verwendet (unter 10 %). Außermathematische Kontextbezüge kamen ebenso wie grafische Lö-

⁵³ In der Studie wurden jeweils nur die Werte für die einzelnen Standorte gesondert voneinander aufgeführt. Ein Gesamtwert für die Aufgaben, die in diese Kategorie fallen, wurde nicht angegeben. Jedoch lag der Anteil an entsprechenden Standorten bei allen drei Standorten über 80 Prozent.

sungen nur bei einzelnen Aufgaben vor. Eine ähnliche Befundlage zu Aufgabencharakteristika ist auch aus dem Schulkontext bekannt. Jordan et al. (2008) berichten für den Bereich der Schulmathematik, basierend auf einer Analyse 45.000 Aufgaben, die von Lehrkräften erstellt wurden, ebenfalls von einer sehr homogenen Aufgabekultur. Die analysierten Aufgaben wiesen durchgängig (und schulartübergreifend) ein niedriges kognitives Aktivierungspotenzial auf und nutzten wenig anspruchsvolle Aufgabentexte ohne dabei außermathematische Bezüge herzustellen. Mathematische Darstellungsformen wurden kaum variiert.

Da die Lehrenden keine Lehrziele benutzten, wurde auch keine an den Lehrzielen ausgerichtete inhaltliche Überprüfung der Aufgabenanforderung durchgeführt (Teilschritt 3). Ohne Lehrziele fehlte dafür auch ein inhaltlicher Referenzpunkt, um eine inhaltliche und kognitive Validierung von Aufgaben vorzunehmen. Bei den Fällen MW 2 und MW 5 ließen die Lehrenden die Prüfung von Kolleginnen und Kollegen probeweise bearbeiten. Dadurch konnten Probleme hinsichtlich der Bearbeitbarkeit, nicht eindeutigen Aufgabenstellungen, oder fehlender Angaben aufdeckt werden. Außerdem gaben die Lehrenden an, dass bei der probeweisen Bearbeitung der Prüfungen das Anforderungsniveau durch die Kolleginnen und Kollegen eingeschätzt wurde. Von den Lehrenden der anderen Fälle wurden diesbezüglich keine Maßnahmen berichtet. Eine umfassende und systematisch-inhaltsbezogene Analyse der Aufgabenpotenziale kann mit diesem Vorgehen ohne explizierte Zielvorgaben, die in Lehrzielen festgehalten sind, nicht erreicht werden. Im Rahmen des Qualitätsentwicklungsprogramms wurde daher zunächst versucht, ein Problembewusstsein bei den Lehrenden bezüglich der Aufgabenvalidität zu vermitteln. Außerdem wurden Verfahren aufgezeigt, um die inhaltliche Passung von Aufgaben und Lehrzielen zu überprüfen.

Wie eingangs berichtet, wurden von den Lehrenden bei allen Fällen verschiedene Maßnahmen durchgeführt, um die Validität und Reliabilität bei der Bewertung von Prüfungsantworten (Teilschritt 4) zu sichern. Die Lehrenden gaben zu fast allen Aufgaben (95 %) Ankerbeispiele in Form von Musterlösungen an. Bei knapp 80 Prozent der Aufgaben wurde anhand der Musterlösung angegeben, auf welche Aspekte in den Antworten Punkte zu vergeben sind. Regelwerke, mit denen beschrieben wird, welche Merkmale eine Antwort aufweisen muss, um einer Bewertungskategorie (z.B. richtig oder falsch) zugeordnet werden zu können, wurden von den Lehrenden dagegen kaum angefertigt.

Neben der Angabe von Ankerbeispielen und Hinweisen zur Vergabe von Punkten führten die Lehrenden jedoch noch weitere Maßnahmen durch, um die Validität und Reliabilität bei der Bewertung von Prüfungsantworten sicherzustellen. Bei allen Fällen wurden mehrere Korrektoren eingesetzt. Jeder Korrektor korrigiert jeweils nur eine Prüfungsaufgabe für alle Studierenden. Dadurch sollte die Konsistenz der Bewertung gesichert werden. Bei den Fällen MW 2, MW 3 und

MW 4 fand außerdem eine Schulung von Korrektoren statt, bei der sie mit den Aufgaben und Musterlösungen vertraut gemacht wurden. Wie Untersuchungen zeigen, kann die Validität und Reliabilität bei der Bewertung von Prüfungsantworten durch dieses Vorgehen deutlich erhöht werden (Gorman & Rentsch, 2009).

Außerdem bestand bei drei Fällen (MW 1, MW3 und MW 4) die Möglichkeit, dass sich die Korrektoren bei unklaren Antworten an den Aufgabenersteller beziehungsweise die Aufgabenerstellerin wenden. Somit wurden bereits umfassende Maßnahmen getroffen, um die Qualität bei der Korrektur sicherzustellen. Im Rahmen des Qualitätsentwicklungsprogramms wurde deshalb zusätzlich auf die Anfertigung von Regelwerken eingegangen, die bisher noch nicht angefertigt worden waren. Möglicherweise, so die Überlegung, könnte dadurch dem von den Lehrenden geschilderten Problem begegnet werden, bei offen gestellten Fragen in der Bewertung nicht zwischen substantiellen Antworten und solchen Antworten, die zwar sinngemäß richtig sind, jedoch bestimmte Merkmale nicht aufweisen, differenzieren können. Studien zeigen außerdem, dass Korrektoren durch die Benutzung von Regelwerken, die klar formulierte Deskriptoren zur Bewertung von Prüfungsantworten enthalten, die Validität und Reliabilität der Bewertung deutlich erhöhen können (Graham et al., 2012).

Die Punktevergabe hatte sich bei allen Fällen nach der Bearbeitungsdauer, die Studierende laut Schätzung der Lehrenden zur Lösung einer Aufgabe benötigen, gerichtet. Pro Minute Bearbeitungsdauer wurde ein Punkt vergeben. Aspekte, wie die Relevanz von Anforderungen, eine Gewichtung nach den Lehrzielen oder den erwarteten Aufgabenschwierigkeiten, wurden bei der Punktevergabe nicht zu Grunde gelegt. Die Annahme der Lehrenden hierbei war es, dass die Aufgabenschwierigkeit mit der Bearbeitungsdauer korreliert. Die schwierigeren Anforderungen wurden von den Lehrenden dabei auch als relevanter bezeichnet. Die Studierenden erhalten laut Aussagen der Teilnehmenden außerdem einen Anhaltspunkt, wie viel Zeit sie für eine Aufgabe aufwenden können. Mit der Vergabe von Punkten nach Zeit ist zwar ein klares und nachvollziehbares Kriterium gesetzt. Besteht der Zusammenhang zwischen Schwierigkeit und Bearbeitungsdauer jedoch nicht, werden durch dieses Vorgehen aber möglicherweise periphere Anforderungen stärker gewichtet als tatsächlich relevante Anforderungen. Die auf Grundlage der Punkte gebildeten Noten würden in diesem Fall keine lehrzielvaliden Unterschiede im Leistungsniveau von Studierenden widerspiegeln (Von Davier, 2012). Die Vorgaben zur Gestaltung von Studiengängen im Zuge der Modularisierung schreiben auf Ebene von Modulen ebenfalls eine Vergabe von Leistungspunkten gemessen am geschätzten Arbeitsaufwand, der für die Ableistung eines Moduls aufgewendet werden muss, vor (Brändle, 2010). Diese Vorgabe wäre mit dem gleichem Argument zu hinterfragen wie die Punktevergabe nach Zeit auf Prüfungsebene, insofern Arbeitsaufwand nicht notwendigerweise mit inhaltlichem Anspruch korreliert.

Die Bildung von Noten erfolgte vorrangig bezugsgruppenorientiert. Eine an Lehrzielen ausgerichtete (kriteriumsorientierte) Festlegung von Notenstufen fand nicht statt. Das entscheidende Kriterium bei der Notenvergabe war für die Lehrenden das Erreichen einer Normalverteilung der Noten. Außerdem strebten die Lehrenden an, dass die Durchschnittsnoten über die Studierendenkohorten hinweg in etwa gleich bleiben. Die Erfüllung dieser beiden Aspekte wurde von den Lehrenden als Indikator für eine „gute Prüfungsqualität“ betrachtet. Neben der Anpassung der Notenverteilung und des Notendurchschnitt versuchten die Lehrenden zu vermeiden, dass sich größere Gruppen von Studierenden knapp unter der Punktegrenze zur nächstbesseren Note befinden, um Beschwerden von Studierenden vorzubeugen. Um diese drei Aspekte zu erreichen, wurde an der zu erreichenden Gesamtpunktzahl (indem nicht alle Aufgaben, die in der Prüfung enthalten waren, in die Bewertung mit eingehen), dem Herabsetzen der Bestehensgrenze oder der Veränderung der Spannweite von Notenstufen, variiert.

Müller-Benedict und Tsarouha (2011) stellten bei der Analyse von Hochschulexamensnoten der letzten 15 Jahre fest, dass zwar Unterschiede in den Notendurchschnitten zwischen Fachdisziplinen vorliegen. Innerhalb einer Disziplin sind die Notenschnitte über den Zeitverlauf relativ konstant. Die Autoren verweisen darauf, dass bislang nicht erforscht wurde, mit welchen Mechanismen innerhalb einer Fachdisziplin ein stabiles Notenniveau hergestellt wird. Die in dieser Arbeit beschriebene Ausrichtung an einer Normalverteilung und dem Notendurchschnitt der vorangegangenen Jahre könnte hierfür eine Erklärung liefern. Nach welchen Überlegungen festgelegt wird, welcher Notenschnitt als optimal angesehen wird, konnte jedoch nicht ermittelt werden. Zu hinterfragen ist in diesem Zusammenhang auch die starke Ausrichtung der Lehrenden an einer Normalverteilung der Prüfungsnoten als Indikator für die Prüfungsqualität. Die Bildung von Einzelnoten hängt in diesem Fall vom Leistungsvermögen der aktuellen Prüfungskohorte ab. Die Noten drücken dabei nicht aus, inwieweit die in den Lehrzielen beschriebenen Anforderungen und Kompetenzen erreicht wurden (Webler, 2010). Vor diesem Hintergrund ist auch die Empfehlung der Europäischen Kommission (2009) in ihrem ECTS-Leitfaden zu hinterfragen, wonach bei Prüfungen eine Normalverteilung von Noten anzustreben sei.

Bei allen Fällen wurden Datensätze erstellt, in denen für jeden Studierenden angegeben ist, wie viele Punkte pro Aufgabe erreicht wurden. Damit bestand bereits eine Datengrundlage, um die Prüfungsdaten statistisch auszuwerten und damit Teilschritt 6 des Prüfungsmodells durchzuführen. Die Daten wurden dabei jedoch mit unterschiedlicher Granularität erfasst. Bei Fall MW 5 lagen die Daten beispielsweise nur auf Ebene von Aufgabenblöcken vor, die zum Teil bis zu zehn Teilaufgaben enthielten. Eine Auswertung des Abschneidens bei einzelnen Aufgaben kann mit dieser Form von Daten nicht vorgenommen werden. Die Prüfungsdaten wurden von den Lehrenden bisher vorwiegend dazu benutzt, um die Notenverteilungen zu bestimmen, Studierendengrup-

pen zu identifizieren, die sich knapp unter der nächstbesseren Notenstufe befinden, Aufgaben zu identifizieren, die nicht bearbeitet wurden, und um Noten automatisiert zu bilden. Hinweise auf eine weiterführende Betrachtung von statistischen Kennwerten wurden nicht gefunden. Das Informationspotenzial, das in den Prüfungsdaten steckt, wurde demnach bisher nur ansatzweise ausgeschöpft. Die Daten könnten aber genutzt werden, um die Qualität einzelner Aufgaben zu überprüfen oder Hinweise auf die Passung der gestellten Anforderungen (sowohl inhaltlich als auch bezogen auf das Anforderungsniveau) zu gewinnen (Wilson, 2005). Die Studierenden könnten ebenfalls von einer differenzierten Rückmeldung über das Abschneiden bei einzelnen Aufgaben profitieren, wie verschiedene Studien zeigen (Krause, Stark, & Mandl, 2004; Shute, 2008). Die Ergebnisse von Prenzel, Schindler, et al. (2012) zur Prüfungspraxis an der Technischen Universität München deuten darauf hin, dass Studierende über die Note hinaus kaum Feedback über das Abschneiden bei Prüfungen erhalten, wodurch sich über die Prüfung hinaus kaum nachhaltige Lerneffekte einstellen (Epstein et al., 2002).

Auf Grundlage dieser Ergebnisse kann der Eindruck aus der Betrachtung der Programmziele bestätigt werden: Bei der Prüfungserstellung und -auswertung liegt der Fokus der Lehrenden vorwiegend auf Aspekten, bei denen erfahrungsgemäß am häufigsten Beschwerden von Studierenden auftreten. Die genannten Probleme waren dabei über die Fälle hinweg vergleichbar, da sich das Vorgehen der Lehrenden, trotz zum Teil erheblichen Unterschieden hinsichtlich der Prüfungserfahrung sowie in den personellen Konstellationen, in der bestehenden Praxis kaum voneinander unterschieden hatte. Johannes und Seidel (2010) führen diese Konsistenz im Lehrhandeln von Hochschullehrenden mitunter darauf zurück, dass sich Lehrende, gerade auch beim Einstieg in die Lehre, oftmals im Sinne eines *learning by doing* einarbeiten und sich dabei zum einen daran orientieren, was sie selbst als Studierende im Rahmen der Lehre erfahren haben und sich zum anderen an den etablierten Vorgehensweisen und Lehrkonzepten an ihrer Fakultät, beziehungsweise ihrem Lehrstuhl anlehnen (ebd.). Dadurch werden tradierte Praktiken weitergegeben, wodurch sich bestimmte Strukturen und Methoden im Lehrhandeln etablieren (Seidel & Hoppert, 2011).

Diese Erkenntnisse verdeutlichen zum einen die Notwendigkeit, die Inhalte des Qualitätsentwicklungsprogramms an der etablierten Praxis der Lehrenden und deren Wissens- und Erfahrungsstand auszurichten, um die darin bestehenden Probleme aufzugreifen (Desimone, 2009). Zum anderen sollte bei den Lehrenden ein Problembewusstsein dafür geschaffen werden, dass es mit dem etablierten Vorgehen nicht ausreichend gelingt, Merkmale von Kompetenzorientierung unter Einhaltung der zentralen Gütekriterien, umzusetzen. Bezüglich der Validität ist insbesondere die unzureichende Passung zwischen Lehrzielen und Aufgaben problematisch. Die Lehrenden sollten deshalb dazu angeregt und angeleitet werden, ihr Handeln entlang professioneller Standards auszurichten (Ostermeier et al., 2010). Das in dieser Arbeit entwickelte Prüfungsmodell bildet diese

professionellen Standards ab und wurde daher als inhaltlicher Rahmen für das Qualitätsentwicklungsprogramm verwendet. Da die empirischen Analysen zur bestehenden Prüfungspraxis ebenfalls entlang des Modells durchgeführt wurden, konnten die dabei identifizierten Probleme zusammen mit den Lehrenden aufgegriffen und unter Bezugnahme auf wissenschaftlich fundierte Kriterien überarbeitet werden. Den Lehrenden wurde dabei das entsprechende theoretische Hintergrundwissen problembezogen vermittelt, um sie in die Lage zu versetzen, ihre Prüfungspraxis vor dem Hintergrund professioneller Standards zu reflektieren. Im folgenden Abschnitt wird dargestellt, welche Veränderungen die Lehrenden entlang dieses Modells an ihrer Prüfungspraxis vornahmen und worin Barrieren für eine Übernahme bestimmter Aspekte bestanden.

6.1.3 Diskussion der Befunde zur veränderten Praxis

Aufgrund der Befundlage zur bestehenden Praxis (Forschungsfrage 2), wonach ein Mangel an explizitem und an testtheoretischen Kriterien orientiertem Wissen bei den Teilnehmenden vorlag, wurde den Lehrenden im ersten Arbeitstreffen zunächst die Grundidee des Prüfens erläutert, wonach aus dem Verhalten von Studierenden, dass sie bei der Lösung von Prüfungsaufgaben zeigen, auf deren Wissens- beziehungsweise Kompetenzstand geschlossen werden kann (Rost, 2004). Außerdem wurde das Prinzip einer Prüfung als Messinstrument sowie die Bedeutung der Gütekriterien der pädagogischen Diagnostik erörtert, da diesbezüglich in der bestehenden Praxis nur wenig Bewusstsein vorherrschte. Dies äußerte sich unter anderem darin, dass die in den Lehrzielen beschriebenen Anforderungen bisher nicht in den Aufgaben abgebildet waren. Im Folgenden wird zunächst auf Aspekte eingegangen, die von den Lehrenden im Verlauf des Qualitätsentwicklungsprogramms erfolgreich umgesetzt werden konnten. Dabei werden außerdem Unterschiede zwischen den Fällen betrachtet. Abschließend wird dargestellt, welche Aspekte von den Lehrenden nicht verändert wurden und mögliche Gründe dafür diskutiert.

Im Anschluss an den theoretischen Input zur Grundidee des Prüfens, lag der Schwerpunkt im ersten Arbeitstreffen bei den Fällen MW 1, MW 2, MW 3 und MW 4 auf der Festlegung des Prüfungsgegenstands und der Formulierung von Lehrzielen, um das vermittelte Wissen direkt auf die Materialien der Lehrenden anzuwenden. Bei Fall MW 5 erfolgte im ersten Arbeitstreffen keine gemeinsame Arbeitsphase, worauf zu einem späteren Zeitpunkt eingegangen wird. Bei den Fällen MW 1, MW 2, MW 3 und MW 4 kamen während der Überarbeitung der Lehrziele konstruktive Diskussionen sowohl zwischen den Teilnehmenden, als auch mit den Moderatoren zustande. Mit der inhaltlichen Diskussion der Lehrziele entwickelten die Lehrenden ein Problembewusstsein dahingehend, dass zum Teil kein gemeinsames Verständnis über die Bedeutung der Prüfungsanforderungen unter den Kolleginnen und Kollegen vorlag (siehe Zitat 2, S. 264, Zitat 9, S. 268, Zitat 13, S. 269, Zitat 14, S. 270, Zitat 20, S. 273). Dies zeigt, dass im ersten Treffen des Quali-

tätsentwicklungsprogramms bereits ein erster Anstoß für Veränderungen gegeben werden konnte, indem für die Lehrenden der nötige Raum für eine konstruktive und kollegiale Auseinandersetzung mit der eigenen Praxis geschaffen wurde. Die Moderatoren konnten dabei auf Probleme hinweisen und Lösungsmöglichkeiten aufzeigen. In der Professionalisierungsforschung werden die Bildung einer kollegialen Arbeits- und Lerngemeinschaft (community of practice) (Little, 2002), sowie eine wertschätzende und kritische Begleitung durch externe Moderatoren (Gröschner, Seidel, Pehmer, & Kiemer, 2014), als wichtige Maßnahme betrachtet, um Lernprozesse bei Lehrenden anzustoßen.

In der darauffolgenden Arbeitsphase setzten die Lehrenden die Arbeit an den Lehrzielen eigenständig fort, um die in den Lehrzielen genannten Anforderungen begrifflich zu klären. Den Lehrenden ist es mit der Arbeit an den Lehrzielen gelungen, ein gemeinsames Verständnis über die Bedeutung der in den Lehrzielen beschriebenen Anforderungen zu erreichen. In diesem Zusammenhang erlebten sie die Übertragung der Lehrziele in eine Lehrzielmatrix als besonders hilfreich, da dadurch Themen- und Anforderungsbereiche identifiziert werden konnten, die nicht durch die bestehenden Lehrziele abgedeckt wurden. In der Folge begannen die Lehrenden damit, ihre Prüfungsanforderungen zu modellieren, indem übergeordnete Anforderungen in Teilanforderungen zerlegt und sequenziert wurden, um verschiedene Anforderungsniveaus abzubilden. Damit wurden in der veränderten Praxis die wesentlichen Schritte zur Festlegung des Prüfungsgegenstands umgesetzt.

Durch die Überarbeitung der Lehrziele und die Rückbindung von Aufgaben an den Prüfungsgegenstand (Teilschritt 3 des Prüfungsmodells) anhand einer rationalen Aufgabenanalyse, konnten die Lehrenden eigenständig eine fehlende Passung zwischen Lehrzielen und Aufgaben feststellen. Die rationale Aufgabenanalyse war der Auslöser für die Lehrenden sowohl Lehrziele, als auch Aufgaben zu überarbeiten, beziehungsweise neu zu erstellen, um eine bessere Passung der Anforderungen zu erreichen. Die Lehrenden schätzten durch die rationale Aufgabenanalyse die in den Prüfungsaufgaben umgesetzten Anforderungen realistischer ein. Bei Fall MW 2 wurden zudem die Veranstaltungsinhalte vor dem Hintergrund der Lehrziele hinterfragt. Die Lehrziele bildeten damit für die Lehrenden den entscheidenden Ausgangspunkt, um ursächlich an Problemen zu arbeiten, die in den nachgelagerten Schritten, und speziell bei der Aufgabenerstellung, aufgetreten waren (anstatt zu versuchen isoliert Symptome bei einzelnen Teilschritten zu beheben). Als besonders unterstützend und hilfreich für den Prozess der rationalen Aufgabenanalyse und die Überarbeitung von Aufgaben und Lehrzielen wurde wiederholt der Einsatz der Lehrzielmatrix genannt, die als gemeinsamer Bezugspunkt für die Lehrenden fungierte.

Schaper (2012) weist in diesem Zusammenhang ebenfalls auf die Bedeutung konkret ausgearbeiteter Lehrziele und den Einsatz einer Lehrzielmatrix für die Umsetzung einer kompetenzorientierten Prüfungspraxis hin. Die Arbeit an Lehrzielen sollte daher bei Professionalisierungsmaßnahmen zur Entwicklung der Prüfungsqualität im Mittelpunkt stehen, indem die Lehrenden auf die Relevanz von Lehrzielen für die Erstellung und Auswertung von Prüfungen aufmerksam gemacht werden und sie diese Relevanz in der Folge auch selbst feststellen können. Dabei ist der beschriebene Perspektivenwechsel von einer operationalen Festlegung von Prüfungsinhalten und -anforderungen (implizit geleitete Erstellung von Aufgaben beim Sichten von Veranstaltungsunterlagen), wie sie in der bestehenden Praxis vorgefunden wurde, hin zu einer lehrzielorientierten Prüfungs- und Aufgabenerstellung anzubahnen. Die Durchführung einer rationalen Aufgabenanalyse durch die Lehrenden stellt dazu eine effektive Maßnahme dar, wie die dargelegten Befunde verdeutlichen.

Die Einschätzung der Lehrenden zur Passung zwischen Lehrzielen und Aufgaben deckte sich mit den Ergebnissen der in dieser Arbeit durchgeführten Analysen, bei denen ebenfalls eine Reduzierung der Diskrepanz zwischen Lehrzielen und Aufgaben in der veränderten Praxis festgestellt werden konnte (siehe Abbildung 16). Bei den neu erstellten Aufgaben konnten, insgesamt betrachtet, Merkmale einer Kompetenzorientierung etwas häufiger beobachtet werden, insofern sich bei den Wissensarten, (5,2 % mehr Aufgaben zu konzeptionellem Wissen), den Aufgabenkontexten (7,2 % mehr Aufgaben mit Kontexten als Teil der Problemstellung) und den Aufgabenoffenheiten (5 % mehr definierte und divergente Aufgaben) leichte Veränderungen ergaben. Nicht erhöht hat sich dagegen der Anteil von Aufgaben mit einer höheren sprachlogischen Komplexität. Die Lehrenden nahmen eine hohe sprachlogische Komplexität vor dem Hintergrund ihrer Lehrziele als nicht bedeutsam wahr. Jedoch wurden die Anteile an Kontextbeschreibungen reduziert, die nicht für die Lösung der Aufgabenstellung relevant waren. Dies ist als positiv zu bewerten, da dadurch eine unnötige Leselast in den Aufgaben reduziert werden konnte (Wilson, 2005).

Die Lehrenden von Fall MW 1, die bereits mit einer guten Ausgangsbasis starteten, insofern die Merkmale von Kompetenzorientierung in der bestehenden Prüfungen bereits angelegt waren, konnten die umfassendsten Verbesserungen im Hinblick auf die Kompetenzorientierung erzielen. Vermunt und Endedijk (2011) berichten in diesem Zusammenhang davon, dass Lehrkräfte mit unterschiedlichen Voraussetzungen, Bedürfnissen aber auch Aspirationen an Professionalisierungsmaßnahmen teilnehmen und deshalb unterschiedlichen Input und auf sie zugeschnittene Anweisungen benötigen. Damit wird auch nochmal bekräftigt, dass im relativ autonomen und fachlich hochspezialisiertem Hochschulkontext derartige Programme entsprechend flexibel und fallbasiert stattfinden sollten.

Bei Fall MW 1 war die Aufgabenerstellung zudem am konsequentesten an den Lehrzielen ausgerichtet, die differenziert und spezifisch ausgearbeitet waren. Außerdem wurde die Passung zwischen Aufgaben und Lehrzielen unter Kolleginnen und Kollegen diskutiert und bei Bedarf überarbeitet. Die Lehrende von Fall MW 3 betonte außerdem, dass es ihr durch die Nutzung der vorgestellten Techniken zur Gestaltung und Öffnung von Aufgaben gelungen ist, Anforderungen auf unterschiedlichen Niveaus in ihren Aufgaben anzulegen, um ihre Lehrziele entsprechend in Aufgaben abzubilden. Eine konsequente Ausrichtung der Aufgabenerstellung an differenziert ausgearbeiteten Lehrzielen, der Einsatz von verschiedenen Techniken zur Öffnung von Aufgaben, sowie eine kollegiale Diskussion und Reflexion von Aufgaben vor dem Hintergrund der Lehrziele, erscheinen daher als vielversprechende Maßnahmen, die Lehrende umsetzen können, um die Anforderungsheterogenität in ihren Prüfungen zu erhöhen. Diese Erkenntnisse bestätigen außerdem die Relevanz, im Rahmen von Professionalisierungsmaßnahmen Raum für Diskussionen in einer vertrauensvolle Atmosphäre unter den Lehrenden zu schaffen (Desimone, 2009; van Es, 2012), ihnen gleichzeitig aber auch Instrumente an die Hand zu geben, die entsprechend niederschwellig in der Praxis umsetzbar sind (Van Merriënboer & Kirschner, 2012).

Bei den Fällen MW 1, MW 2, MW 3 und MW 4 nahmen die Lehrenden zum Teil umfassende Veränderungen an ihren Prüfungen vor. Die Lehrenden von Fall MW 5 hingegen konnten anteilig nur unwesentlich mehr Aufgaben auf höheren Anforderungsniveaus erstellen, als in der bestehenden Prüfung. Dies hatte zur Folge, dass auch die Diskrepanz zwischen Aufgaben und Lehrzielen nicht verringert werden konnte. Eine mögliche Ursache dafür könnte sein, dass der Ablauf des Qualitätsentwicklungsprogramms aufgrund eines vorgezogenen Prüfungstermins verändert werden musste. Im ersten Arbeitstreffen konnten daher die Lehrziele nicht gemeinsam überarbeitet werden. Die Moderatoren konnten so auch nicht auf bestehende Probleme in den Lehrzielen hinweisen und eine Diskussion unter den Lehrenden bezüglich der in den bestehenden Lehrzielen beschriebenen Anforderungen initiieren. Die Lehrenden erhielten diesbezüglich, neben dem theoretischen Input, lediglich Handreichungen und Checklisten zur Erstellung von Lehrzielen. Anders als die Lehrenden der anderen Fälle, führten die Lehrenden bei Fall MW 5 in der Folge keine der Aufgabenerstellung vorgelagerte Festlegung des Prüfungsgegenstands durch. Die Lehrziele wurden von den Lehrenden nicht überarbeitet und auch im weiteren Verlauf des Prüfungsprozesses nicht verwendet. Dies ist möglicherweise darüber zu erklären, dass bei Fall MW 5 wesentliche Aspekte, die sich bei den anderen Fällen als wirksam herausstellten, um Veränderungen in der Praxis anzustoßen, im Rahmen der Arbeitstreffen nicht umgesetzt werden konnten.

Während die Lehrenden der Fälle MW 1, MW 2, MW 3 und MW 4 bei den ersten drei Schritten des Prüfungsmodells zum Teil umfassende Veränderungen an ihrer Praxis und den Prüfungen vornahmen, änderte sich das Vorgehen bei keinem der Fälle weder bei der Bewertung von Prü-

fungsantworten (Teilschritt 4) noch bei der Punkte- und Notenvergabe (Teilschritt 5). Im Folgenden werden mögliche Barrieren, die eine Umsetzung der vorgestellten Maßnahmen verhinderten, diskutiert. So wurde in den Arbeitstreffen zu den benannten Teilschritten von den Moderatoren in einem theoretischen Input zwar wesentliche Aspekte angesprochen - aus zeitlichen Gründen musste jedoch jeweils auf die gemeinsamen Arbeitsphasen verzichtet werden, im Rahmen derer die Moderatoren auf konkrete Probleme in den Materialien der Lehrenden hätten hinweisen können. Die Lehrenden nahmen die während des theoretischen Inputs angesprochenen Aspekte in der Folge für ihre eigenen Prüfungen nicht als problematisch wahr und setzten sie entsprechend während der Arbeitsphasen auch nicht um. Opfer, Pedder und Lavicza (2011) verweisen darauf, dass Lehrende ihr Wissen und ihre Überzeugungen durch praktische Erfahrungen erlangen. Mit Lernaktivitäten, die auf der Vermittlung theoretischen Wissens basieren, ohne dass dabei auch auf konkrete Probleme hingewiesen und Möglichkeiten zu deren Lösung aufgezeigt werden, ist es dagegen kaum möglich Handlungsweisen von Lehrenden, die sich über einen längeren Zeitraum in einem bestimmten Kontext etablierten, zu verändern (Opfer & Pedder, 2011).

Erst bei der Betrachtung statistischer Kennwerte zur Prüfung konnten den Lehrenden Schwachstellen in den Prüfungen aufgezeigt werden, die bei der Bewertung von Prüfungsantworten und der Punktevergabe entstanden waren (siehe Abschnitt 5.6.3). So wurde beispielsweise die Vergabe von Punkten nach der geschätzten Bearbeitungsdauer von den Lehrenden hinterfragt, nachdem die Moderatoren anhand der Aufgabenschwierigkeiten aufzeigen konnten, dass auf eine sehr einfache Aufgabe insgesamt ein Drittel der Punkte vergeben wurde. Ein weiteres Problem, das aus der Betrachtung der statistischen Kennwerte hervorging, bezog sich auf die Bewertung von Prüfungsantworten. Eine Aufgabe fiel unerwartet einfach aus, da aufgrund eines fehlenden Bewertungskriteriums zur Differenzierung qualitativ unterschiedlicher Antworten, beinahe alle Antworten als richtig gewertet werden mussten. Die inhaltliche Relevanz bestimmter Anforderungen wurde hinterfragt, nachdem eine von den Lehrenden als einfach eingestufte Aufgabe kaum gelöst werden konnte.

Diese Erkenntnisse bestätigen den oben dargelegten Befund, wonach über einen theoretischen Input zu testtheoretischen Kriterien und Konzepten kaum Veränderungen in der Praxis angestoßen werden konnten. Den Lehrenden musste an den eigenen Materialien aufgezeigt werden, welche Probleme darin existieren. Die statistischen Kennwerte zur Prüfung konnten dazu effektiv eingesetzt werden, da die auf den Prüfungsdaten beruhenden Informationen von den Lehrenden als belastbar und aufschlussreich wahrgenommen wurden. Borko (2004) beschreibt in diesem Zusammenhang, dass der Einsatz von Materialien oder Artefakten aus der eigenen Praxis der Lehrenden eine wirksame Komponente sein kann, um Veränderungsbedarfe aufzuzeigen. Diese Ergebnisse decken sich zudem mit dem Modell von Guskey (2002), der beschreibt, dass Lehrende

ihre Überzeugungen und ihr Handeln nur dann ändern, wenn ihnen Evidenz darüber vorgelegt wird, dass in den eigenen Materialien Probleme bestehen. Außerdem müssen die Lehrenden erleben, dass die im Rahmen von Professionalisierungsmaßnahmen aufgezeigten Maßnahmen und Instrumente zum einen erfolgreich in die Praxis implementierbar sind und zum anderen dazu führen, dass die Probleme erfolgreich gelöst werden können (Guskey, 2002). Dafür spricht auch, dass die Lehrenden insbesondere die Aspekte weiter umsetzten, die bereits im Rahmen der Arbeitstreffen an den eigenen Materialien besprochen und erarbeitet wurden.

Zusammenfassend kann festgehalten werden, dass die Lehrenden im Verlauf des Qualitätsentwicklungsprogramms zum Teil erhebliche Verbesserungen an den Prüfungen erzielen konnten. Die vorgestellten Befunde verdeutlichen die Notwendigkeit, dass das Lernen von Lehrpersonen im Rahmen von Professionalisierungsmaßnahmen auf die Probleme der täglichen Praxis bezogen sein und außerdem aktiv und über einen längeren Zeitraum erfolgen sollte, in dem sich Input- mit eigenständigen Arbeits- sowie gemeinsamen Reflexionsphasen abwechseln (Desimone, 2009). Für die Durchführung und Moderation von Arbeitstreffen bleibt demnach festzuhalten, dass die Moderatoren den Lehrenden gezielt Probleme in den Materialien der Teilnehmenden aufzeigen und diese vor dem Hintergrund des theoretischen Inputs diskutieren müssen. Außerdem sollten sie in gemeinsame Arbeitsphasen mit den Lehrenden eintreten, um den theoretischen Input direkt auf deren Materialien anzuwenden (Gröschner et al., 2014).

Mit den sechs Schritten des in dieser Arbeit eingeführten Modells zur Erstellung und Auswertung von Prüfungen konnten die Bedürfnisse beziehungsweise Zielsetzungen der Lehrenden vollständig abgedeckt werden. Die Festlegung des Prüfungsgegenstands und die Arbeit an und mit Lehrzielen bildeten für die Lehrenden den zentralen Bezugspunkt, um bestehende Probleme in den Prüfungen zu überarbeiten. Bei Professionalisierungsmaßnahmen zur Prüfungspraxis von Hochschullehrenden sollte die Arbeit an und mit Lehrzielen deshalb im Mittelpunkt stehen (siehe dazu auch Schaper, 2012). Die vorgestellten Instrumente und Techniken, wie die Lehrzielmatrix oder das Verfahren der rationalen Aufgabenanalyse, konnten von den Lehrenden effektiv umgesetzt werden und wurden als unterstützend wahrgenommen.

Ein weiterer Aspekt, der bislang nicht angesprochen wurde, sind die gemeinsamen Reflexionsphasen während der Arbeitstreffen. Die Lehrenden erhielten dabei sowohl von den Moderatoren, als auch von den Kolleginnen und Kollegen Feedback zu den vorgenommenen Veränderungen. Dabei konnten zum einen ungelöste Probleme geklärt und zum anderen entstandene Fehlvorstellungen bezüglich bestimmter Konzepte korrigiert werden. Welche Auswirkungen diese Reflexionsphasen auf die Prüfungen hatten, konnte in dieser Arbeit jedoch nicht ermittelt werden, da Zwischenprodukte nicht analysiert wurden. Basierend auf Erkenntnissen der Professionalisierungsfor-

schung ist jedoch davon auszugehen, dass solche Reflexions- und Feedbackphasen einen positiven Einfluss auf das Lernen von Lehrpersonen haben (Borko, 2004).

6.2 Methodische Reflexion und Limitationen

Nach Darstellung und Diskussion der zentralen Befunde wird die Arbeit in diesem Abschnitt aus methodischer Perspektive diskutiert und Limitationen beleuchtet. Die empirischen Erhebungen wurden im Rahmen des beschriebenen Qualitätsentwicklungsprogramms durchgeführt, in dessen Verlauf die Lehrenden mit externen Moderatoren in vier Arbeitstreffen an Lehrzielen, Prüfungen und Prüfungsaufgaben arbeiteten, um dabei gemeinsam Probleme und Schwachstellen zu identifizieren und Lösungsmöglichkeiten zu entwickeln. Gegenstand der Untersuchungen waren schriftliche Prüfungen mit hohen Studierendenzahlen, da dieses Format an Hochschulen am weitesten verbreitet ist und somit die größte Reichweite aufweist (Prenzel, Schindler, et al., 2012). Für die Lehrenden entstand mit diesem Vorgehen ein praktischer Nutzen, da im Rahmen des Programms direkt an den Lehrzielen, Prüfungen und Prüfungsaufgaben gearbeitet wurde, die zum Ende des Semesters eingesetzt werden konnten. Auf wissenschaftlicher Seite bestand der Vorteil darin, dass Veränderungen der Prüfungspraxis und deren Bedingungen in ihrer Komplexität prozessnah und damit authentisch betrachtet und analysiert werden konnten.

Eine weitere Stärke der Arbeit liegt im Zugang über zwei verschiedene Datenquellen und der systematischen Beschreibung der Ausgangsbedingungen zur Prüfungspraxis von Hochschullehrenden, bevor diese im Rahmen des Qualitätsentwicklungsprogramms aufgegriffen und verändert wurden. In die Analysen gingen sowohl materialbasierte Daten in Form von Lehrzielen, Prüfungen und Prüfungsaufgaben, als auch Verbaldaten, die auf Audioaufnahmen der Arbeitstreffen basieren, mit ein. Die bestehende Prüfungspraxis wurde auf Grundlage der Betrachtung von Programmzielen der Lehrenden (Forschungsfrage 1) und der Analyse der Lehrziele und Prüfungen, die im Semester vor Beginn des Qualitätsentwicklungsprogramms von den Lehrenden eingesetzt wurden, analysiert. Aus den Verbaldaten, die Auskunft über das bisherige Vorgehen der Lehrenden gaben, konnten dazu ergänzend prozessbezogene Informationen gewonnen werden (Forschungsfrage 2). Die Analyse der materialbasierten Daten erfolgte mit deduktiv gebildeten Kategoriensystemen (Mayring, 2000), deren Grundlage aktuelle Standards zu den Gütekriterien der pädagogischen Diagnostik (Abschnitt 2.1.2) und der Kompetenzmessung (Abschnitt 2.1.1) bilden, um damit standardisierte Erkenntnisse zur Prüfungsqualität und Kompetenzorientierung in den betreffenden Prüfungen zu erhalten. Für die Auswertung der Verbaldaten wurde hingegen ein induktives Vorgehen gewählt, um den Ist-Stand der Prüfungspraxis möglichst unvoreingenommen zu charakterisieren (Flick, 2010). Prinzipiell wäre hierbei auch eine deduktive Auswertung der Verbaldaten denkbar gewesen, jedoch könnte damit lediglich festgestellt werden, ob die aus der

Theorie abgeleiteten Kategorien auf die beobachtete Praxis zutreffen oder nicht. Eine möglichst praxisnahe Beschreibung, in Form bereits etablierter Vorgehensweisen und bewährter Strategien und Problemlösungen, wäre so nicht gewährleistet gewesen. Durch das induktive Vorgehen konnten dann auch Erkenntnisse zum bisherigen Vorgehen der Lehrenden bei der Prüfungserstellung und -auswertung sowie Einblicke in die damit verbundenen Überlegungen erbracht und mit idealtypischen Vorgehen, das in dem entwickelten Prüfungsmodell nahegelegt wird, verglichen werden. Mit diesem Vorgehen konnten Ursachen für bestehende Probleme aufgeklärt werden. Mit der Betrachtung der Programmziele und der bestehenden Praxis wurde, unter Nutzung verschiedener methodischer Zugänge und Perspektiven, eine systematische empirische Basis geschaffen, an der die Inhalte des Qualitätsentwicklungsprogramms ausgerichtet werden konnten (Seidel & Hoppert, 2011). Der Forderung aus der Professionalisierungsforschung, Maßnahmen zur Professionalisierung an die Probleme und Schwachstellen in der bestehenden Praxis sowie auf die Bedarfe der Lehrenden anzupassen (Desimone, 2009), konnte so erfolgreich nachgekommen werden.

Die Erkenntnisse zur veränderten Prüfungspraxis wurden zum einen über eine Analyse der neu erstellten Lehrziele und Prüfungen gewonnen (siehe Abbildung 5). Zum anderen berichteten die Lehrenden ab dem zweiten Arbeitstreffen im Rahmen der Reflexion vorangegangener Treffen und den eigenständig durchgeführten Arbeitsphasen über die Veränderungen, die sie bei der Erstellung und Auswertung ihrer Prüfung vorgenommen hatten (siehe Abschnitt 2.3.4). Die Reflexionen waren zum einen als lernunterstützende Maßnahme angedacht, andererseits bot sich dadurch auf wissenschaftlicher Seite die Möglichkeit, Wissen darüber zu generieren, welche Aspekte von den Lehrenden umgesetzt werden konnten und an welchen Stellen im Transfer Probleme auftraten. Die vorliegende Studie liefert damit, neben einer Beschreibung der bestehenden Prüfungspraxis, Erkenntnisse darüber, an welchen Stellen und unter welchen Bedingungen Lehrende Veränderungen an ihrer Prüfungspraxis im Rahmen eines Qualitätsentwicklungsprogramms umsetzen können.

Vor dem Hintergrund der Besonderheiten akademischer Lehre, die geprägt ist durch eine hohe Autonomie der Lehrenden und der Vermittlung von spezialisiertem Wissen auf einem hohen fachlichen Niveau (Braun et al., 2014), wurde die Prüfungspraxis an einer relativ kleinen Stichprobe an einigen Lehrstühlen aus der Fakultät für Maschinenwesen untersucht, um so möglichst authentische Einblicke in die Prüfungspraxis zu erhalten und die Durchführbarkeit der zeitintensiven Maßnahme zu gewährleisten. Deshalb sind an dieser Stelle im Hinblick auf die Generalisierbarkeit der Ergebnisse aber auch Einschränkungen zu machen. So beschränkt sich die Untersuchung auf die Fachdisziplin des Maschinenwesens und fünf Fälle. Innerhalb der Fälle ergaben sich wiederum Unterschiede, insofern die Lehrenden unterschiedliche Themenbereiche vertraten und die jeweiligen Lehrveranstaltungen sowohl in Bachelor- (MW 2, MW 3 und MW 4) als auch in Mas-

terstudiengängen (MW 1 und MW 5) verortet waren. Die Fälle unterschieden sich außerdem in der personellen Zusammensetzung und der Prüfungserfahrung der Teilnehmenden. Dennoch bestanden in der bestehenden Prüfungspraxis nur geringfügige Unterschiede zwischen den Fällen, weder im Hinblick auf die identifizierten Probleme und Schwachstellen in Lehrzielen und Prüfungen, noch beim Vorgehen und den Annahmen der Lehrenden bei der Erstellung und Auswertungen von Prüfungen. Die Lehrenden aller Fälle legten den Fokus auf Aspekte im Prozess der Prüfungserstellung und -auswertung, bei denen sie Probleme und Beschwerden von Seiten der Studierenden erwarteten (siehe Abschnitt 6.1.1). Die von den Lehrenden vorgenommenen Veränderungen während des Qualitätsentwicklungsprogramms waren ebenfalls nicht auf die beschriebenen Unterschiede zwischen den Fällen zurückzuführen. Vielmehr waren Merkmale des Qualitätsentwicklungsprogramms dafür entscheidend, ob die Lehrenden ihre Prüfungspraxis veränderten und die Qualität von Prüfungen verbessern konnten (siehe Abschnitt 6.1.3). Seidel und Hoppert (2011) berichten ebenfalls von einer hohen Konsistenz im Handeln von Lehrenden, unabhängig von deren Erfahrungsstand. Sie stellten in ihren Untersuchungen außerdem fest, dass zwischen drei verschiedenen Fachdisziplinen kaum Unterschiede in der Gestaltung von Hochschulseminaren sowie bezüglich der Lehrorientierung auftraten. Von daher ist es plausibel, anzunehmen, dass die in dieser Arbeit berichteten Erkenntnisse zur bestehenden Prüfungspraxis von Hochschullehrenden sowie die beschriebenen Probleme und Schwachstellen, auch bei anderen Fachdisziplinen und Hochschulen in ähnlicher Weise beobachtet werden können. Inwieweit die beschriebenen positiven Veränderungen im Verlauf des Qualitätsentwicklungsprogramms auch mit anderen Gruppen von Lehrenden erzielt werden können, ist andererseits aber auch mit davon abhängig, ob die Motivation zur Teilnahme an einem entsprechenden Programm und die damit verbundenen Aspirationen (Vermunt & Endedijk, 2011) ähnlich ausgeprägt sind wie bei den Teilnehmenden dieser Studie, die sich freiwillig und eigeninitiativ für eine Teilnahme entschieden hatten.

Neben den beschriebenen Stärken der Arbeit hinsichtlich des methodischen Vorgehens sind auch Limitationen zu nennen. Für das Kriterium der Einhaltung der Gütekriterien der pädagogischen Diagnostik wurde zum Beispiel analysiert, inwieweit die Lehrenden bestimmte Maßnahmen umsetzen, die dazu beitragen können, die Validität und Reliabilität von Prüfungen zu sichern (z.B. die Durchführung einer rationalen Aufgabenanalyse oder die Verwendung von Bewertungsrastern bei der Korrektur). Die in der Testtheorie üblichen statistischen Verfahren zur Bestimmung von Validität und Reliabilität konnten jedoch nicht durchgeführt werden. Die Reliabilität einer Prüfung könnte beispielsweise durch die Testhalbierungs- oder Paralleltestmethode bestimmt werden (Bühner, 2011). Die Prüfung müsste dazu zwei Prüfungsteile enthalten, in denen jeweils vergleichbare Aufgaben zu den gleichen Lehrzielen abgebildet sind. Die Prüfungen der Lehrenden waren jedoch nicht entsprechend angelegt, um einen solchen Test durchzuführen. Deshalb müssen

die Ergebnisse hinsichtlich der Maßnahmen zur Sicherung der Reliabilität, die in den Ergebnissen berichtet wurden, eben auch als Maßnahmen betrachtet werden, die eine Erhöhung der Reliabilität zwar wahrscheinlich machen, aber nicht sicherstellen.

Zur Bewertung von Prüfungsantworten erfolgte eine Betrachtung, inwieweit die Lehrenden bestimmte Maßnahmen durchführten, die laut den Erkenntnissen verschiedener Studien (Graham et al., 2012) zu einer besseren Konsistenz bei der Korrektur führen können. Eine statistische Bestimmung von Übereinstimmungswerten zwischen Korrektoren oder der Konsistenz, mit der ein einzelner Korrektor Prüfungsantworten bewertet, wurde auch hier nicht durchgeführt. Die tatsächlich erreichte Konsistenz beziehungsweise Reliabilität, die von den Lehrenden bei der Bewertung von Prüfungsantworten erreicht wird, müsste deshalb in weiteren Untersuchungen geklärt werden.

Zur Bestimmung der Passung zwischen Lehrzielen und Aufgaben wurden in dieser Arbeit die prozentualen Verteilungen von Wissensarten in den Lehrzielen und den Prüfungsaufgaben gegenübergestellt. Diese Gegenüberstellung ist als ein relatives grobes Maß anzusehen, da sich die Grundhäufigkeiten zwischen Lehrzielen ($n=24$) und Aufgaben ($n=353$) deutlich unterscheiden. Für die Gültigkeit der mit diesem Vorgehen gewonnenen Ergebnisse spricht jedoch, dass die Lehrenden bei den eigens durchgeführten Analysen von Aufgaben und Lehrzielen zu ähnlichen Ergebnissen kamen. Hier ist jedoch weiterer Klärungsbedarf angezeigt, um die berichteten Ergebnisse zu überprüfen.

In diesem Zusammenhang ist außerdem anzumerken, dass die verwendeten Kategoriensysteme zur Analyse der materialbasierten Daten überfachlich angelegt waren. Hier wäre deshalb zu klären, ob bei einer domänenspezifischen Analyse der Materialien vergleichbare Ergebnisse zustande kämen. Einen ersten Hinweis auf eine Vergleichbarkeit der Ergebnisse zwischen dem hier gewählten überfachlichen Zugang und einer domänenspezifischen Analyse von Aufgaben liefert die bereits oben beschriebene Studie von Stefanica (2013). Sie berichtet, basierend auf einer domänenspezifischen Analyse von Aufgaben in der Höheren Mathematik, von Aufgabencharakteristika, die mit den hier berichteten Ergebnissen zu den Merkmalen von Aufgaben vergleichbar sind. Für den Einsatz des hier verwendeten überfachlichen Kategoriensystems spricht außerdem, dass die Lehrenden die in dieser Arbeit verwendeten Kategorien (Wissensarten, Funktion von Kontextbezügen, Aufgabenoffenheit und sprachlogische Komplexität) effektiv zur Analyse und Strukturierung ihrer Fachinhalte und der Differenzierung von Aufgabenanforderungen einsetzen konnten. Maier et al. (2013) berichten in diesem Zusammenhang davon, dass überfachliche Kategoriensysteme gerade von Lehrkräften, die verschiedenen Fächern und verschiedenen Fächerverbänden angehören, gemeinsam genutzt werden können.

Auf Ebene des Designs der Studie sind insbesondere drei Aspekte einschränkend anzumerken: (1) Es wurde nicht betrachtet, welche Auswirkungen die von den Lehrenden vorgenommenen Veränderungen an den Prüfungen auf Seiten der Studierenden erzeugten. Außerdem konnte (2) nicht ermittelt werden, inwieweit die Lehrenden das im Rahmen der Reflexionsphasen erhaltene Feedback zu ihren Prüfungsentwürfen umsetzen konnten, da nur die fertig gestellten Prüfungen analysiert wurden. Das Projekt war (3) nicht in einem Kontrollgruppendesign angelegt, sodass keine kausal nachweisbaren Effekte, die auf das Qualitätsentwicklungsprogramm zurückgeführt werden können, nachgewiesen werden konnten. Vor diesem Hintergrund erschien es deshalb auch vertretbar, auf eine inferenzstatistische Überprüfung von Veränderungen zwischen bestehenden und neu erstellten Prüfungen hinsichtlich bestimmter Merkmale zu verzichten. Diese Einschränkung ist vor dem besonderen Erkenntnisinteresse der Arbeit zu verstehen, die das Ziel verfolgte, basierend auf einzelnen Fallbetrachtungen ein fundiertes Beschreibungswissen zur bestehenden Prüfungspraxis von Hochschullehrenden und den Veränderungen, die sie im Laufe eines Qualitätsentwicklungsprogramms an ihrer Praxis vornahmen, zu generieren.

6.3 Implikationen

6.3.1 Praktische Implikationen

Die Erkenntnisse zur bestehenden Prüfungspraxis zeigten auf, dass in den betrachteten Prüfungen zahlreiche Probleme und Schwachstellen existierten, denen jedoch von den Lehrenden durch die Teilnahme am vorgestellten Qualitätsentwicklungsprogramm konstruktiv begegnet werden konnte. Aus Perspektive der Hochschulen erscheint insbesondere die fehlende Passung zwischen den Anforderungen in Lehrzielen und den Prüfungsaufgaben als problematisch. Immerhin werden Studierenden mit dem Bestehen von Prüfungen Qualifikationen und Kompetenzen bescheinigt, die in den Lehrzielen beschrieben sind. Entsprechend valide sollten die Prüfungen als Bewertungsinstrumente dann auch sein. Die Bemühungen, im Zuge der Modularisierung von Studiengängen eine kompetenzorientierte und an Lehrzielen ausgerichtete Lehr- und Prüfungspraxis an Hochschulen zu etablieren, scheinen sich bislang nicht durchgängig auf die Handlungsebene der Lehrenden auszuwirken. Umso wichtiger erscheint es auch vor dem Hintergrund der berichteten Ergebnisse, die Lehrenden durch entsprechende Maßnahmen zu unterstützen, um einen Praxis-transfer von externen Vorgaben und Anforderungen anzubahnen (Coburn, 2004).

Die Hochschulen sollten daher insbesondere auch Programme aufsetzen und unterstützen, die Lehrende zur Verbesserung der Prüfungsqualität anregen und sie dabei fachlich begleiten. Mit dem in dieser Arbeit vorgestellten Qualitätsentwicklungsprogramm konnten die Lehrenden zum Teil erhebliche Verbesserungen an ihren Prüfungen erzielen und dadurch auch eine deutliche bes-

sere Passung zwischen den Anforderungen in Aufgaben und Lehrzielen erreichen. Auch Merkmale von Kompetenzorientierung konnten von den Lehrenden besser realisiert werden. Insbesondere ist es gelungen, Lehrenden den Sinn und Zweck von Lehrzielen zu verdeutlichen und sie dazu zu motivieren, die Prüfungserstellung und -auswertung systematisch an diesen Lehrzielen auszurichten. Das beschriebene Programm könnte für Hochschulen daher als Vorbild dienen, um die Qualität von Prüfungen zu entwickeln und die Reformvorgaben über eine kompetenzorientierte und lehrzielorientierte Lehr- und Prüfungspraxis auf der Handlungsebene der Lehrenden zu verankern.

In Übereinstimmung mit der Forschung zur Professionalisierung von Lehrpersonen konnte gezeigt werden, dass entsprechende Programme über einen längeren Zeitraum angelegt sein und abwechselnde Input- als auch eigenständige Arbeits- sowie gemeinsame Reflexionsphasen beinhalten sollten (Desimone, 2009). Dabei ist eine kollektive Teilnahme von Lehrenden vorzusehen, um gemeinsame Lernprozesse zu organisieren. Externe Moderatoren schaffen dabei in gemeinsamen Arbeitstreffen den Raum für kollegiale Diskussionen und eine vertrauensvolle Atmosphäre, damit sich konstruktive Arbeits- und Lerngemeinschaften bilden können (Little, 2002). Die Lehrenden erhalten einen theoretischen Input, der auf wissenschaftlichen Prinzipien und Kriterien beruht und der dann zusammen mit den Moderatoren direkt auf die eigenen Materialien angewendet wird, um darin bestehende Probleme, aber auch Lösungsmöglichkeiten, aufzuzeigen und zu erarbeiten. Wie sich zeigte, nahmen die in dieser Arbeit untersuchten Lehrenden nur bei jenen Aspekten Veränderungen vor, zu denen ihnen an den eigenen Materialien bestehende Probleme aufgezeigt werden konnten, was die Vorteile einer aktiven Arbeit an Materialien aus der Praxis der Lehrenden unterstreicht. Die Befunde aus dem Programm untermauern damit die Ergebnisse aus der Professionalisierungsforschung, wonach mit einmaligen Maßnahmen, die auf der Vermittlung von theoretischem Wissen basieren (z.B. klassische Formate wie eintägige Workshops), nur selten weitreichende Veränderungen in der Praxis von Lehrenden erzielt werden können (Borko, 2004). Zwar ist die Durchführung eines solch umfassenden Programms aufwändiger und zeitintensiver – im Sinne der Nachhaltigkeit und vor dem Hintergrund der dargestellten Befunde, kann den Hochschulen diese Investition in entsprechende Maßnahmen zu zentralen Themen der Lehre jedoch nahegelegt werden.

In dieser Arbeit wurde außerdem ein Kompetenzkonzept entwickelt, das von Hochschulen zur Gestaltung von Studiengängen genutzt werden kann. Zentral ist dabei die Unterscheidung zwischen drei Testarten: (1) Kompetenztests, (2) kompetenzorientierten Tests und (3) kompetenzorientierten Wissenstests (siehe Abschnitt 2.1.1). Vor allem kompetenzorientierte Wissenstests versprechen einen großen Mehrwert für Hochschulen, da sie die „klassische Form“ der schriftlichen Prüfung um Aspekte der Kompetenzorientierung erweitern und letztlich auch die Relevanzwahrnehmung von Prüfungen bei Studierenden günstig beeinflussen können. Für die Gestaltung von

Studiengängen erscheint dabei eine Kombination der drei Prüfungsformen vielversprechend. Die Erreichung von Kernkompetenzen eines Studiengangs kann mittels Kompetenztests besonders zuverlässig und mit hoher Gültigkeit festgestellt werden, indem die Handlungsfähigkeit von Studierenden in einer realen Situation beobachtet wird (Shavelson, 2010b). Mit einem vorgelagerten kompetenzorientierten Test können die Studierenden dagegen sukzessive an die Komplexität realer Situationen herangeführt werden. In der Medizin werden mit den sogenannten „Objective Structured Clinical Examinations“ schon seit Längerem solche situationsnahe Testverfahren eingesetzt (Rushforth, 2007). Kompetenzorientierte Wissenstests können vor allem dann eingesetzt werden, wenn es zu überprüfen gilt, ob Studierende die Voraussetzungen für kompetentes Handeln, im Sinne von kontextbezogenen Wissensseinheiten und kognitiven Fähigkeiten, besitzen. Hierbei gilt es zu klären, welcher Entwicklungsaufwand und Unterstützungsbedarf nötig wäre, um die skizzierte kompetenzbasierte Prüfungspraxis umzusetzen.

Eine weitere, kaum genutzte, Ressource stellen die Prüfungsdaten dar. In einer systematischen Auswertung von Prüfungsergebnissen liegt großes Potenzial, insofern mit der Prüfung die in den Lehrzielen beschriebenen Anforderungen und Kompetenzen valide und reliabel erfasst wurden. Auf Hochschulebene können Prüfungsergebnisse für ein hochschulinternes Monitoring, aber auch zur externen Rechenschaftslegung über die von den Studierenden erworbenen Qualifikationen genutzt werden (Shavelson, 2010a). Mit einem internen Monitoring kann auf Studiengangsebene die Passung des Lehrangebots und die Sequenzierung verschiedener Module überprüft werden. Die Lehrenden erhalten Rückmeldung, sowohl zur Prüfungsqualität, als auch zur Lernwirksamkeit ihrer Lehrveranstaltung. Auf Basis von Prüfungsdaten, aus denen Stärken und Schwächen bei Studierenden hervorgehen, können lernwirksame Rückmeldungen für Studierende gegeben werden. Letztlich wäre sogar eine an die Praxis angeschlossene Lehr-Lernforschung denkbar, insofern mit den Prüfungen belastbare Daten erzeugt werden (Secolsky & Denison, 2012). In diesem Zusammenhang eröffnen computerbasierte Testverfahren vielfältige Möglichkeiten (Wannemacher, 2007). Diese Form des Testens wird in Zukunft immer weiter an Bedeutung erlangen, da sie das Prüfen räumlich und zeitlich entkoppelt und eine differenziertere Leistungsdiagnostik, zusammen mit Möglichkeiten zu einer effektiven Leistungsrückmeldung an Studierende, bei zudem sinkendem Korrekturaufwand, erlaubt. Viele Nachteile des klassischen Testens können durch computerbasierte Verfahren ausgeräumt werden (ebd.).

6.3.2 Implikationen für weitere Forschungsvorhaben

Die in dieser Arbeit vorgestellten Befunde eröffnen die Möglichkeit zu weiteren Forschungsvorhaben, unterliegen aber auch bestimmten Einschränkungen, die in zukünftigen Forschungsarbeiten aufgegriffen werden sollten. Das in dieser Arbeit eingeführte Modell der sechs Schritte zur

Prüfungserstellung und -auswertung bietet dafür einen theoretischen wie analytischen Rahmen, um gezielt weiterführende Fragen zu einzelnen Teilschritten zu stellen und zu vertiefen. Nachfolgend werden dazu einige Möglichkeiten skizziert.

In Bezug auf die Festlegung des Prüfungsgegenstands erscheint eine Operationalisierung der in dieser Arbeit induktiv gebildeten Kategorien (siehe Tabelle 18) in einen Fragebogen zur Selbsteinschätzung vielversprechend, um zu untersuchen, inwieweit sich Schwerpunktsetzungen in Prüfungsanforderungen und das Vorgehen bei der Auswahl und Modellierung des Prüfungsgegenstands standort- und disziplinübergreifend unterscheiden. Dadurch könnten Maßnahmen zur Professionalisierung der Prüfungspraxis bedarfsgerecht zugeschnitten werden. Eine disziplinübergreifende Analyse von Prüfungsaufgaben, auf Grundlage der in dieser Arbeit eingesetzten Kategoriensysteme, wäre in diesem Zusammenhang ebenfalls aufschlussreich, um festzustellen, inwieweit die an der Fakultät für Maschinenwesen der Technischen Universität München gewonnenen Ergebnisse auch auf andere Fachdisziplinen und Standorte übertragbar sind. Dafür wäre es aber notwendig, das eingesetzte Kategoriensystem hinsichtlich seiner Übereinstimmung mit Ergebnissen einer domänenspezifischen Auswertung zu vergleichen. Mit einer domänenspezifischen Analyse von Prüfungsaufgaben ist es möglich, die Anforderungspotenziale von Aufgaben differenziert zu erfassen, was für die Überprüfung der Lehrzielvalidität von Aufgaben einen zusätzlichen Mehrwert bringen könnte. Dieser Ansatz ist jedoch ungleich zeitaufwendiger als das hier gewählte Vorgehen, da fachdidaktische Überlegungen auf Grundlage fachspezifischer Besonderheiten angestellt werden müssen (Maier et al., 2013). Die Durchführung von kognitiven Aufgabenanalysen, bei denen Studierende ihre Denkprozesse beim Lösen von Aufgaben verbalisieren, könnten aber dabei helfen (Ayala et al., 2002).

Weiter abzuklären wäre außerdem, welche Reliabilitätswerte die Lehrenden mit dem etablierten Vorgehen bei der Bewertung von Prüfungsantworten erzielen, indem Übereinstimmungsmaße zwischen Korrektoren gebildet werden. Die Konsistenz, die eine Person bei der Bewertung von Prüfungsantworten erreicht, kann sichtbar gemacht werden, indem eine Person in einem bestimmten zeitlichen Abstand die dieselben Prüfungsaufgaben erneut korrigiert (Graham et al., 2012). Bezogen auf die Reliabilität erscheint auch eine Analyse von Prüfungsdaten vielversprechend, um festzustellen, ob ein Zusammenhang zwischen der Umsetzung der in dieser Arbeit betrachteten Maßnahmen, die zur Erhöhung der Reliabilität beitragen können, aber nicht müssen, und statistisch bestimmten Reliabilitätskoeffizienten besteht. Weiterhin wären mit den Prüfungsdaten Analysen hinsichtlich der Validität vorstellbar, indem theoretisch angenommene Zusammenhänge zwischen Aufgaben statistisch überprüft werden. Dazu müssten die Prüfungen jedoch entsprechend angelegt und die Prüfungsdaten nach bestimmten Kriterien aufbereitet sein (Bühner, 2011).

Hinsichtlich des Einsatzes verschiedener Forschungsdesigns könnten zum einen längerfristig angelegte Untersuchungen Hinweise liefern, inwieweit die Lehrenden die im Rahmen von Professionalisierungsprogrammen angestoßenen Veränderungen dauerhaft in ihre Praxis etablieren konnten. Zum anderen könnten durch den gezielten Einsatz von experimentellen Designs Wirkungszusammenhänge zwischen Gestaltungsmerkmalen von Professionalisierungsmaßnahmen und deren Auswirkungen auf das Handeln von Lehrenden bestimmt werden. Im Sinne des aktuellen Forschungsparadigmas des „Angebots- Nutzen Modells“, nach dem Lehrende für ein Lehrangebot sorgen, das von Studierenden wahrgenommen werden kann (Seidel & Reiss, 2014), sollte in diesen Designs auch die Seite der Studierenden berücksichtigt werden. Fragebogenuntersuchungen könnten wertvolle Informationen liefern, wie die Studierenden mit veränderten Prüfungsformaten umgehen, ob sie diese zum Beispiel als fair wahrnehmen und welche Auswirkungen sich auf das Lernverhalten, idealerweise hin zu tiefenorientierten Lernstrategien, zeigen (Gijbels et al., 2008).

7 Schlusswort

Diese Arbeit leistet einen Beitrag, die Prüfungspraxis von Hochschullehrenden empirisch fundiert zu beschreiben. Es wurden Möglichkeiten aufgezeigt, wie Lehrende dazu angeleitet werden können, die Qualität von Prüfungen zu entwickeln. Die vorliegende Studie versteht sich daher sowohl als Grundlage für weiterführende Forschungsvorhaben zur Prüfungspraxis an deutschen Hochschulen, als auch als Handlungsempfehlung für die Hochschuldidaktik im Sinne eines Best-Practises. Die Bedeutung von Prüfungen für die Akteure an Hochschulen wird weiter zunehmen. Entsprechend sollten auch weiterhin systematisch Entwicklungsmöglichkeiten geschaffen werden.

Literatur

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational & Psychological Testing*. Washington, DC: American Educational Research Association.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman, Inc.
- Ayala, C. C., Yin, Y., Shavelson, R. J., & Vanides, J. (2002). *Investigating the cognitive validity of science performance assessment with think alouds: Technical aspects*. Paper presented at the American Educational Research Association, New Orleans, LA.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469-520.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International journal of educational research*, 35(5), 463-482.
- Blomberg, G., Stürmer, K., & Seidel, T. (2009). How Pre-Service Teachers Observe Teaching on Video: Effects of Viewers' Teaching Subjects and the Subject of the Video. *Teaching and teacher education*, 27, 1131-1140.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3-13.
- Bloom, B. S., Enghart, M., Furst, E., Hill, W., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. New York: David McKay.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher*, 33(8), 3-15.
- Brändle, T. (2010). *10 Jahre Bologna-Prozess. Chancen, Herausforderungen und Problematiken*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school: Expanded Edition*. Washington, DC: National Academy Press.
- Braun, E., Weiss, T., & Seidel, T. (2014). Lernumgebung Hochschule. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie*. Weinheim: Beltz.
- Brown, N. J., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review*, 23(2), 221-234.
- Büchter, A., & Leuders, T. (2005). *Mathematikaufgaben selbst entwickeln: Lernen fördern-Leistung überprüfen*. Berlin: Cornelsen Scriptor.
- Bücker, S., Deimling, M., Durduman, J., Holzhäuser, J., Schnieders, S., Tietze, M., Sayeed, S., & Schneider, M. (2015). Prüfung. In M. Schneider & M. Mustafic (Eds.), *Gute Hochschullehre: Eine evidenzbasierte Orientierungshilfe. Wie man Vorlesungen, Seminare und Projekte effektiv gestaltet*. Berlin: Springer.
- Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of research on computing in education*, 28(3), 282-299.
- Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. München: Pearson Studium.
- Cilliers, F. J., Schuwirth, L. W., Adendorff, H. J., Herman, N., & Van der Vleuten, C. P. (2010). The mechanism of impact of summative assessment on medical students' learning. *Advances in health sciences education*, 15(5), 695-715.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211-244.

-
- Dany, S., Szczyrba, B., & Wildt, J. (Eds.). (2008). *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen*. Bielefeld: Bertelsmann.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher*, 38(3), 181-199.
- Dubs, R. (1978). *Aspekte des Lehrerverhaltens*. Sauerländer: Aarau.
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52(2), 187-201.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3), 363-406.
- Europäische Kommission. (2009). ECTS-Leitfaden. Luxemburg: Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften.
- Flick, U. (2010). *Qualitative Sozialforschung. Eine Einführung*. Reinbek: Rowohlt.
- Fricke, R. (1974). *Kriteriumsorientierte Leistungsmessung*. Stuttgart: W. Kohlhammer GmbH.
- 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 educational research journal*, 38(4), 915-945.
- Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, K., Möller, G. E., Wiesbeck, A. B., & Prenzel, M. (2015). Fostering Professional Communication Skills of Future Physicians and Teachers: Differential Effects of E-Learning and Role-play. *Instructional Science*, 43(4), 443-462.
- Gijbels, D., Segers, M., & Struyf, E. (2008). Constructivist learning environments and the (im)possibility to change students' perceptions of assessment demands and approaches to learning. *Instructional Science*, 36(5-6), 431-443.
- Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology*, 94(5), 1336-1344.
- Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. Westat: Center for Educator Compensation Reform.
- Gronlund, N. E. (1974). *Die Anlage von Leistungstests*. Frankfurt am Main: Diesterweg.
- Gröschner, A., Seidel, T., Pehmer, A. K., & Kiemer, K. (2014). Facilitating collaborative teacher learning: The role of "mindfulness" in video-based teacher professional development programs. *Gruppendynamik und Organisationsberatung*, 45(3), 273-290.
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *The Teachers College Record*, 111(9), 2055-2100.
- Gschwendtner, T., Geißel, B., & Nickolaus, R. (2010). Modellierung beruflicher Fachkompetenz in der gewerblich-technischen Grundbildung. Projekt Berufspädagogik. In E. Klieme, D. Leutner & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes (Zeitschrift für Pädagogik, 56. Beiheft)* (pp. 258-269). Weinheim: Beltz.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: theory and practice*, 8(3/4), 381-391.

-
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333.
- Herbig, M. (1976). *Praxis lehrzielorientierter Tests*. Düsseldorf: Schwann.
- Ingenkamp, K. H., & Lissmann, U. (2007). *Lehrbuch der pädagogischen Diagnostik (6. Auflage)*. Weinheim und Basel: Beltz.
- Jahn, G., Stürmer, K., Seidel, T., & Prenzel, M. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46(4), 171-180.
- Johannes, C., & Seidel, T. (2010). Professionelles Lernen von Anfängern in der Hochschullehre – Erwartungen und Vorstellungen über Hochschullehre im Rahmen des Projekts LehreLernen. *Personal- und Organisationsentwicklung in Einrichtungen der Lehre und Forschung (P-OE)*, 5(2+3), 2+3.
- Johannes, C., & Seidel, T. (2012). Professionalisierung von Hochschullehrenden: Lehrbezogene Vorstellungen, Wissensanwendung und Identitätsentwicklung in einem videobasierten Qualifikationsprogramm. *Zeitschrift für Erziehungswissenschaft*, 15(2), 233-251.
- Jordan, A., Krauss, S., Katrin, L., Blum, W., Neubrand, M., Brunner, M., Kunter, M., & Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematik-Didaktik*, 29(2), 83-107.
- Katz, I. R., Bennett, R. E., & Berger, A. E. (2000). Effects of Response Format on Difficulty of SAT-Mathematics Items: It's Not the Strategy. *Journal of Educational Measurement*, 37(1), 39-57.
- Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7(3), 255-277.
- Klauer, K. J. (1974). *Methodik der Lehrzieldefinition und Lehrstoffanalyse*. Düsseldorf: Schwann.
- Klauer, K. J., & Leutner, D. (2007). *Lehren und Lernen - Einführung in die Instruktionspsychologie*. Weinheim u. Basel: Beltz.
- Klauer, K. J., & Leutner, D. (2012). *Lehren und Lernen. Einführung in die Instruktionspsychologie*. Weinheim: Beltz.
- Kleinknecht, M., Maier, U., Metz, K., & Bohl, T. (2011). Analyse des kognitiven Aufgabenpotentials. Entwicklung und Erprobung eines allgemeindidaktischen Auswertungsmanuals. *Unterrichtswissenschaft*, 39(4), 328-344.
- 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.
- Kordts-Freudinger, R. (2013). Die Einstellung Hochschullehrender zum kompetenzorientierten Prüfen Zusammenfassung. In I. v. d. Berk, Merkt, M., Salden, P. & Scholkmann, A. (Ed.), ZHW - Almanach. Einzelbeitrag Nr.: 2013-2. Retrieved from <http://www.zhw.uni-hamburg.de/almanach>.
- Krause, U.-M., Stark, R., & Mandl, H. (2004). Förderung des computerbasierten Wissenserwerbs durch kooperatives Lernen und eine Feedbackmaßnahme. *Zeitschrift für Pädagogische Psychologie*, 18(2), 125-136.
- Kultusministerkonferenz. (2005). *Qualifikationsrahmen für Deutsche Hochschulabschlüsse*. Bonn: Kultusministerkonferenz, Hochschulrektorenkonferenz & Bundesministerium für Bildung und Forschung, Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_04_21-Qualifikationsrahmen-HS-Abschluesse.pdf.

-
- Kultusministerkonferenz. (2010). *Rahmenvorgaben für die Einführung von Leistungspunktsystemen und die Modularisierung von Studiengängen. Anlage der Ländergemeinsamen Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengängen*. Bonn: Kultusministerkonferenz.
- Kultusministerkonferenz. (2011). *Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengängen vom 04.02.2010 - Auslegungshinweise vom 25. März 2011*. Bonn: Kultusministerkonferenz.
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen–Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematik-Didaktik*, 35(1), 7-48.
- Li, M., Ruiz-Primo, M. A., & Shavelson, R. J. (2006). Towards a science achievement framework: The case of TIMSS 1999. In S. Howie & T. Plomp (Eds.), *Contexts of learning mathematics and science: Lessons learned from TIMSS* (pp. 291-311). London: Routledge.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse*, 6. Auflage, Beltz Psychologie Verlags Union. Weinheim: Beltz.
- Little, J. W. (2002). Locating learning in teachers' communities of practice: Opening up problems of analysis in records of everyday work. *Teaching and teacher education*, 18(8), 917-946.
- Mager, R. F. (1962). *Lernziele und programmierter Unterricht* (2. Auflage). Weinheim: Beltz.
- Maier, U., Bohl, T., Kleinknecht, M., & Metz, K. (2013). Allgemeindidaktische Kategorien für die Analyse von Aufgaben. In M. Kleinknecht, T. Bohl, U. Maier & K. Metz (Eds.), *Lern- und Leistungsaufgaben im Unterricht*. Bad Heilbrunn: Klinkhardt.
- Mayring, P. (2000). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Deutscher Studien Verlag.
- Mayring, P. (2001). Kombination und Integration qualitativer und quantitativer Analyse [31 Absätze]. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research*, 2(1). <http://nbn-resolving.de/urn:nbn:de:0114-fqs010162>
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741-749.
- Müller-Benedict, V., & Tsarouha, E. (2011). Können Examensnoten verglichen werden? Eine Analyse von Einflüssen des sozialen Kontextes auf Hochschulprüfungen. *Zeitschrift für Soziologie*, 40(5), 388-409.
- National Assessment Governing Board. (2007). *Science Assessment and Item Specifications for the 2009 National Assessment of Educational Progress (Prepublication Edition)*. Washington, DC: National Assessment Governing Board.
- Nickolaus, R., Behrendt, S., Dammann, E., Stefanica, F., & Heinze, A. (2013). Theoretische Modellierung ausgewählter ingenieurwissenschaftlicher Kompetenzen. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Eds.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschaftswissenschaften und der Ingenieurwissenschaften (Lehrerbildung auf dem Prüfstand, Sonderheft)* (pp. 150-176). Landau: Verlag Empirische Pädagogik.
- Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of educational research*, 81(3), 376-407.
- Opfer, V. D., Pedder, D. G., & Lavicza, Z. (2011). The role of teachers' orientation to learning in professional development and change: A national study of teachers in England. *Teaching and teacher education*, 27(2), 443-453.
- Ostermeier, C. (2004). *Kooperative Qualitätsentwicklung in Schulnetzwerken. Eine empirische Studie am Beispiel des BLK-Programms "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts" (SINUS)*. Münster: Waxmann.

-
- Ostermeier, C., Prenzel, M., & Duit, R. (2010). Improving science and mathematics instruction: The SINUS project as an example for reform as teacher professional development. *International Journal of Science Education*, 32(3), 303-327.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academies Press.
- Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48(1), 4-11.
- Prenzel, M., Schindler, C., & Schulz, F. (2012). Prüfungskultur an der Technischen Universität München. Erste Ergebnisse einer Befragung der Studierenden. http://www.lehren.tum.de/fileadmin/w00bmo/www/Downloads/Themen/Pruefungen/p_royktbericht_pruefungskultur_studierende.pdf
- Prenzel, M., Schulz, F., & Schindler, C. (2012). Prüfungskultur an der Technischen Universität München. Erste Ergebnisse einer Befragung der Dozierenden http://www.lehren.tum.de/fileadmin/w00bmo/www/Downloads/Themen/Pruefungen/p_royktbericht_pruefungskultur_dozierende.pdf
- Reisse, W. (2008). *Kompetenzorientierte Aufgabenentwicklung: Ein Lehrerhandbuch für die Sekundarstufen*. Köln: Aulis-Verlag Deubner.
- Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion (2., vollständig überarbeitete und erweiterte Aufl.)*. Bern: Huber.
- Rushforth, H. E. (2007). Objective structured clinical examination (OSCE): Review of literature and implications for nursing education. *Nurse education today*, 27(5), 481-490.
- Schaper, N. (2012). Fachgutachten zur Kompetenzorientierung in Studium und Lehre. Bonn: Hochschulrektorenkonferenzen - nexus.
- Schulz, F., Zehner, F., Schindler, C., & Prenzel, M. (2014). Prüfen und Lernen im Studium: Erste Schritte zur Untersuchung von Prüfungsanforderungen und Lerntypen. *Beiträge zur Hochschulforschung*, 36(2).
- Secolsky, C., & Denison, D. B. (2012). Improving Institutional Decision Making through Educational Measurement, Assessment, and Evaluation. In C. Secolsky & D. B. Denison (Eds.), *Handbook on Measurement, Assessment, and Evaluation in Higher Education* (pp. xvii-xx). New York, NY: Routledge.
- Sedlmeier, P., & Renkewitz, F. (2007). *Forschungsmethoden und Statistik in der Psychologie*. München: Pearson Studium.
- Segers, M., Nijhuis, J., & Gijssels, W. (2006). Redesigning a learning and assessment environment: The influence on students' perceptions of assessment demands and their learning strategies. *Studies in Educational Evaluation*, 32(3), 223-242.
- Seidel, T. (2003). *Lehr-Lernskripts im Unterricht*. Münster: Waxmann.
- Seidel, T., & Hoppert, A. (2011). Merkmale von Lehre an der Hochschule. Ergebnisse zur Gestaltung von Hochschulseminaren mittels Videoanalysen. *Unterrichtswissenschaft*, 39(2), 154-172.
- Seidel, T., & Reiss, K. (2014). Lerngelegenheiten im Unterricht. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie* (pp. 253-275). Weinheim: Beltz.
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American educational research journal*, 51(4), 739-771.
- Seidel, T., Stürmer, K., Schäfer, S., & Jahn, G. (2015). How Preservice Teachers Perform in Teaching Events Regarding Generic Teaching and Learning Components. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 1-13.
- Shaeiwitz, J. A. (1996). Outcomes assessment in engineering education. *Journal of Engineering Education*, 85(3), 239-246.

-
- Shavelson, R. J. (2010a). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.
- Shavelson, R. J. (2010b). On the measurement of competency. *Empirical research in vocational education and training*, 2(1), 41-63.
- Shavelson, R. J. (2012). Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. *Empirical research in vocational education and training*, 4(1), 77-90.
- Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215-232.
- Shavelson, R. J., & Ruiz-Primo, M. A. (1999a). Leistungsbewertung im naturwissenschaftlichen Unterricht. *Unterrichtswissenschaft*, 27, 102-127.
- Shavelson, R. J., & Ruiz-Primo, M. A. (1999b). On the psychometrics of assessing science understanding. In J. J. Mintzes, J. H. Wamhersee & J. D. Novak (Eds.), *Assessing science understanding. A human constructivist view* (pp. 304-338). New York: Academic Press.
- Shavelson, R. J., Ruiz-Primo, M. A., Li, M., & Ayala, C. C. (2003). *Evaluating new approaches to assessing learning*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.
- Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (1999). Note on sources of sampling variability in science performance assessments. *Journal of Educational Measurement*, 36(1), 61-71.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1). Newbury Park, CA: Sage Publications.
- Sherin, M. G. (2004). New perspectives on the role of video in teacher education. *Advances in research on teaching*, 10, 1-28.
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology*, 43(1), 583-626.
- Stefanica, F. (2013). Modulbeschreibungen - Deskriptionen realer Ansprüche oder realitätsferne Lyrik? Eine qualitative Analyse am Beispiel Höhere / Angewandte) Mathematik I/II im Rahmen des Maschinenbaustudiums an ausgewählten Hochschulstandorten Baden-Württembergs. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 109(2), 286-303.
- Strietholt, R., & Terhart, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*, 55(4), 622-645.
- Stürmer, K., & Seidel, T. (2015). Assessing Professional Vision in Teacher Candidates. *Zeitschrift für Psychologie*, 223(1), 54-63.
- Terhart, E., Baumgart, F., Meder, N., & Sychowski, G. v. (2009). Standardisierte Prüfungsverfahren in der Erziehungswissenschaft: Kontext, Formen, Konsequenzen. *Erziehungswissenschaft*, 20(38), 9-36.
- van Es, E. A. (2012). Examining the development of a teacher learning community: The case of a video club. *Teaching and teacher education*, 28(2), 182-192.
- Van Merriënboer, J. J. G., & Kirschner, P. A. (2012). *Ten steps to complex learning: A systematic approach to four-component instructional design*. New York: Routledge.
- Vermunt, J. D., & Endedijk, M. D. (2011). Patterns in teacher learning in different phases of the professional career. *Learning and individual differences*, 21(3), 294-302.
- Von Davier, M. (2012). Schachprogramme, Testaufgaben, Populationsunabhängigkeit und das Rasch Modell - oder: Es ist Zeit für das Rost-Modell. In W. Kempf & R.

-
- Langeheine (Eds.), *Item-Response-Modelle in der sozialwissenschaftlichen Forschung* (pp. 15-31). Berlin: Irena Regner.
- Wannemacher, K. (2007). Computergestützte Prüfungsverfahren. In M. H. Breitner, B. Bruns & F. Lehner (Eds.), *Neue Trends im E-Learning* (pp. 427-440). Heidelberg: Springer.
- Webb, N. M., Shavelson, R. J., Kim, K.-S., & Chen, Z. (1989). Reliability (generalizability) of job performance measurements: Navy machinist mates. *Military Psychology, 1*(2), 91-110.
- Webler, W. D. (2010). Internationale Vergleichbarkeit von Noten im Hochschulbereich? Problematik der Notenvergabe, Referenzgrößen und der Verwendung der Gauß'schen Normalverteilung. *Qualität in der Wissenschaft, 4*(1), 20-23.
- Wieberg, H.-J. W. (1983). Probleme kriteriumsorientierter Leistungsmessung: Sicherung der Kontenvalidität. In R. Horn, K. H. Ingenkamp & R. S. Jäger (Eds.), *Tests und Trends 3. Jahrbuch der Pädagogischen Diagnostik* (pp. 29-52). Weinheim: Beltz.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah: Lawrence Erlbaum Associates.
- Wilson, M. (2008). Cognitive diagnosis using item response models. *Zeitschrift für Psychologie/Journal of Psychology, 216*(2), 73-87.
- Wirtz, M. (2004). Bestimmung der Güte von Beurteilereinschätzungen mittels der Intraklassenkorrelation und Verbesserung von Beurteilereinschätzungen. *Die Rehabilitation, 43*(6), 384-389.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wirtz, M., & Kutschmann, M. (2007). Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Maße. *Die Rehabilitation, 46*(6), 370-377.
- Wissenschaftsrat. (2012). Empfehlungen zur Akkreditierung als Instrument der Qualitätssicherung. Köln.
- Wollersheim, H.-W., März, M., & Schminder, J. (2011). Digitale Prüfungsformate. Zum Wandel von Prüfungskultur und Prüfungspraxis in modularisierten Studiengängen. *Zeitschrift für Pädagogik, 57*(3), 363-374.

Abbildungsverzeichnis

Abbildung 1: Heuristische Kategorisierungshilfe zur Einteilung verschiedener Ansätze und Instrumente zur Kompetenzmessung	13
Abbildung 2: Konzeptionelle Darstellung der Wissensarten zur Einteilung von Anforderungsdimensionen in Prüfungen	21
Abbildung 3: Vorgehensmodell für die Prüfungserstellung und -auswertung	40
Abbildung 4: Ablauf der Fortbildung	73
Abbildung 5: Untersuchungsdesign	79
Abbildung 6: Zusammensetzung der Teams	81
Abbildung 7: Anzahl der bereits existierenden Lehrziele pro Lehrstuhl	114
Abbildung 8: Verteilung der Wissensarten in den bereits existierenden Lehrzielen	114
Abbildung 9: Verteilung der Wissensarten in Lehrzielen nach Fällen.....	115
Abbildung 10: Spezifität der Lehrziele	116
Abbildung 11: Spezifität der Lehrziele nach Lehrstühlen.....	116
Abbildung 12: Anzahl der Lehrziele in der veränderten Praxis	118
Abbildung 13: Spezifität der bestehenden (MZP 1) und der neu erstellten (MZP 2) Lehrziele ..	120
Abbildung 14: Veränderungen in der Spezifität der Lehrziele nach Fällen	121
Abbildung 15: Vergleich der Wissensarten in bestehenden und neu erstellten Lehrzielen	123
Abbildung 16: Unterschiede in den Wissensarten in bestehenden und neu erstellten Lehrzielen	124
Abbildung 17: Formale Gestaltungsmerkmale von Aufgaben.....	132
Abbildung 18: Verteilung der Wissensarten in Lehrzielen und Aufgaben	133
Abbildung 19: Prozentuale Verteilung der Wissensarten in Lehrzielen und Prüfungsaufgaben .	134
Abbildung 20: Anzahl der Aufgaben und Bearbeitungsdauer in den bestehenden Prüfungen	135
Abbildung 21: Eingesetzte Antwortformate in den bestehenden Prüfungsaufgaben	136
Abbildung 22: Art der Antwort in Kurzantwortformaten	136
Abbildung 23: Verwendete Operatoren in den bestehenden Prüfungsaufgaben.....	137
Abbildung 24: Anteil der bestehenden Aufgaben in denen Illustrationen verwendet werden....	138
Abbildung 25: Verteilung der Wissensarten nach Aufgabenanzahl und nach der Anzahl der vergebenen Punkte.....	139
Abbildung 26: Fallweise Verteilung der Wissensarten anhand der vergebenen Punkte	140
Abbildung 27: Funktion von Aufgabenkontexten in den bestehenden Prüfungsaufgaben	141
Abbildung 28: Offenheit der bestehenden Prüfungsaufgaben.....	142
Abbildung 29: Sprachlogische Komplexität der Aufgaben.....	142
Abbildung 30: Lehrzielpyramide mit einer exemplarischen Zuordnung von Aufgaben.....	145
Abbildung 31: Formale Gestaltungsaspekte in den neu erstellten Prüfungen.....	146
Abbildung 32: Passung zwischen Lehrzielen und Aufgaben in der bestehenden und in der veränderten Praxis	147

Abbildung 33: Unterschied zwischen den Differenzen in Aufgaben und Lehrzielen in der bestehenden und veränderten Praxis	148
Abbildung 34: Verteilung der Wissensarten in den bestehenden und neu erstellten Aufgaben...	150
Abbildung 35: Wissensarten in den bestehenden und neu erstellten Prüfungsaufgaben pro Fall	150
Abbildung 36: Verteilung der Wissensarten in der veränderten Prüfung von Fall MW 3;	151
Abbildung 37: Aufgabenoffenheiten pro Fall in den bestehenden und der veränderten Praxis...	152
Abbildung 38: Angabe von Ankerbeispielen, Regelwerken und Punkteverteilung bei den bestehenden Prüfungsaufgaben.....	167
Abbildung 39: Angabe von Ankerbeispielen, Regelwerken und Punkteverteilung bei den bestehenden und neu erstellten Prüfungsaufgabe	169
Abbildung 40: Verteilung der Programmziele auf die sechs Schritte des Prüfungsmodells	189

Tabellenverzeichnis

Tabelle 1: Ursachen und Art der dadurch erzeugten Varianz bei Leistungstests	33
Tabelle 2: Beispiel für eine Lehrzielmatrix	46
Tabelle 3: Auszug aus der Beispielmatrix mit kognitiven Anforderungsniveaus	47
Tabelle 4: Exemplarisches Bewertungsraster	51
Tabelle 5: Beispiel für einen Datensatz zur Erfassung von Prüfungsergebnissen	55
Tabelle 6: Datenquellen zugeordnet zu den sechs Schritten der Prüfungserstellung und - durchführung	56
Tabelle 7: Informationen zu den Fällen aus dem Maschinenwesen	83
Tabelle 8: Anzahl der Prüfungsaufgaben und Lehrziele in den bereits existierenden und den neu erstellten Materialien	86
Tabelle 9: Dauer der Arbeitstreffen pro Fall in Stunden	87
Tabelle 10: Aufbau und Darstellung der Kategoriensysteme	90
Tabelle 11: Einteilung der Übereinstimmungsgüte von Kodierungen	93
Tabelle 12: Erzielte Übereinstimmungswerte bei der Kodierung der Lehrziele	94
Tabelle 13: Erzielte Übereinstimmungswerte bei der Kodierung der Prüfung im Gesamten	95
Tabelle 14: Erzielte Übereinstimmungswerte bei der Kodierung der Einzelaufgaben	95
Tabelle 15: Auszug aus dem Kategoriensystem zur Segmentierung der Verbaldaten	98
Tabelle 16: Beispiel für eine Tabelle zur Darstellung der induktiv gebildeten Kategorien	102
Tabelle 17: Programmziele der Lehrenden zur Festlegung des Prüfungsgegenstands	104
Tabelle 18: Kategorien zur Beschreibung der angestrebten Prüfungsanforderungen	106
Tabelle 19: Referenzpunkte und Auswahlvorschriften bei der Festlegung von Prüfungsinhalten und -anforderungen	110
Tabelle 20: Nutzung der Lehrzielmatrix bei Fall MW 1	125
Tabelle 21: Auszug aus der Lehrzielmatrix von Fall MW 2	127
Tabelle 22: Erkenntnisse zur Festlegung des Prüfungsgegenstands	128
Tabelle 23: Projektziele zur Erstellung von Prüfungsaufgaben	130
Tabelle 24: Wesentliche Erkenntnisse zur Erstellung von Prüfungsaufgaben	155
Tabelle 25: Projektziele zur Rückbindung von Aufgaben an den Prüfungsgegenstand	157
Tabelle 26: Wesentliche Erkenntnisse zur Rückbindung von Aufgaben an den Prüfungsgegenstand	163
Tabelle 27: Projektziele zur Bewertung von Prüfungsantworten	166
Tabelle 28: Maßnahmen bei der Bewertung und Korrektur von Prüfungsantworten	168
Tabelle 29: Wesentliche Erkenntnisse zur Bewertung von Prüfungsantworten	170
Tabelle 30: Exemplarische Notenverteilung bei einer Gesamtpunktzahl von 100	173
Tabelle 31: Wesentliche Erkenntnisse zur Punkte- und Notenvergabe	175
Tabelle 32: Programmziele zur Gewinnung von Rückschlüssen aus den Prüfungsdaten	177

Tabelle 33: Exemplarischer Auszug aus dem Datensatz von Fall MW 5	178
Tabelle 34: Wesentliche Erkenntnisse zu Rückschlüssen aus der Empirie.....	183
Tabelle 35: Kategoriensystem zur Analyse der Lehrziele.....	226
Tabelle 36: Kategoriensystem zur Analyse der Prüfung im Gesamten.....	233
Tabelle 37: Kategoriensystem zur Analyse der Einzelaufgaben	235
Tabelle 38: Kategoriensystem zur Strukturierung der Verbaldaten	259

Anhang

A. Kategoriensysteme	226
a. Kategoriensystem zur Analyse der Lehrziele	226
b. Kategoriensystem zur Analyse der Prüfung im Gesamten	233
c. Kategoriensystem zur Analyse der Einzelaufgaben.....	235
e. Kategoriensystem zur Strukturierung der Verbaldaten.....	259
B. Verwendete Zitate zur Kategorienbildung	264
a. Zitate zur Festlegung des Prüfungsgegenstands	264
b. Zitate zur Erstellung von Prüfungsaufgaben.....	281
c. Zitate zur Rückbindung von Aufgaben an den Prüfungsgegenstand.....	282
d. Zitate zur Bewertung von Prüfungsantworten	283
e. Zitate zur Punkte- und Notenvergabe	287
f. Zitate zu Rückschlüssen aus der Empirie.....	292

A. Kategoriensysteme

a. Kategoriensystem zur Analyse der Lehrziele

Tabelle 35: Kategoriensystem zur Analyse der Lehrziele

Inhaltliche Schwerpunktlegungen von Lehrzielen		
Kategorie 1: Wissensart		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Deklaratives Wissen	<p>„Wissen, dass“</p> <p>Das Lehrziel bezieht sich auf deklaratives Wissen. Deklaratives Wissen beinhaltet Definitionen und Fakten, meistens in der Form von Begriffen, Aufstellungen, Aussagen, Regeln oder Beschreibungen und ist explizit verbalisierbar. <i>Beispiel: Schmelzpunkte verschiedener Materialien bei verschiedenen Temperaturen (Fakt).</i></p> <p>Die Lehrziele können sich beziehen auf Begriffe, Fakten, Definitionen, Aussagen oder Beschreibungen, ohne zu begründen, zu erklären oder Vorhersagen zu treffen, ohne diese in Relation zu anderen Aspekten zu treffen.</p> <p>Es wird nicht verlangt, dass die Studierenden Relationen zwischen Konzepten herstellen oder bestimmte Prinzipien oder Modelle anwenden. In den meisten Fällen kann das Wissen reproduziert werden.</p>	<p><i>Die Studierenden können zentrale Begriffe der Veranstaltung definieren.</i></p>
(2) Prozedurales Wissen	<p>„Wissen, wie“</p> <p>Prozedurales Wissen ist Wissen darüber, wie etwas aus- oder durchgeführt werden kann. Es enthält Wenn-Dann-Produktionsregeln oder eine Sequenz an Schritten, die ausgeführt werden muss. Durch Praxis und zielgerichtetes Üben kann prozedurales Wissen in automatisierte Abläufe überführt werden. Unter prozedurales Wissen fallen typischerweise Berechnungsaufgaben, sofern ein Verständnis dahinterliegender Konzepte nicht angesprochen wird.</p>	<p><i>Die Studierenden können die Dichte von Werkstoffen bestimmen.</i></p> <p><i>Die Studierenden sind in der Lage, gewöhnliche Differentialgleichungen mit einer Veränderlichen zu berechnen.</i></p>
(3) Konzeptionelles Wissen	<p>„Wissen, warum“</p> <p>Konzeptwissen kann verwendet werden, um Probleme zu interpretieren, eine Fehlersuche durchzuführen, zu erklären (was passiert oder passiert ist), zu begründen, Dinge in Relation oder Beziehung zu setzen und um Vorhersagen, die ein logisches Schlussfolgern erfordern, zu treffen, wie zum Beispiel Veränderungen in einem Konzept andere Konzepte beeinflussen (Ursache – Wirkung). <i>Beispiel: Erklärung, warum Tag und Nacht auf der Erde auftreten; Erklärung wie ein bestimmter Virus wirkt, Erklärung welcher Zusammenhang zwischen Aufgaben-Format und kognitiven Prozessen besteht.</i></p>	<p><i>Die Studierenden sind in der Lage, konzeptionelle Schwächen im Produktdesign zu erkennen.</i></p> <p><i>Die Studierenden sind in der Lage, Lösungsvorschläge für Schwächen im Produktdesign zu entwickeln.</i></p>

	Die Lehrziele beziehen sich auf Begründungen, Erklärungen, Argumentationen oder Vorhersagen zu treffen . Konzeptwissen baut auf Fakten- und Prozesswissen auf und stellt Verbindung her und beinhaltet typischerweise die Anwendung wissenschaftlicher Prinzipien, Erklärungsmodellen, Taxonomien, Schemata oder Begriffsnetze oder das argumentieren mit Modellen oder Konzepten. Es müssen z.B. Bedeutungszusammenhänge berücksichtigt werden. Die Studierenden sollen ggf. Relationen zwischen Konzepten herstellen oder bestimmte Prinzipien oder Modelle anwenden. In den meisten Fällen kann das Wissen nicht reproduziert werden. Die Studierenden nutzen zur Lösung entsprechender Aufgaben ihr Wissen über Theorien, Fakten oder Prozesswissen.	
(4) Strategisches Wissen	<p>„Wissen wie, wann, warum“</p> <p>Strategiewissen ermöglicht es, zu entscheiden wann, wo und wie bestimmte Wissensarten in neuen Situationen eingesetzt werden können. Strategiewissen beinhaltet domänenspezifische Strategien wie Planen, Problemlösen und den eigenen Fortschritt zu überprüfen, sowie konditionales Wissen, also Kenntnisse über adäquate Anwendungsbedingungen von Strategien und Wissen.</p> <p>Die Lehrziele erfordern das Interpretieren einer Problemstellung und darauf aufbauend den Vergleich und die begründete Auswählen einer Strategie. Das geeignete Verfahren wird dabei nicht vorgegeben, es müssen selbständig gewählte Verfahren angewendet werden. Personen verwenden strategisches Wissen, um zu erkennen, wann bestimmte Prozeduren ausgeführt werden können, analysieren die Merkmale einer Aufgabe, um zu entscheiden welche Strategie verwendet werden kann, setzen Ziele und überprüfen den Fortschritt bei der Zielerreichung. Allgemeine Anwendungen sind Planungsstrategien, wie das Zerlegen einer Problemstellung, Teilprobleme und deren Kategorisierung und Priorisierung. Strategiewissen kommt vor allem bei neuartigen und unbekanntem Problemstellungen zum Tragen.</p>	<i>Nach der Teilnahme an der Veranstaltung sind die Studierenden in der Lage, sich unter Bezugnahme auf Evidenz aus der Bildungsforschung für bestimmte Vorgehensweisen zur Lösung von Konflikten im Klassenzimmer zu entscheiden.</i>
Kategorie 2: Das Lehrziel enthält einen Kompetenzgrad		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Es ist angegeben, bei welchem Grad des Könnens das Lehrziel als erreicht gilt. Der Grad des Könnens kann z.B. über den Anteil an Aufgaben, der gelöst werden muss, über eine bestimmte Zeitdauer, innerhalb derer die Tätigkeit ausgeführt werden muss, oder einer bestimmten Qualität, in der die Aufgabe ausgeführt werden muss, angegeben werden. Die Dimension Qualität kann sich auf mehrere Dinge beziehen, wie z.B. die Sauberkeit oder Präzision, mit der eine Aufgabe ausgeführt oder ob eine Antwort mit Fachsprache beantwortet wird, anstatt mit Alltagssprache.	<p><i>Die Studierenden sind in der Lage, die gestellten Aufgabenstellungen zu lösen und den gewählten Lösungsweg zu begründen. Zur Begründung verwenden sie ausschließlich treffende Fachbegriffe und Konzepte. Begründungen, in denen Fachbegriffe umschrieben, jedoch nicht explizit verwendet werden, werden nicht mit der vollen Punktzahl gewertet.</i></p> <p><u>Begründung:</u> Dieses Lehrziel hebt klar hervor, dass die Begründungen in einer bestimmten Qualität erfolgen müssen. Die Studieren-</p>

		<p>den benötigen also die Kompetenz nicht nur umgangssprachlich zu begründen, sondern unter Verwendung der richtigen Fachterminologie.</p> <p><u>Beispiel 2:</u> <i>Nach der Teilnahme an den Modulveranstaltungen sind die Studierenden in der Lage, mechanische Fragestellungen der Stereo-Statik, Elasto-Statik, Kinematik und Kinetik unter Zuhilfenahme einer Formelsammlung und eines Taschenrechners (Hilfsmittel) in einer simulierten Problemsituation (Situation) achtzig Prozent der Aufgaben(Grad des Könnens) zu lösen.</i></p> <p><u>Begründung:</u> Bei diesem Lehrziel wird der Kompetenzgrad festgelegt, indem definiert wird, wie viele Problemsituationen erfolgreich bewältigt werden müssen, um einer Person die entsprechende Kompetenz zu attestieren.</p>
(0) Trifft nicht zu	Es ist kein Kompetenzgrad angegeben.	<p><i>Die Studierenden beherrschen die Fremdsprache.</i></p> <p><u>Begründung:</u> Beim Beherrschen einer Fremdsprache könnte das angestrebte Sprachniveau angegeben werden, z.B. mit den Bezeichnungen: Verhandlungssicher, fließend, Schul- bzw. Grundkenntnisse</p>
Formale Gestaltungsmerkmale von Lehrzielen		
Kategorie 3: Lehrziel enthält eine Inhaltskomponente		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Das Lehrziel enthält eine Inhaltskomponente. Durch die Inhaltskomponente können z.B. die Themen einer Veranstaltung benannt werden. Im Ankerbeispiel ist die Inhaltskomponente im Lehrziel unterstrichen.	<i>Die Studierenden können <u>Kostenverläufe</u> unter Vorgabe bestimmter Randbedingungen (z.B. Stückzahl, Bauteilgröße) grafisch darstellen.</i>
(0) Trifft nicht zu	Das Lehrziel enthält keine Inhaltskomponente. Es werden nur Verhaltens- oder Prozesskomponenten aufgeführt.	<i>Die Studierenden sind in der Lage, zu analysieren und Probleme zu lösen.</i>
Kategorie 4: Spezifität der Inhaltskomponente		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Global	Die Inhaltskomponente ist sehr global gefasst. Es werden ausschließlich allgemein gefasste Begriffe verwendet, deren Bedeutung vielfältig sein kann. Wird z.B. der Veranstaltungstitel als Inhaltskomponente eingesetzt (siehe Ankerbeispiel 2), wird diese als global gewertet.	<p><i>Adäquate Aufarbeitung von Inhalten aus sachanalytischer/didaktischer Sicht.</i></p> <p><u>Begründung:</u> Hier ist weder klar, welche Inhalte aufgearbeitet werden,</p>

		<p>noch was „aus sachanalytischer oder didaktischer Sicht“ konkret bedeutet.</p> <p><u>Beispiel 2:</u> <i>Die Studierenden können Fragestellungen der Technischen Mechanik lösen.</i></p> <p><u>Begründung:</u> Zwar ist der Themenbereich der Technischen Mechanik genannt. Die Technische Mechanik beinhaltet jedoch sehr viele Teilbereiche, sodass der Übergriff als global gewertet wird.</p>
(2) Eher unspezifisch	Die Inhaltskomponente enthält übergeordnete Begriffe, die einen größeren Inhaltsbereich bezeichnen.	<p><i>Nach der Teilnahme an den Modulveranstaltungen sind die Studierenden in der Lage, mechanische Fragestellungen der Stereo-Statik, Elasto-Statik, Kinematik und Kinetik zu lösen.</i></p> <p><u>Begründung:</u> Bei diesem Lehrziel ist eindeutig, dass es sich um Fragestellungen aus den mechanischen Teilbereichen Stereo-Statik, Elasto-Statik, Kinematik, und Kinetik handelt. Die Inhaltsbereiche sind klar definiert, innerhalb dieser Inhaltsbereiche gibt es jedoch große Variationsmöglichkeiten, die spezifiziert werden können.</p>
(3) Spezifisch	Die Inhaltskomponente beschreibt die Inhalte spezifisch, indem die zu behandelnden Inhalte konkret benannt werden.	<p><i>Nach der Teilnahme an den Modulveranstaltungen sind die Studierenden in der Lage, mechanische Fragestellungen der Stereo-Statik zu Kräftezerlegung, Kräftepaaren und Momenten, zum Seileckverfahren und Erstarrungsprinzip zu lösen.</i></p> <p><u>Begründung:</u> Der Themenbereich Stereo-Statik wird genauer spezifiziert, indem zu behandelnde inhaltliche Teilbereiche konkret benannt werden.</p>
(4) Sehr spezifisch	Die Inhalte sind konkret benannt und mit weiteren eingrenzenden Informationen versehen. indem z.B. angegeben wird, welcher spezielle Typ von Aufgaben zu einem Inhaltsbereich gelöst werden soll.	<p><i>Nach der Teilnahme an den Modulveranstaltungen sind die Studierenden in der Lage, mechanische Fragestellungen der Stereo-Statik zu Kräftezerlegung, Kräftepaaren und Momenten, zum Seileckverfahren und Erstarrungsprinzip, die ohne Anwendung von Differentialgleichungen bearbeitbar sind, zu lösen.</i></p> <p><u>Begründung:</u> In diesem Beispiel wurden im Vergleich zum Beispiel bei</p>

		der Kategorie 3 spezifisch noch einmal konkretisiert, indem festgelegt ist, dass zu den einzelnen Inhalten nur solche mechanischen Fragestellungen bearbeitet werden, die ohne die Anwendung von Differentialgleichungen beantwortbar sind.
(0) Keine Inhaltskomponente enthalten	Das Lehrziel enthält keine Inhaltskomponente.	Siehe Kategorie 2, Ausprägung 2.
Kategorie 5: Lehrziel enthält eine Verhaltenskomponente		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Verhaltenskomponente vorhanden	<p>Das Lehrziel enthält eine Verhaltenskomponente, die ein Verhalten beschreibt, welches die Studierenden zeigen sollen. Verhalten bezieht sich auf Tätigkeiten, die äußerlich wahrnehmbar sind und unterscheidet sich dadurch von kognitiven Prozessen (siehe Ausprägung 2 in dieser Kategorie). Folgende Verben fallen unter diese Kategorie: nennen, wiedergeben, unterscheiden zwischen, wählen, erkennen, zusammenfügen, anpassen, aufzählen, anwenden, bewerten, analysieren, interpretieren.</p> <p>Substantivierte Formen können ebenfalls eine Verhaltenskomponente enthalten, jedoch nicht in Form eines Verbs, sondern ausgedrückt über ein Nomen. Beispiel: siehe rechts (das Lehrziel wäre in dem Fall nicht als Aussagesatz formuliert, wie gefordert in Item 7).</p>	<p><i>Die Studierenden sind in der Lage, einen Stoffverteilungsplan anzufertigen.</i></p> <p><u>Beispiel 2:</u> <i>Nach der Teilnahme an den Modulveranstaltungen sind die Studierenden in der Lage, mechanische Fragestellungen der Stereo-Statik, Elasto-Statik, Kinematik und Kinetik zu lösen.</i></p> <p><u>Begründung:</u> Bei diesem Beispiel ist klar, dass Studierende in der Lage sind, Aufgaben eines bestimmten Typs zu lösen. Das Lösen von Aufgaben kann direkt beobachtet werden, indem Studierende eine Lösung angeben.</p> <p><u>Beispiel für substantivierte Form:</u> <i>Anfertigen von Stoffverteilungsplänen durch die Studierenden.</i></p> <p><u>Begründung:</u> Das Lehrziel ist zwar nicht als Aussagesatz mit einem Verb formuliert, jedoch ist die Verhaltenskomponente im Verb „anfertigen“ erkennbar.</p>
(2) Kognitive Prozesskomponente	<p>Das Lehrziel ist mit mit Verben beschrieben, die kein beobachtbares Verhalten bezeichnen (verstehen, nachvollziehen).</p> <p>Das Lehrziel enthält eine Prozesskomponente, die einen kognitiven Prozess benennt. Prozesskomponenten sind Vorgänge im Gehirn, die äußerlich nicht unmittelbar beobachtbar sind. Hinweiswörter können z.B. verstehen, nachvollziehen, wissen, verstehen, anerkennen, nachvollziehen, begreifen sein.</p>	<p><i>Die Studierenden sind in der Lage, die psychologischen Grundlagen zur Leistungsbeurteilung zu verstehen.</i></p> <p><u>Beispiel 2:</u> <i>Die Studierenden sind in der Lage, die zentralen Konzepte aus der Lehrveranstaltung verstehend nachzuvollziehen.</i></p> <p><u>Beispiel für substantivierte Form:</u> <i>Die Studierenden haben einen Überblick über die Gütekriterien der Leistungsmessung.</i></p>

(3) Kognitive Prozesskomponente spezifiziert durch Verhalten	Das Lehrziel enthält eine kognitive Komponente, die durch die Angabe einer Verhaltenskomponente spezifiziert bzw. operationalisiert ist.	<i>Die Studierenden haben einen Überblick über die Gütekriterien der Leistungsmessungen, indem sie für eine eigens erstellte Prüfung darlegen, wie die Gütekriterien berücksichtigt wurden.</i>
(0) Weder noch	Keine Verhaltenskomponente bzw. kognitive Prozesskomponente vorhanden.	<i>Folgende Themen werden in der Veranstaltung thematisiert: Validität und Reliabilität.</i> <u>Begründung:</u> Das Lehrziel enthält eine Aufzählung von Inhalten, jedoch fehlt die Angabe, was die Studierenden an diesen Inhalten können sollen.
Kategorie 6: Spezifität der Verhaltenskomponente		
Kodierregel: Das Item wird für Lehrziele verwendet, die bei Kategorie 4 entweder die Ausprägung 1 (Verhaltenskomponente vorhanden) oder 3 (kognitive Prozesskomponente spezifiziert durch Verhalten) erhielten.		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Global	Die Verhaltenskomponente ist sehr global gefasst. Es werden ausschließlich allgemein gefasste Begriffe verwendet. Es kann sehr unterschiedlich interpretiert werden, was der verwendete Begriff bedeutet. Es kann sehr unterschiedlich ausgelegt werden, was am Inhalt gekonnt werden soll. Damit sind auch die Anforderungen interpretierbar, die in entsprechenden Aufgaben abzubilden sind. Das Verhalten ist nicht operationalisiert bzw. messbar. Das Lehrziel eignet sich nicht für eine zielorientierte Aufgabenerstellung.	<i>Adäquate Aufarbeitung von Inhalten aus sachanalytischer/didaktischer Sicht.</i> <u>Begründung:</u> Die Bezeichnung „Aufarbeitung“ ist sehr global, da sehr unterschiedlich interpretiert werden kann, was ein Aufarbeiten ausmacht, bzw. in welchen Handlungen sich die Aufarbeitung niederschlägt.
Eher unspezifisch (2)	Die Verhaltenskomponente enthält übergeordnete Begriffe (z.B. Analysieren, Bewerten, Problemlösen), die zwar einen Verhaltensbereich kennzeichnen, innerhalb die jedoch weiter interpretiert werden können. Das Vorgehen beim Analysieren von Sachverhalten, Produkten, etc. erfordert zwar im Normalfall immer eine Zerlegung von Gesamtzusammenhängen in einzelne Einheiten. Es wird aber nicht eindeutig klar, wie die Analyseschritte aussehen und wie feinkörnig die Analyse erfolgt bzw. auch, ob eine spezielle Analyseform verwendet werden soll. Bei diesen Lehrzielen wird ersichtlich, dass es sich um eine bestimmte Anforderung handelt, die in Aufgaben umgesetzt werden muss. Jedoch bleibt gleichzeitig ein Interpretationsspielraum hinsichtlich der Ausdifferenzierung der Anforderungen bzw. dahingehend bestehen, was am Inhalt konkret gekonnt werden soll. Es erfolgt keine Konkretisierung durch Bezugnahme des Verhaltens auf einen Inhalt.	<i>Lösen von mathematischen Problemstellungen.</i> <u>Begründung:</u> Lösen ist das Überführen eines Ist-Zustandes in einen gewünschten Soll-Zustand durch bewusstes, durchdachtes Handeln. Es herrscht ein allgemeingültiges Verständnis darüber, was unter der Fähigkeit des Lösens gemeint ist. Jedoch gibt es verschiedene Ausprägungen. Anders gesagt: Das Lösen verschiedener Problemstellungen kann verschiedenartige Handlungen erfordern, auch wenn der Ablauf des Lösungsprozesses ein ähnlicher ist (es geht immer darum einen Ist- in einen Soll stand zu überführen). Weitere Beispiele sind: bewerten, modellieren, konzipieren.
Spezifisch (3)	Die Verhaltenskomponente ist dann spezifisch, wenn bekannte Begriffe verwendet werden, die wenig Interpretationsspielraum an der Art des Verhaltens zulassen und über deren Bedeutung ein allgemein anerkanntes Verständnis herrscht.	<i>Die Studierenden sind in der Lage, Begriffe zu gegebenen Definitionen zuzuordnen.</i>

	<p>Es wird erkenntlich, welche konkreten Anforderungen in Aufgaben enthalten sein müssen, um das Lehrziel abzutüpfen, jedoch bleibt ein möglicher Wahlraum an Aufgaben zur Überprüfung möglich. Beispielsweise: Können unterschiedliche Zuordnungsaufgabentypen gewählt werden?</p> <p>Eher unspezifische Verhaltenskomponenten können durch eine konkrete Bezugnahme auf einen Inhalt - bzw. in welcher Form oder wie das Verhalten am Inhalt ausgeübt werden soll - spezifiziert werden.</p>	<p>Begründung: Mit der Anweisung einer Zuordnung ist eine spezifische Handlung bezeichnet, auch wenn die Inhaltskomponente relativ vage bleibt. Die Angabe könnte weiter spezifiziert werden, etwa durch eine weitere Spezifizierung des Prozesses des Zuordnens, z.B.: Die Studierenden sollen zueinander passende Begriffe und Definitionen mit Pfeilen verbinden.</p>
Sehr spezifisch (4)	<p>Die Verhaltenskomponente ist dann sehr spezifisch, wenn kein Interpretationsspielraum darüber existiert, in welchen Handlungen sich das Verhalten niederschlägt. Es ist klar ersichtlich, welche Anforderungen in Aufgaben zur Überprüfung dieses Lehrziels abgebildet sein müssen.</p>	<p><i>Die Studierenden sind in der Lage, Begriffe zu gegebenen Definitionen mit Hilfe von Concept Maps zuzuordnen.</i> <i>Begründung: In diesem Beispiel muss das Lehrziel mittels eines Concept Maps überprüft werden.</i></p> <p>Begründung: Bei diesem Beispiel ist der Verhaltensaspekt des Lehrziels sehr genau definiert, da vorgegeben ist, mit welchem Verfahren Begriffe zu Definitionen zugeordnet werden sollen. Damit ist sehr spezifisch angegeben, welche Anforderungen in entsprechenden Aufgaben abzubilden sind.</p>
(0) Keine Verhaltenskomponente enthalten	<p>Es ist keine Verhaltenskomponente enthalten.</p>	<p>Siehe Kategorie 4, Ausprägung 0</p>

b. Kategoriensystem zur Analyse der Prüfung im Gesamten

Tabelle 36: Kategoriensystem zur Analyse der Prüfung im Gesamten

Instruktionsqualität		
Kategorie 1: Hinweise zur Bearbeitung der Prüfungsaufgaben		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Die Prüfung enthält Hinweise zur Bearbeitung der Aufgaben und welche Dinge beachtet werden sollen.	<i>Antworten müssen eindeutig einer Aufgabe zuzuordnen sein. Beispiel 2: Unlesbare, mehrdeutige oder fachunspezifische (allgemeingültige) Antworten werden nicht gewertet.</i>
(0) Trifft nicht zu	Die Prüfung enthält keine Hinweise zur Bearbeitung.	
Kategorie 2: Hinweise zur Bewertung von Antworten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Auf dem Deckblatt der Klausur sind Hinweise zur Bewertung (<u>nicht</u> Bepunktung) der Aufgaben gegeben, z.B.: welche Antworten werden bewertet oder nicht mehr bewertet, z.B. aufgrund schlechter Leserlichkeit?	<i>Ergebnisse ohne erkennbaren Rechenweg führen nicht zum Erhalt der vollen Punktzahl.</i>
(0) Trifft nicht zu	Auf dem Deckblatt sind keine Hinweise zur Bewertung von Aufgaben gegeben.	
Kategorie 3: Prüfung enthält Überschriften		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Thematische Überschriften	Die Prüfung enthält durchgehend Überschriften zu Themenbereichen (<u>keine</u> Gliederung nach Aufgabentypen wie Kurzfragen oder offene Fragen).	<i>1. Differentialrechnung</i> <ul style="list-style-type: none"> • Aufgabe 1 • Aufgabe 2 • <i>2. Integralrechnung</i> <ul style="list-style-type: none"> • Aufgabe 1 • Aufgabe 2 •
(2) Überschriften zu Aufgabentypen	Die Prüfung enthält durchgängig Überschriften zu Aufgabentypen. Die Bezeichnungen von Aufgabentypen können feststehende Unterscheidungen sein, wie „Wahlantworten“, „Kurzantworten“, „Zuordnungsaufgaben“, oder eigens gewählte Typen von Aufgaben wie „Wissensaufgaben“, „Transferaufgaben“, „Berechnungsaufgaben“, usw.	<i>1. Wissensaufgaben</i> <ul style="list-style-type: none"> • Aufgabe 1 • Aufgabe 2 • <i>2. Transferaufgaben</i> <ul style="list-style-type: none"> • Aufgabe 1
(0) Keine Überschriften enthalten	Die Prüfung enthält keine oder nur teilweise Überschriften zu Themenbereichen	

Kenndaten zur Prüfung		
Kategorie 4: Angabe der Bearbeitungsdauer		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Auf dem Deckblatt der Klausur wird die zur Verfügung stehende Zeit angegeben.	<i>Bearbeitungsdauer 90 Minuten</i>
(0) Trifft nicht zu	Auf dem Deckblatt der Klausur wird die zur Verfügung stehende Zeit nicht angegeben.	
Kategorie 5: Allgemeine Hinweise zur Bearbeitung der Prüfung		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Auf dem Deckblatt der Prüfung werden allgemeine Hinweise zur Bearbeitung gegeben, die sich z.B. auf die erlaubten Hilfsmittel, eine Regelung zum Verlassen des Raums, die Möglichkeiten zur Abgabe vor Ablauf der Prüfungszeit, die Regelungen bzgl. eines Rücktritts von der Prüfung, usw. beziehen	<i>Zugelassene Hilfsmittel sind Schreib- und Zeichengeräte, Fremdsprachen-Wörterbücher ohne Anmerkungen und nichtprogrammierbare Taschenrechner. <u>Beispiel 2:</u> <i>In den letzten 30 Minuten der Prüfung kann nicht mehr vorzeitig abgegeben werden.</i> <u>Beispiel 3:</u> <i>Sollten Sie den Prüfungsraum während der Prüfungszeit verlassen müssen, sind die Prüfungsunterlagen solange bei der Aufsicht abzugeben. Es kann immer nur eine Person den Raum verlassen.</i></i>
(0) Trifft nicht zu	Es sind keine allgemeinen Hinweise zur Bearbeitung der Prüfung vorhanden.	

c. Kategoriensystem zur Analyse der Einzelaufgaben

Tabelle 37: Kategoriensystem zur Analyse der Einzelaufgaben

Differenzierung von Aufgabentypen										
Kategorie 1: Aufgabentyp										
Ausprägung	Kodierregel	Ankerbeispiel								
(1) Zuordnungsaufgabe	In der Aufgabe wird verlangt, Begriffe zu Aussagen, Begriffen zu Grafiken oder Aussagen zu Fakten zuzuordnen.	<p>Verbinden Sie die zueinander passenden Begriffe und Beschreibungen:</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%;">Begriff 1</td> <td style="width: 50%;">Beschreibung 1</td> </tr> <tr> <td>Begriff 2</td> <td>Beschreibung 2</td> </tr> <tr> <td>Begriff 3</td> <td>Beschreibung 3</td> </tr> <tr> <td>Begriff 4</td> <td>Beschreibung 4</td> </tr> </table>	Begriff 1	Beschreibung 1	Begriff 2	Beschreibung 2	Begriff 3	Beschreibung 3	Begriff 4	Beschreibung 4
Begriff 1	Beschreibung 1									
Begriff 2	Beschreibung 2									
Begriff 3	Beschreibung 3									
Begriff 4	Beschreibung 4									
(2) Entscheidungsfrage	Die Aufgabe enthält genau zwei vorgegebene Wahlantworten, aus denen die richtige Antwort(en) ausgewählt werden muss/müssen.	<p>Entscheiden Sie, ob die folgende Aussage richtig ist:</p> <p>Ja Nein</p> <p><input type="checkbox"/> <input type="checkbox"/> Aussage</p>								
(3) Wahlantwortformat mit mehr als zwei vorgefertigten Wahlantworten	Die Aufgabe enthält mehr als zwei vorgegebene Wahlantworten, aus denen die richtige/die richtigen Antwort(en) ausgewählt werden muss/müssen.	<p>Kreuzen Sie die richtige(n) Aussage(en) an:</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 70%;">Aussage 1</td> <td style="width: 30%; text-align: center;"><input type="checkbox"/></td> </tr> <tr> <td>Aussage 2</td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> <tr> <td>Aussage 3</td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> <tr> <td>Aussage 4</td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> </table>	Aussage 1	<input type="checkbox"/>	Aussage 2	<input type="checkbox"/>	Aussage 3	<input type="checkbox"/>	Aussage 4	<input type="checkbox"/>
Aussage 1	<input type="checkbox"/>									
Aussage 2	<input type="checkbox"/>									
Aussage 3	<input type="checkbox"/>									
Aussage 4	<input type="checkbox"/>									
Offene Kurzantwort (4)	<p>Die Antworten sollen in Form einzelner Begrifflichkeiten oder weniger kurze Sätze formuliert werden. Die Studierenden sollen selbstständig eine (kurze) Antwort formulieren (und nicht aus vorgegebenen Antwortmöglichkeiten auswählen). Die Antworten beschränken sich auf das Aufzählen von Begriffen, Aussagen (stichpunktartig), Definitionen oder Fakten (Vor- und Nachteile aufzählen). Das Lösen von Berechnungsaufgaben zählt ebenfalls dazu, sofern der Lösungsweg eindimensional ist, also nur eine Rechnung erforderlich ist und nicht selbstständig mehrere Lösungsschritte konstruiert und/oder kombiniert werden müssen.</p> <p>Weitere Erkennungsmerkmale: Die Aufgabe muss so formuliert sein, dass es nur eine richtige, <u>kurze</u> Antwort gibt. Eventuell ist das Lösungsfeld im Umfang begrenzt. Man erwartet eine bestimmte Antwort.</p>	<p>Nennen Sie drei Vorteile von Multiple Choice-Prüfungen im Gegensatz zu offenen Antwortformaten:</p> <p>Vorteil 1: _____</p> <p>Vorteil 2: _____</p> <p>Vorteil 3: _____</p> <p><u>Beispiel 2:</u> Nach welchen Kriterien entscheiden Sie, welchen Aufgabentyp Sie verwenden? Geben Sie drei Kriterien an.</p> <p>_____</p> <p>_____</p> <p>_____</p> <p><u>Beispiel 3:</u> Berechnen Sie die Lösung der gegebenen Gleichung.</p>								
(5) Offene Langantwort	Die Aufgabenanforderungen sind in der Regel multidimensional, d.h. es werden mehrere Anforderungsbereiche im Sinne zu verarbeitender Inhalte und/oder kogni-	<p>Sie wollen die Hypothese überprüfen, dass Studierende ihr Lernverhalten an den antizipierten Prüfungsanforderungen</p>								

	<p>tiver Anspruchsniveaus angesprochen. Typischerweise müssen Problemsituationen bearbeitet werden, die konzeptionelles Verständnis und/oder strategisches Wissen erfordern. Oftmals müssen Situationen analysiert werden, die sich auf mehrere Inhaltsbereiche beziehen, oder Problemlösungen entwickelt und/oder durchgeführt werden. Zusätzlich zur Lösung von Problemen wird gefordert, den gewählten Lösungsweg zu begründen. Die geforderten Antworten sind in ganzen Sätzen zu formulieren, die eine Argumentationskette bilden. Antwortfelder sind nicht vorstrukturiert, es wird viel Platz zur Bearbeitung gegeben. Die Studierenden müssen selbstständig konstruieren bzw. die Antwort erfolgt nicht in Form kurzer Aussagen, sondern durch eine Aneinanderreihung mehrerer Aussagen. Dies kann auch auf Konstruktions- und Problemlöseaufgaben zutreffen, in denen mehrere Konstruktions- bzw. Lösungsschritte gefunden und berücksichtigt werden müssen, die voneinander abhängig sind.</p>	<p><i>ausrichten. Erstellen Sie einen Forschungsplan und beschreiben Sie das Design ihrer Studie, die Stichprobe, die Erhebungsinstrumente und Auswertungsverfahren. Begründen Sie ihre Entscheidungen.</i></p>
--	--	---

Kategorie 2: Verwendete Handlungsaufforderungen (Operatoren) im Aufgabenstamm

Bei dieser Kategorie sind mehrere Kodierregeln zu beachten:

Regel 1: Bei manchen Aufgaben kommt es vor, dass die Aufgabenbeschreibung bereits ein Verb im Sinne einer Handlungsaufforderung enthält, die jedoch zur Beantwortung der Frage nicht direkt durchgeführt, sondern durch die Angabe weiterer Handlungsaufforderungen spezifiziert wird. In diesem Fall werden nur die tatsächlich auszuführenden Handlungsaufforderungen eingeschätzt.

Beispiel:

Beschreiben Sie die technischen Einschränkungen in der Umsetzung der geforderten Eigenschaftsausprägungen eines Fahrzeugs vom Typ xy im Rahmen der Produktstrategie z. a) Nennen Sie hierzu zwei Aspekte.

b) Geben Sie dazu jeweils eine kurze Erläuterung.

Dieses Aufgabenbeispiel beginnt mit der Handlungsaufforderung „Beschreiben Sie“. Die Studierenden sollen jedoch nicht allgemein beschreiben, sondern es ist vorgegeben, dass die Beschreibung über die Nennung von zwei Aspekten erfolgt, mit einer dazugehörigen Erläuterung. Daher werden nur diese beiden Handlungsaufforderung eingeschätzt, da sie tatsächlich von den Studierenden ausgeführt werden.

Regel 2: Kommen in einer Aufgabe mehrere Handlungsaufforderungen vor, wird die dominantere gewählt.


Beispiel 1: *Skizzieren und beschriften Sie ein Diagramm.*


In dem Fall wird nur skizzieren geratet, da zum Skizzieren eines Diagramms die Beschriftung in aller Regel dazu gehört.

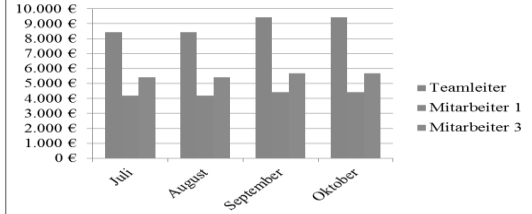

Beispiel 2: *Nennen Sie dazu jeweils zwei Beispiele.*

In dem Fall wird die Handlungsaufforderung mit 9 „Nennen von Beispielen“ gewertet, da der Anspruch der Aufgabe auf Beispiele abzielt.

Ausprägung	Kodierregel	Ankerbeispiel
(1) Nennen	Die Handlungsaufforderung enthält ein „Nennen Sie“.	<i>Nennen Sie drei Vorteile von Schutzgasschweißverfahren im Vergleich zum Elektroschweißen.</i>
(2) Definieren	Die Handlungsaufforderung verlangt nach Definitionen.	<i>Definieren Sie den Begriff....</i>
(3) Beschreiben	Die Handlungsaufforderung verlangt nach Beschreibungen.	<i>Beschreiben Sie,</i>
(4) Berechnen	Die Handlungsaufforderung verlangt nach Berechnungen.	<i>Berechnungen Sie die folgende Gleichung.</i>
(5) Begründen	Die Handlungsaufforderung verlangt nach Begründungen.	<i>Geben Sie eine Begründung für....</i>
(6) Erklären	Die Handlungsaufforderung verlangt nach Erklärungen.	<i>Erklären Sie, warum Ebbe und Flut auftreten.</i>

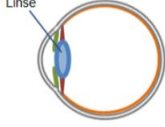
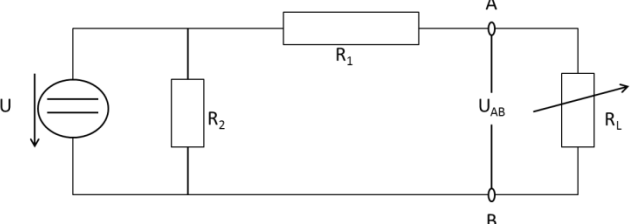
(7) Skizzieren oder zeichnen	Die Handlungsaufforderung verlangt, etwas zu skizzieren oder etwas zu zeichnen.	<i>Skizzieren sie den Kurvenverlauf in einem Diagramm.</i>																		
(8) Nennen von Beispielen	Die Studierenden sollen ein Beispiel angeben.	<i>Nennen Sie ein Beispiel zu...</i>																		
(9) Erklären anhand von Beispielen	Die Studierenden sollen anhand eines Beispiels etwas erklären	<i>Erklären Sie den Ansatz xy anhand eines Beispiels.</i>																		
(10) Andere	Die Aufgabe enthält eine Handlungsaufforderung, die nicht aufgeführt ist.																			
(0) Keine Handlungsaufforderung enthalten	Die Aufgabe enthält keine Handlungsaufforderung oder die Aufgabe ist als Frage formuliert.																			
Kategorie 3: Form der Antwort																				
Ausprägung	Kodierregel	Ankerbeispiel																		
1. Prosa	Die zu gebende Antwort erfolgt durch Wörter, Begriffe, die Angabe von Zahlenwerten (ohne Berechnung) oder die Anfertigung eines Antworttextes. <u>Zusatzregel:</u> Tabellen, in die lediglich in Schriftform Begriffe oder Werte eingetragen werden müssen, zählen zu dieser Kategorie.	<i>Beschreiben Sie den Zweck einer Gewinn- und Verlustrechnung.</i> <u>Beispiel 2:</u> <i>Vervollständigen Sie die folgende Tabelle, indem Sie die fehlenden Begriffe ergänzen.</i> <table style="margin-left: auto; margin-right: auto;"> <tr> <td colspan="2" style="text-align: center;"><i>Bilanz</i></td> </tr> <tr> <td style="text-align: center;"><i>Aktiva</i></td> <td style="text-align: center;"><i>Passiva</i></td> </tr> <tr> <td style="text-align: center;"><i>I. Anlagevermögen</i></td> <td style="text-align: center;"><i>I.</i></td> </tr> <tr> <td style="text-align: center;"><i>Grundstücke</i></td> <td style="text-align: center;"><i>Rücklagen</i></td> </tr> <tr> <td style="text-align: center;"><i>Gebäude</i></td> <td></td> </tr> <tr> <td style="text-align: center;"><i>Fuhrpark</i></td> <td></td> </tr> <tr> <td style="text-align: center;"><i>II.</i></td> <td style="text-align: center;"><i>II. Fremdkapital</i></td> </tr> <tr> <td style="text-align: center;"><i>Vorräte</i></td> <td style="text-align: center;"><i>Hypotheken</i></td> </tr> <tr> <td style="text-align: center;"><i>Forderungen aus Lieferungen u. Leistungen</i></td> <td></td> </tr> </table>	<i>Bilanz</i>		<i>Aktiva</i>	<i>Passiva</i>	<i>I. Anlagevermögen</i>	<i>I.</i>	<i>Grundstücke</i>	<i>Rücklagen</i>	<i>Gebäude</i>		<i>Fuhrpark</i>		<i>II.</i>	<i>II. Fremdkapital</i>	<i>Vorräte</i>	<i>Hypotheken</i>	<i>Forderungen aus Lieferungen u. Leistungen</i>	
<i>Bilanz</i>																				
<i>Aktiva</i>	<i>Passiva</i>																			
<i>I. Anlagevermögen</i>	<i>I.</i>																			
<i>Grundstücke</i>	<i>Rücklagen</i>																			
<i>Gebäude</i>																				
<i>Fuhrpark</i>																				
<i>II.</i>	<i>II. Fremdkapital</i>																			
<i>Vorräte</i>	<i>Hypotheken</i>																			
<i>Forderungen aus Lieferungen u. Leistungen</i>																				
(2) Formalsprache (Formeln und Berechnungen)	Die Antwort erfolgt durch Berechnung (werden Zahlen im Rahmen von Berechnungen verwendet, zählt die Antwort zur Kategorie Formalsprache) oder die Angabe von Formeln bzw. die Angabe von mathematischen Begriffen oder Sätzen in Formalsprache.	<i>Geben Sie den Gauß'schen Integralsatz an.</i> <i>Lösung</i> $\int_V \operatorname{div} \vec{F} \, d^{(n)}V = \oint_S \vec{F} \cdot \vec{n} \, d^{(n-1)}S.$																		
(3) Grafik oder Zeichnung	Die Studierenden fertigen eine Zeichnung oder Grafik (z.B. ein Diagramm oder eine Kurve im Diagramm, oder eine Ergänzung einer vorgegebenen Grafik, nicht aber deren Beschriftung mit Begriffen, Zeichen oder Zahlen) an. Die Handlung, die Studierende ausführen, ist das Zeichnen oder Skizzieren (im Sinne eines Aufzeichnens).	<i>Zeichnen Sie eine S-Kurve in Abhängigkeit der gegebenen Daten in das untenstehende Diagramm.</i> 																		
(4) Ankreuzen	Die Antwort erfolgt durch Ankreuzen.	<i>Kreuzen Sie die richtige(n) Aussage(n) an:</i> Aussage 1 <input type="checkbox"/> Aussage 2 <input type="checkbox"/>																		

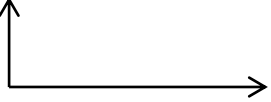

Form der Antwort freigestellt (5)	Es ist freigestellt wie die Fragestellung bzw. die Lösung der Problemstellung erfolgt (z.B. in Schriftform, Formalsprache oder mittels Zeichnung). Bitte Musterlösung beachten. Sollte in der Musterlösung nur eine Form der Beantwortung dargestellt sein (z.B. eine Berechnung) ist anzunehmen, dass die Beantwortung nicht freigestellt ist, sondern lediglich die Aufgabenstellung uneindeutig bleibt.	<i>Geben Sie die Formel oder einen Weg zur Berechnung der Varianz in einer Stichprobe an.</i>																				
Kategorie 4: Aufgabe enthält eine Illustration Regel 1: Diagramme sind immer Illustrationen. Regel 2: Sofern Tabellen lediglich zur Strukturierung der Antwort dienen, werden sie nicht als Illustration gewertet. Tabellen werden als Illustrationen gewertet, sofern die Lösung der Aufgabe von der Tabelle beeinflusst ist (z.B. bei Matrizen bei denen das Ausfüllen der Zellen davon abhängig was in den entsprechenden Spalten und Zeilen der Matrix steht). Regel 3: Wird als Lösung einer Aufgabe eine Illustration verlangt, die jedoch nicht in der Aufgabenstellung angegeben ist, sondern von den Studierenden von Grund auf angefertigt werden muss, enthält die Aufgabenstellung <u>keine</u> Illustration. Dies gilt auch dann, wenn nur ein kleiner Teil der Zeichnung vorhanden ist, z.B. die Achsen eines Diagramms. Regel 4: Beziehen sich mehrere Folgeaufgaben auf eine Illustration in einem übergeordneten Aufgabentext, wird die Ausprägung 5 (schon geratet) kodiert.																						
Ausprägung	Kodierregel	Ankerbeispiel																				
(1) Grafik	Die vorgegebene Illustration ist eine Grafik, z.B. eine Concept Map, eine Blockliste oder ein Schema.	 <pre> graph TD A[Kosten- und Leistungsrechnung] --> B[Kostenartenrechnung] A --> C[Kostenstellenrechnung] A --> D[Kostenträgerrechnung] </pre>																				
(2) Tabelle	Die vorgegebene Illustration ist eine Tabelle. Sofern Tabellen lediglich zur Strukturierung der Antwort dienen, werden sie nicht als Illustration gewertet. Tabellen werden als Illustrationen gewertet sofern die Lösung der Aufgabe von der Tabelle beeinflusst ist (z.B. bei Matrizen, bei denen das Ausfüllen der Zellen davon abhängig ist, was in den entsprechenden Spalten und Zeilen der Matrix steht).	<p><i>Zum September wird eine Tarifierhöhung in Höhe von 3,2 Prozent wirksam. Berechnen Sie die Monatsgehälter für die aufgeführten Mitarbeiterinnen und Mitarbeiter.</i></p> <table border="1" data-bbox="1429 1168 2087 1337"> <thead> <tr> <th>Personalkosten</th> <th>Juli</th> <th>August</th> <th>September</th> <th>Oktober</th> </tr> </thead> <tbody> <tr> <td>Teamleiter</td> <td>8443,60 €</td> <td>8443,60 €</td> <td></td> <td></td> </tr> <tr> <td>Mitarbeiter 1</td> <td></td> <td></td> <td>4433,33€</td> <td>44333,33€</td> </tr> <tr> <td>Mitarbeiter 3</td> <td></td> <td></td> <td></td> <td>5677,44€</td> </tr> </tbody> </table>	Personalkosten	Juli	August	September	Oktober	Teamleiter	8443,60 €	8443,60 €			Mitarbeiter 1			4433,33€	44333,33€	Mitarbeiter 3				5677,44€
Personalkosten	Juli	August	September	Oktober																		
Teamleiter	8443,60 €	8443,60 €																				
Mitarbeiter 1			4433,33€	44333,33€																		
Mitarbeiter 3				5677,44€																		

(3) Diagramm	Die vorgegebene Illustration ist ein Diagramm.	
(4) Bild	Die vorgegebene Illustration ist ein Bild.	 <p data-bbox="1422 614 1570 639">Biodiesel</p> <p data-bbox="1592 419 2080 518">Quelle: http://www.pisa.tum.de/fileadmin/w00bgi/www/Beispielaufgaben/Released_Cognitiv_Items_Nawi_FT_2014_DEU.pdf</p>
(0) Keine Illustration vorhanden	Es sind keine Illustration vorhanden. Diese Kategorie betrift alle Aufgaben, bei denen keine Illustration vorhanden ist.	

Kategorie 5: Verwendung der Illustration

- Regel 1: Diagramme sind immer Illustrationen. Es muss jedoch entschieden werden, ob darin gearbeitet wird, ob daraus Informationen zu entnehmen sind oder ob lediglich die Aufgabenstellung visualisiert wird.
- Regel 2: Sofern Tabellen lediglich zur Strukturierung der Antwort dienen, werden sie nicht als Illustration gewertet. Tabellen werden als Illustrationen gewertet, sofern die Lösung der Aufgabe von der Tabelle dadurch beeinflusst ist (z.B. bei Matrizen, bei denen das Ausfüllen der Zellen davon abhängig ist, was in den entsprechenden Spalten und Zeilen der Matrix steht).
- Regel 3: Wird am Anfang einer Aufgabe, die aus mehreren Teilaufgaben besteht, eine Illustration verwendet, können sich nur Folgeaufgaben (Definition für Folgeaufgabe siehe Kategorie 12) darauf beziehen.
- Regel 4: Bei Folgeaufgaben wird die in der vorherigen Aufgabe vorhandene Illustration betrachtet. Stellt sich heraus, dass die Illustration aus der vorangegangenen Aufgabe für die Beantwortung der aktuellen Folgeaufgabe keine originären Information enthält, wird in dem Fall Kategorie 5 „keine Illustration vorhanden“ geratet. Bei Folgeaufgaben kann demnach die Kategorie 1 nicht gewählt werden, da in dem Fall, wenn keine originären Informationen enthalten sind die Kategorie „keine Illustration vorhanden“ gewählt wird.
- Regel 5: Sobald eine Illustration sowohl originäre Informationen enthält und auch darin gearbeitet werden muss, wird Kategorie 3 (in der Illustration wird gearbeitet) geratet.
- Regel 6: Wird als Lösung einer Aufgabe eine Illustration verlangt, die jedoch nicht in der Aufgabenstellung vorgezeichnet wird, sondern von den Studierenden angefertigt werden muss, enthält die Aufgabenstellung keine Illustration.

Ausprägung	Kodierregel	Ankerbeispiel
(1) Illustration dient reiner Visualisierung	Die Illustration visualisiert die Aufgabenstellung. Sie steht im Bezug zur Aufgabenstellung. Die Illustration enthält keine Informationen, die für die Lösung der Aufgabenstellung relevant sind. Es müssen für die Aufgabenbearbeitung keine Informationen daraus abgelesen werden. Es sind nur die Informationen enthalten, die im Aufgabentext enthalten sind und keine zusätzlichen Informationen. Ein Hinweis könnte auch sein, dass nicht direkt auf die Illustration verwiesen wird.	<p>Das Konzept der verstellbaren Linsen ist nicht neu. Das menschliche Auge besitzt auch eine Linse, die verstellbar ist.</p>  <p>Die Form der Augenlinse wird durch Muskelaktivität verstellt. Warum ist es wichtig, dass die Augenlinse ihre Form verändern kann?</p> <ul style="list-style-type: none"> <input type="radio"/> Um das Sehen von Gegenständen mit unterschiedlichen Helligkeiten zu ermöglichen <input type="radio"/> Um das Sehen von Gegenständen mit unterschiedlichen Farben zu ermöglichen <input type="radio"/> Um das Sehen von Gegenständen in unterschiedlichen Entfernungen zu ermöglichen <input type="radio"/> Um das Sehen von Gegenständen mit unterschiedlichen Größen zu ermöglichen <p>Begründung: Die Fragestellung in dieser Aufgabenstellung wäre auch ohne die Abbildung der Linse zu beantworten, da nicht spezifisch auf einen Aspekt in der Abbildung Bezug genommen wird, der zur Beantwortung der Aufgaben notwendig wäre. Dennoch illustriert die Abbildung den Zusammenhang von Muskelaktivität und Form der Linse. Quelle: http://www.pisa.tum.de/fileadmin/w00bgi/www/Beispielaufgaben/Released_Cognitiv_Items_Nawi_FT_2014_DEU.pdf</p>
(2) Illustration enthält originäre/zusätzliche Informationen	Aus der Illustration sind zusätzliche Informationen zu entnehmen, die nicht Teil der schriftlichen Aufgabenstellung sind. Ohne die Visualisierung bzw. den darin enthaltenen Informationen ist es nicht möglich, die Aufgabenstellung zu lösen. Innerhalb der Visualisierung muss jedoch nicht gearbeitet werden. Die Illustration dient als Informationsquelle, es müssen Informationen entnommen werden. Ein Hinweis könnte sein, wenn direkt auf die Illustration und daraus zu entnehmende Werte verwiesen wird.	<p>Gegeben ist folgende Schaltung:</p>  <p>Stellen Sie den Lastwiderstand R_L so ein, dass in diesem eine Leistung von $P_L = 10 \text{ W}$ umgesetzt wird. Legen Sie die Größe von R_1 und R_2 selbst fest. Geben Sie den entsprechenden Wert für R_L an.</p> <p>Begründung: Diese Aufgabe ist nur zu lösen, indem aus dem Schaltbild Informationen entnommen werden. Es muss jedoch nicht</p>

		direkt im Schaltbild gezeichnet werden.
(3) In Illustration wird gearbeitet	In der Illustration soll direkt gearbeitet werden, in dem z.B. etwas eingezeichnet oder gekennzeichnet oder Werte eingetragen werden sollen. Oftmals wird ein konkreter Hinweis in der Aufgabenstellung gegeben, z.B. zeichnen Sie einen Graphen.	Zeichnen Sie in untenstehendes Diagramm den Kostenverlauf für Produkt xy ein. 
(4) Illustrationen hat keinen Bezug zur Aufgabenstellung	Die Illustration steht in keinem Bezug zur Aufgabenstellung, z.B. ein Smiley oder Glühbirne.	Lösen Sie die folgenden Gleichungen: Gleichung 1 Gleichung 2 
(0) Keine Illustration vorhanden	Der Aufgabenstamm enthält keine Illustrationen (reine Textform).	Lösen Sie die folgenden Gleichungen: Gleichung 1 Gleichung 2
Sprachliche Merkmale von Aufgaben		
Kategorie 6: Aufgabenstamm in Frage- oder Aussageform		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Frage	Der Aufgabenstamm ist als Frage formuliert und enthält keine Handlungsaufforderung, die angibt welche Handlung zur Lösung der Aufgabe auszuführen ist.	<i>Für welches Verfahren würden Sie sich entscheiden?</i>
(2) Aussage	Die Aufgabenstellung umfasst einen Arbeitsauftrag in Form einer Aussage. Dazu zählen auch Aufgaben, die mit einer Frage beginnen und zusätzlich eine Handlungsaufforderung beinhalten.	<i>Nenne ein Verfahren zur Berechnung der Reliabilität eines Tests.</i>
Kategorie 7: Der Aufgabenstamm ist mit einfacher Satzstruktur formuliert		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Die Sprache im Aufgabenstamm ist klar, eindeutig, mit einfacher Satzstruktur (keine Schachtelsätze, doppelte Verneinungen, etc.). Der Satz ist beim erstmaligen Lesen nachvollziehbar.	<i>Bei welchem Aufgabentyp bereitet ein Mangel an plausiblen, jedoch falschen, Antworten Schwierigkeiten bei der Aufgabenerstellung?</i>
(0) Trifft nicht zu	Der Aufgabenstamm ist umständlich und mehrdeutig formuliert (z.B. komplizierter Satzbau mit Schachtelsätzen, doppelten Verneinungen.). Der Satz wird erst durch mehrmaliges Nachlesen verständlich oder auch gar nicht verständlich.	<i>Bei welcher der folgenden Testaufgaben erschwert der Mangel an wahrscheinlichen, aber unrichtigen Aussagen, die sich auf einen zentralen Gedanken beziehen lassen, die Konstruktion?</i> <u>Beispiel 3:</u> <i>Welches Integral ist zu lösen, das laut dem Divergenzentsatz von Gauß denselben Wert hat wie der Fluss des Vektorfeldes $v(x,y,z)=(x,\cos(z)+y,z)$ durch die Oberfläche der Kugel vom Radius 3, um den Ursprung mit nach außen gerichteten Normalvektoren.</i>

Kategorie 8: Der Aufgabenstamm enthält doppelte Verneinungen		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Der Aufgabenstamm enthält doppelte Verneinungen.	<i>Welches der folgenden Beispiele entspricht keinem Negativbeispiel?</i>
(0) Trifft nicht zu	Der Aufgabenstamm enthält keine doppelten Verneinungen.	<i>Welches der folgenden Beispiele entspricht einem Positivbeispiel?</i>
Instruktionsqualität der Aufgaben		
Kategorie 9: Der Aufgabenstamm enthält keine unnötigen Informationen		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Der Aufgabenstamm enthält ausschließlich Informationen, die zur Lösung des gestellten Problems erforderlich sind (keine unnötigen Informationen, die zur Lösung der Aufgabenstellung nicht wichtig sind; außer sie sind wichtig zur Erfassung des Konstrukts). Aufgabenkontexte oder Situationen, die bei der Lösung der Aufgabenstellung beachtet werden müssen, sind davon auszunehmen.	<i>„Gegeben ist der xy-Entwicklungsprozess. Benennen Sie die gekennzeichneten Meilensteine und Freigaben.“</i>
(0) Trifft nicht zu	Der Aufgabenstamm enthält belehrende bzw. unwichtige Informationen oder Hinleitungen, die zur Lösung der Aufgabenstellung nicht relevant sind. Diese Kategorie umfasst außerdem Aufgaben, bei denen Kontexte geschaffen wurden, die nicht notwendig sind, um die Aufgabenstellung zu lösen.	<p><i>„Ein noch in der Konzeptentwicklung befindliches Elektrofahrzeug besitzt auf Grund der Form seiner Batterie im Fahrzeugunterboden eine große Fahrzeughöhe. Von den Entscheidungsträgern im Unternehmen wird das Fahrzeugkonzept kritisiert, da es nicht der sportlich flachen Designsprache der Marke entspricht. Welche designbeeinflussenden Elemente können Sie ändern, um den offensichtlichen Nachteil der großen Fahrzeughöhe mit Blick auf die Fahrzeugproportionen zu korrigieren? Nennen Sie drei gestalterische Lösungsansätze.“</i></p> <p>Beispiel 2: <i><u>Durch die Besetzung von Nischen im Fahrzeugmarkt ist es für die Hersteller möglich, innovative Fahrzeugkonzepte zu positionieren.</u> Als Beispiel sind der Renault Espace (siebensitzige Großraumlimousine, 1984) und der Renault Twizy (zweisitziges, elektrisches Kleinstfahrzeug, 2011) zu nennen. Nennen Sie zwei wesentliche Chancen und zwei Risiken, welche sich durch die Einführung von Nischenprodukten für die Fahrzeughersteller ergeben. Nehmen Sie dabei Bezug auf die genannten Beispielfahrzeuge. Fotos von Autos sind gegeben.</i></p>

Kategorie 10: Der Erwartungshorizont ist klar definiert		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	<p>Es ist erkennbar, welche Aspekte zur vollständigen Lösung in der Antwort enthalten sein müssen. Es wird in der Aufgabenstellung deutlich, welche Aspekte in der Antwort enthalten sein müssen, damit diese vollständig gelöst ist. Der Umfang der Antworten wird gelenkt und eingeschränkt durch spezifische Angaben oder Vorgaben eines Erwartungshorizontes. Dies kann durch Verweise auf die erwartete Länge oder Detailliertheit der richtigen Antwort erfolgen.</p> <p>Bei Fragen, bei denen aufgrund der Art der Anforderung ersichtlich wird, wann die Antwort bzw. der Vorgang vollständig ist, wie zum Beispiel „Definieren Sie den Begriff“, ist der Erwartungshorizont auch ohne weitere Einschränkungen klar definiert.</p>	<p>a) <i>Nennen sie drei Aspekte der GuV-Rechnung</i> b) <i>Erläutern sie die Aspekte kurz in einem Satz</i></p> <p><u>Begründung:</u> In beiden Beispielen wird der Erwartungshorizont eingeschränkt. Im ersten Beispiel wird konkret angegeben, dass drei Aspekte verlangt sind. Im zweiten Fall wird die erwartete Länge der Erläuterung über das Wort „kurz“ und die Angabe „in einem Satz“ eingeschränkt.</p>
(0) Trifft nicht zu	<p>Es ist nicht eindeutig ersichtlich, welchen Umfang eine Antwort haben soll. Es werden keine Einschränkungen vorgenommen.</p>	<p>a) <i>Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind.</i> b) <i>Begründen Sie ihre Antwort</i></p> <p><u>Begründung:</u> Bei der Handlungsaufforderung zur Teilaufgabe wird nicht klar, in welchem Umfang oder Detailliertheit die Begründung erfolgen muss, um vollständig gelöst zu sein. <i>Ist es für das Unternehmen lohnenswert, die überzähligen Produkte, die nicht selbst gefertigt werden können, beim Lieferanten produzieren zu lassen?</i></p> <p><u>Begründung:</u> Bei diesem Beispiel wird auf den ersten Blick nicht ersichtlich, ob die Antwort durch eine Berechnung oder einfach durch die Antwort Ja/Nein erfolgen kann</p>
Kategorie 11: Vorstrukturierung des Antwortfelds		
<p>Die Antworten werden vorstrukturiert, indem das Antwortfeld vorgegeben wird. Sind Grafiken bzw. Teile davon, in denen etwas eingezeichnet oder ergänzt werden muss, in der Aufgabenstellung vorgegeben, werden diese nicht als Vorstrukturierung gewertet, außer, innerhalb der Grafik ist genau vorgesehen, an welcher Stelle die Beschriftungen einzufügen sind (siehe Beispiel unten bei „Trifft zu“)</p>		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	<p>Innerhalb eines Antwortfeldes (z.B. leerer Platz oder leerer Kasten) wird gezielt strukturiert, z.B. durch die Angabe einer bestimmten Anzahl an Zeilen und die Aufteilung nach Teilantworten. Dazu zählt auch die Angabe von Bulletpoints. im</p>	<p><i>Sie wollen ein neues Bauteil für einen Prototypen entwickeln.</i> a) <i>Welcher zwei Simulationsmethoden bedienen Sie sich</i></p>

Falle von Zeilen (auch als gestrichelte Linie) muss entschieden werden, ob diese tatsächlich das Antwortverhalten lenken. Sind z.B. für eine relativ kurze Antwort fünf Zeilen vorgesehen, wird keine Lenkung des Antwortverhaltens erreicht. Auch bei Aufgaben, bei denen die Angabe mehrerer Punkte erforderlich ist, ist lediglich die Angabe von Zeilen nicht ausreichend. In diesem Fall müsste weiter strukturiert werden, an welcher Stelle ein bestimmter Antwortteil stehen soll (siehe Beispiel 1, rechts).

jeweils um Kinematik und mechanische Festigkeit der Bauteile auszulegen?

b) Erläutern Sie diese kurz (evtl. hinzufügen: in einem Satz.)

Antwortfeld

a) Simulationsmethode für Kinematik:

Erläuterung: _____

b) Simulationsmethode für mechanische Festigkeit: _____

Erläuterung: _____

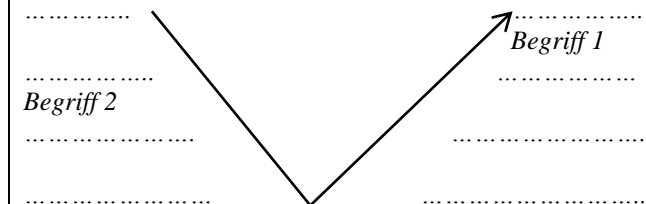
Beispiel 2:

Definieren Sie den Begriff Kundenmanagement.

Begründung:

Bei dieser Aufgabe sind zwei Zeilen vorgegeben. Das Antwortfeld wird als vorstrukturiert gewertet, da die Definition, je nach Schriftgröße, maximal zwei Zeilen benötigt (gemessen an der Länge der Musterlösung). Der Platz für die Antwort ist also eingeschränkt.

Vervollständigen Sie untenstehende Grafik. Markieren Sie anschließend die Einsatzgebiete von Simulation xy und Simulation yz.



Begründung:

Im ersten Teil der Aufgabe muss die Grafik vervollständigt werden. Für die Vervollständigung sind gestrichelte Linien vorgesehen, die das Antwortfeld vorstrukturieren. Für die Markierung der Einsatzgebiete im zweiten Teil der Aufgabe

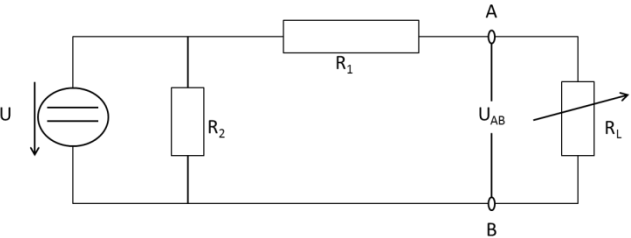
		wurde keine weitere Strukturierung vorgenommen, ggf. aus inhaltlichen Erwägungen. Der zweite Teil der Aufgabe wird mit 0 „Trifft nicht zu“ kodiert.
(0) Trifft nicht zu	Das Antwortfeld ist offen und ohne Vorstrukturierung. Wird das Antwortfeld lediglich eingerahmt, findet jedoch innerhalb des Rahmens keine weitere Strukturierung statt, wird das Item mit „Trifft nicht zu“ bewertet. Zur Vorstrukturierung zählt nicht, wenn der Platz zwischen zwei Aufgaben beschränkt ist, sofern der freie Platz nicht weiter strukturiert ist.	<i>Aufgabenstellung 1.</i> <i>Antwortfeld unstrukturiert</i> <i>Aufgabenstellung 2.</i>
Kategorie 12: Aufgabenstamm enthält nur eine Handlungsaufforderung		
Kodierhinweis: Durch die Segmentierung der Aufgaben wurden die Aufgaben, in denen zwei oder mehrere Handlungsaufforderungen enthalten sind, identifiziert und in Einzelaufgaben zerlegt. Bei der Kodierung derartig zerlegter Aufgaben wird nur bei der ersten Teilaufgabe mit „(1) Trifft zu“ kodiert. Für die darauffolgenden Teilaufgaben wird „(2) bereits kodiert“ gewählt		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Die Angabe enthält genau eine Handlungsaufforderung.	<i>Erläutern Sie das Verfahren xy.</i>
(0) Trifft nicht zu	Die Angabe enthält mehrere Handlungsaufforderungen. Oftmals sind die Aufforderungen über ein „und“ verknüpft.	<i>Erläutern Sie das Verfahren xy und geben Sie ein Beispiel dazu an.</i>
(2) bereits kodiert	<i>Wurde bei vorgehender segmentierter Prüfungsaufgabe angegeben.</i>	
Kategorie 13: Anzahl der vergebenen Punkte auf Ebene von Einzelaufgaben		
Regel 1: Einzelaufgaben stehen für sich alleine oder fungieren als Teilaufgaben, sofern sie unter einem Aufgabenstamm mit fortlaufender Nummerierung aufgeführt werden. Beispiel: Aufgabe 1 beginnt mit der Darstellung einer Situation, zu der mehrere Teilaufgaben zu lösen sind. Die Teilaufgaben sind normalerweise fortlaufend nummeriert, zum Beispiel mit 1.1, 1.2, 1.3, et cetera, oder a, b, c, oder ähnlichem.		
Regel 2: Durch unsere Segmentierung der Aufgaben werden Teilaufgaben erzeugt, die vom Aufgabenersteller nicht explizit als Teilaufgaben ausgewiesen sind. Für diese Kategorie werden die durch die Segmentierung erzeugten Teilaufgaben, die innerhalb eines Aufgabenblocks entstehen, als Teilaufgaben gewertet.		
Regel 3: Wenn auf der Deckseite Angaben zur Punkteverteilung hinsichtlich der Teilaufgaben gemacht werden, so finden diese beim Rating Berücksichtigung.		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Punktezahl	Ermittlung der Punktezahl für die Aufgabe, entweder aus der Angabe, oder der Musterlösung bzw. dem Bewertungsraster. Es kann auch vorkommen, dass übergeordnet eine Angabe gemacht wird, dass z.B. die folgenden zehn Aufgaben mit jeweils 2 Punkten versehen sind. Punkte (per Hand eintragen): _____	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Angaben zur Punktevergabe im Bewertungsraster:</i> <i>Es wird jeweils ein Punkt auf die richtige Formel und ein Punkt für das richtige Ergebnis vergeben.</i> <u>Begründung:</u> Bei dieser Aufgabe ist bei der Aufgabenstellung nicht angegeben, wie viele Punkte vergeben werden, jedoch kann aus den Angaben zur Korrektur und Punktevergabe entnommen

		werden, dass bei der Aufgabe 2 Punkte erreicht werden können.
(0) keine Punkte vorhanden	Für die Aufgabe kann nicht ermittelt werden, wie viele Punkte vergeben werden - weder in der Angabe, noch im Bewertungsraster..	
Stochastische Unabhängigkeit von Aufgaben		
Kategorie 14: Die Aufgabe ist vom Ergebnis der vorangegangenen Aufgabe abhängig		
<p>Folgeaufgaben beziehen sich auf die vorangegangene ursprüngliche Aufgabenstellung/Geschichte, indem sie beispielsweise auf deren Daten, Zwischenergebnisse oder Kontext Bezug nehmen. Sobald sie sich nicht mehr auf den vorangegangenen Kontext, die Daten, oder die Zwischenergebnisse beziehen, sind es keine Folgeaufgaben mehr, da sie sich hiervon losgelöst beantworten lassen. Teilaufgaben sind nicht zwingend Folgeaufgaben, sondern nur dann, wenn sie die beschriebenen Eigenschaften erfüllen.</p> <p><u>Beispiel:</u> <i>Das Gesetz von Hinz und Kuntz weist auf ein Problem hin, dass sich <u>durch die geänderte Arbeitssituation</u> ergeben könnte. Beschreiben Sie das Gesetz und unterstützen Sie Ihre Ausführungen durch eine Grafik.</i></p> <p><u>Begründung:</u> Die Aufgabe war eingebettet in eine Situation, in der sich Arbeitssituationen verändert hatten. Durch den Bezug auf diese gegebene Arbeitssituation wird diese Teilaufgabe als Folgeaufgabe bewertet. Ohne den Bezug könnte die Aufgabe auch losgelöst von diesem Szenario bearbeitet werden. Würde der Bezug auf die „geänderte Arbeitssituation“ fehlen, wäre es keine Folgeaufgabe mehr sondern nur eine Teilaufgabe, da sie sich losgelöst bearbeiten lässt und auch der Bezug zu einem möglichen vorangegangenen Kontext nicht ersichtlich ist.</p>		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Bei Folgeaufgaben, zu deren Bearbeitung und Lösung das Ergebnis der vorherigen Aufgaben vorausgesetzt bzw. benötigt wird.	<p>a) Erstellen Sie zum gegebenen Beispiel einen Vorgangspfeil-Netzplan. b) Führen Sie in ihrem Netzplan eine Vorwärts- und Rückwärtsterminierung durch.</p> <p><u>Begründung:</u> Teilaufgabe b) kann nur bearbeitet und richtig gelöst werden, wenn in Teilaufgabe a) der Vorgangs-Netzplan richtig erstellt wurde. Die Bearbeitung der Teilaufgabe b) ist somit abhängig von Teilaufgabe a).</p>
(0) Trifft nicht zu	Bei Folgeaufgaben, deren Bearbeitung und Lösung das Ergebnis der vorherigen Aufgaben nicht voraussetzt bzw. nicht benötigt.	
(2) keine Folgeaufgabe vorhanden	Aufgaben, die inhaltlich für sich alleine stehen und losgelöst und ohne die Informationen eines vorangegangenen Angabentextes bzw. Lösungen vorheriger Teilaufgaben beantwortet werden können.	<p><i>Die LMM unterscheidet zwei unterschiedliche Personenkreise. Nennen Sie diese und erläutern Sie die unterschiedlichen Eigenschaften der beiden Gruppen. “</i></p> <p><u>Begründung:</u> Diese Aufgabe ist Teilaufgabe einer größeren Aufgabe, in der eine Arbeitssituation dargestellt ist. Die Aufgabe wird jedoch nicht als Folgeaufgabe gewertet, da sie auch allein-</p>

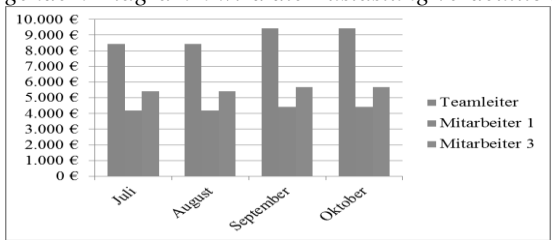
		stehend, ohne die dargestellte Situation bzw. den übergeordneten Aufgabenkontext, beantwortet werden kann. Darüber hinaus sind in der Fragestellung keine Verweise auf die in der Gesamtaufgabe dargestellte Arbeitssituation gegeben.
Kategorie 15: Es werden Zwischenlösungen angegeben		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Es werden Zwischenlösungen angegeben, sofern dies für die Bearbeitung von Folgeaufgaben nötig ist. Dies gilt für alle Aufgabentypen und nicht nur bei Rechenaufgaben.	<i>Falls Sie keine Gemeinkostenzuschlagssätze ermitteln konnten, rechnen Sie mit folgenden Zuschlagssätzen weiter: Material 10 %, Vertrieb 5%.</i>
(0) Trifft nicht zu	Es werden keine oder nicht alle benötigten Zwischenergebnisse angegeben, obwohl diese für eine weitere Bearbeitung von Aufgaben notwendig wären. Außerdem bei Aufgaben, bei denen mit den Lösungen aus der vorherigen Aufgabe weiter gearbeitet oder gerechnet werden soll.	<i>a) Erstellen Sie zum gegebenen Beispiel einen Vorgangspfeil-Netzplan.</i> <i>b) Führen Sie in ihrem Netzplan eine Vorwärts- und Rückwärtsterminierung durch.</i> <u>Begründung:</u> Teilaufgabe b) kann nur bearbeitet und richtig gelöst werden, wenn in Teilaufgabe a) der Vorgangs-Netzplan richtig erstellt wurde. Somit müsste eine Zwischenlösung angegeben werden, evtl. ein etwas veränderter Netzplan. Dieser ist jedoch bei diesem Beispiel nicht vorhanden.
(2) keine Folgeaufgabe vorhanden	Aufgaben, die inhaltlich für sich alleine stehen und die losgelöst - und ohne die Informationen eines vorangegangenen Angabentextes bzw. Lösungen vorheriger Teilaufgaben - beantwortet werden können. Siehe Kodierregel weiter oben.	<i>„Die LMM unterscheidet zwei unterschiedliche Personengruppen. Nennen Sie diese und erläutern Sie die unterschiedlichen Eigenschaften der beiden Gruppen.“</i> <u>Begründung:</u> Diese Aufgabe ist Teilaufgabe einer größeren Aufgabe, in der eine Arbeitssituation dargestellt ist. Die Aufgabe wird jedoch nicht als Folgeaufgabe gewertet, da sie auch allein stehend, ohne die dargestellte Situation bzw. den übergeordneten Aufgabenkontext, beantwortet werden kann. Darüber hinaus sind in der Fragestellung keine Verweise auf die in der Gesamtaufgabe dargestellte Arbeitssituation gegeben.
(3) Keine Zwischenlösung benötigt	Bei Folgeaufgaben, die jedoch unabhängig von den Ergebnissen aus vorherigen Aufgaben gelöst werden können. <i>Für die Lösung der Aufgabe sind die Zwischenlösungen aus den vorangegangenen Aufgaben nicht notwendig. Die Aufgabe lässt sich mittels des erworbenen Wissens zu beantworten.</i>	<i>a) „Skizzieren Sie den Verlauf der Pulsfrequenz für den Fall, dass die mittlere gemessene Pulsfrequenz während der Arbeitsphase 141 1/min betragen würde.“</i> <i>b) Welche weiteren Kennwerte ändern sich bei einer erhöhten mittleren Pulsfrequenz während der Arbeitsphase von 141 1/min</i> <u>Begründung:</u> In Teilaufgabe b) wird nach weiteren Kennwerten gefragt, die sich bei Erhöhung der Pulsfrequenz ändern. Die Aufgabe

		ist eine Folgeaufgabe von Teilaufgabe a), da sie sich auf den gleichen Aufgabenkontext bezieht. Um diese Frage richtig zu beantworten ist es jedoch nicht notwendig, Teilaufgabe a) richtig zu lösen.
Inhaltliche Gestaltungsmerkmale von Aufgaben		
Kategorie 16: Wissensarten in den Aufgaben		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Deklaratives Wissen	<p>Wissen, „dass“</p> <p>Die Aufgabe bezieht sich auf deklaratives Wissen. Sie beinhaltet Definitionen und Fakten, meistens in der Form von Begriffen, Aufstellungen, Aussagen, Regeln oder Beschreibungen, und ist explizit zu verbalisieren.</p> <p>Beispiel: Schmelzpunkte verschiedener Materialien bei verschiedenen Temperaturen (Fakt).</p> <p>Die Antworten zu den entsprechenden Aufgaben sind in Form von Begriffen, Fakten, Definitionen, Aussagen oder Beschreibungen zu geben, ohne diese zu begründen, zu erklären oder Vorhersagen zu treffen oder ohne diese in Beziehung zueinander zu setzen.</p> <p>Die Studierenden haben keinen Spielraum, Relationen zwischen Konzepten herzustellen oder bestimmte Prinzipien oder Modelle anzuwenden. In den meisten Fällen kann das Wissen reproduziert werden.</p>	<p><i>Nennen Sie sechs Kriterien, die sich für die Erstellung Unterrichtskonzepte als wirksam erwiesen haben.</i></p>
(2) Prozedurales Wissen	<p>Wissen, „wie“</p> <p>Prozedurales Wissen ist Wissen darüber, wie etwas aus- oder durchgeführt werden kann. Es enthält Wenn-Dann-Produktionsregeln oder eine Sequenz an Schritten, die ausgeführt werden muss. Durch Praxis und zielgerichtetes Üben kann prozedurales Wissen in automatisierte Abläufe überführt werden.</p> <p>Mögliche Beispiele sind: das Vorgehen zur Messung der Dichte eines Objekts; die Durchführung der ersten Ableitung, das Ausführen einer Differentialgleichung, die Berechnung der Geschwindigkeit eines Autos, das Zeichnen eines Graphs, der die Geschwindigkeit eines Autos abbildet oder das Anwenden von Algorithmen zum Aufstellen chemischer Reaktionsgleichungen.</p> <p>Die Antwort erfordert nicht notwendigerweise das Verstehen des übergeordneten Konzeptes.</p>	<p><i>Berechnen Sie die folgenden Gleichungen</i></p> <p><i>Gleichung 1</i></p> <p><i>Gleichung 2</i></p> <p><u>Beispiel 2:</u></p> <p><i>Beschreiben Sie, wie Sie beim Target-Costing – vom Marktpreis ausgehend – die Zielkosten ermitteln können.</i></p> <p><u>Begründung:</u></p> <p>Es muss eine Vorgehensweise (also wie) für einen bestimmten Prozess (Target-Costing) dargestellt werden.</p>
(3) Konzeptionelles Wissen	<p>Wissen, „warum“.</p> <p>Konzeptwissen kann verwendet werden, um Probleme zu interpretieren, eine Fehlersuche durchzuführen, zu erklären (was passiert oder passiert ist), zu begründen, in Relation oder Beziehung zu setzen, und um Vorhersagen, die ein logisches Schlussfolgern erfordern, zu treffen, wie zum Beispiel Veränderungen in einem Konzept andere Konzepte beeinflussen (Ursache – Wirkung).</p> <p>Beispiele: Erklärung, warum Tag und Nacht auf der Erde auftreten; Erklärung wie</p>	<p><i>Welche Auswirkungen hat das sogenannte Warndilemma auf den Einsatz und die Nutzung von Assistenzsystemen in Fahrzeugen?</i></p> <p><u>Begründung:</u></p> <p>Es müssen Vorhersagen über das Konzept Warndilemma getroffen werden und Auswirkungen auf die Nutzung von Assistenzsystemen eingeschätzt werden.</p>

	<p>ein bestimmter Virus wirkt; Erklärung welcher Zusammenhang zwischen Aufgaben-Format und kognitiven Prozessen besteht.</p> <p>Die Antworten zu den entsprechenden Aufgaben sind in Form von Begründungen, Erklärungen, Argumentationen oder Vorhersagen zu treffen. Konzeptwissen baut auf Fakten- und Prozesswissen auf und stellt Verbindung her und beinhaltet typischerweise die Anwendung wissenschaftlicher Prinzipien, Erklärungsmodellen, Taxonomien, Schemata oder Begriffsnetzen oder das Argumentieren mit Modellen oder Konzepten. Es müssen z.B. Bedeutungszusammenhänge berücksichtigt werden.</p> <p>Die Studierenden haben Raum Relationen zwischen Konzepten herzustellen oder bestimmte Prinzipien oder Modelle anzuwenden. In den meisten Fällen kann das Wissen nicht reproduziert werden. Die Studierenden nutzen zur Lösung entsprechender Aufgaben ihr Wissen über Theorien, Fakten oder Prozesswissen. Die Aufgabe verlangt von Studierenden voraussagen zu treffen oder unter Rückgriff auf Theorien oder Modelle zu argumentieren.</p>	
(4) Strategisches Wissen	<p>Strategiewissen beinhaltet Wissen, wann, wo und wie bestimmte Wissensarten in neuen Situationen eingesetzt werden können. Strategiewissen beinhaltet domänen-spezifische Strategien wie Planen, Problemlösen und den eigenen Fortschritt zu überprüfen, sowie konditionales Wissen, also Kenntnisse über adäquate Anwendungsbedingungen der Strategien und des Wissens. Aufgaben erfordern das Interpretieren einer Problemstellung und darauf aufbauend den Vergleich und begründete Auswählen einer Strategie. Das geeignete Verfahren wird dabei nicht im Vorfeld in der Aufgabenstellung vorgegeben, sondern selbst gewählt werden.</p> <p>Personen verwenden strategisches Wissen, um zu erkennen, wann bestimmte Prozeduren ausgeführt werden können, analysieren die Merkmale einer Aufgabe, um zu entscheiden, welche Strategie verwendet werden kann, setzen Ziele und überprüfen den Fortschritt bei der Zielerreichung. Allgemeine Anwendungen sind Planungsstrategien, wie das Zerlegen in Teilaufgaben, Kategorisieren und Priorisieren.</p> <p>Strategiewissen kann vor allem bei Aufgaben hervorgerufen werden, die neuartige, unbekannt Problemstellungen beinhalten.</p>	<p>Die Leistungen einer Schülerin verschlechtern sich zur Mitte des Schuljahres deutlich. Welche Maßnahmen ergreifen Sie als Lehrkraft?</p> <p><u>Begründung:</u> Bei dieser Aufgabe wird eine Situation geschildert, auf die auf ganz unterschiedliche Weise reagiert werden könnte. Zur Beantwortung können pädagogische, fachliche oder psychologische Ansätze verwendet werden, die selbstständig gewählt und priorisiert werden. In einem ersten Schritt wäre es vermutlich sinnvoll, mögliche Herangehensweisen festzulegen und zu begründen.</p>
Kategorie 17: Aufgabenoffenheit		
Bei der Offenheit der Aufgabe kommen zwei Merkmale zum Tragen_ Zum einen, ob der Anfangszustand des zu lösenden Problems klar definiert ist, oder nicht. Hinzu kommt die Unterscheidung zwischen konvergenten Aufgaben (eine Lösung) und divergenten Aufgaben (mehrere Lösungen). Aufgaben werden dabei nach Klarheit von Anfangs- und Zielzustand sowie nach der Transformation eingeteilt. Dadurch ergeben sich für die Analyse drei Stufen der Offenheit.		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Definierte und konvergente Aufgabe	Definiert-konvergente Aufgaben haben einen klar definierten Anfangszustand und es ist genau eine Lösung richtig. Die Aufgabe umfasst einen eindeutigen Arbeitsauftrag bzw. eine klar identifizierbare Fragestellung. Die Gliederung der Auf-	„Gegeben ist der Entwicklungsprozess xy . Benennen Sie die gekennzeichneten Meilensteile und Fragen.“ <u>Begründung:</u>

	<p>gabe ist typischerweise kleinschrittig. Es ist genau eine Lösung gesucht bzw. richtig (ist nicht zwingend explizit angegeben). Es gibt z.B. nur einen Lösungsweg bei Berechnungen. Neben einer einzigen richtigen Lösung kann ein weiterer Hinweis sein, wenn nicht freigestellt ist, wie eine Aufgabe, z.B. durch Berechnung vs. Prosa, bearbeitet wird. Die Aufgabe kann z.B. nur mit einem Ansatz (z.B. Berechnung oder Aufzählung von) gelöst werden. Die Aufzählung von Faktenwissen, z.B. Vor- und Nachteilen, ist ebenfalls konvergent. Speziell bei der Aufzählung von Vor- und Nachteilen, etc. gibt es oft mehrere Möglichkeiten, die jedoch von der Anzahl her in der Aufgabenstellung nicht vollständig abgefragt werden, z.B. 3 von 10. Welche der 10 Möglichkeiten ausgewählt werden ist hier dem Studierenden überlassen. In dem Fall wird die Aufgabe auch als konvergent gewertet. Multiple-Choice-Aufgaben mit mehreren möglichen Wahlantworten gehören zu dieser Art Aufgaben, da die Lösung über Ankreuzen erfolgt und keine individuellen Lösungswege ermöglicht werden.</p>	<p>Der Anfangszustand ist gegeben. Außerdem ist eine richtige Lösung gesucht.</p>
<p>(2) Definierte und divergente Aufgabe</p>	<p>Die Aufgabe umfasst einen eindeutigen Arbeitsauftrag bzw. eine eindeutige Fragestellung, der Ausgangszustand ist klar definiert. Es sind jedoch mehrere richtige Lösungen möglich. Es zählen Aufgaben dazu, bei denen es nicht nur eine eindeutig richtige Lösung gibt. Es können unterschiedliche Lösungen richtig sein (ausgenommen sind Faktenwissensaufgaben, z.B. mit nennen sie 4 von 10 Vorteile.). Soll jedoch ein Beispiel gefunden werden, das frei zu konstruieren ist, kann dieses sehr unterschiedlich ausfallen, was für eine divergente Aufgabe spricht. Ist zwar nur eine Art der Bearbeitung zulässig, sind dabei jedoch mehrere richtige Lösungen möglich, ist die Aufgabe als divergent zu werten. Die unterschiedlichen Möglichkeiten für eine richtige Antwort sind als dominantes Merkmal zu werten.</p>	<p>Gegeben ist folgende Schaltung:</p>  <p>Stellen Sie den Lastwiderstand R_L so ein, dass in diesem eine Leistung von $P_L = 10 \text{ W}$ umgesetzt wird. Legen Sie die Größe von R_1 und R_2 selbst fest. Geben Sie den entsprechenden Wert für R_L an.</p> <p><u>Begründung:</u> Der Ausgangszustand ist bekannt, es sind jedoch mehrere Lösungen möglich, in Abhängigkeit der Festlegung der Widerstände R_1 und R_2.</p>
<p>(3) undefinierte und divergente Aufgabe</p>	<p>Der Ausgangszustand ist vage beschrieben. Es sind mehrere richtige Lösungen denkbar. Eine Problemsituation ist die einzige Handlungsaufforderung. Studierende erhalten Informationen über ein Problem, Situation, etc. Unterschiedliche Fragestellungen sind denkbar und damit auch mehrere Lösungen und Lösungswege. Die Aufgabenstellung ist dabei unscharf definiert. Die Lösungen müssen nicht exakt sein, sondern können auch Schätzungen beinhalten. Zur Lösung bedarf es der Integration verschiedener domänenspezifischer Kenntnisse. Evtl. besteht die Notwendigkeit der Erweiterung der Wissensbasis. Benötigte Daten sind ggf. nicht</p>	<p><i>Eine Arbeiterin trägt im Rahmen ihrer Arbeit in einer Art Rucksack Metallteile am Rücken und muss sich in leicht gebückter Haltung fortbewegen. Wie analysieren Sie als Ergonome diese Arbeitssituation.</i></p> <p><u>Begründung:</u> Die Handlungsaufforderung ist gegeben, jedoch bleibt der Ausgangszustand vage, da wichtige Informationen für eine Analyse fehlen. Der Studierende muss selbstständig ent-</p>

	angegeben sondern müssen geschätzt werden. Ein weiteres Merkmal ist die methodische Reflexion zum Vorgehen bei der Problemlösung.	scheiden, welche Informationen für eine Analyse notwendig sind und diese ggf. auch schätzen. Außerdem muss er schließen, dass im ersten Schritt einer solchen Analyse der Risikowert zu bestimmen ist, was ebenfalls nicht vorgegeben ist. Weiterhin muss er auf zu beachtende Bedingungen Rücksicht nehmen, ohne dass darauf hingewiesen wäre.
Kategorie 18: Kontext- und Situationsbezug		
Regel 1: Bei Teil- und Folgeaufgaben, die sich auf den gleichen Kontext beziehen, kann sich die Funktion des Kontextes verändern. Es ist möglich, dass für eine Teilaufgabe der Kontext einen Teil der Aufgabenstellung darstellt für eine weitere Aufgabe jedoch nur vorgeschaltet ist und umgekehrt.		
Ausprägung	Kodierregel	Ankerbeispiel
(0) Kein Kontextbezug	Es gibt keine Bezüge zwischen Fachwissen und Kontexten /Situationen. Die Probleme sind rein innerfachlicher Natur und ohne Bezüge zu möglichen Berufsfeldern herzustellen.	<i>Nennen Sie drei Vorteile einer LAN-Verbindung im Vergleich zu einer WLAN-Verbindung.</i>
(1) Kontextbezug schwach ausgeprägt	In der Aufgabenstellung wird keine Situation oder Kontext beschrieben, jedoch wird in einem Wort oder durch einen Beisatz eine mögliche Praxisanwendung angedeutet. Die Aufgabenstellung ist aber unabhängig davon und rein innerfachlich zu lösen.	<p><i>Eine punktförmige Lichtquelle beleuchtet eine Anzeigentafel [...]. Berechnen Sie die Beleuchtungsstärke der Anzeigentafel.</i></p> <p><u>Begründung:</u> Es wird ein konstruierter Bezug zu einer realen Anwendung der verlangten Berechnung hergestellt, indem auf die Anzeigentafel Bezug genommen wird. Die Aufgabe wäre auch ohne die Anzeigentafel zu bearbeiten. Es müssen keine Bezüge zu außerfachlichen Kontextbedingungen hergestellt werden.</p> <p><u>Beispiel 2:</u> <i>Der Elektromotor einer Waschmaschine hat eine Nutzleistung von 300 Watt. Berechnen Sie die Nennleistung des Motors, wenn der Wirkungsgrad 90% beträgt.</i></p> <p><u>Begründung:</u> Die Berechnung der Nennleistung eines Elektromotors ist eine innerfachliche Aufgabenstellung. Der Bezug auf die Waschmaschine ist für die Bearbeitung der Aufgabe nicht relevant, zeigt jedoch ein mögliches Einsatzgebiet für Elektromotoren auf. Daher wird der Kontextbezug dieser Aufgabe als schwach ausgeprägt bewertet.</p>
(2) Kontextbezug vorgeschaltet	Es wird ein Kontext oder eine Situation beschrieben. Die Aufgabenstellung kann jedoch losgelöst von dem konstruierten Kontext gelöst werden. Der Kontext bzw. die Situation spielt keine Rolle für die Lösung der Aufgabe. Ein Indiz ist, wenn sich die Aufgabenanforderungen nicht auf die Situation beziehen bzw. auch ohne Berücksichtigung der in der Situation gegebenen Bedingungen gelöst werden kann.	<i>Ein Bauarbeiter hat die Aufgabe auf einer Baustelle mit einem Schlaghammer Schlitzlöcher in einer Mauer anzubringen, damit Elektrokabel unter dem Putz verlegt werden können. Die Schlitzlöcher sind auf einer Höhe von 1,10 m anzubringen. Das Gewicht des Schlaghammers beträgt 8kg. Die gesamte</i>

	<p>Ein weiterer Hinweis ist, ob der Kontext beliebig austauschbar ist, z.B. ob anstatt dem möglichen Kontext „Tennis“ der Kontext „Fussball“ gewählt wird, ohne dass sich die Lösung verändert. Es dominiert das fachliche Problem gegenüber der Einkleidung in den Kontext. Sind Informationen, die zur Lösung der Aufgabe benötigt werden (z.B. die Angabe bestimmter Werte, wie Temperatur, Geschwindigkeit) in einen Kontext eingebettet, könnten aber auch ohne diesen Kontext gelöst werden, wird der Kontext ebenfalls als vorgeschaltet gewertet (und nicht als Teil der Aufgabenstellung).</p>	<p><i>Länge der Schlitzte belüftet sich auf 15m. Pro Meter benötigt der Arbeiter 10 min.</i> <i>Berechnen Sie den Arbeitsumsatz für die Tätigkeit „Arbeiten mit dem Schlaghammer.</i> <u>Begründung:</u> Der Aufgabenstellung geht zwar der Kontext mit dem Bauarbeiter, der Schlitzte stemmen muss, voraus, jedoch ist eine Bezugnahme zu diesem Kontext nicht notwendig, um die Aufgabenstellung zu bearbeiten. Der Kontext ist beliebig austauschbar. Anstatt eines Bauarbeiters mit einem Schlaghammer hätte man beispielsweise auch einen Bandarbeiter bei einem Automobilhersteller nehmen können, der aus ergonomischer Sicht eine ähnliche Tätigkeit ausführt. Der Lösungsweg zur Berechnung des Arbeitsumsatzes würde unverändert bleiben. Der Kontext muss bei der Bearbeitung der Aufgabe also nicht beachtet werden. Aus dem Kontext sind keine Bedingungen zu entnehmen, die bei der Lösung berücksichtigt werden müssen.</p>
<p>(3) Kontextbezug als Teil der Aufgabenstellung</p>	<p>In der Aufgabenstellung wird eine Verknüpfung zwischen Fachwissen und außerfachlichen Kontexten, z.B. Berufsfeldern, hergestellt. Dies kann durch die Einbettung der Aufgabenstellung in Situationen oder Kontexte erfolgen. Die Aufgabenstellung ist direkt in einer Situation oder einem Kontext verortet. Die Situationen bzw. Kontexte und darin enthaltene Informationen spielen bei der Bearbeitung bzw. Lösung der Aufgabe eine Rolle, insofern bestimmte Bedingungen daraus hervorgehen oder gestellt werden, die bei der Lösung der Aufgabe berücksichtigt werden müssen. Der Umfang bzw. Länge des beschriebenen Kontexts spielt dabei keine Rolle.</p>	<p><i>In Ihrem Unternehmen haben Sie eine Produktionsmaschine, die momentan nur schwach ausgelastet ist. Aktuell ist ein längerfristig laufender Auftrag auf der Maschine eingelastet. Der Auftrag wird stets an den gleichen drei Tagen auf der Maschine gefertigt. Es ist jedoch möglich, dass die Fertigungstage innerhalb der Woche verschoben werden. In folgendem Diagramm wird die Auslastung verdeutlicht.</i></p>  <p><i>Sie möchten nun einen weiteren Auftrag auf der Maschine einlasten. Nennen Sie zwei mögliche Lösungsmaßnahmen für den Fall, dass die derzeitigen Kapazitäten nicht für die Bearbeitung des Auftragspektrums ausreichen.</i> <u>Begründung:</u> Der Kontext der Aufgabe beschreibt die Situation in einer Firma. Es müssen Lösungsmöglichkeiten für die im Kontext</p>

		beschriebene Situation gefunden werden. Dabei müssen die im Kontext dargestellten Bedingungen bei der Aufgabenbearbeitung berücksichtigt werden und liefern hierfür Informationen.
Kategorie 19: Komplexität von Aufgaben		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Einfache Darstellung	Neben der Handlungsaufforderung werden <u>keine</u> oder <u>nur wenige</u> , einfach abzulesende, Informationen gegeben, wie z.B. (Raumtemperatur = 20°C) oder z.B. eine übersichtliche und einfache Tabelle oder Grafik, die zusammen mit der Handlungsaufforderung positioniert ist. Es muss nichts gefiltert werden.	<i>Für eine dynamischen Muskelarbeit mit einer Dauer von 20 min. ist eine Nutzleistung von 30 W aufzubringen. Der Wirkungsgrad beträgt 10 %. Bestimmen Sie, wie oft diese Tätigkeit durch eine männliche Person an einem siebenstündigen Arbeitstag ausgeführt werden kann unter Einhaltung notwendiger Erholzeiten.</i> <u>Begründung:</u> Diese Aufgabe wird bzgl. der Komplexität als einfach gewertet. Neben der Handlungsaufforderung müssen zwar mehrere Werte betrachtet werden, jedoch sind die Werte kompakt und zusammen mit der Handlungsaufforderung angeordnet. Der Umfang der zu beachtenden Bedingungen ist überschaubar und schnell zu erfassen.
(2) Weniger komplex	Es ist möglich, sich einen schnellen Überblick über die bereitgestellten Informationen zu verschaffen, jedoch muss genauer bzw. gezielter in den Informationen gelesen werden, um diese zu sichten und zu verarbeiten, als es bei einer einfachen Darstellung der Fall ist. Es müssen keine Informationen gefiltert werden. Ein weiterer Hinweis kann sein, wenn die Anordnung der Informationen in etwa der Chronologie der Aufgabenbearbeitung entspricht. Es sind vorwiegend nur solche Informationen gegeben, die auch benötigt werden. Die Informationen sind übersichtlich dargestellt und dadurch gut erfassbar. Es müssen keine irrelevanten Informationen gefiltert werden, z.B. keine Entscheidung, welche von mehreren Tabellen relevant ist. (wenn jedoch innerhalb einer Tabelle nur einzelne Informationen entnommen werden müssen, also nicht alle Werte oder Informationen benötigt werden, würde dies noch in diese Kategorie fallen). Eine weitere Komplexitätssteigerung und damit Hinweis auf die Kategorie 2 „weniger komplex“ kann vorliegen, wenn in der Aufgabenstellung nicht konkret darauf verwiesen wird, an welcher Stelle die benötigten Informationen hinterlegt sind.	<u>Beispiel 1 (aufgrund des Umfangs der Materialien wird an dieser Stelle auf eine Aufgabe in einer der betrachteten Prüfungen verwiesen).</u> <i>Aufgabe 7, aus Prüfung 1</i> <u>Begründung:</u> Es muss sowohl aus Text als auch Bild Information herangezogen werden. Es ist jedoch sehr leicht, sich einen Überblick zu verschaffen. Informationen müssen nicht gefiltert werden.
(3) Komplex	Die Aufgabe enthält typischerweise „mehrere“ Informationsquellen. Es sind zu viele Informationen verfügbar aus denen Relevantes gefiltert werden muss. Ein weiterer Hinweis kann sein, wenn die Anordnung der Informationen nicht der Chronologie der Aufgabenbearbeitung entspricht. In der Aufgabenstellung wird nicht konkret auf die zu nutzenden Informationsquellen verwiesen.	<u>Beispiel 1 (aufgrund des Umfangs der Materialien wird die Aufgabe an der Stelle verlinkt).</u> http://www.cuny.edu/academics/initiatives/cla/sample-tasks.html <u>Begründung:</u>

		Bei dieser Aufgabe müssen Informationen aus mehreren Tabellen gefiltert und entnommen werden. Ein Überblick ist nicht ohne eine grundlegende Einarbeitung in die Tabellen zu erreichen. Die Anordnung der Informationen entspricht nicht notwendigerweise der Chronologie der Bearbeitung.
Bewertungsraster für die Korrektur von Aufgaben		
Ein vollständiges Bewertungsraster enthält die Bewertungskategorien (richtig, teilweise richtig, falsch) und ein Regelwerk zu jeder Kategorie, in dem spezifiziert und beschrieben wird, welche Aspekte in einer Aufgabe enthalten sein müssen bzw. fehlen, um einer Kategorie zugeordnet werden. Zu jeder Kategorie müssen ein oder mehrere Ankerbeispiel(e) enthalten sein. Außerdem werden Angaben zur Vergabe der Punkte innerhalb einer Aufgabe gemacht. Mit den folgenden Kategorien wird festgestellt, ob die einzelnen Bestandteile eines Bewertungsrasters vorhanden sind. Es kann zum z. B. sein, dass bei einer Aufgabe zwar ein Ankerbeispiel angegeben ist, jedoch Angaben zur Punkteverteilung oder die Beschreibung der Bewertungskategorie fehlen. Werden gar keine Angaben zur Korrektur gegeben, wird die Ausprägung (2) „kein Bewertungsraster“ gewählt.		
Kategorie 20: Ankerbeispiel für richtige Antwort enthalten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Das Bewertungsraster enthält Ankerbeispiele für richtige Lösungen.	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Ankerbeispiel richtige Antwort:</i> $y=mx+t$ $y=0,5 \bullet 3+2=3,5$
(0) Trifft nicht zu	Das Bewertungsraster enthält keine Ankerbeispiele für richtige Lösungen.	Es ist kein Ankerbeispiel für die richtige Antwort angegeben.
(2) Kein Bewertungsraster	In der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
Kategorie 21: Ankerbeispiel für falsche Antwort vorhanden		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Das Bewertungsraster enthält Ankerbeispiele für falsche Lösungen.	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Ankerbeispiel falsche Antwort:</i> $y=m+x+t$ $y=0,5+3+2=5,5$ oder: $y=mx+t=0,5 \bullet 3+2=5$
(0) Trifft nicht zu	Das Bewertungsraster enthält keine Ankerbeispiele für falsche Lösungen.	Es ist kein Ankerbeispiel für mögliche falsche Antworten angegeben.
(2) Kein Bewertungsraster	In der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	

Kategorie 22: Ankerbeispiel für teilweise richtige Antworten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Das Bewertungsraster enthält Ankerbeispiele für teilweise richtige Lösungen.	<p><i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i></p> <p><i>Ankerbeispiel teilweise richtige Antwort: Formel richtig, Ergebnis falsch $y=mx+t=0,5 \bullet 3+2=5$ (1 Punkt für richtige Formel, 1 Punkt für richtiges Ergebnis)</i></p> <p><u>Begründung:</u> Bei diesem Beispiel ist lediglich aufgrund der Angabe zur Vergabe der Punkte ersichtlich, dass teilweise richtige Antworten gewertet werden, da konkret angegeben ist, dass es für die richtige Formel sowie das richtige Ergebnis einen Punkt gibt. Wäre diese Angabe nicht vorhanden, so müsste man bei der Kodierung davon gehen ausgehen, dass lediglich Punkte vergeben werden, wenn sowohl die Formel als auch das richtige Ergebnis vorhanden sind.</p>
(0) Trifft nicht zu	Das Bewertungsraster enthält keine Ankerbeispiele für teilweise richtige Lösungen.	Es ist kein Ankerbeispiel für teilweise richtige Antworten angegeben.
(2) Kein Bewertungsraster	In der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
(3) Teilweise richtige Antworten sind nicht vorgesehen	Diese Kategorie wird bei den Aufgaben gewählt, bei denen kein Verweis auf teilweise richtige Lösungen gegeben wird; weder durch das Ankerbeispiel, noch durch das Regelwerk oder die Punktevergabe. Die Kategorie „teilweise richtig“ darf nur gewählt werden, wenn in den besagten Angaben ein Verweis gefunden wird, dass teilweise richtige Antworten gewertet werden.	<p><i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i></p> <p><i>Ankerbeispiel teilweise richtige Antwort: Formel richtig, Ergebnis falsch $y=mx+t=0,5 \bullet 3+2=5$</i></p> <p><u>Begründung:</u> Rein aus dieser Angabe kann nicht geschlossen werden, ob teilweise richtige Antworten, wie z.B. „Formel ist richtig vorhanden, jedoch das Endergebnis ist falsch“, gewertet werden, da nicht weiter spezifiziert ist, wie man unvollständigen Lösungen umgegangen wird.</p>

Kategorie 23: Regelwerk für richtige Antworten enthalten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Es ist ein Regelwerk gegeben, aus dem hervorgeht und beschrieben ist, in welchen Fällen die Antwort als richtig bewertet wird.	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Bewertungshinweis: Die Antwort wird als richtig gewertet, wenn sowohl die Formel richtig angegeben ist, als auch das richtige Ergebnis angegeben ist.</i>
(0) Trifft nicht zu	Es ist kein Regelwerk in der Musterlösung vorhanden, aus dem hervorgeht, in welchen Fällen die Antwort als richtig bewertet wird.	Es ist kein Regelwerk vorhanden, das beschreibt, welche Aspekte eine richtige Antwort aufweisen muss.
(2) Kein Bewertungsraster	Bei der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
Kategorie 24: Regelwerk für falsche Antworten enthalten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Es ist ein Regelwerk vorhanden, aus dem hervorgeht, in welchen Fällen die Antwort als falsch bewertet wird.	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Bewertungshinweis: Die Antwort wird als falsch gewertet, wenn weder die Formel, noch das richtige Ergebnis angegeben ist. Antworten bei denen zwar das Ergebnis richtig, jedoch die Formel falsch ist, werden ebenfalls als falsch gewertet.</i>
(0) Trifft nicht zu	Es ist kein Regelwerk vorhanden, aus dem hervorgeht, in welchen Fällen die Antwort als falsch bewertet wird.	Es ist kein Regelwerk vorhanden, das beschreibt, welche Aspekte eine falsche Antwort aufweisen muss.
(2) Kein Bewertungsraster	In der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
Kategorie 25: Regelwerk für teilweise richtige Antworten enthalten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	Es ist ein Regelwerk vorhanden, aus dem hervorgeht, in welchen Fällen die Antwort als teilweise richtig bewertet wird.	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Bewertungshinweis: Als teilweise richtig werden Antworten bewertet, die zwar die richtige Formel, jedoch ein falsches Ergebnis angegeben enthalten. Ist umgekehrt zwar das richtige Ergebnis angegeben, jedoch die Formel falsch, werden keine Punkte vergeben und die Antwort als falsch bewertet.</i>

(0) Trifft nicht zu	Es ist kein Regelwerk vorhanden, aus dem hervorgeht, in welchen Fällen die Antwort als teilweise richtig bewertet wird.	Es ist kein Regelwerk für teilweise richtige Antworten vorhanden.
(2) Kein Bewertungsraster	In der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
(3) Teilweise richtige Antworten sind nicht vorgesehen	Diese Kategorie wird bei solchen Aufgaben gewählt, bei denen kein Verweis auf teilweise richtige Lösungen gegeben ist; weder durch das Ankerbeispiel, noch durch das Regelwerk oder die Punktevergabe. Die Kategorie „teilweise richtig“ darf nur gewählt werden, wenn in den besagten Angaben ein Verweis gefunden wird, dass teilweise richtige Antworten gewertet werden.	<p><i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i></p> <p><i>Ankerbeispiel teilweise richtige Antwort: Formel richtig, Ergebnis falsch $y=mx+t=0,5 \bullet 3+2=5$</i></p> <p>Begründung: Rein aus dieser Angabe kann nicht geschlossen werden, ob teilweise richtige Antworten, gewertet werden, da nicht weiter spezifiziert ist, wie unvollständigen Lösungen umgegangen wird.</p>
Kategorie 26: Angaben zur Punktevergabe bei richtigen Antworten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	In der Musterlösung (entweder im Regelwerk oder den Ankerbeispielen) ist gekennzeichnet, auf welche Aspekte der Antwort Punkte vergeben werden.	<p><i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i></p> <p><i>Punktevergabe: Es werden jeweils ein Punkt auf die richtige Formel und ein Punkt für das richtige Ergebnis vergeben.</i></p>
(0) Trifft nicht zu	In der Musterlösung (entweder im Regelwerk oder den Ankerbeispielen) ist nicht gekennzeichnet, auf welche Aspekte der Antwort Punkte vergeben werden.	Es werden keine Angaben zur Punktevergabe bei richtigen Antworten gemacht.
(2) Kein Bewertungsraster	In der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
Kategorie 27: Angaben zur Punktevergabe bei falschen Antworten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	In der Musterlösung (entweder im Regelwerk oder den Ankerbeispielen) ist gekennzeichnet, auf welche Aspekte der Antwort keine Punkte vergeben werden.	<p><i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i></p> <p><i>Punktevergabe Es werden keine Punkte vergeben, wenn weder Formel noch Ergebnis richtig sind. Auf das richtige Ergebnis bei falscher Formel werden ebenfalls keine Punkte vergeben.</i></p>
(0) Trifft nicht zu	In der Musterlösung (entweder im Regelwerk oder den Ankerbeispielen) ist nicht gekennzeichnet, wann auf eine Aufgabe keine Punkte gegeben werden.	Es werden keine Angaben zur Punktevergabe bei falschen Antworten gemacht.

(2) Kein Bewertungsraster	Bei der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
Kategorie 28: Regelwerk für teilweise richtige Antworten enthalten		
Ausprägung	Kodierregel	Ankerbeispiel
(1) Trifft zu	In der Musterlösung (entweder im Regelwerk oder den Ankerbeispielen) wird gekennzeichnet, auf welche Aspekte der Antwort teilweise Punkte vergeben werden.	<i>Berechnen Sie für eine lineare Gleichung den Wert von y wenn $x=3$, $m=0,5$ und $t=2$. Geben Sie den Rechenweg an.</i> <i>Bewertungshinweis: Ein Punkt wird vergeben, wenn zwar die Formel richtig, jedoch das Ergebnis falsch ist.</i>
(0) Trifft nicht zu	In der Musterlösung wird nicht gekennzeichnet, auf welche Aspekte der Antwort teilweise Punkte vergeben werden.	Es werden keine Angaben zur Punktevergabe bei teilweise richtigen Antworten gemacht.
(2) Kein Bewertungsraster	Bei der Aufgabe sind keinerlei Angaben wie Beschreibung der Bewertungskategorien, Ankerbeispiele oder Angaben zur Punktevergabe vorhanden.	
(3) Teilweise richtige Antworten sind nicht vorgesehen	Diese Kategorie wird bei den Aufgaben gewählt, bei denen kein Verweis auf teilweise richtige Lösungen gegeben wird; weder durch das Ankerbeispiel, noch durch das Regelwerk oder die Punktevergabe. Die Kategorie „teilweise richtig“ darf nur gewählt werden, wenn in den besagten Angaben ein Verweis gefunden wird, dass teilweise richtige Antworten gewertet werden.	<i>Ankerbeispiel teilweise richtige Antwort: Formel richtig, Ergebnis falsch $y=mx+t=0,5 \bullet 3+2=5$</i> Begründung: Aus dieser Angabe ist nicht ersichtl. ob teilweise richtige Antworten gewertet werden, da nicht weiter spezifiziert ist, wie mit unvollständigen Lösungen umgegangen ist.

e. Kategoriensystem zur Strukturierung der Verbaldaten

Tabelle 38: Kategoriensystem zur Strukturierung der Verbaldaten

Festlegung des Prüfungsgegenstands		
Kategorie 1: Prüfungsanforderungen		
Ausprägung	Kodierregel	Ankerbeispiel
Abzuprüfende Fähigkeiten und Kompetenzen	Aussagen darüber, was Studierende wissen oder können sollen, welche Fähigkeiten oder Kompetenzen geprüft werden sollen.	<i>In unserer Veranstaltung lernen die Studierenden die verschiedenen Konzepte der Unternehmensführung kennen. Wir wollen, dass die Studierenden Zusammenhänge zwischen den behandelten Themenbereichen herstellen können.</i>
Kategorie 2: Konzeptionelle Analyse des Prüfungsgegenstandes		
Ausprägung	Kodierregel	Ankerbeispiel
Vorgehen und Bezugspunkte bei der Auswahl und Festlegung von Lehrzielen und Prüfungsinhalten	Aussagen darüber, wie die Lehrziele und Prüfungsinhalte ausgewählt werden, warum die angestrebten Kompetenzen und Fähigkeiten wichtig sind bzw. warum diese Kompetenzen und Fähigkeiten angestrebt werden. Wie wird entschieden, was Studierende am Ende einer Lehrveranstaltung können sollen?	<i>Ich nehme das Skript der Veranstaltung und überlege mir dann, welche Frage ich zu den Inhalten stellen könnte.</i>
Kategorie 3: Bildung und Strukturierung von Teilkompetenzen, Festlegung Bearbeitungsniveaus		
Ausprägung	Kodierregel	Ankerbeispiel
Strukturierung von Kompetenzen und Inhalten	Aussagen darüber, wie Kompetenzen aufgebaut sind, ob gewisse Inhalte als Grundlage für andere Inhalte betrachtet werden.	<i>Damit die Studierenden in der Lage sind, den Prozess zu erläutern, benötigen sie Wissen über die Bestandteile des Prozesses.</i>
Unterscheidung von Kompetenz- und inhaltlichen Niveaus	Aussagen darüber, dass Anforderungen auf bestimmten Niveaus ausgeführt werden können bzw. Abstufungen in der Qualität der Bewältigung von Aufgaben beachtet werden sollten-.	<i>Wir unterscheiden, ob Studierende die Begründung in Fachsprache formulieren und auf wissenschaftliche Theorien zurückgreifen. oder ob lediglich Allerweltsaussagen kommen.</i>
Kategorie 4: Formulierung von Lehrzielen		
Ausprägung	Kodierregel	Ankerbeispiel
Vorgehen und beachtete Kriterien bei der Formulierung von Lehrzielen	Aussagen, die sich auf das Vorgehen bei der Formulierung von Lehrzielen beziehen und welche Kriterien bei der Formulierung beachtet werden	<i>Bei der Formulierung von Lehrzielen achten wir darauf, dass die verwendeten Begrifflichkeiten eindeutig sind.</i>
Kategorie 5: Lehrzielmatrix		
Ausprägung	Kodierregel	Ankerbeispiel
Nutzung einer Lehrzielmatrix	Aussagen darüber, ob das Konzept der Lehrzielmatrix genutzt wird und, falls ja, auf welche Weise	<i>Mit der Matrix kann man Übersichtlich darstellen, wie viele Punkte pro Lehrziel in der Prüfung vergeben werden sollen.</i>

Erstellung von Prüfungsaufgaben		
Kategorie 6: Ausgangspunkt für die Aufgabenerstellung		
Ausprägung	Kodierregel	Ankerbeispiel
Lehrziele als Ausgangspunkt	Aussagen darüber, dass Lehrziele als Grundlage bzw. Ausgangspunkt für die Prüfungs- bzw. Aufgabengestaltung verwendet werden.	<p><i>Bei der Erstellung von Prüfungsaufgaben orientieren wir uns an den Anforderungen, die in den Lehrzielen beschrieben sind.</i></p> <p><i>Wir achten darauf, dass die Prüfungsinhalte die Lehrziele abbilden.</i></p> <p><i>Die Lehrziele helfen dabei, die Aufgaben zu konkretisieren.</i></p> <p><i>Durch die Lehrziele wissen wir, welche Anforderungen in einer Aufgabe vorkommen sollen.</i></p>
Vorlesungs-skript	Aussagen darüber, dass das Vorlesungsskript als Grundlage bzw. Ausgangspunkt für Prüfungs- bzw. Aufgabengestaltung verwendet wird.	<i>Zur Erstellung von Aufgaben blättere ich das Vorlesungsskript durch und überlege mir, zu welchen Seiten ich eine Aufgabe erstelle.</i>
Altklausuren	Aussagen darüber, dass vorhandene Altklausuren als Grundlage bzw. Ausgangspunkt für Prüfungs- bzw. Aufgabengestaltung verwendet werden.	<p><i>Ich entnehme die Prüfungsaufgaben aus bestehenden Klausuren.</i></p> <p><i>Als erstes sehe ich mir die Altklausuren an.</i></p>
Übungen	Aussagen darüber, dass die Übungen bzw. die darin bearbeiteten Aufgaben als Grundlage bzw. Ausgangspunkt für Prüfungs- bzw. Aufgabengestaltung verwendet werden.	<i>Ich lehne mich der Gestaltung der Prüfungsaufgaben an die Übungsaufgaben an.</i>
weitere	Aussagen zu weiteren Ausgangspunkten zur Aufgabengestaltung.	
Kategorie 7: Vorgehen bei der Aufgabenerstellung		
Ausprägung	Kodierregel	Ankerbeispiel
Generelle Angaben zum Vorgehen	Generelle Aussagen dazu, wie bei der Aufgabenerstellung vorgegangen wird, woher Ideen für Aufgaben kommen, was Überlegungen dabei eine Rolle spielen.	<p><i>Wenn ich eine Aufgabe erstelle, fange ich zuerst damit an, mir mögliche Szenarien zu überlegen.</i></p> <p><i>Ich erstelle zuerst den Angabentext.</i></p> <p><i>Ich gehe vom Ergebnis aus und überlege mir dazu einen Angabentext.</i></p> <p><i>Ich achte immer darauf, dass möglichst kurze Antworten möglich sind.</i></p>
Kategorie 8: Aufgabenanforderungen		
Ausprägung	Kodierregel	Ankerbeispiel
allgemein	Aussagen dazu, welche Anforderungen in Aufgaben enthalten sind bzw. was Studierende zur Lösung einer Aufgabe tun müssen.	<i>In dieser Aufgabe müssen die Studierenden erstmal erkennen, welche Werte sie brauchen..</i>

Kategorie 9: Qualitätsmerkmale von Aufgaben		
Ausprägung	Kodierregel	Ankerbeispiel
allgemein	Angaben dazu, auf welche Aspekte und Standards bei der Aufgabenerstellung geachtet wird, z.B. dass alle Aufgaben eindeutig sind. Gemeinsamer Standard, den Aufgaben erfüllen sollen.	<i>Ich achte darauf, dass die Aufgaben eindeutig gestellt sind. Es ist wichtig, dass jede Aufgabe nur eine Handlungsaufforderung enthält.</i> <i>Wir achten darauf, die Antwortfelder vorzustrukturieren.</i> <i>Jede Aufgabe soll einen Kontextbezug haben.</i>
Kategorie 10: Aufgabentypen		
Ausprägung	Kodierregel	Ankerbeispiel
Aufgabentypen	Aussagen zur Verwendung verschiedener Aufgabentypen, z.B. Wissensaufgaben, Rechenaufgaben, etc.	<i>Ich versuche immer, einen Mix aus Wissens-, Anwendens-, und Begründungsaufgaben hinzubekommen</i>
Rückbindung der Aufgaben an den Prüfungsgegenstand		
Kategorie 11: Maßnahmen zur Überprüfung der kognitiven und inhaltlichen Validität von Aufgaben		
Ausprägung	Kodierregel	Ankerbeispiel
Maßnahmen zur Sicherung der Aufgabengültigkeit	Aussagen, die darauf hindeuten, dass die Inhaltsvalidität der Aufgaben beachtet und betrachtet wird, indem z.B. Überlegungen angestellt werden, welche Fähigkeiten ein Student benötigt bzw. welche Prozesse er durchläuft, um eine Aufgabe zu lösen. Mit dieser Kategorie kann festgestellt werden, ob das Prinzip der Inhaltsvalidität unter Dozierenden bekannt ist und eine Rolle spielt. Allgemein werden hier alle Aussagen eingeordnet, die Auskunft darüber geben, ob in irgendeiner Form die Inhaltsvalidität von Aufgaben betrachtet bzw. überprüft wird.	<i>Ich schaue mir die Aufgaben schon an und überlege, ob ich unterschiedliche Anforderungen abgedeckt habe.</i> <i>Man sieht ja relativ schnell, was eine Aufgabe abprüft und ob es das ist, was wir wollen.</i> <i>Ich schaue schon, dass die Prüfungsaufgaben von der Schwierigkeit her ähnlich sind wie die Übungsaufgaben.</i>
Bewertung von Prüfungsantworten		
Kategorie 12: Schulung von Korrektoren		
Ausprägung	Kodierregel	Ankerbeispiel
Trifft zu	Aussagen, die zeigen, dass die Korrektoren vor der Korrektur instruiert werden.	<i>Bei uns werden die Korrektoren eingewiesen.</i>
Trifft nicht zu	Aussagen, die darauf hinweisen, dass keine Schulung von Korrektoren erfolgt.	<i>Die Korrektoren erhalten eine Prüfung mit Musterlösung und legen dann los.</i>
Kategorie 13: Konsistenz der Bewertung		
Ausprägung	Kodierregel	Ankerbeispiel
Allgemein	Aussagen, die sich auf Maßnahmen beziehen, um die Konsistenz bei der Bewertung von Prüfungsantworten zu beeinflussen.	<i>Bei uns korrigiert jede Person immer nur eine Aufgabe, damit es gleichwertig wird.</i>

		<p><i>Wenn nicht klar ist, wie eine Aufgabe bewertet werden sollte, können wir beim Ersteller der Aufgabe nachfragen.</i></p> <p><i>Ich schreibe mir immer die Antworten raus, auf die ich Punkte vergeben habe, obwohl sie von der Musterlösung abweichen, damit ich es bei allen einheitlich bewerten kann.</i></p>
Kategorie 14: Vorgehen bei der Bewertung von Prüfungsantworten		
Ausprägung	Kodierregel	Ankerbeispiel
allgemein	Aussagen, die beschreiben, wie bei der Korrektur vorgegangen wird. Dazu zählen Aussagen, welche Personen beteiligt sind, wer welche Aufgaben korrigiert.	<p><i>Bei uns korrigieren immer alle am Lehrstuhl mit.</i></p> <p><i>Wir machen Haufen und jeder nimmt sich einen Stapel weg.</i></p> <p><i>Jeder Korrektor korrigiert immer nur eine Aufgabe komplett durch.</i></p>
Punkte- und Notenvergabe		
Kategorie 15: Vergabe von Punkten		
Ausprägung	Kodierregel	Ankerbeispiel
allgemein	Aussagen, die sich auf die Vergabe von Punkten beziehen, also nach welchen Kriterien Punkte auf Aufgaben verteilt werden.	<p><i>Wir vergeben die Punkte nach geschätzter Bearbeitungsdauer.</i></p> <p><i>Wir haben keine halben Punkte.</i></p> <p><i>Wir vergeben pro Minute Bearbeitungsdauer einen Punkt.</i></p>
Kategorie 16: Notenvergabe		
Ausprägung	Kodierregel	Ankerbeispiel
allgemein	Aussagen dazu, wie Noten vergeben werden, z.B. wie der Notenschlüssel festgelegt wird, und welche Rolle die Notenverteilung spielt.	<i>Erfahren wie durch Prüfungsaufgaben das Lernverhalten gesteuert werden kann.</i>
Rückschlüsse aus den Prüfungsdaten		
Kategorie 17: Betrachtung von statistischen Kennwerten		
Ausprägung	Kodierregel	Ankerbeispiel
Minima / Maxima	Aussagen dazu, dass die minimal und maximal erreichte Punktzahl bei Aufgaben betrachtet wird.	<i>Dass durch die Vergabe von Punkten eine Gewichtung vorgenommen wird, war mir so nicht bewusst.</i>
Aufgabenschwierigkeit	Aussagen dazu, dass die Aufgabenschwierigkeit betrachtet wird.	<i>Wenn wir sehen, dass eine Aufgabe von sehr wenigen gelöst wird, überlegen wir, ob wir die Aufgabe raus nehmen.</i>
Trennschärfe	Aussagen dazu, dass die Trennschärfe betrachtet wird.	<i>Wir möchten, dass alle Aufgaben eine bestimmte Trennschärfe aufweisen.</i>
Kategorie 18: Rückschlüsse aus den Daten		
Ausprägung	Kodierregel	Ankerbeispiel

allgemein	Aussagen dazu, welche Rückschlüsse aus den Daten gezogen werden.	<i>Wenn bei einer Aufgabe von niemandem alle Punkte erreicht wurden, schauen wir uns an, ob irgendetwas an der Aufgabenstellung falsch ist.</i>
Kategorie 19: Aufbereitung von Datensätzen		
Ausprägung	Kodierregel	Ankerbeispiel
allgemein	Aussagen dazu, wie die Daten der Studierenden erfasst und in Datensätzen angelegt werden.	<i>Wir legen zu jedem Student in einer Excel-Liste an, wie viele Punkte er oder sie pro Aufgabe erzielte.</i>

B. Verwendete Zitate zur Kategorienbildung

a. Zitate zur Festlegung des Prüfungsgegenstands

Zitat 1 Fall MW 1

Interviewer: Was sollen die Studierenden am Ende der Veranstaltung können?

Akademischer Oberrat: Okay, was ist das, was wir hier in unserer VL vermitteln wollen? Natürlich geht das auch in die Richtung von den Ingenieuren, weil hoffentlich haben wir irgendwo ein gesamtheitliches Bild, das stimmig ist und wir haben Professoren aus der Industrie und damit hat man natürlich auch die Anforderungen aus der Industrie. Oder wir haben ein Industrieprojekt, wo man sich austauscht, ja, was ist denn jetzt das Problem von den aktuellen Ingenieuren, was ist denn gut von den neuen Ansätzen und dann versucht man das natürlich in ein stimmiges Bild zu geben. Und ich sage jetzt mal: ganz weit weg sind wir ja davon nicht, aber man muss ja ständig noch besser werden und ich sage jetzt mal das Resümee aus diesen Feedbacks, das kommt dann letztendlich auch wieder in die Vorlesung hinein. Wo man merkt: okay, was fehlt, es gibt sehr gute Spezialisten, aber wir wollen jetzt mit unserer, *mit dieser Vorlesung, mehr den Allrounder, also die gesamtheitliche Betrachtung, fördern. Deswegen eben der Fokus auf zentrale Konzepte anstatt die Auslegung von kleinen Teilprozessen.* Sondern da geht es ein bisschen größer und mit dieser Vorlesung adressieren wir das.

Zitat 2 Fall MW 1

Doktorand 1: Das ist interessant mit dem Wissen jetzt [zu den Lehrzielen]. Allein, wenn man sich den Satz durchliest, dann kann man sich schon gleich die Frage stellen, was heißt das eigentlich? Ich meine, es ist ja grundsätzlich nicht falsch, es ist halt nur sehr grob.

Akademischer Oberrat: Also, [liest Lehrziel vor] „nach der Teilnahme an dieser Veranstaltung können die Studenten den Ablauf des Prozess xy darlegen“. So, also der Prozess xy, ich komme jetzt sozusagen von der Inhaltsseite her, das ist ja das, was wir vermitteln wollen, oder hat sich das geändert mittlerweile?

Doktorand 1: [...] Genau stimmt, der Prozess xy ist einer der Hauptpunkte.

Akademischer Oberrat: [...]Ok, also für mich ist es jetzt erstmal wichtig: ist es das auch wirklich, was wir erzählen wollen. Also, ich meine: ja. Also ich weiß noch, vor allem früher, bei Professor x [dem vorherigen Lehrstuhlleiter], war das genau, was wir erzählen wollen. Jetzt weiß ich nicht, hat sich das bei Professor y [neuer Lehrstuhlleiter] geändert oder ist das weiterhin wirklich ein Inhalt, der relevant ist [den Prozess xy darzulegen]? Wir wollen jetzt, dass die Leute wissen,

wie man das Produkt xy baut [...] anstatt, dass sie für einen Teilprozess eine bestimmte Rechnung durchführen können.

Doktorand 2: Ja, auf jeden Fall. Das ist einer der Hauptschwerpunkte eigentlich. Und sie [die Studierenden] müssen dafür die [Produktentwicklungs-] Prozesse verstanden haben.

Doktorand 1: Ja, die Prozessphasen von Projektstart bis zur Serienentwicklung.

Doktorand 2: Aber wir wollen auch etwas mehr als verstehen, glaube ich. Wir wollen doch schon, es geht fast Richtung Anwendung, oder? Wenn ich da nämlich an die Prüfung denke, an den Teil B, da müssen die Studierenden das selber eigentlich teilweise schon anwenden.

[...] Und sie sollten das auch später im Beruf als Ingenieure anwenden können.

Interpretatives Vorgehen 1 zu Zitat 2

Aus diesem Dialog lassen sich zwei Erkenntnisse ableiten. Zum einen wird die Wahrnehmung der Teilnehmer darüber, was relevante Anforderungen für die Prüfung sind, von den Lehrzielen des Lehrstuhlleiters beeinflusst, in diesem Fall, dass die Studierenden in der Lage sein sollen, ein bestimmtes Produkt auszulegen und herzustellen unter Berücksichtigung des Gesamtprozesses anstatt einer Unterteilung in viele Teilprozesse mit sehr speziellen Berechnungen. Dieser Anspruch ist zudem in der Struktur der Prüfung abgebildet. Die Prüfung des Lehrstuhl MW 1 besteht seit 2009 aus zwei Teilen: Im ersten Teil wird vorwiegend Wissen abgefragt; im zweiten Teil wird mittels eines Fallbeispiels überprüft, ob die Studierenden ein ganzheitliches Verständnis der behandelten Konzepte entwickeln konnten und ihr Wissen und Verständnis einsetzen können, um das Fallbeispiel richtig zu lösen. Den Mitarbeitern wurde über das bereits vorab festgesetzte Format der Prüfung vorgegeben, dass in der ersten Hälfte der Prüfung die Wiedergabe der behandelten Inhalte im Vordergrund steht und im zweiten Teil das Anwenden der Inhalte auf ein Fallbeispiel.

In den Aussagen sind Überlegungen zur theoretischen Modellierung des abzu prüfenden Konstrukts enthalten. Die geäußerten Annahmen darüber, dass ein Verständnis über den Produktentstehungsprozess und weiteren Konzepten Grundlage für die Entwicklung eines Produkts sind, wurden jedoch durch den Input der Moderatoren getriggert und werden daher der veränderten Praxis zugeordnet. Für die bestehende Praxis konnten keine Hinweise über derartige Überlegungen gefunden werden. Die Gruppe arbeitete zudem nicht mit den bestehenden Lehrzielen. Die im theoretischen Rahmen vorgesehen Schritte der theoretischen Modellierung, der Konkretisierung von Lehrzielen und der Nutzung einer Lehrzielmatrix wurden in der bestehenden Prüfungspraxis am Lehrstuhl MW 1 nicht genutzt.

Zitat 3 Fall MW 1

Interviewer: Wie legen Sie die Inhalte für die Prüfung fest?

Doktorand 2: *Wir haben uns die Vorlesungstermine angesehen und haben dann jeweils eine bestimmte Anzahl an Prüfungsaufgaben erstellt, angelehnt daran, wie lange ein bestimmtes Kapitel, also wie viele Stunden es in der Vorlesung in Anspruch genommen hat. Wir versuchen, dass wir jedes Kapitel abdecken und dass die Inhalte der Prüfung in etwa anteilmäßig den Inhalten der Vorlesung entsprechen. Das war so unser Vorgehen.*

Zitat 4 Fall MW 1

Interviewer: Wie sind Sie bislang bei der Auswahl der Prüfungsinhalte vorgegangen?

Akademischer Oberrat: *Prinzipiell schon irgendwo am Skript entlang, um festzustellen: was sind die Inhalte. Dann aber auch, ich meine bei den Inhalten gibt es schon so Sachen mit Wissensvermittlung, da kann ich dann auch nicht mehr abfragen als „nennen Sie“. Und dann gibt es natürlich die Sachen, die ein bisschen tiefergreifend sind, wo man aber dann auch im Allgemeinen mehr Fokus drauf legen möchte. Da gibt es dann aber auch mehr Begründen und mehr von diesen Fragen. Wenn wir die Prüfung erstellen, in gewisser Weise machen wir dann schon sowas [zeigt auf die Lehrzielmatrix], aber bei weitem nicht so strukturiert. Man hat da irgendwo im Hinterkopf: ja, okay. das sind schon die schwierigen Punkte und die wollen wir ihnen [den Studierenden] beibringen und dann fragen wir das auch ab. Aber, wie gesagt, halt bei weitem nicht so strukturiert. Und dann ergibt es sich meistens, dass wir häufig zu viel Auswendiglern-Fragen haben, die wir nicht mögen, aber manchmal kommt man nicht drum rum. Sobald wir einen Inhalt haben, zu dem wir den Studierenden bloß eine blöde Auswendiglernfrage stellen, versuchen wir in der Regel auch was anderes, um etwas anderes zu fragen. Und dann schaut man das man in Summe ein vernünftiges Konstrukt zusammen bringt. Nicht so strukturiert, aber ein ähnliches Vorgehen.*

Zitat 5 Fall MW 2

Interviewer: Wie legen Sie die Prüfungsinhalte fest?

Doktorand 2: *Also bei der Prüfungserstellung gehen wir in gewisser Weise ähnlich vor [wie Lehrstuhl MW 1: Ableitung von Prüfungsaufgaben aus den Vorlesungskapiteln und Materialien]. Nur, dass es mehr nach unserer Überzeugung ist, dass die Studenten sozusagen jedes Kapitel lernen und nicht dadurch, [...] dass die jetzt sagen, ok, das ist jetzt das einzig Relevante, weil das ist so lange behandelt worden. Deswegen picken wir eigentlich im Grunde eine Zufallsauswahl, jedes Mal eine neue Auswahl von Aufgaben aus und das kann halt dann verschiedene Dinge abprüfen,*

weil wir dann davon ausgehen, wenn der Student nicht weiß, sage ich mal, es kommt Kapitel 2 und 3 dran, dass er dann auch gleichmäßig lernen und vorbereiten muss.

Zitat 6 Fall MW 2

Doktorand 2: Wir haben eine klare Trennung gemacht zwischen einem VL-Teil und einem Übungs-Teil. Das heißt, das Wissen, das in der VL vermittelt wird, oder das im Skript steht, wird mehr über MC-Fragen und offene Fragen abgefragt. Und das, was wir in den Übungen machen, dazu haben wir in der Regel Rechenaufgaben, die auch einen Transfer erfordern oder in vielen Fällen auch eine Reflexion des Ergebnisses. Aber nicht so sehr ein Wissensabfragen, das findet ja im ersten Teil statt.

Zitat 7 Fall MW 2

Interviewer: [bezieht sich auf die Erkenntnisse aus dem Informationsgespräch]: Bei euch ist es ja so: ihr sagt in der Veranstaltung tauchen einfach Grundbegriffe auf, die den Studierenden auch ständig begegnen [im weiteren Studienverlauf und der beruflichen Praxis] und die müssen sie einfach wissen und das ist der Grund, warum ihr die Aufteilung so habt [mit Wissens- und Rechenanteil], oder?

Doktorand 2: *Im Grunde ist das Ziel, dem Ingenieur Grundlagen der technischen Betriebsführung, so ähnlich hieß das Fach früher auch mal, zu vermitteln. Sozusagen eigentlich einen Einblick in betriebswirtschaftliche Inhalte zu geben, die eben auch für den heute praktizierenden Ingenieur von Bedeutung sind, dass er sie weiß- also ein bisschen auch von Finanzen zu verstehen, ein bisschen von Führungsverhalten zu verstehen, ein bisschen was von Auftragsabwicklung zu verstehen, ein bisschen was von Arbeitsrecht und Betriebsorganisation. Also wie ist so ein Unternehmen überhaupt aufgebaut? Was gibt es da für verschiedene Formen? Ein bisschen was Rechtliches ist auch noch mit dabei. Es geht auch darum, ein Bewusstsein zu wecken, dass man eben nicht nur seine technische Qualifikation im Maschinenbaustudium ausbildet, sondern wenn es später eben in die Praxis geht, es sehr förderlich ist, wenn man eben auch mitreden kann in anderen Dingen. Dass man einmal eben Interesse schafft, aber auch eben ein Grundverständnis. Mehr können wir uns natürlich in dem breiten Überblick nicht leisten. Aber dass man von allem eben ein bisschen was mitnimmt, um dann später gerüstet zu sein.*

Zitat 8 Fall MW 2

Doktorand 1: Ich wollte nur noch ergänzen, dass die Veranstaltung eigentlich eines unserer Grundlagen-Basisfächer ist, wo dann eben andere Fächer auch wieder darauf aufbauen.

Doktorand 3: *Es geht ja auch noch, ich meine, es ist ja eine wissenschaftliche Veranstaltung und da muss es halt dann eben exakt sein [das begriffliche Wissen]. [...]* Wenn man in der Praxis, später im Unternehmen, Begriffe erklären soll, dann spüren die Leute das schon, wenn man das ein bisschen umschreibt. Das ist ja auch richtig dann. Aber wenn man jetzt hier ist und sagt: wir machen jetzt eine Veranstaltung im Bereich xy und wir machen jetzt hier Wissenschaft, dann muss das eben exakt sein und dann muss das auch das sein [die verwendeten Begriffe in den Prüfungsantworten], was letztendlich so auch in der Veranstaltung vorgestellt wurde. Es geht nicht um Praxiswissen dann direkt, das hören vielleicht viele nicht gerne, aber das ist halt letztendlich so.

Zitat 9 Fall MW 2

Doktorand 1: Ich wollte gerade sagen, also das Problem ist, es heißt zwar so, die Veranstaltung und es werden verschiedene Methoden auch vorgestellt. Das ist halt ein bisschen wenig konkret, ja. Die Methoden und Konzepte, weil: was ist das Konzept, oder was ist ein Konzept?

Doktorand 2: Wir präsentieren sozusagen keine Methoden, wie man die Zusammenhänge versteht oder so, sondern wir präsentieren aus den Einzelteilbereichen Methoden. Und das sind dann eben Ergo-Methoden der Unternehmensführung.

Doktorand 1: Genau, aber für die Anwendung der Methoden, also gerade jetzt zum Beispiel in der Kostenrechnung oder so, brauchst du teilweise auch Wissen aus den angrenzenden Bereichen.

Doktorand 3: Was ist mit angrenzenden Bereichen gemeint? Sind die fachlich?

Zitat 10 Fall MW 2

Doktorand 2: Vielleicht noch eins, das würde ich gerne reinbringen mit dem, dass man sozusagen diese breite Übersicht gewinnt, das steht ja irgendwie jetzt schon drinnen [in den Lehrzielen], aber ein Bewusstsein für die Wichtigkeit dieser Themen für den praktizierenden Ingenieur im produzierenden Betrieb. Oder wie auch immer. Weil das ist dem Lehrstuhlleiter wichtig, dass wir...

Doktorand 1: Wie würdest du das formulieren?

Doktorand 2: Ja, ich weiß nicht, ob man sagen kann, dass das Ergebnis ist, dass die Bedeutung dieser vermittelten Thematiken im Rahmen der praktischen Anwendung bewusst geworden ist, oder so was.

Zitat 11 Fall MW 3

Doktorandin: Eigentlich müsste unser Ziel sein [für die Lehrveranstaltung], dass sie durch die Vorlesung und *durch die Prüfung Sachen gelernt bekommen, die sie dann in Studienarbeiten an unserem Lehrstuhl anwenden können*. Also, dass man nicht bei null anfangen muss und denen erklären muss, was das jetzt genau für ein Tool ist, sondern ihnen vielleicht die Anwendung dann mal zeigt, aber sie jetzt dann schon Bescheid wissen, was man mit dem [Tool] eigentlich machen kann. Also und das ist ja jetzt auch so, dass meine Lehrveranstaltung *als Grundlagenveranstaltung* gilt und also ich habe da jetzt auch am Anfang des Semesters darauf hingewiesen. *Es werden halt die Themenbereiche xy, die werden halt theoretisch hier haarklein besprochen*. Da werden Modelle gezeigt und die erklärt der Professor genau und *die sollen dann halt eigentlich in den anderen weiterführenden Veranstaltungen einfach bekannt sein und vorausgesetzt werden*. Also, dass das halt nicht mehr von vorne erklärt wird. [...] *Es wäre schon unser Ziel gewesen, dass man in dieser Lehrveranstaltung grundlegendes bespricht*.

Zitat 12 Fall MW 3

Doktorandin: Ich habe die letzte Prüfung gestellt, das war im Sommersemester. Aber halt noch zusammen mit meinem Vorgänger praktisch. Also ich habe sie zusammengestellt und er hat es sich angeschaut, ob das so okay ist. Aber ich denke, *das war ein bisschen sehr wissenslastig*. Also die Rechenaufgaben haben halt da gefehlt. Und sonst habe ich halt für das letzte Jahr eher nur so aus den vorherigen Jahren was übernommen.

Zitat 13 Fall MW 3

Moderator: Also, wenn man das [die Prüfungsanforderungen] jetzt mal mit den Lehrzielen [der Veranstaltung] vergleicht, dann ist hier ja ein Lehrziel mit Erinnern und zwei Lehrziele, die sich auf Analysieren bezieht. Wenn man dann aber in die Prüfung hineingeht (unterbrochen durch Doktorandin) /

Doktorandin: Also das wäre auch schön, wenn sie was Neues entwickeln können, aber ich denke, ich muss mir das nochmal kurz anschauen. Also und da geht es ja eher darum, wie kann ich das bewerten, eher nüchtern, die Entwicklung. Also und das würde halt einen Überblick geben, beispielsweise hier Vor- und Nachteile, aber ich gehe ja nicht mehr drauf ein. Also beispielsweise das wären jetzt Analyseverfahren, die man halt zur Bewertung von Arbeitsplätzen hernimmt. Die sollen das mal gehört haben, sollen halt das abwägen können, für welche Ergebnisse ich halt dann welche einsetzen kann.

Moderator: Aber ist dann in dem Fall nicht wieder bei der Reproduktion.

Doktorandin: Das ist wieder Reproduktion, ja. Wenn ich eigentlich denen schon genau vorher zeige: okay, das sind die Pros und Kontras und jetzt reproduziere mal kurz deine Bewertung, die ich euch gezeigt habe.

Zitat 14 Fall MW 3

Doktorandin: *Anwenden, denke ich, ist halt ein bisschen schwierig* [für die Studierenden in der Prüfung], weil so detailliert werden die [verschiedene Methoden] *in der Vorlesung nicht durchgenommen, dass die sich da hinsetzen können und das sofort können.* Also ihnen wird halt praktisch nur grob erzählt, was man [mit den besprochenen Methoden] damit machen kann. Aber wie das genau das vonstattengeht, erfolgt halt in der Vorlesung eigentlich nicht. Und das ist halt dann eher die Sache, die dann im Praktikum noch gemacht wird. Dass man ihnen zeigt, wie man das Messsystem anwenden kann.

Zitat 15 Fall MW 3

Moderator [bezieht sich auf ein Lehrziel]: Aber ist dann in dem Fall nicht wieder bei der Reproduktion, wenn ich eigentlich denen schon genau vorher zeige: okay, das sind die Pros und Kontras zu einem Verfahren und jetzt reproduziere mal kurz die Bewertung dieses Verfahrens, die ich euch in der Veranstaltung gezeigt habe.

Doktorandin: Ja, wie gesagt, bei mir ist halt echt sehr viel Reproduktion.

Zitat 16 Fall MW 4

Doktorand [zur Frage, ob die Projektziele noch aktuell sind]: Also, bei mir passt es. Vor allem dieses Bulimie lernen und oben [am Anfang der Prüfung] diese großen Aufgaben, das geht Hand in Hand. Weil mit diesen großen Aufgaben, wenn sie die bewältigt haben, dann zeigen die, dass sie es verstanden haben und damit dieses Bulimielernen irgendwo umgehen. *Es wird sich aber nicht vermeiden lassen, da weiß ich nicht, 40-50% an Fragen zu haben, die einfach platt Wissen abfragen. Weil teilweise fragen wir Wissen ab.* Was für eine DIN, was für ein Modell, was für eine Tools kann man nutzen, um das zu bewerten. *Also da wird Wissen abgefragt.* Es wird beides zu machen sein.

Doktorandin MW 3: Man kann das auch nicht, also manche Thematiken kann man irgendwie nicht anwenden. Oder? Das meinst du ja.

Doktorand: Richtig, also es wird so eine Mischklausur, wird es bleiben auf jeden Fall. Also Anwendung plus die Frage: Wisst ihr denn, was es für Tools gibt, um Arbeitsplätze zu bewerten beispielsweise. Und da wird Wissen abgefragt.

Moderator: Ja, das ist auch genau die spannende Frage, die wir heute mit Ihnen besprechen wollen. Was sollen die Studierenden am Ende [der Lehrveranstaltung] können? Und wenn es zum Teil das ist, dass die ein paar Tools auswendig können oder verstanden haben, dann ist es auch okay, dass Wissen abgefragt wird in der Klausur. Deswegen ist heute genau der Fokus darauf, dass wir das mit Euch durchsprechen.

[...]

Doktorand: Also das ist, ja, die sollen Arbeitsplätze in der Produktion bewerten können und du siehst, die erste Aufgabe [in der Prüfung] ist nicht direkt in der Produktion, sondern so ein Waldarbeiter und die [Studierenden] bewerten den Arbeitsplatz anhand der Methoden, die sie in der Vorlesung bekommen haben. Also die Prüfungen sind nicht vorlesungsfremd. Und, wie gesagt, teilweise diese einzelnen kleineren Fragen, fragen halt einfach plattes Wissen ab.

Zitat 17 Fall MW 4

Moderator [weiterführende Frage zu den Prüfungsanforderungen]: Genau, jetzt versetze ich mich in die Rolle dieses Studenten und überlege mir, dass ein Student der in der Lage ist, diese Aufgabe zu lösen, was muss der jetzt Können, Wissen und Tun, um die Aufgabe richtig zu bearbeiten. Und in dem Fall muss er [bei einer Aufgabe, die besprochen wurde] diese Informationen erstmal finden, er muss diese Formel finden und muss da einsetzen und muss dann diesen Wert ausrechnen.

Doktorand: Ok, in der Maximalkraft. Er muss checken, okay, diese Formel darf ich nur anwenden, wenn die die Grenze xy, das sind diese 15%, überschritten wird, ist [...] Die gilt aber nur, wenn die Säge schwerer ist, wie gesagt, als diese Grenze von 15%, wenn die überschritten wird. Das muss der vorher prüfen. Setze ich ein: der Mann hat eine Maximalkraft, das habe ich aus der Tabelle. [...] Wir geben die Tabellen ja vor. Die Tabellen sind für bestimmte Fälle gemacht worden, mit bestimmten Kohorten. Und so läuft das auch in der Praxis. Du kannst ja nicht für jeden Arbeitsplatz, dann kannst du ja eine Million von Tabellen, also so viele Arbeitsplätze wie es gibt, kannst du da erfassen.

Zitat 18 Fall MW 4

Doktorand: Die [Prüfungsaufgaben] lehnen sich schon stark an die Übungsaufgaben an. Also gerade das Rechnen in der Produktionsergonomie. Und diese frei zu beantwortenden Fragen kommen aus dem Skript. Also das muss dann aus dem Skript beantwortet werden.

[...]

Moderator: Okay, also das heißt, das machst komplett du jetzt [die Erstellung der Prüfung]? Also es gibt zwar die ganzen alten Prüfungen, die du auch reingestellt hast auf's Laufwerk, aber die waren alle vom Vorgänger?

Doktorand: Genau, und selbst der hat sich bei seinem Vorgänger bedient. Gerade bei den *großen Aufgaben gucken, was war vorher [in den Altklausuren]*. Und wie gesagt, die kleinen, die gehst du anhand des Skriptes an.

Zitat 19 Fall MW 3 und MW 4

Moderator: Von der Gewichtung der Aufgaben in der Prüfung, legt ihr da irgendwie fest, zu welchen Themenbereichen wie viele Aufgaben dran kommen sollen? Oder wie werden die Aufgaben zusammengestellt?

Doktorand MW 4: *Weiß ich noch nicht*, also wie gesagt, ich....

Doktorandin MW 3: *Ach so, bei mir ist das so, eigentlich sollte 1/3 Wissensfragen, 1/3 Rechenfragen, 1/3 Transferaufgaben sein. So in der Richtung. Aber aus welchen Kapiteln die jetzt drankommen...*

Doktorand MW 4: Wer hat das gesagt?

Doktorandin MW 3: *Kollege x hat das immer gesagt. Aber ich glaube, das war nicht immer so. Weil 1/3 Rechenaufgaben...*

Doktorand MW 4: *Also hier ist, glaube ich, weiß nicht, 50% Rechnen. Wenn nicht sogar noch mehr....*

Moderator: Also 1/3 Rechnen, 1/3 Wissen...

Doktorandin MW 3: *Wissen und 1/3 Transfer. Ja, dass man das halt irgendwie noch auf ein Beispiel anwendet.*

Moderator: Ist das das Ziel oder ist das der Ist-Zustand?

Doktorandin MW 3: Also der Ist-Zustand ist das nicht. Also das ist das Ziel und das sagen wir auch immer, aber das ist nicht der Ist-Zustand.

Moderator [Frage an Doktorand von MW 4]: Was ist dann euer Ziel von der Verteilung?

Doktorand MW 4: *Wie gesagt, da müsste ich mit dem Chef nochmal reden, was er für ein Ziel hat. Ich habe es übernommen und die Klausur steht noch nicht an. Wurde jetzt vom Vorgänger gemacht und der hat mich jetzt auch nicht dahingehend gebrieft, okay: 1/3, 1/3, 1/3.*

Zitat 20 Fall MW 4

Doktorand: Er [der oder die Studierende] nimmt sich dann hier die Formel. Wir fragen hier nach der Maximalkraft und der maximalen Haltezeit. Das ist Einsetzen, das ist nicht mal... also, das ist zwar Analysieren, aber auch nur Einsetzen.

Moderator: Die Frage ist, ob es Analysieren ist, oder ob es die Lösung von dieser Matheaufgabe ist.

Doktorand: Es ist das Lösen der Matheaufgabe.

Moderator: Dann messe ich hier eigentlich ganz was anderes, wie das, was ich hier intendiert habe. Und deswegen müsste ich es hier vielleicht noch ein bisschen ausführen. Wenn ich sage: Analysieren. Analysieren heißt aber in unserem Fall, dass er einfach die einzelnen Analyseschritte, die sind vorgegeben, dass er die ausführen kann. Das haben wir ja hier. Oder ich sage: Ja, analysieren bedeutet bei uns, ich stelle mir das vor, hinterher im Betrieb, der kriegt eine Situation vorgestellt und der muss das dann selber entscheiden, welche Methode nimmt er, er muss dann festlegen, was sind die einzelnen Analyseschritte und muss sie dann noch ausführen können. Dann hätte ich vorgeschaltet noch zwei andere Sachen. Und das muss ich hier in den Lehrzielen ebenfalls festlegen. Dann kann ich zielgerichtet die Prüfung erstellen. Das müsste man hier jetzt noch herausarbeiten.

Doktorand: Geht da hervor, dass das funktioniert, das Analysieren. Wir geben aber,- also so ist das hier in der Klausur gemacht worden. Das ist schon Analyse, aber die Schritte sind schon vorgegeben. Das ist eine schrittweise vorgegebene Analyse.

Moderator: Jetzt aber nochmal die Frage, was tut der Student: Analysiert der tatsächlich, oder rastet der ein Schema runter, wo dann in der Gesamtschau, das war dann eine Analyse, aber selber hat er eigentlich nicht analysiert, sondern diese Einzelaufgaben ausgeführt? Das muss ich mir hier schon überlegen. Weil dann weiß ich, wie müssen die Aufgaben aussehen. Wenn ich analysiere in diesem breiteren Sinne verstehe, dann müsste ich die Aufgaben anders machen. Und von dem her müsste ich das jetzt soweit definieren, dass dann einer dieses Verständnis hat. Aber worum geht es euch denn? Dass sie die einzelnen Schritte ausführen können in diesem vorgegebenen Schema, oder dass sie tatsächlich diese Situationen erstmal erfassen und strukturieren und dann entscheiden können, welches Analyseschema nehme ich und das dann aufstellen können? Und die Rechnung macht dann ein anderer, ich weiß es nicht.

Zitat 21 Fall MW 5

Doktorand 1: *Ich sehe es jetzt nicht zwangsläufig so, dass die Prüfung unser finales Ziel ist. Das ist es für mich nicht. Es ist für mich die Ausbildung der Studenten für die industrielle Praxis. Deshalb bringen wir Praxisbeispiele. Im Maschinenbau sind wir sehr praxisorientiert, das ist ja auch ein Vorteil. Die Prüfung ist für mich die letzte Qualitätskontrolle, bevor die Studenten rausgehen. Nichts anderes eigentlich.*

Zitat 22 Fall MW 5

Akademischer Oberrat: Das ist in den anderen Fächern, da heißt es: berechnen Sie eine Welle-Nabe-Verbindung. Und man klappt ein Buch auf und dann nehme ich die Formel her und man setzt die Kennwerte ein, weil die Welle ist so groß und ich habe den Werkstoff und dann rechne ich die Formel aus und habe das Ergebnis. *Und das kann ich in der Industrie später genauso machen.* Wir zeigen halt unterschiedliche Vorgehensweisen und dann kommt es darauf an, *bin ich bei einer Firma, die macht immer nur Anlagenbau und jede Maschine ist unterschiedlich, oder bin ich bei Bosch-Siemens-Hausgeräte und da fallen die Kaffeemaschinen nur so vom Band runter, also große Stückzahl. Und dann habe ich eine andere, im Detail eine andere Vorgehensweise, wie ich da versuchen würde, Kosten zu senken.* Die Grundstruktur der Methode ist die Gleiche. Ich brauche ein Team, ich muss meine Kostenstruktur des Produktes kennen, von der Kaffeemaschine das Gehäuse und die Elektronik, bei einer großen Maschine muss ich auch die Kostenstruktur kennen. Wie viele Zukaufteile sind da drinnen, wie viel Eigenfertigungsanteil habe ich da? Also die Grundmethodik ist die Gleiche, aber ich muss dann unterschiedliche Ideen entwickeln, wie ich bei der Maschine die Kosten senke und wie ich bei der Kaffeemaschine die Kosten senke. Bei der Kaffeemaschine schaue ich vielleicht auf die Materialkosten und bei der großen Maschine schaue ich auf das grundsätzliche Konzept. Welche Elektromotoren wende ich da an oder wie baue ich die Maschine insgesamt auf?

Zitat 23 Fall MW 5

Akademischer Oberrat: Wir haben fünf Vorlesungskapitel. *Dann haben wir verteilt, wer zu welchem Kapitel Aufgaben stellen soll.* Das war verpflichtend. Ich hatte das erste Kapitel. Ich habe mir Aufgaben rausgesucht aus meiner Tabelle, die ich von früher noch habe, welche Aufgaben passen da rein, und habe die zusammengeschrieben. Ich mache es in der Regel so, dass wir vom Grunddokument, nehme ich die vorhergehende Prüfung her, weil da habe ich bei mir immer vorne meinen Text drauf mit Geheimhaltung und darf nicht kopiert werden und so weiter. Da ist copyright drauf und dann habe ich *meine Exceltabelle, wo ich sag:, die Aufgabe nehme ich her und die kopiere ich dann da rein. Die nehme ich aber aus zwei Jahren vorher aus einer Prüfung.*

Oder ich nehme meine ursprüngliche Variante, wo ich nur eine Datei habe. Aufgabe Nummer 148 die kopiere ich dann da rein.

Doktorand 2: *Der Umfang [der Aufgaben pro Kapitel] sollte in erster Linie passen, die Punktzahl, dass das Kapitel einen entsprechenden Anteil in der Prüfung hat.*

Zitat 24 Fall MW 5

Moderator: Sie gehen dann so vor, dass Sie sagen: Wir haben einen Rechenteil, wir haben einen Teil über Begriffsdefinitionen, oder?

Akademischer Oberrat: Ja, eine gesunde Mischung. *Ich habe die Aufgabentypen und da kommt in jeder Prüfung zu jedem Typ was dran.* Jetzt fragen mich auch Studenten, wieviel Anteil die Rechenaufgaben ausmachen. Wenn es da zwischen 130 und 140 Punkte gibt, dann machen Rechenaufgaben vielleicht so 30, 40, 50 Punkte aus. Beim Rest lege ich mich jetzt nicht fest. Aber so zur Orientierung.

Zitat 25 Fall MW 5

Doktorand 1 [zur Frage nach der Auswahl von Prüfungsinhalten]: *Ich hab das Anwendungsbeispiel verändert, also Anwendungsbeispiel und Zahlen habe ich modifiziert für das eine und, diese eine Aufgabenstellung von mir ist sowieso, wo sie selber was entwickeln sollen, ich weiß jetzt nicht, ob das in der neuen Version auch noch so ist, aber das ist Aufgabe [sucht nach der Aufgabe], die Anwendungsaufgabe, Aufgabe 5, und da dann die mit den Vorschlägen. Im Prinzip ist es, man muss nicht sehr viel dran ändern. Da steht dann im Text die paar Angaben drin und dann muss jeder Student da selber Lösungsvorschläge entwickeln und entsprechend argumentieren. Von dem her ist die von der Erstellung her einfach und auch schnell gemacht und schwer zu kopieren, weil das jedes Mal ein anderes Beispiel ist mit anderen Zahlen, mit anderen Sachverhalten wo man dann schon jedes Mal überlegen muss. Allerdings ist es zum Korrigieren nicht so angenehm, es sind halt zehn Felder, die alle offen sind wo der Student eintragen kann, was er will. Das heißt, man muss das alles Lesen und auf Richtigkeit prüfen, wo es auch wiederum schwierig wird, eine hundertprozentige Konsistenz zu gewährleisten.*

Zitat 26 Fall MW 5

Akademischer Oberrat [zur Frage welche Anforderungen mit der Prüfung geprüft werden]: Kann der Student *Formeln, Rechenwege anwenden. Rechnet er richtig.* Dazu auch *graphische Auswertung*, das heißt, wie haben sich bestimmte Anteile verändert, wenn ich zum Beispiel eine Variable verändere. [...] Das muss er [der Student] auch *graphisch darstellen* und das dürfen nicht nur solche Geraden sein, sondern das wird oben eine Kurve. Dass er das auch verstanden

hat. Das ist bei Rechenaufgaben. Da gibt es unterschiedliche Formeln, die sie anwenden können müssen.

Doktorand 1: Die müssen *wissen und anwenden* können.

Akademischer Oberrat: Ja, ne, das Gesetz xy, das gebe ich ihnen in der Prüfung an.

Doktorand 1: Ja, es gibt ja auch andere Aufgaben, *wo du die Formel wissen* musst und die *dann natürlich auch anwenden musst*.

Zitat 27 für die Fälle MW 1 bis MW 5

Akademischer Oberrat MW 1 [zur Frage, ob Lernergebnisse bei der Prüfungserstellung verwendet werden]: In den Modulbeschreibungen müssen wir was reinschreiben, wobei bei dieser Vorlesung ist das relativ kurz geworden, bei manchen anderen haben wir es etwas ausführlicher gemacht. Also, das heißt, sogar hier hat man noch Luft nach oben. [...] Für die Modulbeschreibung hat jeder im Maschinenwesen die Modulbeschreibung gemacht und dementsprechend Lernergebnisse formuliert. Ich weiß gar nicht, habt ihr beide [bezieht sich auf seine Kollegen] konkret noch gearbeitet.

Doktorand 1 MW 1: Das was in TUMOnline steht...

Akademischer Oberrat MW 1: Theoretisch sollte, war das die Vorlage zu dem was in TUMOnline kommt, wobei gerade diese Lernergebnisse im Allgemeinen nicht in TUM Online drin sind. Wir selber haben es auf dem Server liegen [das Modulhandbuch]. Es kann aber sein, dass das noch von Euren [bezieht sich auf die Doktoranden] Vorgängern erstellt wurde.

Doktorand 1 MW 1: Ich hatte damit noch nie was zu tun [mit der Modulbeschreibung und den Lernergebnissen].

Akademischer Oberrat MW 1: Und da gabs eben eine Handreichung, mit zwei, drei DIN A4 Blättern und nach denen hat man dann etwas zusammengehext, nach bestem Wissen und Gewissen, aber deswegen muss das noch lange nicht perfekt sein.

Doktorand 1 MW 2: *Für uns war es nie besonders wichtig, das toll auszuformulieren, weil das einfach nur irgendwo in TUMOnline hinterlegt ist und für uns keine direkte Auswirkung hat.* Darum ist das bei uns bestimmt auch nicht mit der höchsten Qualität erstellt worden.

Doktorandin MW 3 [zur Frage nach einer Verwendung von Lehrzielen]: Ja, also bei mir, bei meiner Modulbeschreibung, die ist halt von meinem Vorgänger geschrieben worden und das ist

eigentlich nur, das sind die Kapitelüberschriften, die da drinnen stehen. *Also und ich weiß nicht, ob das so die Lernziele sind.*

Doktorand MW 4 [zur Frage, ob er die Lehrziele der Veranstaltung kennt und verwendet]: Peinlicherweise nein. [...] Also bei mir steht da gar nichts.

Moderator zu akademischen Oberrat MW 5: Sie haben Lernergebnisse formuliert in der Modulbeschreibung. Das haben wahrscheinlich Sie selber gemacht, Herr xy, oder? Und Sie haben auch die Prüfung beschrieben. Was in der Prüfung drankommt. Inwiefern haben Sie denn die Lernergebnisse und das, was in der Prüfung drankommt, abgestimmt?

Akademischer Oberrat MW 5: Mhm, also da habe ich mir jetzt aktiv nicht Gedanken drüber gemacht. [...]Die [Lernergebnisse] in der Vorlesungsbeschreibung drin steht. Ne, die haben wir nicht abgeglichen [mit der Prüfung].

Zitat 28 Fall MW 1

Akademischer Oberrat [zur Frage nach dem Vorgehen bei der Erstellung der Lehrziele]: Wir hatten im Endeffekt vier Lernergebnisse. Das fünfte kam dann am Schluss dazu. Wir haben gemerkt irgendwas passt nicht beziehungsweise fehlt noch und: ja stimmt, "ausgewählte Fahrzeugkomponenten und -systeme beschreiben", das ist eigentlich so ein Wald und Wiesen Lernergebnis, deswegen haben wir es vorher nicht benannt. Aber eigentlich müssen wir es fast nennen. Damit haben wir jetzt diese fünf. Wobei man dann schon *sieht, die ersten sind mehrere, deswegen sind es hier jetzt auch mehr, weil praktisch der Bestandteil des Hauptprozesses erläutern, die zugehörigen Prozesse darstellen und Ziele samt Ergebnisse erklären.* Das ist dann jetzt auch das, was in deinen Anmerkungen war [in einer E-Mail des Moderators]. Wir haben jetzt hier ein Lernergebnis so genannt, aber eigentlich sind es *drei eigene und hier haben wir es dann auch aufgeteilt, damit wir es hier auch detaillierter beschreiben können.* Das hat uns dann schon sehr gut geholfen bei der Arbeit hier und das, glaube, hilft uns auch bei der Prüfungserstellung, sodass wir genau sagen können, was wollen wir abfragen, da sehen wir da haben wir einen großen Fleck, den haben wir in der Prüfung nicht drin, wollen wir das und dann können wir gezielt unsere Prüfungsaufgaben stellen.

Zitat 29 Fall MW 1

Akademischer Oberrat [zum Vorgehen bei der Erstellung der Lehrziele]: Also das lief bei uns insofern ganz schön, da wir den Schwung [aus den Arbeitstreffen] gleich mitnehmen konnten und gesagt haben: jetzt setzen wir uns zusammen. Haben schon drei, vier Stunden gebraucht bis wir unsere Lernergebnisse halbwegs zusammen hatten. Viel Diskussion bis man das gleiche Ver-

verständnis hatten, was ist jetzt ein Lernergebnis und vor allem: wie sind die von unserer Vorlesung? Also da wirklich fachlich reingehen und sagen: ist es das jetzt wirklich was wir vermitteln wollen oder ist es was anderes? Das war natürlich der allergrößte Teil.

Zitat 30 Fall MW 2

Doktorand 1 [Überarbeitung der Lehrziele im Rahmen des ersten Arbeitstreffens]: Ich bin am Überlegen, ob wir den dritten Satz [eines bestehenden Lehrziels] aufspalten in mehrere Sätze und dafür den ersten und zweiten mehr oder weniger weglassen. Weil die ersten und zweiten, die dritte ist nur die Konkretisierung von dem ersten und dem zweiten.

Doktorand 2: Deswegen ich würde vorschlagen, dass man das *ein bisschen hierarchisch aufbaut*. Sozusagen den ersten Punkt, erstmal dass die dieses Verständnis besitzen. Dafür kann man den dritten Satz dann einfach als ersten stellen und sagen: nach der Lehrveranstaltung besitzen die Studierenden grundlegendes Verständnis in wichtigen Bereichen der Unternehmensführung. , nämlich: Organisationsgestaltung, so und so. Wenn es erstmal dieser Wissensaspekt abgefragt und dann müssten wir uns vielleicht eine zweite, dritte, vierte Frage ausdenken, wo, in welchem Bereich wir wollen, dass sie das dann anwenden können.

Doktorand 1: Also das dritte praktisch, dass sie aus den Bereichen das Wissen präsent haben und wiedergeben können. Ok, also setze ich hier als erstes und zwar, Wissen, auswendig wiedergeben, präsent haben.

Doktorand 2: Dass sie das wiedergeben, reproduzieren, wie sagt man dazu?

Moderator: Wiedergeben, genau. Könnt man da reinschreiben. Also ein typisches Wort wäre wiedergeben.

Doktorand 2: Okay, so. Wie sagt man denn zum Beispiel. Sie müssen es ja nicht nur wiedergeben können, sondern sozusagen, wenn man Vor- und Nachteile kennt und das evaluieren kann, denn diskutieren sollte man das auch können.

Doktorand 1: Genau, aber das wäre dann der nächste Satz. Der erste Satz wäre einmal, aus diesen Themenbereichen *sollen sie die Grundlagen präsent haben und wiedergeben können*. Das wäre der erste Satz. Und der zweite Satz wäre dann, *das Verständnis der Zusammenhänge und Vorgänge zwischen diesen Themenbereichen ist innerhalb der Themenbereiche...*

Doktorand 2: Innerhalb der Themenbereiche auch das Diskutieren Können. Und da halt die Zusammenhänge kennen und innerhalb diskutieren können.

Moderator: Kennen und dann auch benennen können, oder?

Doktorand 2: *Ja, benennen auf jeden Fall auch. Wir haben so ein, am Anfang jeder VL ist wie so ein Haus und das zeigt sozusagen, wir sind gerade im Kapitel drei und das- in diesem Haus des produzierenden Unternehmens sind wir sozusagen in der Wertschöpfung oder Testung oder so was und dann kann man den Zusammenhang erkennen und sagen, das sind irgendwie primäre Prozesse und solche Sachen.*

Moderator: Ja, klar, weil dann habt ihr ganz zentral die Zusammenhänge zu verstehen. Also es reicht nur, die einzelnen Bereiche zu kennen und nennen zu können und zu beschreiben, sondern auch die Zusammenhänge zwischen den...

Doktorand 1: *Genau, Kennen, Verstehen und Wiedergeben. Das wäre dann schon die zweite Stufe, auch Verstehen. Jetzt wollen wir, dass auch noch angewendet werden kann. Das wäre im Grunde,- der bisher erste Satz würde dann zum dritten. Methoden und Konzepte zu diesen...*

Doktorand 2: Aus diesen Teilbereichen sozusagen, aus diesen Themenbereichen. Ne, aus dem gesamten Komplex. *Also es geht ja auch um die Zusammenhänge. Methoden und Konzepte für Zusammenhänge.*

Zitat 31 Fall MW 3

Doktorandin [berichtet über ihr Vorgehen beim Erstellen der Lehrziele]: Also oben stehen bei mir die Lernergebnisse, die halt vorhanden waren. Es waren vier Lernergebnisse. *Die habe ich jetzt an Inhalt auch ergänzt. Deswegen habe ich jetzt hier sieben Lernergebnisse. Ich habe es so geordnet, dass es auch an die Vorlesung angepasst ist; also auch von der Reihenfolge her. Zunächst war mir wichtig, dass Grundlagen vermittelt werden. Und dass die Studierenden diese dann auch darstellen können und wiedergeben können. Das zweite Lernergebnis ist gleich geblieben. Das dritte Lernergebnis „Grundprinzipien der ...“ habe ich noch hinzugefügt. Beim nächsten Lernergebnis „Die Prozesse...“ war zunächst nur Verstehen. Wenn ich das jetzt aber analysiere und mit meiner Inhalt- und Verhaltensmatrix prüfe, sehe ich, dass ein Bereich noch vorhanden ist, indem Aufgaben gestellt werden und sie das auch anwenden müssen in Aufgaben. Dementsprechend habe ich es dann auch ergänzt.*

Moderator: Kommt das so auch in der Vorlesung dran?

Doktorandin: Normalerweise werden diese Aufgaben schon so durchgenommen. Man zeigt ihnen, wie man vorgehen muss. Es sind Transferaufgaben, bei denen sie gelerntes Wissen anwenden müssen. *Ein Lernergebnis habe ich noch hinzugefügt, da es in der Vorlesung auch um die Messung und um Evaluationsmethoden geht. Das sollte als Lernergebnis hinzugefügt werden, weil es noch nicht in den Lernergebnissen enthalten ist. Allerdings wurde in der Vorlesung noch*

nicht speziell darauf eingegangen. *Das sollte hinzugefügt werden und Aufgaben dazu erstellt werden.*

Zitat 32 Fall MW 4

Doktorand [berichtet über Vorgehen bei Erarbeitung der neuen Lehrziele]: [...] Jetzt habe ich Konzepte noch mit aufgenommen. *Weil hier in der Matrix kommt irgendwann Konzeptwissen vor.* Schwierig war es, einzelne Aufgaben dem Faktenwissen oder Prozesswissen zuzuordnen, weil es eigentlich Konzeptwissen war. *Der dritte Punkt war in der Ausgangssituation eigentlich der Vierte, aber wenn man die Matrix/ Pyramide heranzieht, ist es so logischer, denn „Erinnern, Verstehen, Anwenden, Analysieren“ sind die hierarchischen Punkte, wie sie in der Matrix vorkommen.* In der Ausgangslage war es umgedreht. Und da habe ich versucht, den Bezug zum zweiten Punkt herzustellen [liest Lehrziel vor]: „die gewonnenen xy Erkenntnisse verstehen“, die wir im dritten Punkt anwenden wollen. Und den ehemals dritten Punkt habe ich auf den vierten gesetzt, weil wir da in das Analysieren und Bewerten gehen. Die „Belastung von xy, da haben wir auch eine ganz bestimmte Vorlesung zu entsprechenden Themenbereichen. Der Begriff ist hier neu. Der war in der ursprünglichen Version „yz“. So finden die Studierenden aber genau diesen Begriff wieder in den Lernergebnissen und in den Folien der Vorlesung. Und wissen, sie müssen nur den xy analysieren und bewerten können. Und yz ist so geblieben. Nicht nur analysieren, wie in der Ursprungsversion, sondern auch bewerten, weil sie sollen auch bewerten. *Und deswegen tauchen diese jetzt fünf (und nicht mehr vier) Handlungskomponenten jetzt da mit auf.* Das geht eher in Richtung der Klausur und der Vorlesung und spiegelt auch wieder, was wir machen. Wenn wir einen Haken an das Projekt machen, ist das genau ein Produkt, das ich meinem Chef gebe: Ursprüngliche Fassung der Lernergebnisse und überarbeitete Version. Dann werden wir die Modulbeschreibung anpassen.

Zitat 33 Fall MW 5

Akademischer Oberrat [Bericht zum Vorgehen bei der Erstellung von Aufgaben]: Wir haben fünf Vorlesungskapitel. Dann haben wir verteilt, wer zu welchem Kapitel Aufgaben stellen soll. Das war verpflichtend. Ich hatte das erste Kapitel. Ich habe mir Aufgaben rausgesucht aus meiner Tabelle, die ich von früher noch habe, welche Aufgaben passen da rein, und habe die zusammengeschrieben und hab danach die Pyramide versucht auszufüllen beziehungsweise die Aufgaben zuzuordnen. Ich war dann nicht so ganz zufrieden, ein bisschen einseitig ist unter Erinnern und nur zwei Teilaufgaben bei Entwickeln und Bewerten steht. Ich habe meine Aufgaben dahingehend nochmal überprüft, ob ich es nicht irgendwie anders machen könnte und bin da aber auf keinen grünen Zweig gekommen.

Akademischer Oberrat [zur Frage, wie Kolleginnen und Kollegen über das veränderte Vorgehen informiert wurden]: Das war nach unserer letzten Sitzung [Arbeitstreffen zur Festlegung des Prüfungsgegenstands], da habe ich ziemlich schnell die E-Mail rumgeschrieben, damit die das Schema [die Lehrzielpyramide], die Einordnung anwenden. Und dann noch einen Link auf eine andere Datei, aber sagen mir, war das erstmal wichtig [die Einordnung in die Pyramide], dass man da erstmal was sieht.

Moderator: Inwieweit habt ihr auch die Lernergebnisse berücksichtigt, die formuliert sind, also neben der Pyramide? Spielte das auch eine Rolle?

Akademischer Oberrat: Die in der Vorlesungsbeschreibung drin stehen. Ne, die haben wir nicht abgeglichen.

Doktorand 1 [zur Frage, ob er ebenfalls die Pyramide zur Erstellung von Aufgaben verwendete]: Wir haben es nicht durchgängig gemacht, aber das ist im Prinzip auch der Zeit geschuldet gewesen, weil wir das noch vor Weihnachten beziehungsweise gleich nach Weihnachten fertig stellen wollten. Zum Thema wie ich mich dabei gefühlt hab: Im Prinzip habe ich mich daran orientiert, dass ich mehr Aufgaben aus unterschiedlichen Bereichen zu stellen, ähm, das ist aber natürlich begrenzt aufgrund des Pools an Aufgaben, den wir auch schon haben. Ich hab dann eben versucht, Aufgaben zu nehmen, die auf unterschiedlichen Ebenen [der Lehrzielpyramide] waren, sodass man nicht nur sehr einseitig das Repetieren von Studenten verlangt, sondern auch das selber Entwickeln. Ich glaube ich habe auch eine Aufgabe mit drin, wo sie auch selber was entwickeln müssen. Ähm, genau von dem her habe ich eher versucht, alles ein bisschen mit rein zu nehmen. Das hab ich zumindest für mich als Ziel gesetzt. Das hat dann auch ganz gut funktioniert. Wenn ich noch mehr Zeit hätte, würde ich noch viel schönere Aufgaben stellen können. Also, das ist ja das, was ich letztes Mal auch schon gesagt habe. Irgendwo ist halt ein begrenzender Faktor die Zeit, die wir aufwenden können

b. Zitate zur Erstellung von Prüfungsaufgaben

Zitat 34 Fall MW 1

Akademischer Oberrat: Wir haben Sachen, wo etwas analysiert wird. Wenn das aber genauso in der Vorlesung gezeigt haben, ist es eigentlich nicht mehr Analysieren, sondern nur noch Erinnern und Hinschreiben. Genauso ist es bei eurem Beispiel auch. Bewertet er wirklich, wenn der Student nur ein Schema durchrattert und eigentlich nicht denken muss. Eigentlich nicht, oder? Wenn wir das genauso in der Vorlesung gemacht haben, dann ist es halt eher wiedergeben. Es sei denn,

jemand war nicht in der Vorlesung, dann hat er tatsächlich eine Analyse gemacht, aber dann hat er halt Pech gehabt, wäre er halt in der Vorlesung gewesen. Für die anderen ist es nur Wiedergeben.

Zitat 35 Fall MW 4

Doktorand: Die Lehrziele standen, als wir anfangen. Die Taxonomie besteht und daran hatte sich offensichtlich schon jemand gehalten. Jetzt kommt noch die Arbeit mit der Matrix und der Klausur. Vielleicht der falsche Weg. Weil ich jetzt eigentlich die Lehrziele an die Klausur anpasse, eigentlich sollte die Klausur ja an die Lehrziele angepasst sein. Aber man hat ein Gefühl dafür, was machen wir in der Übung, in der Vorlesung und was prüfen wir ab; und ganz oben stehen die Lernergebnisse. Macht für mich also schon Sinn.

c. Zitate zur Rückbindung von Aufgaben an den Prüfungsgegenstand

Zitat 36 Fall MW 2

Doktorand 2: Nach der Erstellung der Prüfung geht sie an zwei junge Kollegen bei uns am Lehrstuhl. Die kriegen erst mal nicht die Musterlösung. Nur die Klausur und müssen diese komplett durchrechnen und durchbeantworten mithilfe des Skripts oder mithilfe der Übungsaufgabe als Stütze, aber eben ohne Musterlösung. Idealerweise haben sie die Prüfung schon als Studenten geschrieben und haben ein Gefühl dafür, ob man die Aufgaben hätte schaffen können. Sie müssen halt ungefähr abschätzen, ob die Prüfung von den Anforderungen so passt oder die Aufgaben ein bisschen heftig sind. Das hat bisher eigentlich immer sehr gut geklappt. Wenn sie das ganz fertig haben, kriegen sie die Musterlösung und können das auch noch mal kontrollieren. Danach geht es an unseren Abteilungsleiter, der die auch noch mal komplett durchschaut. Also vor allem dahingehend: stimmt die Gesamtsumme der Punkte? Ist formal alles perfekt? Bis zu einem Punkt, der irgendwo fehlt, wird alles durchgesehen. Ist die Einheit korrekt? Ist die Zusammenstellung der Aufgaben angemessen vom Schwierigkeitsgrad? Und danach geht sie wieder an uns zurück. Wir arbeiten ein, was einzuarbeiten ist. Dann geht es zum Professor, der schaut auch nochmal drüber. Da kann ich nicht genau sagen, was er genau damit macht. Bisher habe ich es noch nicht erlebt, dass er danach gesagt hat: das müsste man noch ändern. Weil wir es eigentlich so machen, dass er praktisch nur noch drüber schauen muss.

Zitat 37 Fall MW 5

Doktorand 1: Als zusätzliche Qualitätssicherungsmaßnahme haben wir im Vorfeld nochmal eine Proberechnung mit dem gesamten Korrekturteam angesetzt. Das heißt, da sitzen dann vier bis sechs wissenschaftliche Mitarbeiter da, die auf Zeit die Aufgaben rechnen und im Anschluss da-

ran, die sowohl die Bepunktung, als auch die Fragestellungen, als auch, wie es von der zeitlichen Einteilung her ist, bewerten. Also das machen wir jetzt zum zweiten Mal und das hat eigentlich sehr gut funktioniert. Wir versuchen, mehr Aufwand in die Erstellung zu investieren, um dann bei der Korrektur uns das mit gutem Gewissen sparen zu können.

Zitat 38 Fall MW 5

Akademischer Oberrat: Ich hatte das erste Kapitel. Ich habe mir Aufgaben rausgesucht aus meiner Tabelle [einem bestehenden Aufgabenpool], die ich von früher noch habe, welche Aufgaben passen da rein, und habe die zusammengeschrieben und hab danach die Pyramide versucht auszufüllen, beziehungsweise die Aufgaben zuzuordnen. *Ich war dann nicht so ganz zufrieden, ein bisschen einseitig ist unter Erinnern und nur zwei Teilaufgaben bei Entwickeln und Bewerten steht. Ich habe meine Aufgaben dahingehend nochmal überprüft, ob ich es nicht irgendwie anders machen könnte und bin da aber auf keinen grünen Zweig gekommen.*

d. Zitate zur Bewertung von Prüfungsantworten

Zitat 39 Fall MW 1

Akademischer Oberrat: Wir machen es grundsätzlich so: bei der Korrektur hilft der gesamte Lehrstuhl mit. Das heißt wirklich jeder. Ich erstelle dann immer eine Liste, in der sich jeder eintragen kann, was er bewerten möchte. Und wenn er es schlaun macht, trägt er sich genau bei den Aufgaben ein, die er selber erstellt hat. Das funktioniert dann schon ganz gut. Es gibt natürlich haufenweise Leute, die selber keine Prüfung erstellt haben, aber dennoch mitkorrigieren. Dafür gibt es die Musterlösung und man weiß, wer die erstellt hat und kann sich mit demjenigen absprechen. Und dann muss man eben eine Aufgabe korrigieren.

Zitat 40 Fall MW 1

Akademischer Oberrat: [...] Wie argumentiere ich aber dann dem lieben Studenten gegenüber, der natürlich jetzt in dem Moment in der Prüfungseinsicht nicht glücklich ist mit seiner Situation und sagt, *oh, das steht doch da*. Also, wenn man da einfach ein gutes Argument hat, oder so zu sagen, einfach eine klare Formulierung, die nicht nur lautet: jetzt pass mal auf, 90 Prozent deiner Kommilitonen haben es verstanden, bloß du bist sozusagen zu doof dazu. Dass bei ihm natürlich ankommt, auch wenn ich sage, 90 Prozent haben es verstanden, dann fühlt er sich natürlich zurückgesetzt. Irgendwie brauche ich da natürlich eine gute Argumentation, wo ich sage: okay, pass auf, eigentlich war das schon klar für jeden, der in der Vorlesung ist.

Zitat 41 Fall MW 2

Doktorand 3: Wir müssen immer darauf achten, dass die Korrigierbarkeit gewährleistet ist. Wir sitzen so schon mit 6-7 Leuten zwei Wochen da dran.

Zitat 42 Fall MW 2

Doktorand 2: Also wir [Doktorand 2 und 3] hatten die Musterlösung erstellt für alle Aufgaben. Wir haben dann gesagt: bei strittigen Fällen... *also wir haben es vorher mit den Kollegen durchgesprochen einmal, und dann aber bei strittigen Fällen ist es denn Leuten auch selber überlassen.* Weil man kann von vornherein nicht alle strittigen Fälle absehen. Wir hatten auch zwei, drei... wir hätten noch besser die Aufgabe einengen können, glaube ich. *Jeder korrigiert eine Aufgabe.* Ich hatte zum Beispiel Aufgabe 3 und 4 und dann für alle korrigiert. Man nimmt sich da halt einen Stapel mit ins Büro. *Aber dadurch, dass wir Aufgabe für Aufgabe korrigieren, war das dann jedem Korrektor irgendwie selber überlassen auf die Konsistenz zu achten, da hatten wir uns auch so drauf geeinigt, dass wer die Aufgabe korrigiert, der muss eben sich eine Richtlinie überlegen, die stringent ist.*

Zitat 43 Fall MW 2

Doktorand 2: [...] Wir hatten aber eine andere Prüfung, da war das auch der Fall. Es war eine offene Frage, wo die Studenten einfach irgendwohin argumentiert haben. Also eine Frage war zum Beispiel: Warum hat ein Schneckengetriebe einen schlechteren Wirkungsgrad als andere Getriebe? Und da haben meine Studenten halt geschrieben, weil es weniger Verluste gibt und das ist halt eine Antwort, die mir auch ein Grundschüler geben kann. Darauf gab es einfach keinen Punkt, es ist zwar richtig, aber es hat keinen technischen Bezug zu der Vorlesung und ist eine triviale Antwort. *Und an mancher Stelle kommen wir dann schon in Diskussion nach der Einsicht. Ja, ich habe ja dieses Wort hingeschrieben, das ist doch auch irgendwie richtig, das kann man doch so und so auch interpretieren und das ist manchmal auch richtig erklärt.* Und da wir das auf dem Deckblatt haben, waren wir dann bei solchen Sachen relativ streng. Das wird vorgelesen und unterschrieben. Das war jetzt ein Probeversuch, den wir bei zwei Vorlesungen gemacht haben. Das werden wir jetzt auch einführen für alle Prüfungen, weil da einfach ganz klar geregelt ist, wie korrigieren wir, was ist ein Unterschleif, was ist erlaubt und so weiter.

Zitat 44 Fall MW 3

Doktorandin: Also ich muss gestehen, dass ich nicht selber mitkorrigiert hab, sondern, dass wir das halt aufgeteilt haben und ich hab halt danach die Punkte, also, eingegeben und hab halt dann noch einen Überblick bekommen, wie die Aufgaben so bearbeitet wurden. *Aber im Endeffekt hab*

ich meinen Korrektoren halt die Musterlösung gegeben und wenn sie eine Frage hatten, haben sie die halt an mich gestellt. Also bei uns werden immer Teile eingeteilt, dass komplett die ganzen Teile von einer Person durchkorrigiert werden. [...] Wir haben auch eine Sitzung einberufen mit allen Korrektoren und sind mit denen die Klausur und die Musterlösung durchgegangen.

Zitat 45 Fall MW 3

Doktorandin: Im Endeffekt wollten wir nur vermeiden, dass wir so Endlosdiskussionen in der Einsicht haben. *Also manche sitzen, also bei manchen Studenten, die sitzen zwei Stunden drin dann und diskutieren um jeden halben Punkt und das wollten wir halt vermeiden.* Es gab halt dieses Mal auch keine Notenänderungen direkt an dem Einsichtstermin, sondern durch den Zettel haben wir uns das halt nachher nochmal angeschaut. Also es ist zwar dann nochmal Zeitaufwand...

Zitat 46 Fall MW 4

Doktorand: Wir haben die Klausur aufgeteilt. Ich *hatte sechs oder sieben Korrektoren*, die mich da unterstützt haben, weil es über 170 Klausuren á 22 Seiten waren. *Und teilweise kamen die Leute und haben nachgefragt, also nach dem ersten Packen, wir haben so 30er Packen gemacht und die dann aufgeteilt, und nach dem ersten 30er Packen kamen dann so ein zwei Leute, die und die Fragen zu Aufgaben und Musterlösung, kann ich das noch werten, kann ich das nicht werten.* Wir haben festgestellt, dass es auch seitens der Korrektoren notwendig ist, sich in das Thema irgendwie ein zu denken, sprich, die Vorlesung zur Hand zu nehmen, die Übung zur Hand zu nehmen und nicht einfach nur Musterlösung und Klausur und ich hak das ab, was drin steht und wenn es nicht drin steht, trag ich es nicht ein oder es ist falsch. [...] *Es gibt auch ein Meeting, da werden 5-6 Kollegen gebrieft. Das ist die Musterlösung, die gehen wir dann zusammen durch und dann werden die Aufgaben verteilt. Du machst 1-8, du 9-15 wird aufgeteilt und dann eine Woche, zwei Wochen.*

Zitat 47 Fall MW 4

Doktorand: Also, wir haben keine Lust, 4 Wochen lang Klausuren zu kontrollieren. Und dann danach noch die Einsicht. *Letztens waren 3 Mann da, die haben bis auf's Messer diskutiert und zwar zwei Stunden lang. Einer hat sich verbessert, die anderen beiden nicht.*

Zitat 48 Fall MW 5

Akademischer Oberrat: Bei dieser Prüfung ist der Vorteil, *dass jeder Korrektor seine Aufgabe hat und die wird konsequent erstkorrigiert.* Das ist für die Konsistenz, besser geht es eigentlich nicht.

Zitat 49 Fall MW 5

Doktorand 1: Es wurden nicht nur die Musterlösungen aus dem Vorlesungsskript berücksichtigt, sondern eben auch, *was die anderen Kollegen beim Probearbeiten der Prüfungen da mit rein-geschrieben* haben und was wir in der Diskussion auch noch als richtig erachtet haben. Das heißt, wir haben nicht nur die 5 richtigen Antworten aus dem Skript, sondern auch noch 3-4 andere Antworten von jemandem, der den Kontext von dem Ganzen verstanden hat, der aber jetzt offensichtlich nicht wörtlich das Skript wiedergeben kann. Haben wir eben auch noch in Musterlösung mit integriert. Und dann muss man auch sagen, dass die harte Bepunktung auch sehr viel sauberer ist. *Also worauf gebe ich Punkte, das ist durchgängiger durch die Prüfung, und damit auch für die Studenten besser nachzuvollziehen.*

Zitat 50 Fall MW 5

Doktorand 3: *Das wurde wahrscheinlich auch schon angesprochen, dass wenn man in einer Korrektur merkt: okay, da wurden schon mehrmals die und die Antworten [von den Studierenden] gegeben, die in unserer Musterlösung zwar nicht auftauchen, aber dennoch richtig sind, oder richtig gemeint* sein könnten. Also da ist dann schon noch so gut-Wille dabei, dass man das dann mitnotiert und dann dem Zweitkorrekteur mit an die Hand gibt und dann sagt: du, in die bisher angelegte Musterlösung, haben wir noch die und die Begriffe, wenn es so eine Abfrage ist, weil die auch einfach reinpassen.

Zitat 51 Fall MW 5

Doktorand 3: *Also ich kann mir nicht vorstellen, dass man vorher so ein Raster erstellen kann, anhand dem man dann das tatsächlich jede Antwort einordnen kann. Also, das kann man erst machen, wenn man alle Antworten einmal gesehen hat. Weil da, manchmal kommen da total abstruse Antworten, wo es dann aber auch schwierig ist, zu sagen: ist das jetzt teilweise richtig, oder richtig?*

Akademischer Oberrat: Wenn es jetzt bei den konstruktiven Vorschlägen in oberen Teil, da kann man sich ja so viel überlegen, *also da kann man wahrscheinlich 2 Seiten vollschreiben, um zu sagen, das kann man bewerten oder das, sagen wir mal, liegt im Ermessen des Korrektors, dass man sagt, ja das ist, führt wahrscheinlich zum Kosten senken.*

Moderator: Habt ihr da mal, also gerade bei solchen Fragen, wo die Antworten so unterschiedlich sein können, mal die Übereinstimmung gezielt bestimmt und mal geschaut, wie groß die Abweichung ist? Zwischen Erst- und Zweitkorrektor? Also wenn kein Regelwerk niedergeschrieben ist.

Akademischer Oberrat: *Also da ist selten Diskussion nötig. Nicht, dass man nicht selten drüber redet, sondern da ist selten Diskussion nötig.*

e. Zitate zur Punkte- und Notenvergabe

Zitat 52 Fall MW 1

Akademischer Oberrat [zur Frage, ob man mit der Vergabe von Punkten nach Zeit auch Aspekte wie Relevanz oder Schwierigkeit berücksichtigt]: Das ist bei uns identisch. Einerseits, das was uns besonders wichtig ist, davon machen wir mehr Aufgaben wie von anderen oder größere und an sich ist es so angesetzt, *dass Punkte pro Zeit konstant sind und wenn wir dann praktisch einen wichtigen Punkt haben, dann muss der auch mehr Zeit füllen.* Natürlich kann ich sagen: wenn er es einfach weiß dann hat er es schnell beantwortet. Wir passen aber die Prüfung, die Aufgaben, so an, klar gibt es Aufgaben, wenn ich die einfach beantworten kann, dann bin ich schnell und davon gehen wir aus. Und wenn der sich das erst herleiten muss, dann kriegt er trotzdem bloß wenige Punkte. *Aber die schwierigen Aufgaben, die sind dann auch so gestellt, dass ich mehr Zeit brauche, das heißt, ich muss ein Diagramm zeichnen, ich muss begründen und und und. Dadurch kommen wir auch wieder auf die gleichen Zeitpunkte. Es läuft immer wieder darauf hinaus. Natürlich gibt es die Fragen, wo es heißt: nennen sie drei von fünf Punkten, zählen sie auf, und so weiter, und dann gibt es auch: erklären sie anhand von einem Diagramm - und da braucht man einfach, um ein schönes Diagramm, da braucht man Zeit und die muss man ihm geben und entsprechend bepunkten und natürlich kommt man immer wieder zu dem Problem, dass man sagt: eigentlich ist das ein Punkt alles, wo man ein bisschen hin und her schieben muss.*

Zitat 53 Fall MW 1

Akademischer Oberrat: Wir haben sozusagen *die Mindestpunktzahl für eine 4.0 und die Mindestpunktzahl für eine 1.0. und damit gebe ich den Überhang vor. Und wieviel ich meine Grundanforderung, dazwischen wird es linear reingehauen.* Und dann schaut man eben. Auch meine Erfahrung ist genau die, wenn man merkt, es ist eine schöne Prüfung gewesen, dann kommt *ziemlich die Gaußkurve* raus. Das passt dann einfach recht schön. *Wenn man merkt, die kommt eben nicht raus, dann merkt man auch, da hat irgendwas nicht gepasst.* Dann waren irgendwelche Aufgaben nicht sauber, und und und. [...] Das sind dann nochmal die Schiebereien. Die sind bei uns auch nicht ganz unbekannt. Das umgeht man halt, wenn du jetzt praktisch, wenn der halbe Punkt schon drei Punkte sind, *dann tust du dich schon mal deutlich leichter beziehungsweise du hast deutlich weniger Leute, die halt wirklich nur einen Punkt von der nächsten Notenstufe weg sind.* Und deswegen habe ich gesagt, für jedes gibt es notfalls mal drei Punkte, damit man in

Summe einfach viele Punkte haben zum vergeben und nicht alles so auslegen, dass es eh nur schon fünfzig Aufgaben mit einem halben Punkt gibt.

Doktorand 3: Im Prinzip ist unser Schnitt immer etwas um die 2.5. Und dann betreiben wir sozusagen das Justieren auch in Grenzen. Da gibt es schon mal ein Jahr, in dem es mehr auf die 3 zugeht. Das ist ja schon eine Varianz über die Menge an Studenten. Woran wir dann eben spielen ist der Überhang und die Notenschritte, womit die Verteilung dementsprechend angepasst werden kann. [...] Wir machen *das so, also wir können mal rübergehen in den Reiter Auswertung. Das ist unsere Auswertung quasi, als Notenverteilung, und wir können oben den Prozentabstand manipulieren beziehungsweise den Punktabstand.* Und die erreichbaren Punkte abzüglich Überhang und wenn wir das dann eintippen, also wenn man dann eine schöne Verteilung gefunden hat, *dann sieht man Anzahl Prüfungen mit weniger als einen Punkt sind acht Stück und die lässt man sich dann ausgeben und die kriegen einen Extrapunkt.*

Zitat 54 Fall MW 2

Moderator: Wie bepunktet ihr denn bisher? Wie legt ihr fest, ob eine Aufgabe sechs oder sieben Punkte bekommt?

Doktorand 1: Nach dem zeitlichen Aufwand. *Wir haben neunzig Minuten, das sind neunzig Punkte und wir schätzen einfach für die Aufgabe braucht man so und so viel Minuten und so viele Punkte gibt es.*

Doktorand 3: Wobei wir da so ein bisschen top-down und bottom-up gehen. Einmal wird es abgeschätzt und da sagen wir: diese Aufgabe soll am Ende zwölf Minuten dauern, dann soll es auch zwölf Punkte geben. Auf der anderen Seite, bei der Bepunktung, gehen wir so vor, dass wir immer gucken, *die kleinste Einheit ist ein halber Punkt bei uns und das muss am Ende matchen, dass wir so viele bepunktbare Einheiten haben, die sich dann irgendwann zu zwölf Punkten aufaddieren und wenn das funktioniert, dann ist es vernünftig.*

Zitat 55 Fall MW 2

Doktorand 2: Wir wollen eigentlich, dass sie die Lernziele erreichen. Das Lernziel ist auch Bewerten und Entwickeln und wenn sie diese hohe Stufe erreichen, kriegen sie dafür mehr Punkte als wenn sie nur auf der unteren Stufe bleiben. *Das geht über die Zeit, denn zum Entwickeln braucht man einfach mehr Zeit als zum Erinnern.*

Doktorand 1: *Aber die Aufgabe selber ist natürlich, wenn ein langer Text da, ist auch komplizierter. Das heißt, für die komplizierte Aufgabe gibt es mehr Punkte, wenn die Aufgabe länger ist. Also das passt ja auch zusammen.*

Zitat 56 Fall MW 2

Doktorand 1: Aber die kleinste Einheit, die wir vergeben, ist halbe Punkte. *Wenn wir von einem Studenten verlangen, in einer Aufgabe, er soll vier von zehn Punkten aufzählen, dann gibt es dafür vier halbe Punkte egal, wie lange wir meinen, dass er dafür braucht.* Er braucht tendenziell kürzer, um schnell mal vier Stichwörter hinzuschreiben, als zwei Minuten, wenn er es weiß.

Zitat 57 Fall MW 2

Doktorand 3: Ich find da ein wichtiger Fürsprecher für dieses Zeitbepunktungssystem ist, dass der *Student schon auch irgendwie einen Anhaltspunkt kriegen muss, wie viel Aufwand da jetzt wirklich nötig ist. Und das ist Zeitaufwand ja immer.*

Zitat 58 Fall MW 2

Doktorand 2: Ansonsten, bewertet an sich haben wir, da sind uns auch ein zwei Inkonsistenzen aufgefallen, aber auch erst im Nachgang, was die Zeit pro Punkt angeht. Bei manchen Aufgaben haben wir festgestellt: ok, die kann man schnell schön runterrechnen und da haben wir dann verhältnismäßig viel Punkte gegeben. *Es war ja angegeben, wie viel Punkte es gibt, das heißt, da konnte ich mit zwei Schritten mal drei oder vier Punkte sammeln, im Vergleich zu anderen Aufgaben, wo wir festgestellt haben: ok da hätten wir uns besser abstimmen können, dann hätten wir auf andere mehr geben können, wo man eben vier, fünf Schritte braucht, um einen oder zwei Punkte zu erreichen.*

Zitat 59 Fall MW 2

Doktorand 2: Also systematische Fehler gab es. Ich würde vielleicht mal auf zweierlei eingehen, die ich festgestellt hab, also einmal zum Beispiel, dass im Nachgang festgestellt hat, manche Punkte waren eben, was ich auch schon mal angesprochen hab, sehr leicht zu ergattern und manche waren sehr schwierig. Also das war zum Beispiel die eine Rechnung. die so durchgeführt werden musste. und auf die es eben insgesamt 14 Punkte gab. Also wenn wir das hier ganz kurz durchgehen, also man musste hier erstmal folgende Schritte berechnen [...]. Und dafür gab's 14 Punkte. Und im Vergleich dazu hatten wir genau auf die Aufgabe 5. Und das war halt blöd, da haben wir hier 3 Punkte auf jede richtige Rechnung, dann gibt es hier 15 Punkte drauf und dann sieht man schon, also das ist immer hier richtig eintippen und hinschreiben.

Moderator: Also von der Anforderung her niedriger einzustufen?

Doktorand 2: Also gut, du musst schon wissen, von der Anforderung her ist das halt Wissen und dann anwenden können, aber es ist auf jeden Fall wesentlich billiger und vor allem zeitlich we-

sentlich aufwandsärmer zu lösen. Also da hätten wir entweder hier wesentlich weniger Punkte oder auf der anderen Seite, hätten wir da 1,5-fache davon geben können.

Zitat 60 Fall MW 2

Doktorand 2: *Also was der beste Student erreicht ist die Maximalpunktzahl, alles drüber hinaus ist Überhang. Da hatten wir eben 20 Punkte Überhang am Schluss.* Und klar: wenn der eine des bearbeitet, der andere des, dann hat man unterm Strich etwas mehr Korrekturaufwand.

Zitat 61 Fall MW 2

Doktorand 3: [...] Die Studierenden werden ja nicht dümmer. Deswegen schauen wir, dass wir immer den gleichen Schnitt haben. Und die Klausur ist mal schwerer und mal leichter. [...] Bei so einer großen Anzahl an Studenten, vor allem da es ja eine Grundlagenvorlesung ist, die viele Studenten schreiben müssen. Es sind nicht nur die, die besonders motiviert sind sondern es ist wirklich ein recht repräsentativer Querschnitt durch die Studenten. Deswegen kann man so eine Annahme auch treffen. Es sind wirklich 300-400 Studenten. Die werden nicht von einem Jahr auf das nächste deutlich blöder oder intelligenter.

Zitat 62 Fall MW 2

Doktorand 1: Wir schauen, dass wir dem Studenten dann den halben Punkt noch irgendwo geben, damit ein Mindestabstand von einem ganzen Punkt zur besseren Notenstufe ist.

Doktorand 3: Also wir haben da noch einen weiteren Aspekt, wir haben also ein ähnliches Diagramm, eine ähnliche Excel-Auswertung, allerdings haben praktisch diese Punkte, ab wieviel Punkten gibt es praktisch die bessere Note, die können wir von Hand auch noch bearbeiten. Das heißt praktisch, wenn ich auch sehe, was weiß ich, jetzt 105,0 Punkte ist jetzt die Grenze, aber mir würde jetzt 105,5 ein besseres Bild geben, dann geh ich rein und ändere das. Dann ist der Notenschlüssel sozusagen, nicht komplett linear, sondern, also bei einer großen Prüfung geht das nicht, weil da praktisch eigentlich jede halbe Punktzahl vertreten ist, aber habe ich nur 50 Kandidaten, dann sind immer wieder Lücken zwischendrin und dann kann ich schon mal sagen, dann setz ich die Grenze um einen halben Punkt nach unten. Dann gibts keinen mehr, der in der nächsten da ist. Die haben automatisch die bessere Note. Ich muss keinem die Punkte schenken auf die Aufgabe, sondern ich setz halt einfach den Notenschlüssel einen halben Punkt daneben.

Zitat 63 Fall MW 3

Doktorandin: Also bei uns ist es so, das hat der Kollege zu mir gesagt, dass ich 90 Punkte auf 90 Minuten auslegen soll, plus 10 Prozent Überhang.

Zitat 64 Fall MW 4

Doktorand: Ich müsste die Punkte in der Prüfung noch anpassen, um die 90 Punkte zu erreichen und 10 % Überhang. Da müssten noch ein paar Aufgaben rein. Bei mehr Aufgaben, weniger Zeit pro Aufgabe. Ob ich es einfach so lass.

Zitat 65 Fall MW 5

Akademischer Oberrat [auf die Frage nach welchen Kriterien Punkte auf Aufgaben vergeben werden]: Nein, nicht Wichtigkeit, sondern das ist der Zeitaufwand. *Pro Minute gibt es 2 Punkte.* 60 Minuten dauert die Prüfung und dann gibt es 120 plus X Punkte. *Wenn wir die Prüfung proberechnen, dann schätzen wir ab, wie lange wird man da brauchen.* Wenn es eine Rechenaufgabe ist, dann muss der Student ja auch schon viel lesen, damit er die Aufgabe versteht. *Das muss man auch irgendwie in der Punktezahl berücksichtigen,* wenn man sagt, sie haben als Pi-mal-Daumen-Wert zwei Punkte pro Minute. Ich hab es dann so gemacht, wenn es jetzt für den Rechenweg, das hinzuschreiben, das hat er [der Studierende] eigentlich in zwei Minuten erledigt, des wären vier Punkte, das passt dann nicht, wenn ich so einen langen Angabentext habe. *In dem Fall erfordert es, dass ich erstmal überlege und [den Angabentext] verstehe und dann ist das Rechnen ganz einfach, dann kriegt er auch mehr Punkte, damit dieses Durchlesen und Verstehen, es braucht ja auch Zeit, in der Punktezahl berücksichtigt wird.*

Zitat 66 Fall MW 5

Akademischer Oberrat: Es ist eine Mischung: erst gebe ich eine Standard 4,0 Grenze, anschau ich wie viele fallen durch. Wenn es dann 20 % sind würde ich sagen, oh da war die Prüfung wohl schwerer und dann täte ich mich auch darauf einlassen die 4,0 Grenze runterzusetzen, dass da noch mehr durchkommen. In punkto Abstand, das ist so ein Thema, wie viel 1,0er gibt es, oder auch Gauß: ist das gleichmäßig. Wenn es zu viele 1,0er gibt, dann war die zu leicht, dann schieb ich da noch ein bisschen. Aber ich mache es nicht so extrem, dass ich sage, der Notendurchschnitt muss immer gleich sein. Das nicht. Da habe ich schon Flexibilität und sage halt: dann war die halt diesmal ein bisschen leichter, dann ist der Durchschnitt besser.

Zitat 67 Fall MW 5

Akademischer Oberrat: Also es gibt, sagen wir mal, zwischen 130 und 140 Punkte hat die Prüfung immer und dann ab 4,0 gibt es ab, sagen wir mal, das steht auch in der Exeldatei drinnen. Ich weiß nicht, zwischen, b jetzt die Spanne so groß ist, zwischen 60 und 65 Punkten, gibt es eine 4,0, so groß ist die Spanne glaube ich nicht. Zwischen 62 und 65 Punkten gibt es in der Regel eine 4,0. Und der Sprung zur nächsten. Der 0,3er Sprung ist dann zwischen, sagen wir mal, zwischen 5 und

7 Punkte ist der Sprung dann zur nächsten. Das entscheidet sich dadurch, ob es eine Gaußverteilung ist, ob man vielleicht Grenzfälle hat, die so gerade auf der Kippe stehen zur nächsten Note und dann sagt man, ja, dann mache ich nicht 62 Punkte für eine 4,0, sondern mache 61 Punkte 4,0 und dann rutschen viele in die nächste bessere Note rein. Dann nehme ich die Diskussion bei der Noteneinsicht heraus. Oder wenn ich sage, es ist jetzt festgelegt, aber jetzt habe ich trotzdem noch 1 oder 2, die einen halben Punkt vor der nächstbesseren Note sind, dann hebe ich es manuell auf die nächste hoch. Also die, die dann gerade so nicht über die Schwelle kommen. Also wenn ich das schon einmal optimiert habe mit Gauß und dass nicht so viele genau knapp unter der Grenze sind, sondern dass ich die schon rübergehoben habe in die nächstbessere Note. Automatisiert nicht manuell, sondern dann für alle. Eben dass ich sage, nicht ab 62 Punkte gibt es 4,0, sondern ab 61 und damit rutschen viele in die bessere Note rein. Aber dann kommen halt andere wieder knapp an die Notengrenze ran. Ja und wenn das dann mal zwei sind, die hebe ich dann und sage: ja wegen einem Punkt haben sie es jetzt nicht geschafft, oder wegen einem halben Punkt haben sie jetzt die bessere Note nicht geschafft, dann habe ich die auch schon mal manuell hochkorrigiert.

f. Zitate zu Rückschlüssen aus der Empirie

Zitat 68 Fall MW 1

Doktorand 3: Wir haben ja quasi eine Exceltabelle, in der jede Teilaufgabe drinnen steht. Also Aufgabe 1 von a) bis irgendwas. *Da wird von jedem Studenten die Punktzahl eingetragen.* Das sind dann so 16.000 Felder pro Prüfung. Jeder der die Prüfung korrigiert, muss die Summe draufschreiben. Wir machen praktisch eine Prüfsumme. Wenn wir das in die Exceltabelle eingetragen haben, dafür brauchen wir 2 Tage für das Eintragen, und wenn das mit der Klausur übereinstimmt, dann passt es. *Auf Grund dieser Daten, die wir haben, können wir die Notenverteilung genau wiedergeben.* Da haben wir so ein Histogramm. Eine schöne Klausur erzeugt praktisch, ohne dass wir was drehen, eine Normalverteilung. *Dazu noch, kann man bei der Bewertung sehen, welche Aufgabe gar nicht bearbeitet wurde.* Das war dann eine schlechte Aufgabe. *Wenn wir es schaffen, dass fast alle Aufgaben bearbeitet wurden und dann automatisch die Normalverteilung vorliegt, dann kann man sagen: das war eine schöne Klausur.* Das ist glaube ich das, was mein Ziel ist.

Zitat 69 Fall MW 1

Moderator: Also ich meine, hier gibt es jetzt auch eine nicht so wirklich schwierige Aufgabenstellung, die 87% der Studierenden lösen konnten und da gibt es 24 Punkte. Und das heißt man hat viele Punkte erreicht, oder man hat selbst das nicht hingekriegt und hat dementsprechend wenige Punkte bekommen. Also wirklich dieses entweder oder. Weil 24 Punkte ist schon wirklich viel, dafür, dass die Aufgabe jetzt nicht unbedingt mittelmäßig sondern eher einfach ist.

Doktorand 2: Also sie ist vor allem deswegen einfach, in dem Sinne, dass es viele gelöst haben, weil das meiste dann zählt. Das ist auch das was ich ganz am Anfang gesagt hab. *An der Stelle ein wenig schwierig, da eine saubere Trennung hinzukriegen* [bei der Bewertung der Prüfungsantworten]. *Weil der eine schreibt halt eine saubere Begründung mit ein oder zwei Sätzen und sagt okay, dem glaube ich, der hat verstanden um was es geht und der andere schreibt halt nur ein Wort, der kriegt aber den Punkt halt auch.*

Zitat 70 Fall MW 1

Moderator: Ja, was uns bei der Aufgabe aufgefallen ist, dass auch diejenigen, die eigentlich sehr wenig Punkte in der Gesamtprüfung erzielt haben, auch hier teilweise 24 Punkte erzielt haben.

Akademischer Oberrat: Ich mein, wenn wir Aufgaben haben, dass wir da keine Differenzierbarkeit oder so haben, das ist ja eigentlich okay, aber wenn wir natürlich eine Frage haben, die 24 Punkte hat, dann ist natürlich die Frage, ob das so eine Fünferbremse sein sollte oder ob wir jetzt mit 24 Punkten jetzt schon eine deutlich bessere Differenzierung erreichen sollten. Ich glaub das wäre halt hier tatsächlich die Idee, oder?

Doktorand 2: Genau, da drauf wollte ich nämlich gerade auch raus. Weil das einfach zu machen ist halt nicht so einfach. Also wir haben das ja immer so gemacht, dass das mit der Zeitkurve geht, die ich brauche. Also erstmal hier diese ganze Seite runterzuschreiben, ist halt ein wenig Zeitaufwand, aber auch das zu lesen und das auch zu verstehen. Was ist denn da überhaupt diese Situation. *Das ist ja auch eine ganze DIN A4 Seite und das braucht halt auch einfach Zeit und wenn man da jetzt einfach nur zehn Punkte hergeben würde, statt 30, dann korreliert das nicht mehr.*

Zitat 71 Fall MW 1

Doktorand 1: Also übrigens Durchschnitt ist 65%. Gut, da haben wir noch recht gute Werte, das ist das mit der Ergodynamik. Das war denk ich offen genug. Da sollte man das halt einzeichnen, gut da könnte man überlegen, dass die meisten richtig haben. Fühlt sich vielleicht ein wenig so an, als ob es eventuell zu leicht gewesen ist. Vielleicht kann man dann einfach weniger Punkte geben, weil 8 Punkte ist halt schon recht viel.

Zitat 72 Fall MW 1

Akademischer Oberrat: Genau, weil was wir auf jeden Fall auch machen werden, ist mit der Matrix weiterarbeiten, das ist so der erste große Punkt. [...] Wenn man jetzt so langfristig denkt und man könnte sozusagen nach jeder Klausur die Aufgaben ablegen in die Aufgabendatenbank mit Schwierigkeit und Empirie hinterlegen, also mit allen möglichen Kennzahlen und mir dann später mal eine tolle Klausur zusammenziehen, dann hab ich den zeitlichen Faktor halt noch mit

drin, also dass ich dann Aufgaben zusammenziehe, die alle ganz am Schluss standen und denk, dass die schwierig war, weil die kaum einer gelöst hat. Ich würde jetzt quasi nach diesem Schema Aufgaben nach schwer ablegen, weil sie kaum einer bearbeitet hat zum Beispiel und wenn ich jetzt 5 Jahre sammel, dann hab ich einen Pool von schweren Aufgaben und die Hälfte war wirklich schwer und die andere Hälfte stand immer am Schluss.

Zitat 73 Fall MW 2

Doktorand 2: Das ist die Summe der erreichten Punkte. Genau, wenn man jetzt nach rechts rüberzieht [in der Excel Tabelle], sieht man halt Teil 1, Teil 2 [der Prüfung] als Summe. Und die Punktabstände *der Studierenden zur besseren Note geben die immer noch aus. Wir schauen, dass niemand weniger als ein Punkt von der besseren Note entfernt ist, vor der Einsicht. Da schauen wir dann nochmal sehr gründlich, ob die nicht den einen Punkt noch von uns bekommen.*

Zitat 74 Fall MW 3

Doktorandin: Ich habe extra mein Sheet umgebaut, dass wenn ich zu viele Punkte bei einer Aufgabe eingebe, dass es rot wird.

Zitat 75 Fall MW 4

Doktorand: Und dann hab ich noch eine Aufgabe reingenommen. Dazu gab es in der Vorlesung ein Diagramm. Die DIN Norm wurde mehrfach genannt und das war die Aufgabe, die nur 3 % der Leute beantwortet haben. Es gibt 10000 DIN Normen und jetzt guckst du alle tausend nach oder wie. Die sollte auch ohne nachschauen sitzen. Da gab es nur 1 Punkt, aber das war die, die gnadenlos in die Hose gegangen ist.

Zitat 76 Fall MW 4

Doktorand: Diese Aufgabe und Aufgabe 5a, wo sie die beiden Begriffe aus der Biomechanik ja einfach mal definieren sollten. Das sind halt die zwei Aufgaben, die eben rot aufgefallen sind.

Moderator: Was hast du erwartet bei der Aufgabe. Bei der 5a?

Doktorand: Bei der 5a. Was ich erwartet habe. Ja die zwei gelben Begriffe. Das ist eine Folie und das sind zwei Definitionen und die muss man halt genauso lernen wie sie im Skript stehen. Die lernst du auswendig. Grad diese Aufgabe 5a. Also da hast du gelernt oder nicht gelernt. Also das ist so, die zwei Möglichkeiten gibt es. Und so halb gelernt. Also was schwierig ist, also von der Bewertung her, ist motorische Einheit. Und da gibt es so Anzahl und Schlagwörter. Und das wurde oft umschrieben, aber teilweise konnten wir den Punkt nicht geben. Da gab es halt einen halben Punkt, weil du weißt schon, um was es geht, aber die Fachtermini fehlen. Und da würde ich halt

nächstes Mal in die Musterlösung reinschreiben, okay, volle Punktzahl gibt es bei der Definition und wenn diese Fachtermini auftauchen. Wird es umgangssprachlich beschrieben, gibt es nur einen Punkt.

Zitat 77 Fall MW 5

Doktorand 1: Wo es dann interessant ist, ist bei vielen Aufgaben wo man [eine Hälfte der Studierenden] fast komplett volle Punktzahl hat und der andere Teil der Studenten hat fast gar keine Punkte. Bei sowas sind wir dann immer etwas kritischer, also wenn das Delta relativ groß ist. *Das könnte halt darauf hinweisen, dass man Folgefehler hätte geben sollen und wir das eben nicht gemacht haben. Da sind wir eben stutzig geworden.*

Zitat 78 Fall MW 5

Akademischer Oberrat: Also das wundert mich bei der ersten Aufgabe auch [die geringe Lösungsquote]. Weil die Graphik ist einfacher. Bei der Begründung, ja. Unsauber. Unsauber formuliert. Bei der Aufgabe 2 ist die Graphik ja schon kompliziert. Ja, oder sehr aufwendig. Da steckt sehr viel drinnen. Sehr viele Details drinnen. Der Graphenteil hat in der Aufgabe 2 a, macht der 11 Punkte aus. Und das Entwickeln, ja nennen Sie zwei konkrete Maßnahmen, das sind nur 2 Punkte von den 13. Ja. Also haben bei der Aufgabe 2, ist zumindest eine Maßnahme wohl genannt worden, weil maximal 12 Punkte erreicht worden sind.

Moderator: Und 11 Punkte auf die Grafik. Also da ist es halt so, dass man eigentlich eine eher reproduktiv angelegte Anforderung sehr hoch dann belohnt sozusagen, und die Entwicklungsaufgabe eher dann so untergeht. Gut, wie man seinen Schwerpunkt legen will natürlich. Ich habe mich auch gefragt, weil die Grafik so detailliert ist. Wie wichtig ist denn das eigentlich so von dem was ihr, was die können sollen. Dass sie die im jeden Detail können?

Akademischer Oberrat: Es sind unterschiedliche Aspekte drinnen, die in so einer Graphik ja schon abgefragt werden. Es steht ja eigentlich auch drinnen, was sie im Angabentext, so: den Verlauf der Selbstkosten und der Erlöse. Bezeichnen Sie die Diagrammachsen. Ja, ok, was da jetzt reinkommt, schreibt man nicht hin, weil sonst gibt man die Lösung vor. So, aber es steht drinnen, die Abschnitte vor und während der Marktledensdauer. Ja. Kennzeichnen Sie die Lage des Break-Even-Point. Ja, das ist, ja, es gibt eine andere Kurve in der Vorlesung, da hat man auch Selbstkosten drinnen und Deckungsbeitrag und Erlös. Break-Even-Point, wo der Zusammenhang schon erklärt wird, ja. Also, und Gewinn und Erlös, dass das eigentlich, dass man das da in die Graphik auch übertragen kann. Wobei die Graphik gibt es ja so auch 1 zu 1 in der Vorlesung.