

Herausforderung Prüfen

**Entwicklung und Evaluation eines Qualitätsentwicklungsprogramms für
Lehrende an Hochschulen**

Christoph Schindler, Florian Schulz, Dr. Johannes Bauer,
Prof. Dr. Manfred Prenzel (Lehrstuhl für Empirische Bildungsforschung)

Dr. Nadine Schlomske-Bodenstein, Dr. Alexandra Strasser (ProLehre)

München im Juli 2015

Inhalt

1. Einleitung: Prüfungen im Fokus.....	1
2. Hintergrund	2
2.1 Prüfungen als zentraler Stellhebel der Hochschullehre	2
2.2 Herausforderungen bei der Umsetzung zielorientierten Lehrens und Prüfens.....	3
2.3 Prüfungskultur an der Technischen Universität München	4
3. Gutes Prüfen: Ein theoretisches Modell.....	6
4. „Herausforderung Prüfen“ – Ein Weg zu besseren Prüfungen?	10
4.1 Nachhaltige Qualitätsentwicklung	10
4.2 Teilnehmerinnen und Teilnehmer und Ablauf des Programms.....	11
4. Wissenschaftliche Begleitforschung.....	12
4.1 Fragestellungen	12
4.2 Forschungsdesign	13
4.3 Stichprobe und Rücklaufquote.....	14
4.4 Erhebungsinstrumente.....	14
4.5 Auswertungsverfahren.....	15
5. Ergebnisse	16
5.1 Akzeptanz des Projektes und wahrgenommene Kompetenzentwicklung.....	16
5.2 Veränderungen in der Prüfungsqualität.....	22
6. Diskussion und Ausblick	26
Literatur.....	28
Veröffentlichungen	29
Tagungsbeiträge und Vorträge.....	29
Monografien.....	29
Anhang	30

1. Einleitung: Prüfungen im Fokus

Prüfungen erfüllen in der akademischen Lehre mehrere Funktionen. Sie sollen zuverlässig Auskunft über das von Studierenden erlernte Wissen und Können geben. So dienen sie für die Studierenden einerseits zur Lernstandkontrolle und qualifizieren andererseits für den weiteren Bildungs- und Berufsweg. Prüfungen sind aber gleichzeitig auch ein Spiegel für die Lehre. Prüfungen geben Lehrenden eine Rückmeldung über den Erfolg ihrer Veranstaltungen und liefern Hinweise für Verbesserungen, indem sie mit Hinblick auf die angestrebten – und dann tatsächlich erreichten Lernergebnisse dabei helfen, den Lehrstoff gegebenenfalls neu zu strukturieren oder andere Lehrformen zu nutzen.

Obwohl die Bedeutung von Prüfungen unmittelbar einleuchtet, wurde die Qualität der Prüfungen in der akademischen Bildungsdiskussion bisher nur beiläufig thematisiert. Wenn von „guter Lehre“ die Rede ist, wird vorwiegend an faszinierende Veranstaltungen gedacht und nicht an gute Prüfungen. Dabei wird übersehen, dass Prüfungen eine erhebliche zeitliche Investition für die Beteiligten auf beiden Seiten bedeuten. Studierende verbringen sicherlich ebenso viel Zeit mit der Vorbereitung auf Prüfungen wie für die Nachbereitung von Veranstaltungen. Für die Lehrenden kann vor allem der Korrekturaufwand immens sein. „Gutes Prüfen“ verdient deshalb mehr Augenmerk. Deshalb gilt es zu bedenken: Gute Lehre endet nicht mit einer Prüfung, sondern sie beginnt mit einer guten Prüfung. Wenn von Anfang an klar ist, was die Studierenden am Ende einer Veranstaltung tatsächlich sicher beherrschen sollen, dann bestehen beste Voraussetzungen für eine zielbezogene, kohärente und kompetenzorientierte Lehre.

Doch wie viel ist über die Prüfungspraxis an Hochschulen hierzulande überhaupt bekannt? Wie systematisch wird geprüft? Systematisches Wissen über die bestehende, akademische Prüfungspraxis gibt es kaum. Dies legt nahe, dass Prüfen an Hochschulen trotz seiner langen Geschichte bisher vor allem auf individueller Erfahrung beruht. Vor diesem Hintergrund wurde 2011 deshalb an der Technischen Universität München durch den Lehrstuhl für Empirische Bildungsforschung der TUM School of Education eine umfassende Befragung zur Prüfungspraxis durchgeführt. Die Ergebnisse verdeutlichen, dass auch nach der Bologna-Reform und der Einführung gewisser Mindeststandards höchst unterschiedlich geprüft wird. Verständlich ist daher, dass sich Studierende mehr Information und ebenso transparente wie klare Anforderungen wünschen. Verständlich ist ebenfalls, dass die Lehrenden entsprechend nach (konzeptioneller) Unterstützung bei der Prüfungserstellung und –auswertung fragen. Diesen Wünschen trägt das Projekt „Herausforderung: Prüfen“ Rechnung, das 2012 als Kooperationsprojekt zwischen dem Susanne Klatten-Stiftungslehrstuhl für empirische Bildungsforschung an der TUM School of Education und ProLehre gestartet wurde. „Herausforderung Prüfen“ soll Lehrende an der TUM im bei der Prüfungskonzeption unterstützen, um nachhaltig eine auf wissenschaftlich begründeten Standards beruhende Prüfungspraxis zu etablieren.

2013 und 2014 wurde das vom Stifterverband für die Deutsche Wissenschaft ausgezeichnete Programm mit zwei Kohorten, bestehend aus Lehrenden der Technischen Universität, umfassend erprobt und zeitgleich in wichtigen Punkten weiter entwickelt. Mit dem vorliegenden Projektbericht ist diese Phase des Projekts abgeschlossen. Die hier

vorgestellten Ergebnisse, die vorwiegend auf Forschungsfragen aus dem Projektantrag aufbauen, sind dabei auch als Anreiz für Lehrende zu verstehen, sich mit der eigenen Prüfungspraxis systematisch auseinanderzusetzen. Im Laufe der Pilotierung konnten wichtige Eindrücke gesammelt und Wissen über die Chancen und Herausforderungen bei der Erstellung guter Prüfungen generiert werden.

Im ersten Teil des Berichts soll ein Verständnis für die Relevanz von Prüfungen geschaffen werden. Dabei werden auch einige Herausforderungen benannt, die bei der Umsetzung von Standards, wie sie im Zuge der Bologna-Reform an die Universitäten herangetragen wurden, in Bezug auf Prüfungen entstanden sind. Abschließend wird anhand von Ergebnissen aus der Gesamtbefragung zur Prüfungspraxis an der TUM dargelegt, dass wichtige Kriterien guten Prüfens bisher nur ansatzweise realisiert sind (Vgl. Kapitel 2). Systematische Kriterien sind jedoch wichtig, um die Qualität von Prüfungen nachhaltig zu sichern. In Kapitel 3 werden deshalb einige theoretische Überlegungen zu einer guten Prüfung angestellt und anschließend in einem Vorgehensmodell integriert. Dieses bildet die Grundlage für das Projekt „Herausforderung Prüfen“, das in Kapitel 4 vorgestellt wird. Kapitel 5 gibt Informationen zur wissenschaftlichen Begleitung des Programms. In Kapitel 6 werden schließlich Ergebnisse aus der Pilotphase des Projekts vorgestellt. Kapitel 7 dient der abschließenden Diskussion der Befunde und gibt einen Ausblick auf weitere Entwicklungsmöglichkeiten des Projekts an der TUM, aber auch auf Anschlussmöglichkeiten an anderen Hochschulen.

2. Hintergrund

2.1 Prüfungen als zentraler Stellhebel der Hochschullehre

„Prüfen findet am Ende des Semesters statt.“ Jeder, der selbst akademisch lehrt, wird diesen Eindruck so oder in ähnlicher Form bestätigen können. Die Gründe für diese Vorstellung, die sich hartnäckig an Hochschulen hält, sind vielfältig: Genereller Zeitmangel, weshalb die Vorbereitung und Durchführung der Veranstaltungen eine Vorrangstellung einnimmt; der Stellenwert des Prüfens als Hebel für Lehr- wie Lernprozesse wird im Alltag nicht wahrgenommen und rückt allenfalls dann in den Fokus, wenn Studierende schon bereits beim ersten Veranstaltungstermin die Frage aufwerfen, welcher Stoff denn nun eigentlich für die Klausur relevant sei. Ein solches Fragen lässt sich natürlich leicht als latentes Desinteresse an den Fachinhalten interpretieren. Es gibt schließlich in den einzelnen Fächern über die Prüfungsinhalte hinaus so viel Wichtiges und Interessantes zu lernen und zu erfahren. Dieses Semesteranfangsritual sollte man allerdings durchaus ernstnehmen: Es zeigt, dass Prüfungen in erheblichem Maße beeinflussen, wie Studierende die Lehre wahrnehmen. Sie richten nicht nur ihr unmittelbares Lernen an den Anforderungen von Prüfungen aus, sondern erleben Lehrveranstaltungen vor dem Erwartungshorizont der sie abschließenden Prüfungen. Inhalte werden deshalb oft nur selektiv wahrgenommen. Informationen, die durch das Raster „prüfungsrelevant/nicht prüfungsrelevant“ fallen, werden dagegen ausgeblendet. Gelernt wird vor allem ökonomisch. In der Forschung spricht man in diesem Zusammenhang auch von einer „grade point average perspective“ (Becker, Geer & Hughes, 1995).

Nun lässt sich diese Form der Zielstrebigkeit entweder beklagen – man kann durch Appelle natürlich immer versuchen, für Motivation und Interesse zu sorgen -, oder aber konstruktiv im Sinne der Lehre nutzen: Wenn Studierende vor allem die Prüfung im Blick haben, dann ist es umso wichtiger, dass die Prüfungsanforderungen gut elaboriert sind und alle wichtigen Aspekte aus der Lehre abdecken. Gute Prüfungen schlagen zwei Fliegen mit einer Klappe: Sie sind einerseits diagnostische Instrumente und steuern andererseits das Lernen der Studierenden in die gewünschte Richtung (Shavelson, 2010). Dies macht deutlich, warum Prüfungen nicht losgekoppelt von der eigentlichen Lehrveranstaltung betrachtet, sondern im Einklang mit ihr konzipiert werden sollten.

Mit Blick auf die abschließende Prüfung erhält die Lehre zudem eine systematischere Struktur: Wenn klar ist, was geprüft werden soll, ist auch klarer, was gelehrt werden muss. Die gute, alte Lehrbuchstruktur kann in diesem Fall durch entsprechende Zielkonzepte weiter systematisiert werden. Nichts anderes ist im Grunde gemeint, wenn im Zuge der Bologna-Reform von einer „Outcomeorientierung“ der Lehre an Hochschulen und auf Ebene der Konzeption von Studiengängen von „Lehrzielen“ oder „angestrebten Lernergebnissen“, wie etwa im „Wegweiser zur Erstellung von Modulbeschreibungen“¹ des Hochschulreferats Studium und Lehre, gesprochen wird. Diese Zielbeschreibungen stellen nicht mehr und nicht weniger da als ein Bindeglied zwischen Prüfung und Lehre dar. Diese Verbindung muss aber natürlich auch in der Praxis erst einmal hergestellt werden.

2.2 Herausforderungen bei der Umsetzung zielorientierten Lehrens und Prüfens

Eine Auseinandersetzung mit der Prüfungspraxis wird Lehrenden seit der Bologna-Reform explizit abverlangt. Die Anforderungen sind formell an die Modularisierung von Studiengängen gekoppelt. Festgelegt werden sollen Qualifikationsziele für integrative Module. Diese sollen idealerweise nicht nur eine Veranstaltung umfassen, sondern gleich mehrere Lehrangebote in einem Themenkomplex systematisch zusammenbinden. Die Prüfungsanforderungen sollen dabei kohärent aus den Qualifikationszielen für die jeweiligen Module abgeleitet werden. Dazu gibt es als Orientierungshilfe zum Beispiel den „Qualifikationsrahmen für deutsche Hochschulabschlüsse“ (Kultusministerkonferenz, 2005), die Einteilung von Qualifikationen und Kompetenzen nach dem TUNING Project (González & Wagenaar, 2006) oder die Dublin Descriptors, in denen Definitionen von Qualifikationen zur Unterscheidung von Bachelor- und Masterstudiengängen enthalten sind.

Gemeinsam ist diesen Handreichungen, dass sie versuchen, allgemeingültige und fächerunabhängige Orientierungshilfen zu geben, die dann an den einzelnen Hochschulen, Fakultäten und Lehrstühlen entsprechend in die Praxis übersetzt werden sollen. Dies setzt aber voraus, dass die sehr allgemeinen Konzepte durch die Beteiligten entsprechend in ihren Implikationen verstanden und auf einzelne Aspekte der Lehre übertragen werden. Was auf der Ebene von Modulbeschreibungen und Zielformulierungen für einzelne Veranstaltungen noch relativ gut funktioniert – wobei einzelne Forschungsberichte nahelegen, dass vor allem die Modularisierung von Studiengängen nur langsam

¹ Abzurufen unter

https://www.lehren.tum.de/fileadmin/w00bmo/www/Downloads/Themen/Studiengaenge_gestalten/Dokumente/TUM_Wegweiser_Modulbeschreibungen_Stand_November_2014.pdf

voranschreitet und bestehende Veranstaltungen häufig nur als Module deklariert werden – gerät auf der Handlungsebene überaus komplex und erfordert Zeit und Expertise.

Selbst wenn der Schluss von Lehrzielen oder angestrebten Lernergebnissen auf der Studiengangebene zur Relevanz von Prüfungen gezogen wird, wird die eigentliche Umsetzung zu einer Herausforderung. Die Erstellung einer zielbezogenen Prüfung ist alles andere als trivial und darüber hinaus ein entsprechend zeitaufwendiger Prozess. Um dies beispielhaft zu veranschaulichen: Versteht man eine Prüfung als eine Art von Leistungstest, so stehen plötzlich messtheoretische neben lerntheoretischen und fachdidaktischen Überlegungen und sollen miteinander abgestimmt werden. Dies geschieht am besten im Einklang mit der Konzeption der Lehrveranstaltung. Woher soll ein entsprechendes Wissen bei Lehrenden kommen, die im Grunde über ein Lehrdeputat ins kalte Lehrwasser geworfen werden, ohne vorher eine einschlägige pädagogisch-didaktische Ausbildung durchlaufen zu haben? Es ist viel verlangt, von diesen Lehrenden Prüfungen zu erwarten, die allen Gütekriterien gerecht werden.

Vermutlich liegt unter anderem darin ein Grund, weshalb die grundsätzlich erstrebenswerten Zielvorgaben von Bologna auf der Handlungsebene nur langsam umgesetzt werden. Es fehlt nach wie vor an konkreten Entwicklungsangeboten an den Hochschulen, was sicherlich auch auf eine mangelhafte Datenlage zum Thema „Prüfen“ zurückzuführen ist. Wie wird denn nun eigentlich an Hochschulen aktuell geprüft? Gibt es Hinweise darauf, dass die durch Bologna angestoßenen Reformen auch zu einer standardisierten Prüfungspraxis führen? Das Hochschulreferat für Studium und Lehre der Technischen Universität wollte 2011 diesen Fragen an der TU München nachgehen und hat beim Lehrstuhl für empirische Bildungsforschung der TUM School of Education eine Gesamtbefragung zur Prüfungskultur in Auftrag gegeben.

2.3 Prüfungskultur an der Technischen Universität München

An der mittels Online-Fragebogen durchgeführten 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 möglicher Unterstützungsbedarf bei der Erstellung und Korrektur von Prüfungen². Von Interesse für die aufgeworfenen Fragen waren die Ergebnisse der Befragung der Dozierenden, die Auskunft über die Art ihrer Prüfungen gaben. Dabei wurde vor allem auch nach Hinweisen gesucht, inwieweit systematisch geprüft wird und gewisse Standards bei der Vorgehensweise bereits angewandt werden.

² Im Folgenden werden ausgewählte Befunde systematisch in den Zusammenhang mit theoretischen Vorüberlegungen zur Rolle von Prüfungen im Lehr- und Lernprozess gestellt. Weitere Informationen zur Durchführung der Studie und ausführliche Ergebnisse sind in den beiden Projektberichten enthalten, die dem Internetauftritt des Hochschulreferat Studium und Lehre zu entnehmen sind: <https://www.lehren.tum.de/themen/pruefungen/pruefungskultur-an-der-tum-projektberichte/>

Dabei fiel unter anderem der relativ geringe Zeitaufwand für die Erstellung der Prüfung (Mittelwert = 0,5 Stunden) im Vergleich zum immensen Aufwand für die Korrektur (Modus = 21 Stunden und mehr) auf. Dieses Verhältnis ist insofern überraschend, als die systematische Konzeption von Prüfungsaufgaben generell sehr aufwendig ist und viele konzeptionelle Überlegungen erfordert, während die Korrektur dann durch ein passendes Bewertungsraster oft ökonomischer gestaltet werden kann. Falls der niedrige Zeitaufwand bei der der Prüfungserstellung auf die Verwendung von Aufgaben oder gar kompletten Prüfungen aus vorangegangenen Semestern zurückzuführen ist, kann an dieser Stelle dennoch die Systematik hinter dieser Art der Prüfungserstellung hinterfragt werden: Wie robust sind die bestehenden Aufgaben gegenüber Strategien des Auswendiglernens, wenn man das Ergebnis aus der Studierendenbefragung mit einbezieht, wonach Altklausuren das hilfreichste Mittel bei der Prüfungsvorbereitung von Studierenden sind³? Erfolgt eine Randomisierung bei der Ziehung von Aufgaben? Welche anderen Kriterien spielen bei der Auswahl eine Rolle? Ist sichergestellt, dass auch bei neu kombinierten Prüfungen durch die alle Aspekte erfasst werden, die auch als angestrebte Lernergebnisse definiert sind?

Unabhängig davon sprechen der hohe Zeitaufwand, der von den Lehrenden auf 3,6 Stunden pro Woche geschätzt wurde und wohl entsprechend der vorherigen Ergebnisse vor allem als Korrekturzeit anzurechnen ist, sowie die hohe Anzahl an Prüfungsfällen von 200 Studierenden pro Semester, auch dafür, dass relativ selten auf Möglichkeiten zur Standardisierung zurückgegriffen wird. Eine solche könnte etwa durch die Verwendung von Aufgaben im Single- oder Multiple-Choice-Format erreicht werden. Dieses Format wird an der TUM aber derzeit kaum verwendet; die Lehrenden griffen laut eigenen Angaben vornehmlich auf offene Antwortformate zurück. Dahinter steht vermutlich auch die Sorge, dass Wahlantwort-Formate eher oberflächliches Wissen abfragen oder „Tricksen“ begünstigen (im Übrigen lässt sich dies durch sorgfältig konzipierte Aufgabenstellungen und Distraktorwahlen relativ leicht ausräumen).⁴ Gleichzeitig gab über die Hälfte der Befragten an, zu einem großen Anteil in den Prüfungen eine Anwendung des Gelernten auf Problemfälle anzustreben. Eventuell spielt hierbei die Vorstellung eine Rolle, dass mit vorstrukturierten Aufgabenstellungen keine anspruchsvollen Konzepte abgeprüft werden können. Dabei lassen sich auch mit Single- oder Multiple-Choice-Aufgaben systematisch mehrstufige Konzepte wie Kompetenzen⁵ abprüfen. Diesbezüglich sollten also Potenziale verschiedener Aufgabenformate vermittelt und insbesondere standardisierte Auswertungsverfahren nutzbar gemacht werden.

Ebenso zu diskutieren ist der Beginn der Prüfungsvorbereitung. In der vorliegenden Stichprobe gaben 43,3 Prozent an, erst kurz vor dem Prüfungstermin mit der Vorbereitung zu beginnen. Auch hier stellt sich die Frage, inwieweit die angestrebten Lernergebnisse bei einem späten Vorbereitungsbeginn der Prüfung angemessen in Aufgaben übersetzt werden, beziehungsweise die angestrebten Lernergebnisse auch einmal neu formuliert werden, sollten sie nicht überprüfbar sein. Idealerweise findet hier eine Anpassung bereits zu Beginn

³ So ein Ergebnis aus der zeitgleich durchgeführten Studierendenbefragung. Auch hierzu können die ausführlichen Ergebnisse im entsprechenden Projektbericht eingesehen werden.

⁴ Die Allgemeine Prüfungs- und Studienordnung der TUM sieht deshalb auch besondere Regeln bei der Verwendung des Wahlantwortformats in Klausuren vor, was deren Einsatz an der TUM wohl noch unattraktiver macht.

⁵ Zum Kompetenzbegriff, vgl. Kapitel 3.

des Semesters statt, indem geprüft wird, ob die formulierten Lernergebnisse auch in Prüfungsaufgaben operationalisiert werden können.

Abschließend interessierte vor allem die Frage nach dem Unterstützungsbedarf, da dies ein Indikator dafür ist, als wie dringlich Herausforderungen in der Prüfungspraxis von den Lehrenden betrachtet werden. Vor allem der akademische Mittelbau aus wissenschaftlichen Mitarbeiterinnen und Mitarbeitern signalisierte hier einen erhöhten Unterstützungsbedarf bei der Prüfungserstellung. Dies erscheint insofern plausibel, da diese Personengruppe wohl über besonders wenig Erfahrung mit dem Erstellen einer Prüfung verfügt. Gleichzeitig war aber auch eine gewisse Skepsis zu spüren, da ein nicht unerheblicher Anteil Unterstützung in Sachen Prüfung auch ablehnt. Die Ablehnungsquoten könnten ein Hinweis darauf sein, dass die Lehrenden keine drängenden Probleme in ihrer Prüfungspraxis sehen oder sich einfach nicht in die Karten schauen lassen möchten. Für diesen Wunsch nach Autonomie bei der Prüfungsgestaltung bei gleichzeitigem Bewusstsein für Verbesserungsbedarf spricht auch, dass – im Gegensatz zur unmittelbaren Unterstützung – ein Leitfaden zur Prüfungserstellung von den meisten gewünscht wird (M=2,9, auf einer Skala von 1-4). Insgesamt lässt sich daher vor allem vermuten, dass Autonomie und Gestaltungsspielraum bei der Prüfungserstellung für die Lehrenden eine wichtige Rolle spielen. Einstimmiger wurde dagegen Bedarf an Unterstützung bei der Korrektur von Prüfungen geäußert (M=3,0, auf einer Skala von 1-4). Inwieweit die Befragten sich inhaltliche Unterstützung, z.B. zur Objektivierung der Auswertungen, oder lediglich mehr Personal zur Aufteilung der Korrekturlast wünschen, kann auf Basis der Daten nicht beantwortet werden.

Vieles spricht nach Betrachtung der Daten dafür, dass an der TU München durchaus ambitioniert mit hohem Anwendungsbezug geprüft wird. Gleichzeitig sind einzelne Befunde aber auch als Hinweis darauf zu werten, dass wichtige Kriterien, die beim Erstellen einer Prüfung beachtet werden sollten, noch nicht oder nur ansatzweise umgesetzt sind. Die Etablierung solcher Standards könnte als Stellhebel alle Aspekte der Lehre und damit die Qualität von Abschlüssen nachhaltig positiv beeinflussen (Vgl. Kapitel 2.1). Ihr derzeitiges Fehlen darf dabei nicht als Versäumnis der Lehrenden gewertet werden. Prüfen an Hochschulen ist bis jetzt zu einem gewissen Anteil sicherlich auch eine Erfahrungssache, die stark vom individuellen Wissen der Beteiligten abhängig ist. Mit Bologna hat das Prüfen eine neue Verbindlichkeit bekommen. Zusätzlich zu den fächerübergreifenden Handreichungen sind daher nun auch konkrete Entwicklungsangebote notwendig.

3. Gutes Prüfen: Ein theoretisches Modell

Vorstellungen darüber, was eine „gute“ Prüfung auszeichnet, beruhen auf theoretischen Annahmen. Bei einer Prüfung handelt es sich um eine Form der Leistungsdiagnostik. Ziel einer guten Prüfung ist es, ein vorab bestimmtes Konstrukt zu messen und in seiner Ausprägung zu bewerten. Ein Konstrukt ist im Sinne der pädagogischen Diagnostik die Operationalisierung eines Personenmerkmals, also zum Beispiel einer Kompetenz. Eine Kompetenz beschreibt im Grunde die Befähigung einer Person, variable Anforderungssituation in einem bestimmten Lern- oder Handlungsbereich zu bewältigen und nimmt in der Bildungsdebatte bereits seit Längerem einen zentralen Stellenwert ein (Vgl. Klieme & Leutner, 2006). Auch die angestrebten Studiengangziele sollen Kompetenzen benennen (Kulturministerkonferenz, 2005). Generell verweist der Kompetenzbegriff darauf,

dass es nicht nur darauf ankommt, über Wissen oder Fähigkeiten im Sinne kognitiver Potentiale zu verfügen, sondern dieses Wissen und diese Fähigkeiten auch flexibel und zielgerichtet eingesetzt werden können, um kontextualisierte und reale Problemstellungen und Anforderungssituationen erfolgreich zu bewältigen. Im Grunde sind bei der Bestimmung einer Kompetenz also folgende Fragen zentral: Welche Situationen sind relevant? Welche Anforderungen stellen diese Situationen? Und: Über welche Merkmale muss die Person verfügen, um die Situationen bewältigen zu können?

Beim Messen tritt dabei immer ein Grundproblem auf: Menschen sieht man in der Regel nicht an, ob sie über die notwendigen Merkmale, zum Beispiel als Voraussetzung kompetenten Handelns, verfügen. Entsprechend muss versucht werden, aus dem Verhalten bei der Bewältigung bestimmter Anforderungssituationen auf die Ausprägung entsprechender Merkmale zu schließen. Aus dieser Problematik heraus erschließen sich wichtige Gütekriterien der Testtheorie:

Inhaltliche Validität

Rückschlüsse vom Verhalten auf das dahinterliegende Konstrukt müssen gut begründet sein. Zum Beispiel lässt sich nur schwierig argumentieren, dass jemand, der eine einfache lineare Gleichung lösen kann, dazu in der Lage ist, ein komplexes mechanisches Problem zu bearbeiten. Eher würde man wohl vermuten, dass die Person weitere Voraussetzungen benötigt, die – neben den Kenntnissen verschiedener mathematischer Prozeduren - auch das Wissen um deren Anwendung in spezifischen Situationen und auch wirtschaftliche Rahmenbedingungen umfassen. Außerdem muss die Person entsprechend motiviert sein. Es besteht nun entweder die Möglichkeit, die Person zu Testzwecken direkt in der Praxis mit einem solchen Problem zu konfrontieren, oder es wird versucht, die einzelnen Voraussetzungen getrennt voneinander zu erfassen⁶. Die logische Konsistenz dieser Rückschlüsse entspricht der *inhaltlichen Validität*⁷ der Messung. Es geht hierbei also um die bestmögliche Operationalisierung eines Konstrukts.

Reliabilität

Gleichzeitig muss die Messung belastbar, also wiederholbar sein und im idealen Fall aus mehreren Messungen desselben Konstrukts bestehen. Möglicherweise findet ein Studierender ja zufällig die richtige Lösung für das mechanische Problem, da er nach dem Versuch-und-Irrtum-Prinzip die richtigen Schalter findet. Wird er aber in seiner Berufspraxis nun systematisch mit derartigen Problemen konfrontiert, würde eine einmalige Messung hier eine falsche Vorhersage für seine Kompetenz liefern. Messungen müssen daher belastbar sein. Für die Belastbarkeit einer Messung tritt das Kriterium der *Reliabilität* sein.

⁶ Eine Diagnostik der Voraussetzungen für kompetentes Handeln beruht auf Struktur- und Entwicklungsmodellen für Kompetenzen. Der Vorteil und die Relevanz für das Prüfen an Hochschulen liegen darin, dass hierbei keine komplexe Situation geschaffen werden muss, sondern einzelne Merkmale zum Beispiel auch in schriftlichen Prüfungen abgeprüft werden.

⁷ In der Fachliteratur wird nach weiteren Arten der Validität, wie etwa die Konstrukt-, Vorhersage- oder Übereinstimmungsvalidität. Bei der Erstellung einer Prüfung ist aber vor allem der Inhaltsaspekt in besonderem Maße relevant.

Objektivität

Um dies zu erreichen, wird eine größtmögliche Standardisierung im Sinne der *Objektivität* der Messung angestrebt. Es sollten also möglichst keine anderen Aspekte als die gewünschten die Messung beeinflussen. So ist zum Beispiel darauf zu achten, dass es keine ablenkenden Faktoren gibt.

Fehlerquellen lassen sich nun durch die Einhaltung bestimmter Prinzipien bei der Prüfungsgestaltung minimieren. Im Rahmen eines Dissertationsvorhabens wurde für „Herausforderung Prüfen“ deshalb ein theoretisches Vorgehensmodell entwickelt, um die Güte einer Prüfung im Zuge ihrer Konzeption zu sichern (Vgl. Schindler 2015, in Vorbereitung):

- (1) Festlegung des Prüfungsgegenstands:** Dieser Schritt umfasst die Festlegung von angestrebten Lernergebnissen. Hierbei wird eine Passung zwischen den zu überprüfenden Inhalten und dem erwarteten Verhalten hergestellt. Wird zum Beispiel die Überprüfung von Kompetenzen angestrebt, so sollten Anforderungssituationen und zur Bewältigung benötigte Personenmerkmale bestimmt werden. Die Lehrziele bilden den Ausgangspunkt für eine möglichst zielgerichtete Erstellung von Prüfungsaufgaben. Ebenso dienen angestrebte Lernergebnisse der Überprüfung bestehender Prüfungsaufgaben.
- (2) Erstellung von Prüfungsaufgaben:** In diesem Schritt werden adäquate Aufgabentypen zur Abbildung der zuvor definierten Lernergebnisse identifiziert. Die Aufgaben sollen die in den Lernergebnissen beschriebenen Anforderungssituationen möglichst gut abbilden, um Rückschlüsse auf das vorab modellierte Konstrukt zuzulassen. Die Grenzen existierender Prüfungsformate müssen dabei unbedingt mitberücksichtigt werden, da sie unter anderem mitbestimmen, welches Modell den zu bestimmenden Konstrukten zugrunde gelegt werden muss, ob also etwa eine Kompetenz im Ganzen erfasst oder in Teilaspekten zerlegt werden sollte.
- (3) Rückbindung der Aufgaben an den Prüfungsgegenstand:** Bei der Aufgabenentwicklung werden verschiedene Annahmen darüber getroffen, welche Personenmerkmale durch das Lösen der Aufgabe überprüft werden. Die Überprüfung dieser Annahmen entspricht der inhaltlichen Validierung des Tests. Diese kann durch verschiedene Techniken, wie etwa eine logische und kognitive Aufgabenanalyse, sichergestellt werden. Bezugspunkt bleiben auch hier die entsprechenden Lernergebnisse. Problematische Aufgaben sollten in diesem Schritt überarbeitet werden.
- (4) Bewertung von Prüfungsantworten:** Parallel zur Erstellung und inhaltlichen Validierung der Aufgaben sollte ein Bewertungsraster entwickelt werden, das Bewertungskategorien enthält, die den vorab definierten Konstruktbeschreibungen entsprechen. Als Entscheidungshilfe dient ein Regelwerk, das eine eindeutige Zuordnung der Antworten zu den Kategorien erlaubt (z. B. richtig/falsch). Mögliche Techniken zur Qualitätssicherung beinhalten hier die Formulierung von Ankerbeispielen oder die entsprechende Schulung von Korrektoren, um subjektive Verzerrungen bei der Bewertung möglichst gering zu halten.

- (5) Punkte- und Notenvergabe:** Die Punkteverteilung auf die Aufgaben sollte im Regelfall gewichtet erfolgen. Die inhaltliche Gewichtung der Aufgaben durch Punkte kann dabei entweder theoretisch oder empirisch erfolgen. Sie ist immer dann notwendig, wenn bestimmte Antwortaspekte in Bezug auf die überprüften Konstrukte mehr ins Gewicht fallen als andere. So könnten bei einer Vergaberegeln im Stile von „1 Punkt pro richtige Antwort“ periphere Anforderungen in der Endnote ungewollt stärker zu Gewicht fallen als zentrale Anforderungen.
- (6) Rückschlüsse aus der Empirie:** Die gewonnenen Prüfungsdaten lassen sich nicht nur Rückschlüsse auf die Fähigkeiten von Studierenden zu, sondern auch auf die Qualität der Prüfung oder die einzelnen Aufgaben. Einfach zu bestimmende, statistische Kennwerte wie Minima und Maxima, Trennschärfekoeffizienten und die Aufgabenschwierigkeit, als die mittlere bei einer Aufgabe erreichte Punktzahl, liefern hier bereits eine wichtige Rückmeldung über die Prüfungs- und Aufgabenqualität. Zur Überprüfung der zugrundeliegenden Kompetenzmodelle kann außerdem eine ausführliche psychometrische Analyse durchgeführt werden.

Die Prüfungsgestaltung ist dabei kein linearer Prozess, bei dem die einzelnen Schritte streng nacheinander durchlaufen werden, sondern erfordert eine permanente Anpassung der vorangegangenen Schritte aufgrund von Erkenntnissen aus den aktuellen Herausforderungen. Zum Beispiel kann erst aus der Limitierung des Prüfungsrahmens oder der Aufgabenformate ersichtlich werden, dass ein angestrebtes Lernergebnis möglicherweise zu komplex definiert ist und deshalb noch einmal präzisiert oder in Teilziele zerlegt werden muss. Die Prüfungserstellung kann daher auch Prozess betrachtet werden, der bei den jeweiligen Schritten Rückkopplungsschleifen vorsieht.

Das theoretische Modell bildet in gewisser Weise Standards ab, die bei der Prüfungserstellung beachtet werden sollten. Die einzelnen Schritte lassen sich in den Ablauf eines Semesters – von der Veranstaltungsplanung bis hin zur Prüfungsbewertung – integrieren. Gleichzeitig ist damit ein Analysetool gegeben, um die eigene Prüfungspraxis zu analysieren und entsprechend zielgerichtet Veränderungen vorzunehmen.

4. „Herausforderung Prüfen“ – Ein Weg zu besseren Prüfungen?

4.1 Nachhaltige Qualitätsentwicklung

Die bedeutende Rolle von Prüfungen für Lehr-Lern-Prozesse und eine an Hochschulen meist unreflektierte Praxis des Prüfens, die aber durch Bologna zunehmend mit Qualitätsstandards konfrontiert wird – unter diesen Vorzeichen bot sich ein Unterstützungsangebot an, das den vielfältigen Anforderungen der akademischen Lehrpraxis gerecht wird. Wichtige Grundprinzipien sollten daher Praxisnähe und ein unmittelbarer Nutzen für die Lehrenden, aber auch die Etablierung und Sicherung wissenschaftlicher Standards für Prüfungen an der TU München über das Programm hinaus sein. Zu diesem Zweck wurde einerseits ein theoretischer Rahmen erstellt, indem die Prüfungserstellung aus einer theoretischen Perspektive beschrieben und wichtige Gütekriterien in einem Prozessmodell zusammengefasst wurden. Dieses Modell bildete nun die Grundlage für eine Kooperation mit interessierten Lehrenden. Grundsätzlich gilt es zu berücksichtigen, dass das Erstellen von Prüfungen im tertiären Bereich immer eine sehr hohe, oft spezialisierte fachliche Expertise voraussetzt. Deshalb sind kooperative Formate für die Qualitätsentwicklung im Prüfungsbereich absolut notwendig.

Die Lehrenden sollten praxisnah an aktuellen Herausforderungen aus ihrem Prüfungsalltag arbeiten in einen gemeinsamen Dialog über die Prüfungspraxis eintreten. Die Projektverantwortlichen sollten diesen Prozess moderieren und mit theoretischem Input unterstützen. Im Zentrum sollten dabei schriftliche Prüfungen mit hohen Studierendenzahlen stehen, da diese grundsätzlich die größte Reichweite aufweisen. Ziel des Prozesses sollte es sein, die jeweiligen Prüfungen über den Verlauf eines Semesters gemeinsam zu entwickeln und die teilnehmenden Lehrenden und Lehrstühle für kommende Herausforderungen zu stärken. Der praktische Nutzen sollte für die Teilnehmenden unbedingt im Vordergrund stehen. Durch eine wissenschaftliche Begleitung des Projekts sollte gleichzeitig weiteres Wissen über die Prüfungspraxis an der TU München generiert, Bedingungen und Möglichkeiten für Veränderungen derselben nachvollziehbar und beschreibbar zu machen.

Die Teilnehmenden sollten sie am Ende des gemeinsamen Qualitätsentwicklungsprozesses dazu in der Lage sein, (1) Ziele von Lehrveranstaltungen zu modellieren, (2) Aspekte des Wissens und Könnens in Anforderungen von Prüfungsaufgaben zu übersetzen, (3) die Prüfungsaufgaben auf ihre inhaltliche Validität zu überprüfen, (4) geeignete Korrektur und Auswertungsverfahren anzuwenden sowie (5) Punkte und Noten nach testtheoretischen Grundsätzen zu vergeben. Abschließend sollten sie (6) die nach Durchführung der Prüfung gewonnenen Daten in Bezug auf die Qualität der Prüfung analysieren die vorab modellierten Lernergebnisse auf dieser Grundlage bewerten können.

Im Sinne der Nachhaltigkeit der Maßnahme sollte im zweiten Semester der eingetübte Prozess eigenverantwortlich in die Praxis an den Lehrstühlen übernommen werden, idealerweise auch durch eine Weitergabe des erworbenen Wissens an Kolleginnen und Kollegen. Vorschläge, wie eine solche Implementation gelingen kann, sollten deshalb ebenfalls während des Teilnahmesemesters diskutiert werden.

4.2 Teilnehmerinnen und Teilnehmer und Ablauf des Programms

Teilnehmer

Eine Einladung zur Teilnahme am Projekt wurde im Rahmen eines Treffens der Dekane der TUM ausgesprochen. Grundlage war ein Informationsschreiben, das an Kolleginnen und Kollegen und Kollegen innerhalb der Fakultäten weitergereicht werden sollte und in dem Ziele, Inhalte und der Ablauf des Programms näher beschrieben waren. Herausforderung Prüfen stieß in der Folge vor allem an der Fakultät für Maschinenwesen und Mathematik auf Interesse. Insgesamt beteiligten sich fünf Lehrstühle aus dem Maschinenwesen drei Lehrstühle aus der Mathematik. Bei einer Gesamtteilnehmerzahl von 14 Teilnehmern nahmen pro Lehrstuhl 1 bis 4 Lehrende teil. Unter den Teilnehmerinnen und Teilnehmern waren sowohl Professorinnen und Professoren, als auch wissenschaftliche Mitarbeiter mit langer Prüfungserfahrung und Doktoranden mit verhältnismäßig geringer Erfahrung im Erstellen und Auswerten von Prüfungen. Um der kooperativen Ausrichtung des Projekts zu entsprechen, wurden die Lehrstühle bei entsprechender Nähe zu Teams zusammengestellt und die Termine für die Arbeitstreffen gemeinsam vereinbart.

Ablauf

Der Ablauf von „Herausforderung Prüfen“ richtete sich an den in Kapitel 3.2.1 vorgestellten Schritten bei der Prüfungserstellung. Die darin enthaltenen Themen wurden auf insgesamt vier Arbeitstreffen aufgeteilt. In den Treffen wurden bestehende Prüfungen aus dem aktuellen Semester bearbeitet. Entsprechend fanden zwei der Termine zeitlich vor den jeweiligen Prüfungsterminen statt. Ziel dieser Vorbereitung war die Überprüfung und Definition von angestrebten Lernergebnissen und die Arbeit an bestehenden Prüfungsaufgaben vor dem Hintergrund der überarbeiteten Lernergebnisse. Zwischen dem zweiten und dem dritten Termin wurden die Prüfungen schließlich abgehalten. Entsprechend standen beim dritten Arbeitstreffen die Auseinandersetzung mit kriterienorientierten Bewertungsrastern sowie der Vergabe von Punkten und der Benotung im Fokus. Beim abschließenden Termin sollte die Korrektur der Prüfungen abgeschlossen sein und Rückschlüsse aus den gewonnenen Daten gezogen werden.

Die Arbeitstreffen selbst waren in zwei verschiedene Phasen unterteilt: eine Reflexionsphase und eine Inputphase. Die Zeit zwischen den Terminen sollte für die Umsetzung der neuen Inhalte genutzt werden. Die anschließende Reflexion sollte entsprechend auf die praktischen Erfahrungen der Teilnehmenden im Umgang mit den Inhalten aus dem vorangegangenen Arbeitstreffen Bezug nehmen. Sie fand deshalb erst ab dem zweiten Arbeitstreffen statt und bot eine Möglichkeit, an bestehende Problemstellungen der Teilnehmerinnen und Teilnehmer anzuknüpfen und gemeinsam nach Lösungen zu suchen. In der Inputphase wurden jeweils im Anschluss neue Themen von den Moderatorinnen und Moderatoren vorgestellt und gemeinsam auf bestehende Problematiken angewandt. Zum Abschluss jedes Treffens wurden Zielvereinbarungen für die Praxisphase bis zum nächsten Termin getroffen. Abbildung 2 verdeutlicht den idealtypischen Ablauf des Programms.

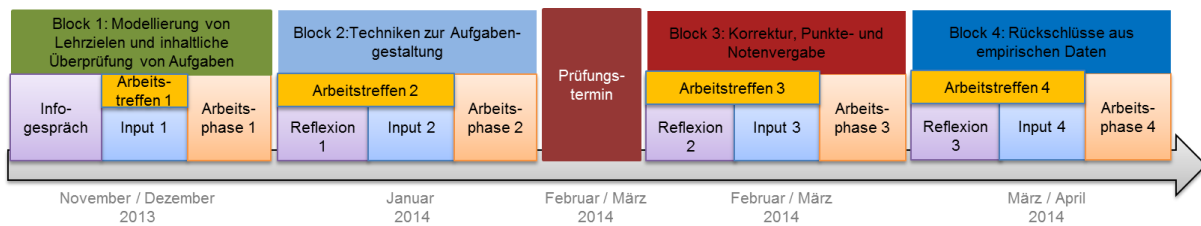


Abbildung 2: Struktur und Ablauf

4. Wissenschaftliche Begleitforschung

Nach der Beschreibung des Ablaufs von „Herausforderung Prüfen“ stehen in den kommenden Kapiteln erste Ergebnisse aus der Begleitforschung im Fokus. Dabei sollen zu Beginn zentrale Fragestellungen skizziert werden, die sich aus der Zielsetzung des Programms und der gewählten Vorgehensweise ergeben und bereits im Projektantrag aufgeworfen wurden. Anschließend werden das zur Beantwortung der Fragen herangezogenen Forschungsdesign skizziert und kurz begründet, die Stichprobe beschrieben und die verwendeten Erhebungsinstrumente und Auswertungsverfahren vorgestellt.

4.1 Fragestellungen

Die wissenschaftlichen Fragestellungen lassen sich in zwei Dimensionen aufgliedern. Die erste Fragegruppe bezieht sich auf die Akzeptanz des Programms in seiner Ausrichtung als Qualitätsentwicklung, die Selbstwirksamkeit der Teilnehmenden im Sinne der wahrgenommenen Kompetenzentwicklung und den durch das Projekt entstandenen Zeitaufwand (Fragen 1-5). Diese Aspekte lassen sich auch als Indikatoren für die Angemessenheit des gewählten Zugangs begreifen. Sie bauen auf den eingangs angestellten Überlegungen zu den Besonderheiten in der Hochschulpraxis auf und überprüfen, inwieweit eine praxisnahe Qualitätsentwicklung ein wirksames Instrument für Veränderung ist. Die zweite Dimension befasst sich konkreter mit den Produkten des Prozesses im Sinne einer veränderten Qualität der Prüfungen und dem veränderten Bewusstsein der Teilnehmenden für den Prüfungsprozess (Fragen 6-9).

1. Wie schätzen die Lehrenden ihre eigene Kompetenzentwicklung durch das Programm ein?
2. Als wie passend empfanden die Teilnehmenden das Programm vor dem Hintergrund persönlicher Herausforderungen?
3. Wie empfanden die Teilnehmenden den zeitlichen Mehraufwand durch das Programm?
4. Wurde das Programm durch die Verantwortlichen professionell moderiert?
5. Wie zufrieden sind die Lehrenden mit der Kooperationskomponente?
6. Welche Produkte wurden von den Lehrenden erstellt? Entstanden im Verlauf des Programms neue reliable und zielvalide Aufgabentypen?

7. Veränderten Dozierende im Verlauf des Workshops ihr Wissen und ihre Annahmen hinsichtlich der Prüfungskonzeption von implizit nach explizit?
8. Wurden durch die Lehrenden grundlegende Standards für die Aufgabenentwicklung und die Aufgabenauswertung entwickelt und für eine Qualitätssicherung nutzbar gemacht?
9. Wurden diese Standards von den Lehrenden in ihrem Handeln routinisiert?
10. Wie ist die Akzeptanz der Studierenden gegenüber den neuen Aufgabentypen?

4.2 Forschungsdesign

Für die Beantwortung der aufgeworfenen Forschungsfragen wurden Formen einer *formativen und einer summativen Evaluation* durchgeführt.

Bei der *formativen Evaluation* handelt es sich um eine prozessbegleitende Evaluation. Die Ergebnisse werden dabei im laufenden Prozess für ein Feedback genutzt, durch das bei den Beteiligten Veränderungen ausgelöst werden sollen, sei es beispielsweise in Bezug auf die Wahrnehmung der eigenen Kompetenzentwicklung oder die Wertschätzung des zeitlichen Aufwands durch die Teilnahme. Dieser Aspekt wurde bei „Herausforderung Prüfen“ vor allem in Form von informellem Feedback durch die Moderatorinnen und Moderatoren im laufenden Prozess realisiert. Dieses wurde auf Basis von im Prozess anfallenden Daten (z.B. Prüfungsaufgaben) oder systematisch miterhobenen Daten (z.B. aufgezeichnete Gespräche) gegeben. Gegenstände formativer Evaluation waren sowohl Aspekte des Programms selbst als auch der inhaltlichen Qualität von Prüfungen. Mit dem Feedback sollte unter anderem der Reflexionsprozess unterstützt und das Wissen gefestigt werden.

Unter einer *summativen Evaluation* versteht man dagegen die abschließende Bewertung eines Prozesses anhand von erhobenen Daten. Sie überprüft die Wirksamkeit von Maßnahmen und Prozessen, greift dabei aber nicht in den laufenden Prozess ein. Auch die summative Evaluation kann für ein abschließendes Feedback genutzt werden, wird aber nicht primär zu diesem Zwecke durchgeführt. Hierbei waren dieselben Aspekte von Bedeutung wie auch schon bei der formativen Evaluation.

Berichtet werden im vorliegenden Projektbericht vor allem Ergebnisse aus der summativen Evaluation des Programms in Bezug auf die beiden skizzierten Fragedimensionen. Die Erhebung der Daten wurde die mit entsprechend standardisierten Instrumenten durchgeführt. Über die Feedback-Prozesse in den Arbeitstreffen existieren zusätzlich Verbaldaten, die im Zuge des Dissertationsvorhabens von Christoph Schindler (2015, in Vorbereitung) ausgewertet werden. Tendenzen aus diesen Daten werden in Kapitel 5.2 dennoch angesprochen.

Das Programm selbst wurde fallbasiert durchgeführt. Die Teilnehmenden, deren Prüfung oder Fachbereich eine inhaltliche Nähe aufweist, wurden einer Arbeitsgruppe zugeteilt. Die Evaluation wurde dagegen personenbezogen durchgeführt, das heißt, jeder Teilnehmer erhielt einen eigenen Evaluationsbogen.

4.3 Stichprobe und Rücklaufquote

Das Projekt wurde 2013 und 2014 über zwei Kohorten angeboten. Entsprechend liegen Daten aus zwei Stichproben vor. Die erste Kohorte umfasste insgesamt 15 Dozierende, darunter 14 Dozenten und eine Dozentin. Die Teilnehmenden kamen von sieben Lehrstühlen aus den drei Fakultäten Maschinenwesen, Mathematik und der School of Education. Vier der Lehrstühle sind an der Fakultät für Maschinenwesen angesiedelt, zwei an der Fakultät für Mathematik und einer an der School of Education.

Für die zweite Kohorte, bestehend aus acht Dozierenden, von vier Lehrstühlen, aus drei Fakultäten, liegen derzeit noch keine vollständigen Daten vor. Auf eine Darstellung dieser Kohorte wird daher vorerst verzichtet.

4.4 Erhebungsinstrumente

Die Beantwortung der Fragen zur Akzeptanz des Projektes und der Selbstwahrnehmung der Dozierenden wurde mittels eines Evaluationsfragebogens in Anlehnung an Ostermeier, Prenzel und Duit (2010) angestrebt (Vgl. Anhang A). Dieser umfasst Fragen zur Einschätzung der Moderation der Arbeitstreffen durch die Projektverantwortlichen, nach der Wahrnehmung der eigenen Kompetenzentwicklung der Lehrenden, die Moderation der Arbeitstreffen, der Nützlichkeit des zusätzlichen Arbeitsmaterials, das zu den einzelnen Themenkomplexen bereitgestellt wurde, sowie nach dem zeitlichen (Mehr-)Aufwand durch die Teilnahme am Projekt.

Zusätzlich zu den standardisierten Items wurden dem Fragebogen offene Fragefelder hinzugefügt. Hierbei sollten die Teilnehmenden kurz ihre „Hauptkenntnisse aus dem Projekt“, Anmerkungen zur „Zufriedenheit hinsichtlich der Inhalte und Techniken“, zur „Umsetzung der Inhalte und Techniken durch die Lehrenden“, zu Grenzen im Sinne von „Inhalten und Techniken, die nicht integriert werden konnten“, ergänzend um ein Feld zu einer möglichen Begründung dieser Grenzen. Diese Items vertieften die entsprechenden standardisierten Fragen zur Akzeptanz des Programms.

Ebenso wurde mit einer offenen Frage erhoben, inwieweit die Kooperationskomponente umgesetzt werden konnte. Hinsichtlich der Funktion von Prüfungen als Stellhebel für die Lehre wurde erfragt, inwieweit durch die Arbeit an den Prüfungen auch die dazugehörige Lehrveranstaltung verändert wurde. Außerdem sollten abschließend positive Aspekte und Verbesserungsvorschläge genannt werden. Auch der Wunsch nach bestimmten Informationen konnte abschließend geäußert werden.

Für die Überprüfung der Ergebnisse des Projekts hinsichtlich von Aspekten der Prüfungsqualität kam ein erweitertes Instrumentarium zum Einsatz. Hierbei wurden unter anderem angestrebte Lernergebnisse, Modulbeschreibungen und Prüfungen gesammelt, Verbaldaten aufgezeichnet und neu erstellte Lernergebnis-Taxonomien untersucht. Im vorliegenden Projektbericht werden zentrale Ergebnisse zur Beantwortung der Fragen zur Prüfungsqualität vorgestellt (zu einer ausführlichen Darstellung siehe Schindler 2015, in Vorbereitung).

4.5 Auswertungsverfahren

Die via Fragebogen erhaltenen Daten wurden anonymisiert und deskriptiv-statistisch ausgewertet. Als Kennwerte wurden arithmetische Mittel, Standardabweichungen und Minimal- und Maximalwerte bestimmt. Die Darstellung erfolgt mittels Item-Tabellen. Antworten aus den offenen Fragefeldern wurden ausgewertet und sinngemäß zusammengefasst. In diesem Schritt wurden die Angaben soweit anonymisiert, dass eine Zuordnung zu einem bestimmten Fachgebiet bzw. Lehrstuhl nicht möglich ist.

Für die Analyse der Entwicklung der Prüfungsqualität kamen je nach Datengrundlage verschiedene inhaltsanalytische Verfahren mit unterschiedlichen Bewertungsrastern zum Einsatz. Bei den Verbaldaten wurde eine induktive Analyseverfahren gewählt, um das Material im Sinne der Schwerpunktsetzung der Teilnehmenden auszuwerten.

5. Ergebnisse

5.1 Akzeptanz des Projektes und wahrgenommene Kompetenzentwicklung

1. Wie schätzen die Lehrenden ihre eigene Kompetenzentwicklung hinsichtlich der Programmziele ein?

Auf die wahrgenommene Kompetenzentwicklung durch die Teilnehmenden entfielen insgesamt 3 Items im Fragebogen. Die Fragen bezogen sich auf eine globale Einschätzung hinsichtlich der Verbesserung der Prüfungsqualität („Mit den in den Arbeitstreffen angesprochenen Inhalten und vermittelten Inhalten konnte ich meine Prüfung verbessern.“), einer Frage nach der persönlichen Sicherheit in Bezug auf Prüfungsfragen durch die Mitarbeit im Projekt („Die Arbeit im Prüfungsprojekt trug dazu bei, dass ich mich bei der Prüfungserstellung sicherer fühle.“) sowie einer Frage nach der Ausrichtung an den vermittelten wissenschaftlichen Standards („Durch die Arbeit im Prüfungsprojekt achte ich stärker als zuvor darauf bestimmte Standards einzuhalten.“) Tabelle 1 zeigt statistische Kennwerte zur wahrgenommenen Kompetenzentwicklung.

Tabelle 1: Wahrgenommene Kompetenzentwicklung

Item	Kompetenzentwicklung	Kennwerte			
		M	SD	Min	Max
b)	Die Arbeitstreffen ließen mich die Stärken und Schwächen in meiner Prüfung besser erkennen.	3.00	0.00	3	3
c)	„Mit den in den Arbeitstreffen angesprochenen Inhalten und vermittelten Inhalten konnte ich meine Prüfung verbessern.“	2.87	0.35	2	3
i)	„Die Arbeit im Prüfungsprojekt trug dazu bei, dass ich mich bei der Prüfungserstellung sicherer fühle.“	2.75	0.46	2	3
j)	„Durch die Arbeit im Prüfungsprojekt achte ich stärker als zuvor darauf bestimmte Standards einzuhalten.“	2.75	0.46	2	3

N = 8 Hochschullehrende; 0= trifft nicht zu; 1= trifft eher nicht zu; 2 = trifft eher zu; 3=trifft zu

Die Lehrenden haben das Programm durchweg als hoch effizient empfunden. Differenziert betrachtet wird sowohl von einem besseren Produkt (=die eigene Prüfung), als auch einer deutlich gestiegenen Sicherheit bei der Prüfungserstellung und einem gestärkten Bewusstsein für wissenschaftliche Standards berichtet. Dabei gibt keiner der Teilnehmenden an, hinsichtlich der eigenen Kompetenzentwicklung nicht von dem Programm profitiert zu haben. Diese Entwicklung fand auf der wichtigen Grundlage statt, dass die Teilnehmenden die Stärken und Schwächen besser erkennen konnten.

2. Als wie passend empfanden die Teilnehmenden das Programm vor dem Hintergrund persönlicher Herausforderungen?

Zu Beginn des Arbeitsprozesses wurden gemeinsam mit den Teilnehmenden individuelle Problembereiche identifiziert. Diese wurden am Ende des Treffens schriftlich festgehalten, um den darauffolgenden Arbeitsprozess entlang dieser individuellen Anforderungen zu gewichten. Die kommunizierten Schwerpunkte wurden anschließend den einzelnen Schritten bei der Prüfungserstellung zugeordnet. Diese Zuordnung erfolgte wissenschaftlich kontrolliert, indem zwei Rater unabhängig voneinander die entsprechenden Ziele den theoretischen Anforderungsbereichen zuordneten. Anschließend wurde die Übereinstimmung mittels Übereinstimmungsmaß für die Interrater-Reliabilität (Cohens Kappa) bestimmt⁸. An dieser Stelle werden die zentralen Ergebnisse dieser Zuordnung nach Häufigkeiten der Problemstellung kurz tabellarisch dargestellt (Vgl. Tabelle 2).

Tabelle 2: Anforderungen und Erwartungen gegenüber dem Projekt

Anforderungsbereich	Arbeitsschritt	Häufigkeit
Prüfungserstellung	Festlegung des Prüfungsgegenstands	6
	Erstellung von Prüfungsaufgaben	10
	Rückbindung der Aufgaben an den Prüfungsgegenstand	4
Prüfungsauswertung	Bewertung von Prüfungsantworten	8
	Punkte- und Notenvergabe	0
	Rückschlüsse aus der Empirie	2

Drei der vier Teams gaben dabei an, Probleme bei der Gestaltung von Prüfungen zu haben, die auf die Überprüfung von konzeptionellem Verständnis statt Faktenwissen ausgerichtet sind. Dabei waren sich die Lehrenden durchaus bewusst, dass die angestrebten Lernergebnisse in der Regel bereits auf konzeptuelles Verständnis abzielten. Das Problem wurde vor allem darin benannt, diese Anforderungen in adäquate Aufgaben zu übersetzen.

In diesem Zusammenhang betonten die Teilnehmenden auch das Bedürfnis, neue Aufgabentypen kennenzulernen. Die bestehenden Aufgaben wurden als zu einförmig beschrieben. Entsprechend sollten im Laufe des Arbeitsprozesses auch die Potenziale von Multiple Choice-Aufgaben und Aufgaben, die die Fähigkeit zum Wissenstransfer überprüfen können, diskutiert werden.

⁸ Werte über 0.7 wurden als zuverlässig betrachtet. Bei Werten zwischen 0.5 und 0.7 fand eine Konsensvalidierung der Zuordnung statt. Items mit einer Interrater-Reliabilität von unter 0.5 wurden aus der Analyse ausgeschlossen. Die Ergebnisse waren Teil eines Vortrags auf der AERA-Konferenz in Chicago, 2015.

Schließlich wurden von Teilnehmenden aus zwei Teams auch Probleme bezüglich formeller Kriterien benannt, wie etwa die Klarheit von Aufgabenstellungen und Leistungserwartungen. Die Teilnehmenden in diesen Teams hatten im Vorfeld noch wenig Erfahrung mit der Erstellung von Prüfungen. Hier wurden auch insbesondere Probleme beim Einhalten von Kohärenzprinzipien bei mehrstufigen Aufgaben geschildert.

Die entsprechenden Fragen im Evaluationsfragebogen zielten darauf ab, inwieweit auf die thematisierten Anforderungen im Laufe des Projekts Bezug vor dem Horizont eigenen Vorwissens genommen wurde und wie praktikabel sich die Lösungen gestalteten. Die Ergebnisse sind in Tabelle 3 dargestellt.

Tabelle 3: Zufriedenheit mit den Inhalten vor dem Hintergrund persönlicher Anforderungen

Item	Arbeitstreffen	Kennwerte			
		AM	SD	Min	Max
a)	In den Arbeitstreffen wurden die für meine Prüfung relevanten Inhalte und Probleme angesprochen.	3.00	0.00	3	3
d)	Die behandelten Inhalte knüpften an meinen Wissensstand an.	2.75	0.46	2	3
e)	Die Programminhalte waren so gestaltet, dass sie in meine bestehende Praxis integriert werden können.	2.50	0.54	2	3

N = 8 Hochschullehrende; 0= trifft nicht zu; 1= trifft eher nicht zu; 2 = trifft eher zu; 3=trifft zu

Auch hier wird ersichtlich, dass die Teilnehmenden die Angemessenheit des Programms vor dem Hintergrund ihrer persönlichen Anforderungen als durchweg positiv beurteilen. Vor allem der inhaltliche Problembezug zur jeweiligen Prüfung konnte scheinbar in vollem Umfang hergestellt werden. Dabei wurde der Wissensstand der Teilnehmenden berücksichtigt. Auch das Prinzip der Umsetzbarkeit scheint erfüllt worden zu sein: Die Teilnehmenden berichten, dass die Programminhalte aus ihrer Sicht praxistauglich waren.

3. Wie empfanden die Teilnehmenden den zeitlichen Mehraufwand durch das Programm?

Besonders wichtig für die Akzeptanz des Programms war außerdem die wahrgenommene zeitliche Mehrbelastung durch das Projekt, da gerade Lehrende an Hochschulen oftmals zwischen verschiedensten Ansprüchen aus Lehre und Forschung vermitteln müssen und entsprechende Maßnahmen vor allem auch zeiteffizient sein müssen. Konkrete Zahlen über die zeitliche Investition durch die Teilnehmer liegen nicht vor, da vor allem die Arbeitsphasen zwischen den Treffen zeitlich nicht erfasst wurden.

So fanden mit den Teams an den Fakultäten für Mathematik und an der TUM School of Education zwei Arbeitstreffen mit einer Dauer von drei Stunden statt und an der Fakultät Maschinenwesen insgesamt vier bis fünf Arbeitstreffen mit einer Dauer von jeweils zwei Stunden. Innerhalb der Arbeitstreffen wandten die Lehrenden insgesamt sechs bis zehn Stunden dafür auf, um gemeinsam mit ihren Kollegen an ihren Prüfungsmaterialien zu arbeiten. Darüber hinaus gaben die Dozierenden aber auch an, sich auch über die Arbeitstreffen hinaus mit ihren Kollegen über die in den Arbeitstreffen behandelten Inhalten ausgetauscht zu haben. Auch dieser Austausch bedeutete sicherlich einen zeitlichen Aufwand, wurde allerdings von den Beteiligten als durchweg lohnend bewertet (siehe Ergebnisse zu Fragestellung 5).

Bei den standardisierten Fragen wurde ein besonderer Fokus auf das Aufwand-Nutzen-Verhältnis gelegt. Ebenso abgefragt wurde die Wahrnehmung des absoluten Mehraufwands durch das Projekt. Abschließend wurde erhoben, inwieweit die Teilnehmenden sich aus den Maßnahmen eine längerfristige Zeitersparnis aus einer systematischeren Prüfungspraxis erwarten. Tabelle 3 gibt eine Übersicht über die Rückmeldung zur Wahrnehmung des zeitlichen Mehraufwands.

Tabelle 4: Wahrnehmung des Zeitaufwands durch die Teilnahme

Item	Zeitaufwand	Kennwerte			
		AM	SD	Min	Max
f)	„Der Aufwand für die Arbeit im Prüfungsprojekt stand in einem angemessenen Verhältnis zum Nutzen.“	2.25	0.46	2	3
g)	„Das vorgeschlagene Vorgehen bei der Prüfungserstellung führte bei der Umsetzung zu einem deutlich erhöhten Zeitaufwand.“	1.38	0.74	0	2
h)	„Die Arbeit im Prüfungsprojekt wird langfristig zu einer Reduzierung des Zeitaufwandes bei der Prüfungserstellung führen.“	1.25	0.46	1	2

N = 8 Hochschullehrende; 0= trifft nicht zu; 1= trifft eher nicht zu; 2 = trifft eher zu; 3=trifft zu

Die Lehrenden gaben an, dass der Aufwand für die Arbeit im Prüfungsprojekt in einem angemessenen Verhältnis zum Nutzen steht. Das in den Arbeitstreffen vorgeschlagene Vorgehen bei der Prüfungserstellung führt ihren Einschätzungen nach in der Umsetzung zu keinem deutlich erhöhten Zeitaufwand, auch wenn vereinzelt von einem erhöhten Zeitaufwand berichtet wurde (Vgl. Maximalwert). Allerdings wird die Teilnahme im Prüfungsprojekt nach Einschätzung der Dozierenden langfristig zu keiner zeitlichen Entlastung bei der Prüfungserstellung führen. Diese Einschränkung sollte abschließend diskutiert noch einmal diskutiert werden.

4. Wurde das Programm durch die Verantwortlichen professionell moderiert?

Zu beurteilen war die Professionalität bei der Begleitung durch die Projektverantwortlichen. Aspekte dieser Professionalität bezogen sich auf die Vorbereitung der Inhalte, das konstruktive Feedback und den Umgang mit Fragen. Auch das Zeitmanagement während der Arbeitstreffen, die Problemfokussierung, die Kommunikation zentraler Konzepte und Begriffe sowie der Anspruch an Vorwissen sollten beurteilt werden. Die Ergebnisse sind in Tabelle 4 zusammengefasst.

Tabelle 5: Wahrgenommene Professionalität bei der Begleitung durch die Projektverantwortlichen

Item	Die ModeratorInnen (MitarbeiterInnen des Projekts)....	Kennwerte			
		AM	SD	Min	Max
k)	waren inhaltlich gut vorbereitet.	3.00	0.00	3	3
l)	gaben mir konstruktive Rückmeldungen zu meinen Prüfungen.	3.00	0.00	3	3
m)	konnten auf meine Fragen angemessen eingehen.	3.00	0.00	3	3
n)	nutzten die gemeinsame Arbeitszeit in den Treffen effektiv aus.	2.63	0.52	2	3
o)	machten auf zentrale Problembereiche aufmerksam.	3.00	0.00	3	3
p)	erklärten zentrale Konzepte und Begriffe verständlich.	3.00	0.00	3	3
q)	setzten ein zu hohes Vorwissen voraus.	0.25	0.46	0	1

N = 8 Hochschullehrende; 0= trifft nicht zu; 1= trifft eher nicht zu; 2 = trifft eher zu; 3=trifft zu

Die Projektverantwortlichen wurden durchweg als professionelle Begleiter in Bezug auf die genannten Aspekte wahrgenommen.

Im weitesten Sinne sind auch die zur Verfügung gestellten Arbeits- und Informationsmaterialien als wichtiger Bestandteil professioneller Begleitung zu betrachten. Sie spielen dabei nicht nur im laufenden Arbeitsprozess eine Rolle, sondern können auch anleitend für eigenverantwortliche Schritte bei der Erstellung zukünftiger Prüfungen oder in der Weitergabe des erworbenen Wissens an Kolleginnen und Kollegen sein. Tabelle 5 gibt das Feedback bezüglich der Qualität und des Nutzens der verwendeten Arbeitsmaterialien wieder.

Tabelle 6: Zufriedenheit der Teilnehmenden mit den Informationsmaterialien

Item	Die Informationsmaterialien...	Kennwerte			
		AM	SD	Min	Max
r)	sind verständlich und nachvollziehbar.	2.75	0.46	2	3
s)	enthalten anschauliche Beispiele.	2.75	0.46	2	3
t)	haben eine angemessene Länge.	2.38	0.52	2	3
u)	tragen zu einem besseren Verständnis der behandelten Inhalte bei.	2.87	0.35	2	3
v)	sind für mich ein wichtiges Nachschlagewerk	2.38	0.52	2	3
w)	enthalten die für mich relevanten Informationen.	2.63	0.74	1	3

N = 8 Hochschullehrende; 0= trifft nicht zu; 1= trifft eher nicht zu; 2 = trifft eher zu; 3=trifft zu

Die Informationsmaterialien und Handreichungen sind für die Teilnehmenden verständlich und nachvollziehbar. Sie enthalten in hohem Ausmaß anschauliche Beispiele und tragen nach Ansicht der Lehrenden zu einem besseren Verständnis der behandelten Inhalte bei. Außerdem enthalten sie für die Mehrzahl an Dozierenden relevante Informationen, wobei hier neben der Zustimmung auch ein Minimalwert von 1 („trifft eher nicht zu“) berichtet werden muss. Der Umfang der Handreichungen scheint dabei für alle Teilnehmenden angemessen gewesen zu sein. Eine wichtige Information bezüglich der Nachhaltigkeit der Maßnahme ist außerdem, dass das zur Verfügung gestellte Material von allen Teilnehmenden als wichtiges Nachschlagewerk angesehen wird.

5. Wie zufrieden sind die Lehrenden mit der Kooperationskomponente?

Ein wichtiger Punkt für die Wirksamkeit des Projektes mit dem Ziel einer nachhaltig verbesserten Prüfungspraxis sollte die Kooperation unter den Teilnehmenden sein – sowohl während der Arbeitstreffen, als auch längerfristig innerhalb der Lehrstühle und Fakultäten. Ausgewertet wurden für die folgenden Ergebnisse die Daten eines offenen Items im Fragebogen. Dabei wurden sinnähnliche Aussagen thematisch zusammengefasst.

Eine wichtige Rolle bei der Kooperation spielt für die Befragten ganz generell – nicht nur hinsichtlich der abgelaufenen Arbeitstreffen - die *fachliche Qualifikation der Kolleginnen und Kollegen*. Dieser Aspekt wurde zweimal betont. Ein Teilnehmer betont dabei vor allem die Inanspruchnahme von Hilfe: Die „Kollegen [sollten] auch fachlich qualifiziert [sein], um Hilfestellungen geben zu können.“ Ein betont ebenso die Bedeutung fachlicher Qualifikation, ergänzt aber, dass diese „zeitaufwendig“ sei. Die „Einarbeitung erfordert Zeit“.

Ein weiterer Kommentar berichtet davon, dass „das Programm den Austausch in dem Lehrerteam angeregt [habe]“. Man habe sich über die Lehrveranstaltungen ausgetauscht. Als Grenze wurde gleichzeitig genannt, dass ein Austausch „über die Grenze des Lehrerteams [...] jedoch nicht initiiert [wurde].“

Als Bedingung für gute Kooperation wird außerdem die fachliche Nähe herausgestellt. Diese wird als „Ähnlichkeit in der „Art“ der Prüfung“ näher benannt. Ähnlichkeit hinsichtlich des fachlichen Hintergrunds wird auch in einem weiteren Kommentar betont, wonach vor allem „Erfahrungen von Kollegen eines anderen Lehrstuhls der gleichen Fakultät“ wertvoll gewesen seien.

Über die Rückmeldungen aus dem Fragebogen hinaus liegt Evidenz für eine gelungene Initiierung von Zusammenarbeit zwischen verschiedenen Lehrstühlen vor. Ein Lehrstuhl führte zum Beispiel eine gesonderte Schulung für alle an der im Projekt behandelten Prüfung beteiligten Mitarbeiterinnen und Mitarbeiter durch, im Rahmen der noch einmal für Aspekte der Aufgabengestaltung und Korrektur sensibilisiert wurde. Zwei Lehrstühle im Maschinenwesen führen mittlerweile eine, einmal im Semester stattfindende, interne Schulung für das gesamte Lehrpersonal am Lehrstuhl durch, in der erlerntes Wissen über und Fähigkeiten bei der Prüfungserstellung eigenverantwortlich weitervermittelt werden. Schwerpunkte setzt der Lehrstuhl dabei auf die Formulierung von erwarteten Lernzielen in Modulbeschreibungen, die Aufgabengestaltung nach Kriterien des Programms und die Erarbeitung von Bewertungsrastern. Die fakultätsinterne Schulung findet lehrstuhlübergreifend einmal im Semester an einem Tag im zeitlichen Rahmen von acht Stunden statt.

Insgesamt zeigen die Rückmeldungen, dass eine Kooperation auch schon während der Arbeitstreffen stattgefunden hat. Die Kooperationskomponente wurde dabei grundsätzlich als positiv wahrgenommen. Grundsätzlich negative Rückmeldung zur Zusammenarbeit gab es nicht. Einwände bestehen in der fehlenden fachlichen Qualifikation und dem damit verbundenen erhöhten Zeitaufwand. Diese Einwände sollten für eine weitere Adaption des Programms im Anschluss noch einmal kritisch reflektiert werden.

Fazit:

Das Projekt kann hinsichtlich seiner Akzeptanz bei den Teilnehmenden und seiner Angemessenheit gegenüber den bestehenden Anforderungen als Erfolg gewertet werden. Die Teilnehmenden gaben an, Fortschritte auf dem Gebiet der Prüfungserstellung gemacht zu haben und sahen die eigenen Anforderungen und bestehende Problemstellungen in vollem Umfang berücksichtigt. Der zeitliche Aufwand wurde im Verhältnis zu seinem Nutzen als angemessen bewertet. Gute Noten wurden der Durchführung durch die Projektverantwortlichen bescheinigt. Auch der Fokus auf Kooperation zwischen den Lehrenden wurde als hilfreich wahrgenommen. Teilweise wurden Impulse für die Zusammenarbeit in Form von Weiterbildungen in die Fakultäten getragen.

5.2 Veränderungen in der Prüfungsqualität

Ergebnisse über Veränderungen in der Qualität der Prüfungen werden im folgenden Kapitel exemplarisch berichtet, da die Analysen des entstandenen Materials zum Zeitpunkt der Erstellung des Projektberichts noch nicht vollständig abgeschlossen waren. Vorgestellt werden dabei erste Zwischenergebnisse zu den einzelnen Forschungsfragen. Die Daten beruhen – im Gegensatz den Evaluationsergebnissen - nicht auf den Daten aus dem Evaluationsfragebogen, sondern sind das Produkt von weiterführenden inhaltlichen Analysen der neu entwickelten Prüfungen (zu ausführlichen Ergebnissen, siehe Schindler 2015, in Vorbereitung).

6. *Welche Produkte werden von den Lehrenden erstellt?* Entstanden im Verlauf des Programms neue reliable und zielvalide Aufgabentypen?

Veränderungen bei den angestrebten Lernergebnissen

Erste Veränderungen konnten empirisch sowohl bei **den angestrebten Lernergebnissen** als auch bei den **Prüfungsaufgaben**⁹ festgestellt werden.

Die *Anzahl an angestrebten Lernergebnissen* wurde bei zwei Teams sukzessive um vier bis acht erhöht. Die Gründe für die zusätzlich aufgenommenen Lernziele waren unterschiedlich. In einem Fall wurde ein Lernergebnis beispielsweise ergänzt, da ein relevanter Themenbereich in den Formulierungen nicht abgedeckt war. In anderen Fällen fand eine Differenzierung übergeordneter Kompetenzen in Teilkompetenzen zur besseren Operationalisierung statt. Durch die gemeinsame Diskussion der Lehrziele mit den Kolleginnen und Kollegen konnte in einem anderen Fall eine Strukturierung der Prüfungsinhalte und -anforderungen erreicht werden, mit dem Ergebnis, dass in den

⁹ An dieser Stelle könne keine der neu entstandenen Prüfungsaufgaben veröffentlicht werden, da diese aktuell in der Prüfungspraxis Verwendung finden.

angestrebten Lernergebnissen nun systematisch zwischen Teilkompetenzen differenziert wird. Grundsätzlich wurde vor allem an einer systematischen Differenzierung bestehender Formulierungen gearbeitet, Inhalte und Anforderungen systematisch in Teilkompetenzen zu zerlegen und aufeinander abzustimmen.

Ein weiterer Aspekt war die *Spezifität der angestrebten Lernergebnisse*. Diese müssen möglichst spezifisch in Bezug auf das erwartete Verhalten sein, sodass sich im Idealfall unmittelbar Aufgabenstellungen daraus ableiten lassen. Für die bestehende Praxis konnten vor Beginn der Qualitätsentwicklungsarbeit lediglich 8,3 Prozent der Lehrziele als ausreichend spezifisch eingestuft werden. Dabei erfolgte eine Einteilung in „unspezifisch“, „auf Modulebene“ und „auf Prüfungsebene“. In die letzte Kategorie fielen angestrebte Lernergebnisse, die als spezifisch genug erachtet wurden, um als Grundlage für die Erstellung von Prüfungsaufgaben zu dienen. Im Laufe der Arbeit an den angestrebten Lernergebnissen konnte vor allem der Anteil an „auf Modulebene“ spezifischen Lernergebnissen deutlich gesteigert werden. Eine Arbeitshypothese ist, dass dieses Niveau für die Lehrenden aus den Fachbereichen ausreichend ist, um auf einem fundierten gemeinsamen Verständnis der Begrifflichkeiten Aufgaben zu konzipieren. Ein positives Ergebnis dieses Schrittes ist auf jeden Fall die deutliche Reduktion „unspezifischer“ Formulierungen.

Darüber hinaus wurde der Versuch unternommen, die didaktischen Ideen hinter den bestehenden Lernergebnissen fächerunabhängig greifbar zu machen. Hierzu wurde auf ein domänenunspezifisches, das heißt fächerübergreifendes, Modell zur *Klassifikation von Wissensarten* zurückgegriffen, wonach erlernbares Wissen in deklaratives, prozedurales, konzeptionelles und strategisches Wissen unterteilt werden kann. Vor Beginn der Maßnahme ließen sich hier besonders viele Lernergebnisse auf Basis von Selbsteinschätzungen durch die Lehrenden dem Bereich des konzeptionellen Wissens zuordnen (66,7 %). Dieser Anteil ging durch die Arbeit an den Lernergebnissen deutlich zurück auf 51,4 Prozent zugunsten der Lernergebnisse, die prozedurales Wissen fokussieren (von 20,8 % auf 34,3 %). Eine vorläufige Erklärung hierfür lautet, dass den Lehrenden bewusst wurde, dass eine Kompetenz sich in verschiedene Stufen gliedern lässt und nicht nur das angestrebte Verhalten in Form der Kompetenz ins Auge gefasst werden sollte, sondern auch die notwendigen Teilschritte im Lernprozess, die eben auch deklaratives und prozedurales Grundlagenwissen einschließen. Diese Differenzierung ist einerseits sinnvoll, um Lernschritte besser zu strukturieren und erleichtert gleichzeitig die Auswertung. Vor allem erhöht sie aber die diagnostische Aussagekraft einer Prüfung, da differenzierter nach Kompetenzstufen unterschieden werden kann.

Insgesamt lässt sich an dieser Stelle festhalten, dass die Praxis der Formulierung von Lernergebnissen durch das Projekt deutlich verändert wurde. Nicht nur wurden angestrebte Lernergebnisse deutlicher differenziert und dadurch handhabbarer gemacht; damit einhergehend wurde auch die Spezifität der Lernergebnisse erhöht. Die Klassifikation der Lernergebnisse nach domänenunabhängigen Wissensarten verschärfte bei den Teilnehmenden außerdem das Bewusstsein für ein erweitertes Verständnis von Kompetenzen, zum Beispiel für deren Voraussetzungen im Bereich des deklarativen und prozeduralen Wissens.

Veränderungen bei den Prüfungsaufgaben

Einheitlich wurde der Wunsch geäußert, im Laufe des Programms neue Aufgabentypen kennenzulernen, um das bestehende Repertoire zu erweitern. Themen waren adäquate Aufgaben für Studierende fremder Fakultäten (z.B. Lehramtsstudierende), Fragen der formellen Aufgabengestaltung und Effizienz bei knappen zeitlichen Ressourcen. Ebenso bestand zu Beginn des Programms eine große Diskrepanz zwischen den Wissensaspekten in den angestrebten Lernergebnissen und den Wissensaspekten in den bestehenden Aufgaben. Die Erstellung von Aufgaben erfolgte in der bestehenden Praxis mehr oder weniger willkürlich auf Basis der Veranstaltungsskripte. Bei 30 Prozent der gesichteten Aufgaben bestanden außerdem Abhängigkeiten zwischen einzelnen (Teil-)aufgaben. Ebenso variierten die Aufgabenstellungen kaum.

Die Verteilung nach Wissensarten in den Aufgaben änderte sich über den Verlauf des Programms im Mittel nur geringfügig. Dennoch konnten Änderungen in den einzelnen Fällen festgestellt werden. Zum Beispiel konnten in einer Arbeitsgruppe die Anzahl an Aufgaben, die auf Konzeptwissen beruhten, deutlich gesteigert werden, während der Anteil an reinen „Wissensaufgaben“ mit Faktenwissen zurückging. Lehrende einer Grundlagenveranstaltung schraubten hingegen den Anteil an Aufgaben mit prozeduralen Wissensselementen zurück und erhöhten den Anteil an Aufgaben, die auf Faktenwissen abzielten. Die Veränderungen haben vielfältige Gründe, vor allem aber wurden die Aufgabenanforderungen mit Blick auf die angestrebten Lernergebnisse geschärft und entsprechend angepasst. Die Ergebnisse zeigen, dass die überarbeiteten und neu erstellten Aufgaben eine bessere Passung zu den angestrebten Lernergebnissen aufweisen.

Auf der anderen Seite wurde ein Bewusstsein für Struktur- und Entwicklungsmodelle von Kompetenzen geschaffen. Das heißt, es wurde verstärkt ein Fokus auf die Überprüfung von Voraussetzungen kompetenten Handelns in kleineren Einzelaufgaben gelegt, anstatt einzelne Aufgaben möglichst komplex und realitätsnah zu gestalten und das Format – und damit auch die Studierenden – zu überfordern. Entsprechend konnte auch deutlich gemacht werden, dass Aufgaben im Kurzantwortformat und Kompetenzorientierung kein Widerspruch sind. Gefördert wurde die Ausrichtung auf Kompetenzen in Aufgaben außerdem durch Techniken wie die Öffnung von Aufgaben oder deren Kontextualisierung mit lebensweltlichem Bezug. Während das Mittel der Öffnung von Aufgaben von den Teilnehmenden häufig aufgegriffen wurde, wurden nur unwesentlich mehr Aufgaben kontextualisiert. Die Umsetzung dieser Maßnahmen hing mitunter auch davon ab, wie systematisch die angestrebten Lernergebnisse bei der Überarbeitung zugrunde gelegt worden waren und wie flexibel sowohl Aufgaben als auch angestrebte Lernergebnisse verändert werden konnten. Waren die Prüfungsaufgaben bereits gesetzt, wie es in einer Gruppe der Fall war, waren die Änderungen nicht so weitreichend wie in Fällen, die einen Zugriff auf Lernergebnisse, Veranstaltungscurriculum und Prüfungsgestaltung hatten.

Des Weiteren konnte die Abhängigkeit zwischen (aufeinander aufbauenden) Aufgaben insofern reduziert werden, als darauf geachtet wurde, dass im Zweifelsfall Zwischenergebnisse mit angegeben wurden.

Fazit: Im Laufe des Programms konnte nicht nur ein Bewusstsein für die Wichtigkeit der Passung zwischen angestrebten Lernergebnissen und Prüfungsaufgaben geschaffen werden. Die Teilnehmenden präzisierten außerdem sowohl Ziele als auch Aufgaben vor diesem Hintergrund, sodass am Ende zielvalidere Prüfungen und genauere Vorstellungen darüber standen, was die Studierenden durch die Lehrveranstaltung können sollten. Vor diesem Hintergrund konnten auch wichtige Aspekte einer Orientierung an Kompetenzen erfolgreich umgesetzt werden. Die Aufgaben wurden dadurch in ihren Anforderungen deutlich vielfältiger.

7. Verändern Dozierende im Verlauf des Workshops ihr Wissen und ihre Annahmen hinsichtlich der Prüfungskonzeption von implizit nach explizit?

Explizites Wissen unterscheidet sich von implizitem Wissen dadurch, dass es unter Rückgriff auf (fachliche) Sprache eindeutig kommuniziert werden kann. Auch wenn für Explikationsfähigkeit im Vorfeld des Projekts keine Indikatoren entwickelt wurden, gibt das Feedback der Lehrenden doch Hinweise darauf, inwieweit sich ein explizierbarer Wissensbestand gebildet hat. In einem offenen Feld des Evaluationsbogens sollten die Lehrenden zum Beispiel angeben, was für sie die Hauptkenntnisse aus dem sind. Hierbei wurde vor allem die systematische Verschränkung von Lernzielen und Prüfungsaufgaben mit Lehrinhalten genannt. Die Frage nach hilfreichen Techniken wurde außerdem sehr präzise unter Bezugnahme auf entsprechenden Fachtermini, wie etwa „Lernergebnis-Taxonomie“, „Taxonomiestufen“ etc., beantwortet. Ebenso sprechen Eigeninitiativen, wie die selbstständig durchgeführte Schulung eines Lehrstuhls, für einen hohen Grad an explizierbarem Wissen. Der Aspekt ist insofern von Bedeutung, als die Etablierung expliziter Standards ein wichtiges Anliegen des Projekts war. Ein explizites Wissen bildet hierfür die Grundlage. Inwieweit dieses Wissen bereits vor dem Programm implizit oder explizit vorhanden war, lässt sich an dieser Stelle jedoch nicht eindeutig beantworten.

8. Werden durch die Lehrenden grundlegende Standards für die Aufgabenentwicklung und die Aufgabenauswertung entwickelt und für eine Qualitätssicherung nutzbar gemacht?

Ein wesentlicher Standard, der an einigen Lehrstühlen im Maschinenwesen eingeführt wurde, ist die logische Aufgabenanalyse zur Überprüfung, inwieweit die in der Prüfung gestellten Aufgaben die in den Modulbeschreibungen formulierten Lernergebnisse abbilden. Bei der logischen Aufgabenanalyse handelt es sich um ein Verfahren, bei dem auf logischer Basis überlegt wird, welche kognitiven Prozesse bei der Lösung einer Aufgabe bei den Studierenden ablaufen und inwieweit diese Prozesse den intendierten Zielen entsprechen. Dies dient einer Einschätzung der kognitiven Validität einer Aufgabe. Eine weitere Arbeitsgruppe erweitert und systematisiert derzeit eine Item-Datenbank, in der nicht nur bestehende Aufgaben für die weitere Verwendung hinterlegt sind, sondern außerdem Daten wie Item-Schwierigkeiten und Trennschärfen gespeichert werden. Wie bereits in den Ergebnissen zu Fragestellung 5 erwähnt, führen zwei der teilnehmenden Lehrstühle aus dem Maschinenwesen mittlerweile eine Schulung für Kolleginnen und Kollegen durch, in der sie die erlernten Standards weitergeben.

9. *Werden diese Standards von den Lehrenden in ihrem Handeln routinisiert?*

Die Frage nach der Etablierung von Routinen auf Basis der vermittelten Standards ist zu diesem Zeitpunkt nicht eindeutig zu beantworten. Einerseits sprechen durch die Teilnehmenden eigenverantwortlich durchgeführte Schulungen und die Fortführung der Qualitätssicherung an den Lehrstühlen dafür, dass diese auch in Bezug auf die eigene Prüfungserstellung in Zukunft automatisch gewisse Standards zugrunde legen. Für die eine ausreichende Beantwortung dieser Fragestellung müsste aber zu späteren Zeitpunkten nachgefasst werden.

10. *Wie ist die Akzeptanz der Studierenden gegenüber den neuen Aufgabentypen?*

Ein im Rahmen des Projekts beteiligter Lehrstuhl aus dem Maschinenwesen erfasste die Akzeptanz der Studierenden mit der Prüfung. Die Ergebnisse können ausführlich in der gerade entstehenden Veröffentlichung nachgelesen werden.

6. Diskussion und Ausblick

Der Ausgangspunkt des Projekts „Herausforderung Prüfen“ lag in der Beobachtung, dass das Prüfen an Hochschulen bisher kaum auf wissenschaftlichen Standards, sondern vor allem auf Intuition und Erfahrung der Lehrenden beruhte. Die mit Bologna angestoßenen Reformen schufen ein Bewusstsein für Standards, die aber vor allem auf der Ebene von Studiengängen, Modulen und in deren Zielbeschreibungen diskutiert und etabliert wurden. Dass gerade auch die Formulierung von angestrebten Lernergebnissen nicht isoliert von Prüfungen zu verstehen ist, sondern im Gegenteil auf Basis ihrer Überprüfbarkeit erfolgen sollte, wird bisher dagegen kaum thematisiert. Dabei sind Prüfungen ein wichtiger Stellhebel für Lehr- und Lernprozesse. Mit „Herausforderung Prüfen“ wurde deshalb ein Versuch unternommen, Standards für die Prüfungserstellung praxisnah und auf Basis bestehender Erfahrung von Hochschullehrenden nachhaltig in der Prüfungspraxis an der TUM zu verankern.

Die Ergebnisse des Projekts legen nahe, dass diese Ziele in einem ersten Schritt erreicht wurden. Das zugrundeliegende theoretische Vorgehensmodell bot eine gute Grundlage, um an relevanten Herausforderungen zu arbeiten. Viele Aspekte, die von den Teilnehmenden geäußert wurden, ließen sich dem Modell thematisch zuordnen und in der Folge systematisch bearbeiten. Im besonderen Fokus stand dabei jederzeit die Passung zwischen angestrebten Lernergebnissen und Prüfungsaufgaben. Herausforderungen wurden nicht isoliert, sondern systematisch als Bestandteil eines Prozesses betrachtet. In der Folge konnten die angestrebten Lernergebnisse und die Prüfungsaufgaben der Teilnehmenden entsprechend optimiert werden. Die Evaluationsergebnisse zeigen, dass die Akzeptanz des Programms auch aufgrund dieser Praxisnähe besonders hoch war. Fragen nach der Nachhaltigkeit der Maßnahmen können zu diesem Zeitpunkt noch nicht abschließend beantwortet werden. Beispiele erfolgreichen Transfers in die Lehrpraxis, wie etwa von den Teilnehmenden initiierte Schulungen oder den systematischen Aufbau eines Aufgabenpools, sprechen jedoch für einen gelungenen Transfer.

Auch der gewählte Zugang in Form einer Qualitätsentwicklungsmaßnahme, die materialbasiert und kooperativ und über einen längeren Zeitraum stattfindet, hat sich bewährt. Die Lehrenden zeigten sich vor allem von den kooperativen Elementen und der problem- und materialzentrierten Arbeit angetan und standen auch dem höheren Zeitaufwand im Vergleich zu Einzelmaßnahmen positiv gegenüber. Projekte in dieser Form erscheinen also auch zukünftig für Veränderungsmaßnahmen im Hochschulbereich als geeignet, um den Autonomievorstellungen der Lehrenden und deren heterogenen Praxen gerecht zu werden. Als nachteilig ist auf jeden Fall der hohe Aufwand bei der Konzeption zu beurteilen. Dieser würde sich aber dann lohnen, wenn die herbeigeführten Änderungen nachhaltig wirkten und Multiplikatoreneffekte entstünden. Erste Hinweise auf die Nachhaltigkeit des Projekts gibt es bereits, etwa in Form diverser Eigeninitiativen durch die Teilnehmenden.

Bisher noch unzureichend beleuchtet – aber natürlich ein zentraler Aspekt – ist die Wirkung der neu konzipierten Prüfungen auf die Studierenden: Wie werden die neuen Anforderungen erlebt? Finden die neuen Prüfungen Akzeptanz? Wie entwickelt sich das Leistungsniveau? Einer der teilnehmenden Lehrstühle hat aus diesem Grund eine Befragung der an der Prüfung teilnehmenden Studierenden initiiert und bereitet momentan eine Veröffentlichung zu den Ergebnissen vor. Darauf wird insbesondere auf die Akzeptanz der neuen Prüfung bei den Studierenden eingegangen. Die Akzeptanz sollte jedoch nicht das einzige Gütekriterium bleiben. Wie beim Forschungshintergrund angesprochen, bereiten sich Studierende mitunter sehr strategisch auf Prüfungen vor. Dazu gehört in besonderem Maße das Sichten von Altklausuren, wie die Befragung an der TUM ergeben hat. Insofern ist es nicht unwahrscheinlich, dass neue Anforderungen diese Strategien zuerst einmal durchkreuzen und bei den Studierenden deshalb Verunsicherung auslösen, ehe das Lernverhalten in Richtung der valideren Prüfungen angepasst wird. Es sollte deshalb unbedingt überprüft werden, inwieweit sich zielorientiertes Prüfen auch auf das Lernverhalten niederschlägt. Die von einem beteiligten Lehrstuhl durchgeführte Studierendenbefragung ist hierfür als wichtiger Schritt zu betrachten. Bereits im Vorfeld von „Herausforderung Prüfen“ wurde zu dieser Frage eine Clusteranalyse des Datensatzes aus der Studierendenbefragung vorgenommen (Schulz, Zehner, Schindler & Prenzel, 2014). Auch diese Analyse könnte ein Ausgangspunkt für weitere Befragungen sein, die an den einzelnen Fakultäten durchgeführt werden könnten.

Insgesamt hat sich der Ansatz einer Qualitätsentwicklungsmaßnahme mit dem Ziel, Prüfungen nachhaltig zu verbessern, bewährt und kann durchaus als „Best Practise“-Beispiel dienen, um ähnliche Projekte auch an anderen Hochschulen zu realisieren. An der TU München wurde das Projekt – wie bereits gesagt – inzwischen bereits mit einer zweiten Kohorte durchgeführt. Möglichkeiten der Verstetigung und systematischen Verbreitung an der TU München sind derzeit in der Diskussion.

Literatur

Becker, Howard S.; Geer, Blanche; Hughes, Everett C. (1995): Making the grade: The academic side of college life. New Brunswick: Transaction.

González, J., & Wagenaar, R. (2006). Eine Einführung in Tuning Educational Structures in Europe - Der Beitrag der Hochschulen zum Bologna-Prozess: Tuning Educational Structures in Europe.

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.

Kultusministerkonferenz (2005). Qualifikationsrahmen für Deutsche Hochschulabschlüsse. Bonn: Kultusministerkonferenz, Hochschulrektorenkonferenz & Bundesministerium für Bildung und Forschung, verfügbar unter:
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_04_21-Qualifikationsrahmen-HS-Abschluesse.pdf; letztes Abrufdatum: 15.5.2015.

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.

Schulz, Florian; Zehner, Fabian; Schindler, Christoph; Prenzel, Manfred (2014). Prüfen und Lernen im Studium: Erste Schritte zur Untersuchung von Prüfungsanforderungen und Lerntypen. In: Beiträge zur Hochschulforschung, 2/2014.

Shavelson, Richard. J. (2010). On the measurement of competency. In: Empirical Research in Vocational Education and Training, 2(1), 41-63.

Veröffentlichungen

Schindler, C.; Bauer, J., Strasser, A.; Schlomske-Bodenstein, N.; Seidel, T. & Prenzel, M. (2015). Prüfungen und Studienerfolg: Herausforderungen an die Prüfungspraxis. In: Handbuch Studienerfolg. In Vorbereitung.

Tagungsbeiträge und Vorträge

Schindler, C.; Bauer, J.; Schlomske-Bodenstein, N.; Strasser, A.; Seidel T. & Prenzel, M. (2015). Assessing and Improving University Exams: A Case Study on Collaborative Professional Development with University Teachers. Paper presented at the AERA, SIG 61: Measurement and Assessment in Higher Education. Paper Session, Chicago, USA.

Schindler, C.; Prenzel, M. & Seidel, T. (2014). Improving university exams through professional development of university teacher – Developing and evaluating a framework. Paper presented at the Earli, SIG 1: Assessment and Evaluation, Paper Session, Madrid, Spain.

Schlomske-Bodenstein, N.; Strasser, A. & Schindler, C. (2015). Kompetenzorientiert Prüfen – Entwicklung und Erprobung eines Weiterbildungskonzepts für Hochschullehrende an der TU München. Beitrag auf der DGHD-Tagung, Paderborn, Germany.

Monografien

Schindler, Christoph (2015, in Vorbereitung). Herausforderung Prüfen: Eine fallbasierte Untersuchung der Prüfungspraxis von Hochschullehrenden im Rahmen eines Qualitätsentwicklungsprogramms. Dissertation.

Anhang

A Auszug aus dem Evaluationsbogen für das Qualitätsentwicklungsprogramm

	trifft zu	trifft	trifft	trifft	
		eher zu	eher	nicht zu	
			nicht zu		
Arbeitstreffen und Zeitaufwand					
a)	In den Arbeitstreffen wurden die für meine Prüfung relevanten Inhalte und Probleme angesprochen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b)	Die Arbeitstreffen ließen mich die Stärken und Schwächen in meiner Prüfung besser erkennen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c)	Mit den in den Arbeitstreffen angesprochenen Inhalten und vermittelten Techniken konnte ich meine Prüfung verbessern.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d)	Die behandelten Inhalte knüpften an meinen Wissensstand an.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e)	Die Programminhalte waren so gestaltet, dass sie in meine bestehende Praxis integriert werden können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f)	Der Aufwand für die Arbeit im Prüfungsprojekt stand in einem angemessenen Verhältnis zum Nutzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g)	Das vorgeschlagene Vorgehen bei der Prüfungserstellung führte bei der Umsetzung zu einem deutlich erhöhten Zeitaufwand.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h)	Die Arbeit im Prüfungsprojekt wird langfristig zu einer Reduzierung des Zeitaufwandes bei der Prüfungserstellung führen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i)	Die Arbeit im Prüfungsprojekt trug dazu bei, dass ich mich bei der Prüfungserstellung sicherer fühle.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j)	Durch die Arbeit im Prüfungsprojekt achte ich stärker als zuvor darauf bestimmte Standards einzuhalten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die ModeratorInnen (MitarbeiterInnen des Projekts)					
k)	waren inhaltlich gut vorbereitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
l)	gaben mir konstruktive Rückmeldungen zu meinen Prüfungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
m)	konnten auf meine Fragen angemessen eingehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
n)	nutzten die gemeinsame Arbeitszeit in den Treffen effektiv aus.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
o)	machten auf zentrale Problembereiche aufmerksam.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
p)	erklärten zentrale Konzepte und Begriffe verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
q)	setzten ein zu hohes Vorwissen voraus.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Informationsmaterialien					
r)	sind verständlich und nachvollziehbar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
s)	enthalten anschauliche Beispiele.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
t)	haben eine angemessene Länge.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
u)	tragen zu einem besseren Verständnis der behandelten Inhalte bei.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
v)	sind für mich ein wichtiges Nachschlagewerk.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
w)	enthalten die für mich relevanten Informationen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>