

Prüfen und Lernen im Studium: Erste Schritte zur Untersuchung von Prüfungs- anforderungen und Lerntypen

Florian Schulz, Fabian Zehner, Christoph Schindler, Manfred Prenzel

Prüfungen sind der entscheidende Steuerungsfaktor für das Lernen im Studium und damit ausschlaggebend für eine Kompetenzorientierung bei Studierenden. Bisher existieren kaum Befunde darüber, wie an deutschen Hochschulen tatsächlich geprüft wird und wie sich Studierende auf die bestehende Prüfungspraxis einstellen. Der vorliegende Beitrag differenziert typische Anforderungsbereiche von Prüfungen und verschiedene Lerntypen an der Technischen Universität München. Die Typenbildung wurde mittels hierarchischer Cluster-Analyse vorgenommen; Grundlage war ein Datensatz aus einer Studierendenbefragung ($n=4615$). Die Ergebnisse zeigen drei typische Anforderungsbereiche von Prüfungen (wiedergabeorientiert, anwendungsorientiert und gemischte Anforderungen). Außerdem wurden fünf Lerntypen identifiziert, die unter anderem hinsichtlich ihrer Prüfungsorientierung beim Lernen weiter charakterisiert werden. Hierbei wurde auch der Zusammenhang zwischen Prüfungsanforderungen und Lernverhalten untersucht.

1 Einleitung

Mehr noch als die Semester gliedern die Prüfungen das Studium. Die Lehrenden vergewissern sich durch Prüfungen, ob die Studierenden im Rahmen der Lehrveranstaltung erfolgreich gelernt haben. Die Ergebnisse verstehen sie vielleicht auch als Rückmeldung über den Erfolg ihres Lehrens und gegebenenfalls als Anstoß, ihr Lehrkonzept im nächsten Semester beizubehalten oder zu modifizieren. Für die Studierenden bedeuten Prüfungen zunächst Hürden, die für den Studienfortschritt übersprungen werden müssen und die ihre Chancen auf dem Arbeitsmarkt beeinflussen. Die oft in der veranstaltungsfreien Zeit zu absolvierenden Prüfungen signalisieren das Ende der Lernphase. Über Prüfungen erfahren die Studierenden, inwieweit sie den Anforderungen der Lehrenden gerecht wurden und wie ihre Leistungen im Vergleich zu anderen Studierenden einzuordnen sind. Je nach Art der Prüfung und Rückmeldung erfahren sie mehr über Stärken und Schwächen in inhaltlichen Teilbereichen oder – genereller – in ihrem Fähigkeitsprofil. Sie sehen möglicherweise auch, ob es ihnen gelungen ist, die Fragen so zu beantworten, wie das von den Lehrenden erhofft und erwartet wurde. In Hinblick auf das weitere Lernen (und die eventuell zu bestehende Wiederholungsprüfung) ziehen sie unter anderem Rückschlüsse über den Lernauf-

wand, den sie bei solchen Prüfungsanforderungen investieren, und wie sie ihr Lernen über das Semester arrangieren und mit Blick auf die Prüfungen konzentrieren sollten.

Folgt man diesen Überlegungen, dann kommt man zu der Erkenntnis, dass letztlich die Prüfungen der entscheidende Steuerungsfaktor für das Lernen im Studium sind. Auch didaktisch höchst mangelhafte Lehrveranstaltungen erzwingen, wenn sie mit Prüfungen verbunden sind, Lernfortschritte. Studierende zeichnen sich durch Intelligenz sowie langjährige Lern- und Prüfungserfahrungen aus, so dass sie notfalls auch auf sich gestellt oder in Gruppen in der Lage sind, sich durch andere Informationsquellen auf die Prüfung einzustellen. Allerdings rechtfertigt dieses Kompensationspotential, über das Lernende im tertiären Bereich verfügen, keinesfalls ein unprofessionelles Lehren. Denn dieses produziert auf Seiten der Studierenden einen nicht zu rechtfertigenden Aufwand (letztlich Kosten) und begünstigt in der Folge eher Cleverness denn fachlich und sachlich begründetes Lernen. Wenn es nun so ist, dass Prüfungen entscheidend das Lernen der Studierenden steuern, dann wird die Qualität der Prüfungen zu dem ausschlaggebenden Faktor. Dieser Faktor bestimmt nicht nur den bescheinigten Studienerfolg, sondern vor allem auch das, was Studierende am Ende eines Semesters können. Noch interessanter ist allerdings die Frage, was die Studierenden am Ende eines Studiums davon (noch) können. Wenn das, was im Verlauf des Studiums geprüft wurde, am Ende des Studiums nicht mehr verfügbar zu sein bräuchte, dann sollte man darüber nachdenken, auf dieses im Studium zu verzichten. Ein kritischer Aspekt der Qualität von Prüfungen im Studium richtet sich somit auch darauf, ob nachhaltiges Lernen getestet wurde.

Spätestens hier rückt der Begriff der Kompetenz oder des „kompetenzorientierten Prüfens“ in das Blickfeld. Auch wenn dieser Begriff im Hochschulbereich in gewisser Weise vorthoretisch (*Koeppe/Hartig/Klieme/Leutner 2008*) verwendet und unterschiedlich verstanden wird (*Schaper 2012*), ist er längst institutionell präsent: Die Anstrengungen, einen europäischen Hochschulraum zu schaffen, verlangen die Explikation von Kompetenzziele, die auf den Europäischen Qualifikationsrahmen bezogen sind. Diese Forderung betrifft die Beschreibung von Lehrzielen in Studienplänen und ist Gegenstand der Betrachtung im Rahmen von Akkreditierungen (*HRK 2005, WR 2008/2012*).

Die zunächst vielleicht harmlos scheinende Frage nach der Art des Prüfens im Studium gewinnt in diesem institutionellen Kontext an Gewicht. Das heißt, die Qualität der Prüfungen kann nicht mehr ignoriert, sondern sie muss gesichert und entwickelt werden. Die Analyse der Qualität von Prüfungen kann aus unterschiedlichen Perspektiven erfolgen. Aus einer testtheoretischen Perspektive können zum Beispiel Kriterien wie die Reliabilität und Validität angesprochen werden. Schon die Sicherung der Reliabilität von Prüfungen im Studium (als notwendige Voraussetzung für Validität) ist

keineswegs trivial. Die Frage nach der (vor allem Konstrukt- und Kriteriums-) Validität von Prüfungen stellt den systematischen Bezug zu den Lehr- oder Kompetenzzielen her, die in Studienplänen festgelegt werden: Erfasst die Prüfung tatsächlich die Kompetenzen, die mit der Lehrveranstaltung oder mit dem Modul angestrebt werden? Und lässt die Prüfungsleistung eine Vorhersage darüber zu, ob und wie die Kompetenz später, etwa in anderen Lern- oder Anwendungssituationen, wieder genutzt werden kann?

Eine andere Analyseperspektive wurde in den Ausgangsüberlegungen skizziert. Sie betrifft die Rückwirkung der Art des Prüfens auf das Lernen der Studierenden. Damit verbinden sich Fragen nach der Transparenz der formalen und inhaltlichen Prüfungsanforderungen sowie nach vermuteten Schwerpunkten und Wissensanforderungen. Angekündigte Prüfungsformate (unterschiedliche Varianten von Multiple choice-Fragen oder Aufgabentypen) beeinflussen das Lernen. Wenn etwa aufgrund früherer Prüfungen erwartet werden kann, dass insbesondere Faktenwissen gefragt ist, dann steht dieses im Zentrum der Prüfungsvorbereitung und des Lernens. In diesem Zusammenhang können auch Beziehungen zur Motivation und Zielorientierung der Studierenden thematisiert werden, bis zur Frage nach Belastungen und Stress im Studium.

Die Relevanz des Prüfens im Studium ist unstrittig. Die bisherigen Ausführungen deuten an, dass auf sehr unterschiedliche Weise geprüft werden kann und dieses Prüfen starke Rückwirkungen auf das Lernen der Studierenden haben kann. Aber wie wird an den Hochschulen in Deutschland tatsächlich geprüft? Und wie lassen sich die Studierenden auf diese Prüfungskultur ein? Hierzu fehlt es an gesichertem Wissen. Bisher beschränken sich Studien auf die geänderte Arbeitsbelastung im Zuge der Reformen (*Schulmeister/Metzger 2011*) oder nehmen am Rand die eine oder andere Frage nach Prüfungsbelastung auf (*Bargel/Ramm/Multrus 2012*).

Vor diesem Hintergrund wurde an der Technischen Universität München (TUM) ein erster Schritt unternommen, durch eine Vollerhebung bei Studierenden und Lehrenden eine Übersicht über Häufigkeiten und Typen von Prüfungen, Art und Weise der Vorbereitung etc. zu erhalten. Die Daten aus dieser Erhebung sollen im Folgenden genutzt werden, um zentrale Fragen zu untersuchen, die auch auf den Steuerungsaspekt von Prüfungen abzielen: Welche Anforderungen stellen Dozierende mit ihren Prüfungen an Studierende und inwieweit decken sich diese mit übergeordneten Kompetenzzielen? Welches Vorbereitungsverhalten auf Prüfungen zeigen Studierende? Inwieweit korreliert dieses Vorbereitungsverhalten mit Prüfungsanforderungen?

2 Theoretischer Hintergrund

2.1 Anforderungsbereiche von Prüfungen

Eine empirische Typisierung von Prüfungsanforderungen im Hochschulbereich entlang grundlegender Kompetenzdimensionen wurde bislang nicht unternommen. Prüfungsanforderungen wurden in der psychologischen Forschung bisher vor allem nach kognitiven Kriterien wie etwa der „task complexity“ (Aufgabenkomplexität) differenziert (*Campbell 1988*). Nach Maßgabe der geforderten Outcome-Orientierung sollten Kompetenzen bei der Beurteilung von Aufgaben und Prüfungen das zentrale Bewertungskriterium sein.

Bisher existiert jedoch kein einheitliches Verständnis über den Kompetenzbegriff im Hochschulbereich, was eine theoriegeleitete Operationalisierung erschwert. Dies ist insbesondere auch auf die hohe Komplexität akademischer Kompetenzen vor dem Hintergrund unterschiedlicher Fach- und Anwendungsdomänen zurückzuführen (*Blömeke/Zlatkin-Troitschanskaia 2013*). Deshalb wird in der aktuellen Hochschulforschung meist von einem mehrdimensionalen Kompetenzverständnis ausgegangen (vgl. *ebd., Schaper 2012*). Eine domänenunabhängige Spezifizierung wird dagegen im Qualifikationsrahmen für deutsche Hochschulabschlüsse (QDH) vorgeschlagen (*HRK, KMK, & BMBF 2005*). Im Folgenden soll erörtert werden, inwieweit sich dieser als wissenschaftliches Konstrukt im Rahmen der präsentierten Untersuchung eignet.

Nach Maßgabe des QDH sollen Studierende im Studium ganz grundsätzlich Lernzuwächse in den drei Bereichen (1) *Wissen*, (2) *Verstehen* und (3) *Können (Anwenden)* erzielen. Die einzelnen Anforderungsdimensionen sind dabei wie folgt näher bestimmt:

1. Der Bereich des *Wissens* bezieht sich auf die Fachkenntnis. Für den Bachelorabschluss wird zum Beispiel ein „breites und integriertes Wissen [...] der wissenschaftlichen Grundlagen ihres Lerngebiets“ vorausgesetzt (*HRK et al., 2005*), welches im Masterstudiengang erweitert werden soll.
2. *Verstehen* beinhaltet über das Wissen hinaus das Prinzip der Wissensvertiefung, insofern zentrale Zusammenhänge der Domäne erschlossen werden müssen und es den Studierenden in der Folge gelingt, die erworbenen Fachkenntnisse horizontal und vertikal zu vertiefen (*HRK et al. 2005*).
3. Bachelorstudierende sollen im Sinne des *Könnens* 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“ (*HRK et al. 2005, S. 2*). Als Zielsetzung von Masterstudiengängen steht hier die Erweiterung dieses Problemlösevermögens auf unbekannte Situationen im Vordergrund.

Als besondere Anforderungsdimension, die im referenzierten QDH als Teilaspekt systemischer Kompetenz im Bereich des Könnens beschrieben wird, lässt sich (4) *das Bewerten von Sachverhalten* ableiten. Die Bewertung von Sachverhalten wird im Sinne einer Urteilskompetenz verstanden und kann als substantieller Teil einer besonderen „akademischen Kompetenz“ interpretiert werden (Schaper 2012, S. 22). Im QDH heißt es dazu:

4. Studierende mit einem Master-Abschluss sollen dazu befähigt sein, „auch auf der Grundlage unvollständiger oder begrenzter Informationen wissenschaftlich fundierte Entscheidungen zu fällen und dabei gesellschaftliche, wissenschaftliche und ethische Erkenntnisse zu berücksichtigen, die sich aus der Anwendung ihres Wissens und aus ihren Entscheidungen ergeben“ (HRK et al. 2005, S. 4).

Das im Qualifikationsrahmen vorgeschlagene Konzept eignet sich zur Beurteilung der Prüfungsqualität im vorgestellten Rahmen, da es grundlegende Lernziele benennt, die als Produkte (Anwenden) wie auch Voraussetzungen (Wissen, Verstehen und Bewerten) nachhaltiger Lernprozesse verstanden werden können. Gleichzeitig lassen sich aus diesen Aspekten manifeste Verhaltensindikatoren ableiten, die auf den Erwerb der jeweiligen Kompetenz hinweisen können: Wissen zeigt sich in dessen adäquater Wiedergabe, Verstehen äußert sich im Erklären von Sachverhalten, Anwenden beinhaltet die Bewältigung von praktischen Problemlagen und Bewerten verweist etwa auf den Umgang mit widersprüchlichen Informationen. Damit ist die Spezifizierung allgemein genug, um Anforderungen über Fächer hinweg abzubilden und gleichzeitig differenziert genug, um einen ersten Eindruck über den Ausprägungsgrad wichtiger Anforderungsbereiche zu vermitteln.

2.2 Die Rolle von Prüfungen für den Lernprozess

Die Vorbereitung auf bestimmte Prüfungsanforderungen hat für Studierende bereits einen nicht zu unterschätzenden Aufgabencharakter. Studierende zeigen Einfallsreichtum und Problemlösevermögen nicht erst in der Prüfung, sondern bereits bei der Prüfungsvorbereitung: Sie sind also kompetente und vor allem zielgerichtete Lerner. Wie Studierende mit der ihnen gestellte Aufgabe in Form einer Prüfung umgehen, ist ein Thema lernpsychologischer Forschung.

Unterschieden wird dabei häufig zwischen *deep* und *surface approaches to learning* (Marton/Säljö 1984). Das Konzept folgt dem *levels of processing*-Ansatz (Craik/Lockhart 1972), der sich mit der Verarbeitungstiefe von Inhalten befasst. Beim sogenannten oberflächlichen Lernen (*surface approach*) findet keine differenzierte Beschäftigung mit den Lerninhalten und damit auch keine Elaborierung von bereits Gelerntem statt. Dementsprechend ist auf diese Weise erlerntes Wissen auch nur bedingt für die

Lösung komplexer Probleme geeignet, weshalb auch von *trägem Wissen* (Renkl, 1994) gesprochen wird. Unter der Maßgabe eines nachhaltigen Lernens muss demnach die Chance minimiert werden, dass sich Studierende oberflächlich auf eine Prüfung vorbereiten. Insofern wird es relevant, ob und auf welche Weise Prüfungen auf Lernprozesse hin wirken, ob sie also den einen oder den anderen Lernansatz evozieren.

Befunde zum Zusammenhang zwischen Lernverhalten und Aufgabenformen stimmen darin überein, dass studentisches Lernen nicht ausschließlich auf persönlichen Präferenzen beruht, sondern *strategisch* ausgerichtet ist. So zeigen Becker, Geer und Hughes (1995), dass Lernprozesse im Studium unter anderem von einer *grade point average perspective* diktiert werden. Demzufolge stehen Noten für Studierende stärker im Vordergrund als der Kompetenzerwerb: Sie lernen also nicht in erster Linie auf Nachhaltigkeit, sondern um die Prüfung mit einer guten Note abzuschließen. Studierende legen aus dieser Perspektive vor allem darauf Wert, die *Prüfungsanforderungen* zu erfüllen. Umso bedeutender ist es, dass in Prüfungen tatsächlich die geforderten Fähigkeiten gezeigt werden müssen, um bestimmte Aufgabenstellungen zu lösen. Dahinter steht die Annahme, dass sich strategisches Lernen nicht verhindern lässt. Studierende wählen in der Regel den ökonomischsten Weg zur Lösung einer Aufgabe, wie etwa Auswendiglernen (Shavelson 2010). Prüfungsinhalte sollten also möglichst so konzipiert sein, dass der wahrscheinliche Lösungsweg den gewünschten Lernzielen entspricht. Dieser Steuerungsperspektive folgend sollte sich diese Passung dann auch in der Qualität des beobachteten Lernverhaltens widerspiegeln.

Im vorliegenden Beitrag wird davon ausgegangen, dass die verwendeten Lernmedien sowie das Ausmaß ihrer Verwendung geeignete Indikatoren für die Qualität von Lernprozessen darstellen. Verwendete Lernmedien stellen erfassbare und vor allem verhaltensnahe Indikatoren dar. Dagegen liegen grundsätzliche und berechtigte Zweifel an der Validität von gängigen Lernstrategieinventaren vor (Artelt, 2000). Lernmedien weisen Unterschiede in Elaboriertheit und Informationsgehalt auf. Sie bringen zudem je nach Person und Lernsituation einen unterschiedlichen Grad an Attraktivität und Adäquatheit für die Verwendung mit sich. So kann zum Beispiel vermutet werden, dass auf Basis von Vorlesungsskripten und Foliensätzen zwar grundsätzlich auch problem- und anwendungsorientiert gelernt werden kann. Wahrscheinlicher ist aber eine Fokussierung auf die Wiedergabe von Wissen, da diese Medien einen hohen Grad an inhaltlicher Vorstrukturierung, Reduktion und Selektion aufweisen. Die Verwendung von (Original-) Literatur in Verbindung mit eigener Literaturrecherche kann dagegen vor allem für komplexere Prüfungsaufgabenstellungen wichtig sein, in denen vielseitiges Wissen hinzugezogen, gegebenenfalls bewertet und in Verbindung gebracht werden muss. Anhand ihres Strukturierungsgrades können die verwendeten Lernmedien damit auch als *veranstaltungspezifisch* und *veranstaltungsunspezifisch* klassifiziert werden. Von besonderem Interesse sind jene Lernsituationen und

-umgebungen, in denen Studierende aus einer Mehrzahl vorhandener Lernmedien schöpfen können, sich letztlich aber auf wenige fokussieren. Diese Fokussierung – *Welche Lernmedien wurden in welchem Ausmaß verwendet und welche nicht?* – lässt sich als Qualitätsmerkmal von Lernverhalten bzw. als Lernstrategien interpretieren und sich auf die von Studierenden wahrgenommenen Prüfungsanforderungen beziehen.

2.3 Zusammenfassung der Forschungsfragen

Die angeführten Überlegungen lassen sich in folgenden Forschungsfragen zusammenfassen:

1. Welche Prüfungstypen lassen sich gemäß wahrgenommener Anforderungen identifizieren und spiegeln diese die im QDH definierten Anforderungsbereiche wider?
2. Welche Lerntypen lassen sich auf Basis der verwendeten und vorhandenen Lernmedien identifizieren?
3. Welche Zusammenhänge bestehen zwischen den Prüfungs- bzw. Lerntypen und weiteren Merkmalen wie etwa Fakultätszugehörigkeit?
4. Inwiefern hängen die gefundenen Prüfungs- und Lerntypen zusammen?

3 Methoden

3.1 Untersuchungsdesign und Stichprobe

Für die Exploration der oben beschriebenen Fragestellungen wurden die Daten einer Studierendenbefragung an der Technischen Universität München (TUM) herangezogen, bei welcher alle Studierenden befragt wurden, die bereits mindestens eine Prüfung im Rahmen ihres Studiums abgelegt hatten. Die Daten wurden mittels Online-Fragebogen gewonnen, der nach allgemeinen Angaben zur Person sowie Kennzahlen und Stressempfinden zu den absolvierten Prüfungen fragte. Außerdem wurden vertiefende Aspekte zu jener Prüfung erhoben, welche für die Studierenden „im letzten Semester am wichtigsten war“, da angenommen werden kann, dass Studierende sich an diese gut erinnern konnten, und die erhobenen Daten auch tatsächlich relevante Prüfungen beschreiben. Neben diesen für den vorliegenden Artikel zentralen Merkmalen wurden Informationen zu relevanten kontextuellen und personenbezogenen Merkmalen abgefragt (z. B. Fakultätszugehörigkeit).

Die Erhebung fand von November 2011 bis Januar 2012 statt. Angeschrieben wurden dafür 19 030 Studierende der TUM. Insgesamt beteiligten sich 4 615, also 24 Prozent, der angeschriebenen Studierenden an der Befragung, von denen 64 Prozent männlich und 36 Prozent weiblich waren. Aufgrund dieser substantiellen (aber für Online-Befragungen üblichen) Nicht-Beteiligung wurde die Selektivität der erfassten Stichprobe

anhand einiger Indikatoren überprüft, zu denen Daten der Grundgesamtheit vorlagen. Diese Analyse zeigte keine substantiellen Verzerrungen in Bezug auf das Geschlecht und die angezielten Studienabschlüsse. Dies schließt jedoch nicht aus, dass Selektionseffekte auf anderen untersuchungsrelevanten Variablen vorliegen, zu denen keine Informationen aus der Grundgesamtheit vorliegen. Zudem fiel die Beteiligung der dreizehn Fakultäten an der Befragung unterschiedlich aus: So beteiligten sich etwa die Studierenden der Mathematik mit 31 Prozent, jene der Architektur nur mit 13 Prozent.¹ Trotz dieser Einschränkungen gehen wir davon aus, dass sich die Stichprobe zur Beantwortung der gestellten Forschungsfragen eignet.

3.2 Erhebungsinstrument

Um die Anforderungsbereiche dieser subjektiv wichtigsten Prüfung zu erheben, wurde die in Kapitel 2.1 beschriebene Differenzierung verwendet. Diese kann im Sinne einer fächerübergreifenden, grundsätzlichen Beurteilung der Prüfungsqualität als einfache Operationalisierung dienen. Dementsprechend wurden die Studierenden um die Angabe gebeten, zu welchem Anteil die Bereiche *Wissen wiedergeben*, *Sachverhalte in eigenen Worten erklären*, *Wissen auf Beispiele und Problemstellungen anwenden* sowie *Sachverhalte bewerten* in der Prüfung eine Rolle gespielt hatte. Dabei konnte der Anteil des jeweiligen Bereichs auf einer fünfstufigen Skala (0, 25, 50, 75 und 100 %) angegeben werden, wobei die resultierende Summe 100 Prozent ergeben sollte.

Ebenfalls im Kontext des Lernprozesses dieser subjektiv wichtigsten Prüfung wurden die Studierenden nach der Verwendung bestimmter Lernmedien gefragt, um somit auf ihr Lernverhalten für die Prüfung schließen zu können. So sollten sie angeben, ob die entsprechenden Lernmedien (*Vorlesungsskript*, *Foliensatz zur Veranstaltung*, *Fragenkatalog*, *Altklausuren*, *Mitschriften*, *Lehrbücher*, *angegebene Begleitliteratur*, *selbst recherchierte Literatur*, *Internet*) grundsätzlich zur Verfügung standen und falls ja, in welchem Ausmaß sie damit gelernt haben (*nie*, *selten*, *häufig*, *sehr häufig*).

3.3 Analysen

Die Typisierungen von Prüfungen und Lernverhalten wurden mittels explorativer, theoriebildender Clusteranalysen durchgeführt, während die Zusammenhänge anhand linearer Modelle und Kennwerte überprüft wurden. Für alle hier angeführten Analysen wurde die Software *R* (V. 2.15.1) verwendet (*R Core Team 2013*).

¹Eine detaillierte, tabellarische Darstellung der Stichprobe kann einem Projektbericht zur Befragung entnommen werden (Vgl. *Prenzel/Schindler/Schulz 2013*).

4 Ergebnisse

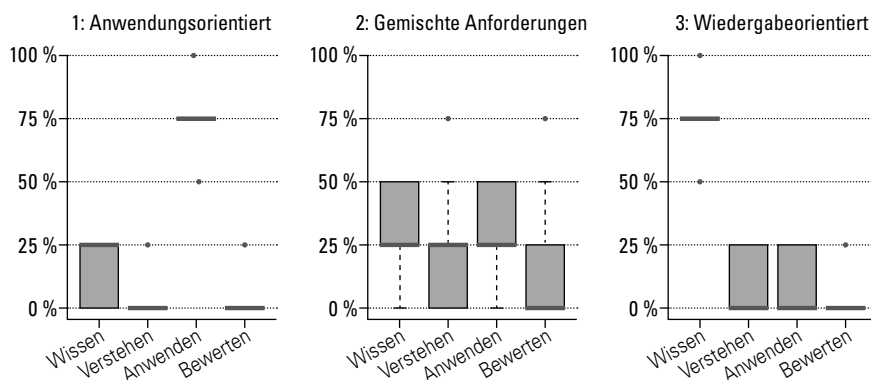
Die Ergebnisdarstellung der in diesem Kapitel dargestellten Analysen orientiert sich an den oben aufgeführten Forschungsfragen. Zuerst werden die von den Studierenden angegebenen Prüfungen anhand der vier Anforderungsbereiche in Clustern kategorisiert, im Folgenden werden dann Typen im Lernverhalten der Studierenden identifiziert und abschließend Zusammenhänge dieser Typisierungen zueinander sowie zu weiteren Merkmalen exploriert.

4.1 Typisierung von Prüfungen nach Anforderungsbereichen

Um Prüfungen danach klassifizieren zu können, wie stark die vier Anforderungsbereiche *Wissen wiedergeben*, *Sachverhalte in eigenen Worten erklären*, *Wissen auf Beispiele und Problemstellungen anwenden* sowie *Sachverhalte bewerten* eine Rolle in der Bearbeitung spielten, wurde eine hierarchische Clusteranalyse durchgeführt. Für die Interpretation der Ergebnisse ist die Abhängigkeit der Variablenausprägungen untereinander zu berücksichtigen, nachdem die Summe der Variablenausprägungen 100 Prozent ergeben musste.

In die Analyse wurden instruktionskonforme (d. h., mit resultierender Summe von 100 Prozent) und vollständige Antworten einbezogen. Mit diesen $n = 3098$ Studierenden-Antworten wurde eine hierarchische Clusteranalyse über die quadrierten euklidischen Distanzen der vier Variablen durchgeführt. Zur Bestimmung einer adäquaten Clusteranzahl wurden die Prüfungen hierzu iterativ nach dem Ward-Verfahren fusioniert. Der Verlauf der Clusterabstände weist dabei auf eine optimale Lösung bei drei Clustern hin.

Abbildung 1: Verteilungen der vier Anforderungsbereiche je Prüfungscluster



Die entstehenden drei Cluster differenzieren zufriedenstellend in Hinblick auf das Verhältnis ihrer Streuung auf den vier Variablen innerhalb der Cluster im Vergleich zur Streuung der Gesamtstichprobe. Ein Merkmal kann dann als typisch beschreibend für einen Cluster gelten, wenn die Standardabweichung der Variablen der im Cluster befindlichen Personen etwa höchstens halb so groß wie jene der Gesamtstichprobe ist (vgl. *Kubinger/Rasch/Yanagida 2011*). Gleichzeitig *kann* ein Verhältnis >1 indizieren, dass die in der Gesamtstichprobe vorhandene Streuung fast ausschließlich durch die in diesem Cluster befindlichen Personen erklärt wird, womit ein Merkmal ebenso als typisch beschreibend gelten kann; in einem solchen Fall stellt sich nicht der Mittelwert, sondern die Streuung als charakteristisch heraus. Die Verhältnisse der Standardabweichung innerhalb der Cluster zur Standardabweichung der Gesamtstichprobe sind in Tabelle 1 für jeden Cluster und jedes Merkmal angeführt (*SD-Q*). Außerdem sind der Tabelle die Verteilungskennwerte² der Cluster zu entnehmen, welche in Abbildung 1 zusätzlich visualisiert werden.

Tabelle 1: Verteilungskennwerte der vier Anforderungsbereiche je Prüfungscluster

	1: Anwendungsorientiert (n = 1148)			2: Gemischte Anforderungen (n = 1524)			3: Wiedergabeorientiert (n = 426)			Gesamt
	M	SD	SD-Q	M	SD	SD-Q	M	SD	SD-Q	
Wissen	14 %	12 %	0,5	35 %	13 %	0,5	79 %	14 %	0,6	25 %
Verstehen	4 %	9 %	0,6	20 %	16 %	1,1	7 %	11 %	0,7	15 %
Anwenden	78 %	13 %	0,5	36 %	15 %	0,5	10 %	12 %	0,4	28 %
Bewerten	4 %	9 %	0,8	9 %	13 %	1,1	4 %	9 %	0,8	12 %

M = Mittelwert; SD = Standardabweichung; SD-Q = Verhältnis SD innerhalb und gesamt

Prüfungen in *Cluster 1* weisen einen hohen Anwendungsbezug mit etwas Wissenswiedergabe auf, während die beiden Bereiche *Verstehen* und *Bewerten* praktisch irrelevant sind. Die *SD-Q*-Werte zeigen, dass die Ausprägungen dieser beiden Variablen als charakteristisch für Cluster 1 und differenzierend gegenüber den anderen Clustern betrachtet werden können. Prüfungen dieser Art sollen an dieser Stelle als *anwendungsorientiert* bezeichnet werden.

²Das Skalenniveau der vier betreffenden Items mit der Antwortskala 0, 25, 50, 75 und 100 Prozent ist strenggenommen als ordinal anzunehmen, dennoch werden Kennwerte und Analysen berichtet, die auf metrische Skalenniveaus ausgelegt sind. Dieser Aspekt wird berücksichtigt, aber zugunsten einer besseren Differenzierbarkeit in der Interpretation geduldet. Zudem sind Anteile durchaus intervallskaliert interpretierbar, die Reduktion auf fünf Antwortkategorien (und damit ordinales Skalenniveau) rührt ausschließlich von der Wahrnehmbarkeitsoptimierung her und (somit) von der Ökonomie des Fragebogens.

Prüfungen in *Cluster 2* lassen sich hingegen als solche mit *gemischten Anforderungen* bezeichnen, da die drei Bereiche *Wissen*, *Verstehen* und *Anwenden* hier etwa zu gleichen Anteilen eine Rolle spielen. Bemerkenswert sind die relativ hohen Streuungsanteile in den Bereichen *Bewerten* und *Verstehen*, was sich in *SD-Q*-Werten > 1 widerspiegelt: Jene (geringe) Streuung, die über den gesamten Datensatz hinweg überhaupt in diesen Variablen vorhanden ist, findet sich vor allem in diesem zweiten Cluster wieder.

Cluster 3 bildet das Gegenstück zum ersten: Darin enthaltene Prüfungen können als *wiedergabeorientiert* bezeichnet werden. Der Anteil der *Wissenswiedergabe* liegt dabei ebenso hoch wie jener der *Anwendung* im ersten Cluster. Analog ist der Anteil der Anwendung in Cluster 3 vergleichbar hoch wie die Wissenswiedergabe in Cluster 1. Die Bereiche *Verstehen* und *Bewerten* stellen wie in Cluster 1 einen vernachlässigbar geringen Anteil dar, wobei sie in Cluster 3 minimal höher sind. Auch hier können die Merkmalsausprägungen als charakteristisch betrachtet werden, da in den Hauptcharakteristika *Wissen* und *Anwenden* nur etwa die Hälfte der Gesamtstreuung durch Personen dieses Clusters ausgemacht wird.

Zusammengefasst werden die Cluster vor allem durch das Wechselspiel der Bereiche *Wissen* und *Anwenden* charakterisiert, wobei ersteres im dritten und letzteres im ersten Cluster deutlich höher ausgeprägt ist. Der jeweils niedrigere Bereich ist dabei dennoch höher ausgeprägt als die weiteren Bereiche *Verstehen* und *Bewerten*, die über alle drei Cluster hinweg eher geringe Ausprägungen aufweisen, wobei sich der Großteil der Streuung und Ausprägungen im Anteilsbereich von etwa 25 Prozent im zweiten Cluster vereint.

Um die Klassifikationsgüte der gefundenen Clusterlösung zu prüfen, wurden die Profile der Medianausprägungen auf den vier Variablen als Prototypen der jeweiligen Cluster herangezogen und mit den im Cluster befindlichen Studierendenangaben verglichen. Mit einem Skalenschritt Toleranz nach oben und unten ausgehend vom jeweiligen Prototypen-Median lassen sich 98,7 Prozent der von den Studierenden angegebenen Prüfungen korrekt dem jeweiligen Cluster zuordnen.³ Wählt man ein noch strengeres Klassifizierungskriterium, indem maximal zwei der vier Bereiche um einen Skalenschritt abweichen dürfen, werden noch immer 76,9 Prozent der Prüfungen dem korrekten Cluster zugeordnet, bei maximal einem abweichenden Bereich sind es noch immer 63,4 Prozent. Die 3-Clusterlösung stellt sich somit als robust und in den vier Anforderungsbereichen hinreichend beschreibend heraus.

³Die Toleranz eines Skalenschritts bedeutet hier Folgendes: Liegt der Median des Clusters in der jeweiligen Variable etwa bei 75 Prozent, so muss die Studierendenangabe bei 50, 75 oder 100 Prozent liegen.

4.2 Typisierung von Lerntypen nach verwendeten Lernmedien

Analog der Vorgehensweise zur Klassifizierung von Prüfungen wurde für die Identifikation von Lerntypen eine hierarchische Clusteranalyse über das Verwendungsausmaß verschiedener Lernmedien durchgeführt. Ob den Studierenden bestimmte Lernmedien von den Dozierenden zur Verfügung gestellt wurden, wird hier als Prüfungsrahmenbedingung in Hinblick auf den korrespondierenden Lernprozess betrachtet. Daher werden für die im Folgenden angeführten Analysen nur die ersten sieben Lernmedien berücksichtigt, da die Literaturrecherche sowie die Internetverwendung nicht als vorgegebene Rahmenbedingung interpretiert werden können. Des Weiteren zeichnet sich Lernverhalten im Sinne eines strategischen Vorgehens vor allem in Umgebungen mit Wahlmöglichkeiten ab, in diesem Fall also mit verfügbarer Lernmedienvielfalt, sodass für die hier präsentierten Analysen nur jene Studierende einbezogen wurden, die aus allen Lernmedien auswählen konnten. Im Folgenden wird daher das Lernverhalten von $n = 1097$ Studierenden berichtet, denen alle sieben Lernmedien zur Verfügung standen und welche diese Items instruktionskonform bearbeitet haben. Sofern mindestens eines der Lernmedien als vorhanden ausgewählt wurde, wurden allfällig fehlende Werte für andere Lernmedien als nicht vorhanden interpretiert.

Über die quadrierten euklidischen Distanzen wurde eine Clusteranalyse mittels Ward-Verfahren durchgeführt. Der Verlauf der Clusterabstände deutet auf eine optimale Clusteranzahl von fünf hin. Die Verteilungskennwerte der resultierenden Cluster sind in Tabelle 2 angeführt und zusätzlich in Abbildung 2 visualisiert⁴. Es wird erkennbar, dass sich die Cluster in der Kombination ihrer Merkmalsmittelwerte unterscheiden, sodass von der Identifikation unterschiedlicher Lerntypen gesprochen werden kann. Gleichzeitig sind die Streuungen innerhalb der Cluster in einigen Merkmalen so groß, dass diese neben den Merkmalsmittelwerten gesondert bei der Clusterbeschreibung berücksichtigt werden müssen. Das heißt außerdem, dass die gefundenen Cluster nicht optimal zwischen Personen differenzieren. Dennoch können Tendenzen im Lernverhalten der Studierenden durch die Zuordnung erkannt werden, vor allem anhand einiger besonders prägnanter und für den jeweiligen Cluster charakteristischer Merkmale.

Studierende in *Cluster 5* zeichnen sich besonders dadurch aus, dass sie sehr häufig mit rein prüfungsorientierten Medien wie dem Fragenkatalog und Altklausuren lernen, während die allermeisten seltener mit Vorlesungsskripten und Foliensatz sowie praktisch nie mit zusätzlicher Literatur wie Lehrbüchern oder Begleitliteratur lernen. Aggregiert über alle Lernmedien lernen Studierende dieses Clusters einseitig und mit geringem Ausmaß. In Hinblick auf die *SD-Q*-Werte zeigt sich vor allem die Kombination aus *hoher Prüfungsorientierung und wenig elaboriertem Lernen* als charakteristisch für das Cluster.

⁴Wie schon im Abschnitt zur Identifikation von Prüfungstypen betont, wird das Ordinalskalenniveau der Merkmale in der Interpretation der Befunde berücksichtigt. Zur besseren Differenzierbarkeit und Veranschaulichung werden dennoch Kennwerte berichtet, die auf Intervallskalierung ausgelegt sind.

In *Cluster 4* geben mehr Studierende als in anderen Clustern an, umfangreichere Lernmaterialien wie Lehrbücher und angegebene Begleitlektur zu nutzen. Aber auch Vorlesungsskripte und Mitschriften werden etwas genutzt. Studierende dieses Clusters nutzen hingegen praktisch keine rein prüfungsorientierten Medien wie den Fragenkatalog und Altklausuren. Im Vergleich zu anderen Clustern weisen sich Studierende dieses Clusters somit besonders als *elaborierte Lerner umfangreicheren Lernmaterials* und als *stark wissensorientiert* aus.

Studierende des *dritten Clusters* geben an, nahezu mit allen Lernmedien viel zu lernen, am wenigsten noch mit Lehrbüchern und Begleitlektur. Ihre sehr häufige Nutzung von Vorlesungsskripten und Foliensätzen, aber auch gleichzeitig prüfungsorientierter Medien wie Altklausuren und Fragenkatalog stellt sich als sehr charakteristisch gegenüber Studierenden anderer Cluster heraus. Somit kann das Lernverhalten dieses Clusters als vielfältig bezeichnet werden, da sowohl Lernmedien zur *elaborierten Kompetenzaneignung* als auch *rein prüfungsorientierte* Medien zum Einsatz kommen.

Das Lernverhalten der Studierenden in *Cluster 2* kann als Vorstufe des dritten Clusters betrachtet werden, da hier allem die sehr häufige Nutzung des Vorlesungsskripts sowie des Foliensatzes charakteristisch ist. Großteils wird auch mittels prüfungsorientierter Medien gelernt, allerdings nur in mittlerer Ausprägung. Außerdem wird von diesen Studierenden seltener zu stark vertiefenden Medien wie Lehrbüchern und Begleitlektur gegriffen. Studierende dieses Clusters lernen also primär anhand *stärker kondensierter und veranstaltungsspezifischer Materialien* und nutzen *rein prüfungsorientierte Materialien in mittlerem Ausmaß*.

Tabelle 2: Verteilungskennwerte der sieben Lernmedien je Cluster auf vierstufiger Skala 1–4: *nie, selten, häufig, sehr häufig*

	Cluster 1 (n = 316)		Cluster 2 (n = 227)		Cluster 3 (n = 241)		Cluster 4 (n = 109)		Cluster 5 (n = 204)		Gesamt SD
	M	SD-Q	M	SD-Q	M	SD-Q	M	SD-Q	M	SD-Q	
Vorlesungsskript	2,3	0,9	3,8	0,4	3,9	0,4	2,9	1,0	2,4	0,8	1,0
Foliensatz	2,0	0,8	3,8	0,4	3,7	0,5	2,3	1,0	2,2	0,8	1,1
Fragenkatalog	1,6	0,7	2,1	0,8	3,6	0,6	1,8	0,9	3,5	0,5	1,2
Altklausuren	3,6	0,7	2,6	1,0	3,6	0,6	1,5	0,7	3,8	0,4	1,0
Mitschriften	2,5	1,0	2,6	1,0	3,2	0,8	2,4	1,0	2,6	1,0	1,1
Lehrbücher	2,1	1,0	1,7	0,7	2,4	1,0	2,5	1,2	1,5	0,7	1,0
angeg. Begleitlet.	1,6	1,0	1,5	0,7	2,0	1,1	2,0	1,3	1,2	0,6	0,9

M = Mittelwert, SD = Standardabweichung, SD-Q = Verhältnis SD innerhalb und gesamt

Cluster 1: stark prüfungsorientiertes Lernen, in insgesamt geringem Ausmaß

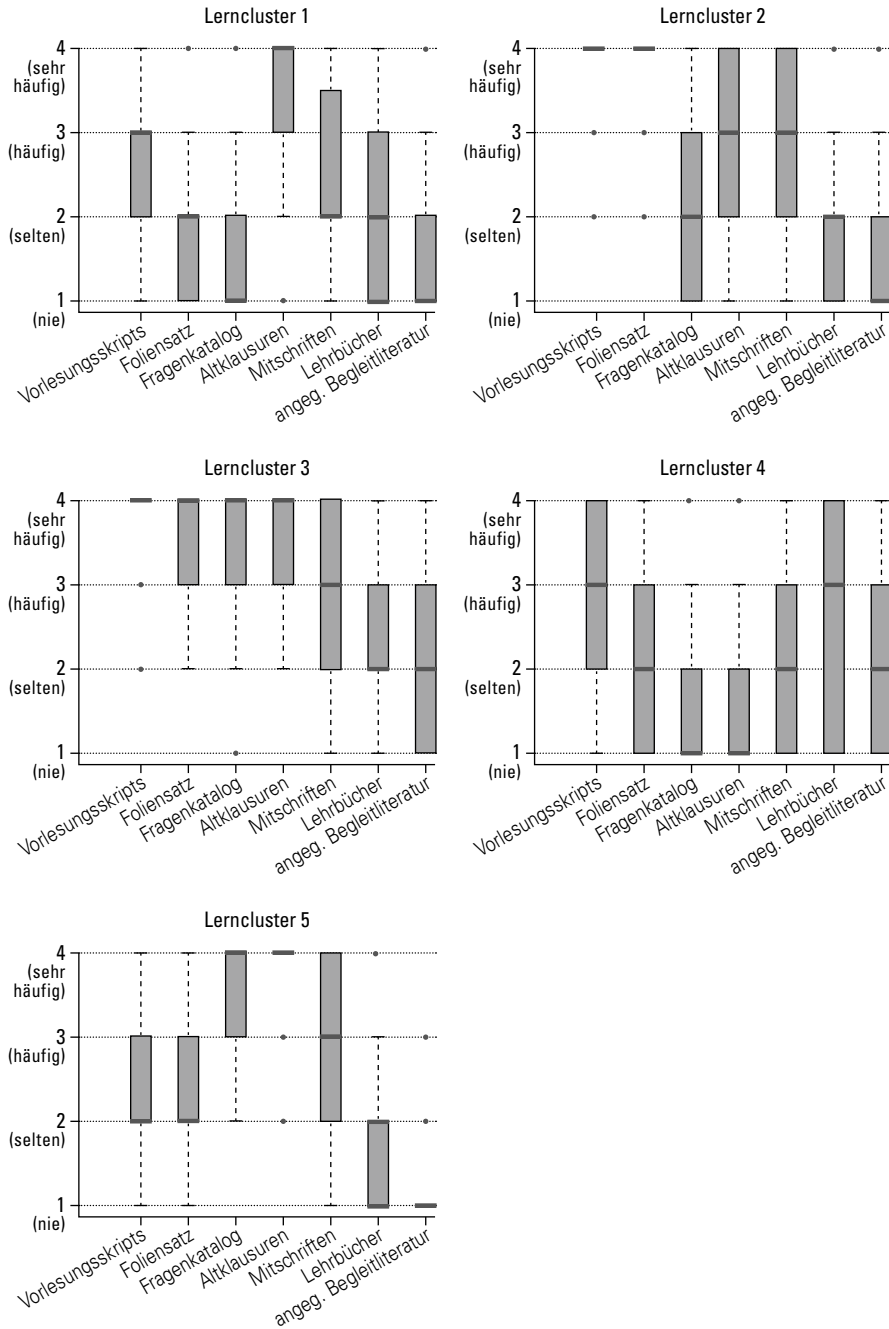
Cluster 2: mittlere Prüfungsorientierung, veranstaltungsspezifische Lernmaterialnutzung

Cluster 3: elaboriertes, prüfungsorientiertes Lernen in hohem Ausmaß

Cluster 4: elaboriertes, stark wissensorientiertes Lernen

Cluster 5: stark prüfungsorientiertes, wenig elaboriertes Lernen

Abbildung 2: Verteilungen des Nutzungsausmaßes der sieben Lernmedien je Cluster.
Für Clusterbeschreibungen s. Tabelle 2.



Cluster 1 stellt sich als sehr undifferenziert mit starker Streuung heraus, sodass dieser als „undifferenzierte Lernverhalten umfassend“ bezeichnet wird. Daneben sind die hohe Nutzung von Altklausuren sowie die nicht-Verwendung des Fragenkatalogs prägnante Merkmale. Gleichzeitig wird vergleichsweise selten zu Vorlesungsskripten und Foliensätzen gegriffen. Neben Cluster 5 geben die Studierenden dieses Clusters das aggregiert niedrigste Lernausmaß an. Die gegenüber anderen Clustern am meisten differenzierenden Merkmale stellen somit das *niedrige Lernausmaß* sowie die *hohe Altklausurnutzung* dar.

Zusammenfassend können die fünf ermittelten Typen des Lernverhaltens anhand vier aggregierter Merkmale beschrieben werden. Erstens unterscheiden sich Studierende in der *Vielfalt und im Ausmaß der Lernmediennutzung* derart, dass etwa jene des ersten und fünften Clusters einseitiger und im Gesamtausmaß weniger lernen, jene des dritten Clusters mittels aller Lernmedien und das in sehr hohem Ausmaß sowie jene des zweiten Clusters in einem Ausmaß zwischen diesen beiden Extrempolen. Zweitens können sich Studierende *Informationen aus veranstaltungsspezifischen Materialien* für die Prüfung aneignen, wie etwa aus Vorlesungsskripten, Foliensätzen und Mitschriften; dies ist über alle Cluster hinweg die gängigste Informationsaneignungsmethode, besonders für die Cluster 2 und 3, für 1 und 5 am wenigsten. Drittens können sich *Informationen auch aus veranstaltungsunspezifischen Materialien* angeeignet werden, die tiefergehend und weniger stark kondensiert sind, wie etwa aus Lehrbüchern und sonstiger Begleitliteratur, was einen höheren Aufwand an Informationsverarbeitung mit sich bringt. Diese Möglichkeit nutzen vor allem Studierende der Cluster 3 und 4, am wenigsten hingegen jene der Cluster 2 und 5. Viertens und letztens können Studierende relevante Informationen aus *rein prüfungsorientierten Materialien* wie dem Fragenkatalog und Altklausuren lernen sowie zuvor erlerntes Wissen abprüfen; dies wird außer von den Studierenden des Clusters 4 generell häufig genutzt, vor allem aber von jenen aus den Clustern 1, 3 und 5. Tabelle 3 gibt eine Übersicht zu allen Clustern anhand dieser Merkmale.

Tabelle 3: Beschreibung der fünf Lerncluster anhand aggregierter Merkmale

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Medienvielfalt & Nutzungsausmaß	+	++	+++	++	+
Veranstaltungsspezifische Informationsaneignung	++	+++	+++	++	++
Veranstaltungsunspezifische Informationsaneignung	++	+	++	++	+
Rein prüfungsorientiertes Lernen	+++	++	+++	+	+++

+ = niedrig; ++ = mittel; +++ = hoch; fett = charakteristisch i. S. niedriger Streuung

Wie bereits die hohen Streuungswerte einzelner Clustermerkmale vermuten lassen, stellt sich die Klassifikationsgüte der Fünf-Cluster-Lösung als weniger zufriedenstellend als jene der Prüfungstypen heraus. So entsprechen mit derselben Vorgehensweise (s. Kapitel 4.1) nur 51,2 Prozent der Studierendenangaben zum Lernverhalten dem Median-Profil ihres jeweiligen Clusters ± 1 als Toleranzbereich. Detailanalysen zeigen, dass die Klassifizierungsgüte der Cluster 2, 3 und 4 in einem tolerablen Bereich von 59, 71 und 59 Prozent liegt, nicht aber für die Cluster 1 und 4 mit je 36 und 20 Prozent. Dieser Befund legt nahe, dass in einem Modell mit nur fünf Clustern einzelne Studierende, speziell im ersten und vierten Cluster, mit unterschiedlichem Lernverhalten gruppiert werden, was für ein Modell mit mehr Clustern spricht. Angesichts des Strebens nach Modellsparsamkeit sowie den trotz alledem interpretierbaren und aufschlussreichen Befunden wird an dieser Stelle das Modell mit fünf Clustern bevorzugt.

4.3 Zusammenhänge der gefundenen Prüfungs- und Lerntypen

Um über die Identifikation von Prüfungs- und Lerntypen hinaus erste explorative Einblicke in Zusammenhänge dieser Typen zu weiteren Merkmalen sowie untereinander zu erhalten, soll an dieser Stelle ein Überblick dazu gegeben werden.

4.3.1 Zusammenhänge mit Prüfungstypen

Prüfungen *mit hoher Anwendungsorientierung* haben meistens ein offenes Aufgabenformat (76 %) und praktisch nie reine *Multiple Choice*-Formate (1 %), während für Prüfungen *mit hoher Wissenswiedergabeorientierung* geschlossene Formate sowie mit offenen Formaten kombinierte eine gewisse Rolle spielen (beide jeweils 23 %). Prüfungen *mit gemischten Anforderungen* werden zumeist mit offenem (67 %) und sehr selten mit geschlossenem Aufgabenformat (4 %) gestaltet. Prüfungen des ersten Clusters werden häufiger nicht bestanden (12 % vs. 5 % bzw. 6 % im 2. bzw. 3. Cluster) und die Studierenden erhalten dort signifikant schlechtere Noten, $F(2, 2494) = 79,145$, $p < ,025$: So liegen die Notenschnitte bei 2,8 im ersten Cluster, im zweiten bei 2,3 und im dritten bei 2,4; die Differenz des ersten und zweiten Clustermittelwerts entspricht einer Effektstärke von $d = 0,53$. Wird darüber hinaus die Abiturnote als Ausgangsfähigkeitsniveau der Studierenden miteinbezogen, zeigt sich, dass diese, $F(3, 2363) = 39,518$, $p < ,05$, zusätzlich mit dem Prüfungstypus, $F(2, 2363) = 7,525$, $p < ,05$, sowie deren Interaktion untereinander, $F(6, 2363) = 3,942$, $p < ,05$, das Abschneiden signifikant erklären. Der Prüfungstyp wirkt dabei auch über eine signifikante Interaktion zwischen Abiturnote und Prüfungstyp: Studierende, die Prüfungen des zweiten Typs beschreiben, schneiden in Abhängigkeit von ihrer Abiturnote besser ab als jene, die Prüfungen des ersten Typs beschreiben. Studierende, die den dritten Prüfungstypus erhielten, schneiden mit relativ schlechterer Abiturnote ≥ 3 jedoch relativ besser ab als Studierende mit gleicher Abiturnote, die mit den anderen Prüfungstypen

konfrontiert waren. Jene mit relativ besserer Abiturnote ≤ 2 schneiden bei solchen Prüfungen relativ schlechter ab als jene mit gleicher Abiturnote, aber anderen Prüfungstypen. Die Fakultätszugehörigkeit moderiert diesen Effekt auf die Prüfungsleistung nicht. Des Weiteren kommen die Prüfungstypen in unterschiedlichen Fakultäten an der TUM zu unterschiedlichen Anteilen vor: In naturwissenschaftlichen und technischen Studiengängen kommen 45 Prozent anwendungsorientierte, 47 Prozent gemischte und nur 8 Prozent wiedergabeorientierte Prüfungen zum Einsatz. Dem stehen die Medizin sowie Sportwissenschaft mit nur 5 Prozent anwendungsorientierten Prüfungen entgegen, dafür aber mit 51 Prozent gemischten und 44 Prozent wiedergabeorientierten. Zwischen diesen beiden Extrempolen liegt eine gemischte Gruppe mit zwei sozialwissenschaftlichen Studiengängen, Architektur und einem Studium für Ernährung, Landnutzung und Umwelt mit der Verteilung 24:55:21 über die drei Prüfungstypen.

Studierende, die sich auf *Prüfungen mit hoher Anwendungsorientierung* beziehen, berichten von signifikant höherem Zeitaufwand für diese, $F(2, 3026) = 34,272, p < ,025$. Der Median des ersten Clusters liegt dabei auf *41–50 Stunden* und für die beiden anderen auf *31–40 Stunden*, was einer Effektstärke von $d = 0,31$ entspricht (und damit übersetzt in die Metrik der Skala etwa +8 Stunden Differenz). Die Prüfungsschwierigkeit wird von allen Studierenden als hoch bis sehr hoch wahrgenommen, gleichzeitig unterscheiden sich die Prüfungstypen jedoch, da Prüfungen des ersten Clusters signifikant als besonders schwierig wahrgenommen werden, $F(2, 3081) = 92,349, p < ,025, d = 0,58$ (zwischen 1 und 3). Studierende, die mit Prüfungen des ersten Typs konfrontiert sind, berichten außerdem häufiger ein höheres Stressempfinden vor und während der Prüfung als die Studierenden der anderen Cluster. Tabelle 7 gibt einen Überblick über clustertypische Merkmale.

Tabelle 4: Überblick Zusammenhänge von Prüfungstypen mit weiteren Merkmalen

	Prüfungstyp 1 (anwendungsorientiert)	Prüfungstyp 2 (gemischt)	Prüfungstyp 3 (wiedergabeorientiert)
Aufgabenformat	Offen	Offen & Kombination	Gemischt & MC
Vorbereitungsaufwand	41–50 Std.	31–40 Std.	31–40 Std.
Subjektive Aufgabenschwierigkeit	Sehr hoch	Hoch	Hoch
TUM-Studiengänge	Technik & Naturwissenschaft	Sozialwissenschaften & Weitere	Medizin & Sportwissenschaft
Prüfungsnote	2,8	2,3	2,4

4.3.2 Zusammenhänge mit Lerntypen

Ob die kommende Prüfung schriftlich oder mündlich ist, hat gemäß der Datenlage nahezu keinen Effekt auf das Lernverhalten der Studierenden. Der größte deskriptive Unterschied zwischen den Lernclustern ist hier, dass Studierende des vierten Clusters vier Prozent mehr schriftliche Prüfungen als jene des ersten und zweiten Clusters hatten, was weder signifikant noch bedeutsam ist. Ebenso zeigt sich kein bedeutender Effekt für das Lernverhalten, ob die kommende Prüfung im offenen oder *Multiple Choice*-Format oder aber einer Kombination daraus gestaltet ist. Lediglich als kleine deskriptive Tendenz zeichnet sich ab, dass der Lerntyp 2 seltener durch geschlossenes Format evoziert wird (15 % seltener als die Lerntypen 3 und 5). Gleichzeitig evoziert ein offenes Format 18 Prozent häufiger Lerntyp 1 als Lerntyp 4. Analog zeigt eine Kombination des offenen und geschlossenen Formats 19 Prozent mehr Lerntyp 1 als Lerntyp 4. Diese Tendenzen sind gemäß χ^2 -Test nicht signifikant, $\chi^2(8) = 9,287$, $p = ,319$, der korrigierte Kontingenzkoeffizient $C_{\text{kor}} = ,11$ indiziert ebenso einen eher geringen Zusammenhang.

Dass die Lerntypen pauschal unterschiedliche Prüfungsbeurteilungen in Form variierender Prüfungsnoten evozieren würden, kann ebenfalls nicht gefunden werden. Auch hier lässt sich ein kleiner, nicht signifikanter, deskriptiver Effekt finden, nämlich dass Studierende mit Lerntyp 2 im Schnitt etwas bessere Noten erhielten (2,3) als die Lerntypen 4 und 5 (2,4) sowie 1 und 3 (2,6), $F(1, 838) = 0,291$, $p = ,590$. Auch das Berücksichtigen des Ausgangsfähigkeitsniveaus der Studierenden in Form der Abiturnote sowie deren Fakultätszugehörigkeit fördert keinen statistisch abgesicherten, signifikanten Zusammenhang zwischen Lerntyp und Prüfungsleistung zutage.

4.3.3 Zusammenhang zwischen Prüfungs- und Lerntypen

Dass sich die gefundenen Prüfungstypen als Rahmenbedingung pauschal auf das Lernverhalten auswirken, kann in den vorliegenden, korrelativen Daten weder inferenzstatistisch noch anhand deskriptiver Tendenzen bestätigt werden (vgl. Tabelle 5), $\chi^2(8) = 7,406$, $p = ,494$. Der korrigierte Kontingenzkoeffizient C_{kor} beträgt ,05, dessen geringer Betrag >0 von der kleinen Tendenz herrührt, dass wiedergabeorientierte Prüfungen sieben Prozent seltener den Lerntyp 2 evozieren als Prüfungen mit gemischten Anforderungen, dafür aber neun Prozent häufiger den Lerntyp 3. Angesichts des sehr geringen Koeffizienten kann aber nicht von einem Zusammenhang gesprochen werden. Auch das Konstanthalten des Ausgangsfähigkeitsniveaus sowie der Fakultätszugehörigkeit der Studierenden führt nicht zu einem signifikant abgesicherten Zusammenhang zwischen Prüfungs- und Lerntypen.

Tabelle 5: Kreuztabelle der Prüfungs- und Lerntypen, mit relativen Häufigkeiten (zu Lerntypen) in Klammern

		Prüfungstyp 1 (anwendungs- orientiert)	Prüfungstyp 2 (gemischt)	Prüfungstyp 3 (wiedergabe- orientiert)
(stark prüfungsorientiert, geringes Lernausmaß)	Lerntyp 1	77 (,36)	107 (,51)	27 (,13)
(mittel prüfungsorientiert, veranstaltungs-spezifische Materialnutzung)	Lerntyp 2	52 (,35)	80 (,54)	16 (,11)
(elaboriert, prüfungsorien- tiert, hohes Lernausmaß)	Lerntyp 3	59 (,40)	62 (,42)	26 (,18)
(elaboriert, wissens- orientiert)	Lerntyp 4	29 (,42)	29 (,42)	11 (,16)
(stark prüfungsorientiert, wenig elaboriert)	Lerntyp 5	49 (,38)	71 (,52)	17 (,12)

5 Diskussion

Welche Prüfungstypen lassen sich identifizieren und spiegeln diese die im QDH definierten Anforderungsbereiche wider?

Ausgehend von dieser Fragestellung konnten drei typische Cluster für Prüfungen identifiziert werden. Neben Prüfungen, die von einzelnen Anforderungen stark dominiert werden (Anwendungs- und Wiedergabeorientierung umfassen jeweils 75:25- bzw. 25:75-Anteile) stellte ein hoher Anteil der Prüfungen an der TUM heterogene Anforderungen, die nahezu alle im QDH aufgeführten Kompetenzfacetten abdecken. Dieses Ergebnis ist insofern positiv zu bewerten, als die verschiedenen Anforderungs-facetten im Lernprozess systematisch aufeinander aufbauen, also etwa Wissen ein notwendiges Fundament für Verstehen und beides wiederum Voraussetzung für das Bewerten und Anwenden des erworbenen Wissens ist. Die beiden Cluster mit stark dominierenden Prüfungsanforderungen erscheinen dagegen als weniger systematisch und daher auf den ersten Blick als weniger ideal. Eine Interpretation dieser beiden Cluster erfolgt im Zusammenhang mit Fragestellung drei.

Welche Lerntypen lassen sich auf Basis der verwendeten und vorhandenen Lernmedien identifizieren?

Fünf Cluster wurden auf Grundlage der verwendeten Vorbereitungsmedien identifiziert: Hohe Prüfungsorientierung und wenig elaboriertes Lernen, wie es die Befunde zu Cluster fünf nahelegen, können als Hinweis auf eine strategische Ausrichtung im Sinne einer Notenfokussierung interpretiert werden (*grade-point-average-perspective; Becker/Geer/Hughes 1995*). Studierende dieses Clusters könnten primär extrinsisch motiviert

sein und die Prüfungsergebnisse anstelle der Kompetenzaneignung im Blick haben, während sie gleichzeitig versuchen, ihren Lernaufwand zu minimieren. Einflüsse dieses Lernverhaltens etwa auf die Motivation oder langfristige Leistungsfähigkeit dieser Personen bleiben an dieser Stelle zu untersuchen.

Studierende des vierten Clusters lernen dagegen mit umgekehrten Vorzeichen: Die Verwendung verschiedenster Lernmedien könnte als Hinweis auf ein eher elaboriertes Lernverhalten (*deep approach to learning* nach Marton/Säljö 1984) bei einer eher geringen Prüfungsorientierung gewertet werden. Dieses Lernverhalten wäre eher intrinsisch durch ein persönliches Interesse an den Inhalten und der Kompetenzaneignung motiviert und weniger auf die Prüfungsleistung ausgelegt. Dafür spricht, dass in dieser Gruppe ein höheres Mindestlernpensum angegeben wurde als in den anderen. Im negativen Sinne könnte dieses Lernverhalten jedoch auch als das Fehlen eines eindeutigen Fokus im Lernverhalten interpretiert werden.

Die Befunde des dritten Clusters legen dagegen ein elaboriertes und gleichzeitig zielorientiertes Lernverhalten nahe. Im Sinne der dargestellten Theorie wären Studierende dieses Clusters damit die idealen Lerner, da sie sowohl die Prüfungsleistung selbst, als auch die Kompetenzaneignung im Blick haben und im Idealfall beides verbinden. Im Hinblick auf die minimale Tendenz zur schlechteren Prüfungsleistung sowie den etwas späteren Vorbereitungsbeginn in dieser Gruppe kann jedoch vermutet werden, dass diese Form des Lernens eventuell nicht ausreichend selektiv im Hinblick auf die Prüfungsinhalte ist, um sich effizient auf die Prüfung vorzubereiten.

Das Lernverhalten der Studierenden im zweiten Cluster orientiert sich hingegen stärker an veranstaltungsspezifischen Materialien wie Vorlesungsskripten und Foliensätzen. Auch dieses Lernverhalten wäre damit als strategisch im Sinne einer lernökonomischen Ausrichtung interpretierbar, insofern relativ eng an den Inhalten der Veranstaltung gelernt wird. Die kleine deskriptive und nicht signifikante Tendenz im Blick haltend, könnte dieser Lerntyp in geringem Ausmaß mit akademischem Erfolg in Prüfungen zusammenhängen. Auch hier besteht somit ein Ansatzpunkt für künftige Untersuchungen (kausaler) Zusammenhänge.

Der erste Cluster beinhaltet schließlich Studierende, die generell eher wenig lernen und dabei etwas häufiger auf Altklausuren zurückgreifen. Dieser Cluster ist aufgrund starker Streuung mit Vorsicht zu interpretieren, könnte jedoch eventuell im Sinne einer *hit-or-miss*-Mentalität verstanden werden: Studierende lernen generell eher wenig und versuchen ihr Glück durch das Repetieren von Altklausuren. Für diese Annahme könnte sprechen, dass Studierende in diesem Cluster Altklausuren tendenziell häufiger als das hilfreichste Medium angegeben haben, dabei aber auch etwas angespannter vor der Prüfung waren.

Welche Zusammenhänge bestehen zwischen den Prüfungs- bzw. Lerntypen und weiteren Merkmalen wie etwa Fakultätszugehörigkeit?

In Bezug auf die Prüfungsanforderungen sind vor allem fachspezifische Unterschiede zu diskutieren. So fällt auf, dass sich der Cluster mit anwendungsorientierten Prüfungen vorwiegend aus Studierenden der naturwissenschaftlichen und technischen Fakultäten zusammensetzt. Dies ist erst einmal wenig überraschend, da beispielsweise ingenieurwissenschaftliche Studiengänge prinzipiell eine stärkere Praxisausrichtung besitzen.

Eine weiterführende Analyse zeigt jedoch auch, dass anwendungsbezogene Prüfungen von den Studierenden als belastender wahrgenommen werden. So berichteten Studierende, die mit diesem Prüfungstyp konfrontiert waren, einen höheren Arbeitsaufwand, eine höher empfundene Schwierigkeit sowie ein höheres Stressempfinden als Studierende, die eher gemischte und wiedergabeorientierte Prüfungen absolvierten. Außerdem fällt die Prüfungsnote von Studierenden, deren Prüfung anwendungsorientiert gestaltet wurde, im Vergleich zu den anderen Gruppen schlechter aus und die Quote derjenigen ist höher, welche die Prüfung nicht bestanden haben.

Ursachen hierfür könnten einerseits in der konkreten Prüfungs- und Aufgabengestaltung zu finden sein: Bauen die betroffenen Prüfungen systematisch genug auf Vorwissen auf? Werden zum Beispiel Aufgaben gestellt, in denen das notwendige sachstrukturelle Wissen vorgegeben und ausschließlich die Fähigkeit der Anwendung dieses Wissens vor dem Hintergrund einer Problemstellung überprüft wird, oder wird beides vorausgesetzt? Wichtig in diesem Zusammenhang ist dann auch die Transparenz der gestellten Anforderungen: Werden Teilanforderungen in Bezug auf die Anwendung von Wissen im Vorfeld der Prüfung verdeutlicht? Und schließlich: Werden die notwendigen Kompetenzen zur Lösung von anwendungsorientierten Prüfungen in der Lehre ausreichend geschult? Die beobachteten negativen Effekte, insbesondere auch das Stressempfinden, könnten in diesem Sinne Zeichen von Unsicherheit sein. Da es sich bei technisch-naturwissenschaftlichen Studiengängen um den Kernbereich der TUM handelt, signalisieren diese Befunde besonderen Handlungsbedarf.

In den anderen beiden Clustern fallen die angesprochenen negativen Effekte dagegen geringer aus. Im Hinblick auf den Cluster mit gemischten Anforderungen könnte dies darauf zurückzuführen sein, dass Teilanforderungen stärker differenziert und kommuniziert werden und Studierende deshalb besser einschätzen können, was in den Prüfungen von ihnen erwartet wird. Dafür spricht auch bereits die Beobachtung, dass Studierende ihre Prüfung als sehr differenziert wahrgenommen haben. Ebenso scheinen Studierende weniger Probleme mit wiedergabeorientierten Prüfungen zu haben.

Auch hier könnte die „Berechenbarkeit“ der Prüfung bei relativ eindimensionalen Anforderungen eine ausschlaggebende Rolle spielen.

Besonderes Augenmerk kommt in diesem Zusammenhang auch dem Interaktionseffekt zwischen dem Ausgangsfähigkeitsniveau (operationalisiert durch die Abiturnote) mit der Prüfungsleistung zu: Besonders bei Studierenden mit wiedergabeorientierten Prüfungen ist die Abiturnote ein schwächerer Prädiktor für das tatsächliche Abschneiden in der Prüfung, während bei den anderen beiden Typen die Abiturnote ein starker Prädiktor ist. Dies weist darauf hin, dass das wiedergabeorientierte Format (mit hohem Multiple Choice-Anteil) auch potentiell leistungsschwächeren Studierenden gelegen kommt. Dies könnte im Sinne einer Chancengerechtigkeit ein Argument für diesen Prüfungstypus sein, wirft aber gleichermaßen die Frage nach der Validität eines solchen Prüfens auf, also ob tatsächlich die Lehr-/Lernziele damit getestet werden. Diese Frage ist vor allem dann bedeutend, wenn die Abiturnote als valider Prädiktor für den Erwerb weiterer Kompetenzen im Studium angenommen wird. Im Zentrum weiterer Forschung könnte hier das Untersuchen der Validität der verschiedenen Prüfungstypen stehen, bei welcher standardisierte Leistungstests als geeignetere Operationalisierung des Ausgangsfähigkeitsniveaus sowie des Lernzuwachses entscheidenden Erkenntnisgewinn liefern könnten.

Inwiefern hängen die gefundenen Prüfungs- und Lerntypen zusammen?

Weiteren Aufschluss über die beschriebenen Zusammenhänge könnten die gefundenen Lernstrategien in Bezug auf die unterschiedlichen Prüfungsanforderungen geben. Hier konnte jedoch kein signifikanter, korrelativer Zusammenhang zwischen den identifizierten Prüfungs- und Lerntypen gefunden werden. Dass Prüfungen das Lernen beeinflussen, wird in der Bildungsforschung inzwischen vorausgesetzt. Insofern müsste im Rahmen einer Folgestudie stärker auf die qualitativen Besonderheiten in den Lernprozessen eingegangen werden. Ebenso sollten bestehende Prüfungen weiter analysiert und differenziert werden. Aus der Forschung ist bekannt, dass Studierende ihr Lernverhalten zwar an den Prüfungsanforderungen ausrichten, dieser Adaptionprozess aber im Sinne (wechselseitiger) Interpretationsprozesse erfolgt (*Sambell/McDowell 1998*). Ebenso differenziert ist auch der Steuerungsaspekt von Prüfungen zu betrachten. Es wird deshalb in Zukunft notwendig sein, das Instrumentarium zur Analyse von Lehr- und Lernprozessen an den Hochschulen um handlungsnahe Methoden (wie etwa Lerntagebücher) sowie experimentelle Designs zu erweitern, um damit auch komplexere Wechselwirkungen zwischen Lernen und Prüfen zu analysieren (*Seidel/Hoppert 2011*).

Insgesamt gewähren die präsentierten Ergebnisse einen Einblick in Anforderungsbereiche von Hochschulprüfungen und die begleitende Verwendung von Lernmaterialien

durch die Studierenden zur Prüfungsvorbereitung, womit erstes Beschreibungswissen generiert werden konnte. Inwieweit sich die Ergebnisse auch auf andere (technische) Universitäten übertragen lassen, müsste überprüft werden. Um die Prüfungsqualität systematisch zu verbessern, sollte die Hochschulforschung in den kommenden Jahren nicht nur Beschreibungs-, sondern auch Erklärungs- und Veränderungswissen zum Thema Prüfungen bereitstellen. Die Ergebnisse dieser Studie können hierzu unter Berücksichtigung der entsprechenden Limitationen als Ausgangspunkt dienen. Bezüglich der Aussagekraft der vorgestellten Befunde ist in diesem Zusammenhang einschränkend anzumerken, dass ein relativ hoher Fallausschluss durch fehlerhafte und unvollständige Angaben die Typenbildung beeinflusst haben könnte. Außerdem beruhen die Ergebnisse auf subjektiven Einschätzungen durch die Studierenden, die retrospektiv getroffen wurden.

Trotz der diskutierten Einschränkungen tragen die vorgelegten Befunde zum Schließen der eingangs formulierten Forschungslücken bei. In weiterführenden Studien sollten neben der Überprüfung der Replizierbarkeit und Generalisierbarkeit unserer Ergebnisse erstens weiterführende Fragestellungen nach den Gründen für die Probleme von Studierenden bei anwendungsorientierten Prüfungen untersucht werden. Solche Untersuchungen wären hilfreich um Möglichkeiten zur Beseitigung der Probleme zu eruieren, die dieses im Sinn einer kompetenzorientierten Prüfungskultur vorteilhafte Format für die Studierenden mit sich bringt. Zweitens wäre die Frage zu beantworten, welche Bedingungen für das Zustandekommen verschiedener Lerntypen identifiziert werden können und inwiefern diese über verschiedene Prüfungssituationen hinweg stabil sind. Zur Beantwortung dieser eher offen angelegten Fragestellung wären unter anderem Tagebuchstudien denkbar, die das Lernverhalten vor Prüfungen im Längsschnitt sichtbar machen.

Literatur

Artelt, Cordula (2000): Wie prädiktiv sind retrospektive Selbstberichte über den Gebrauch von Lernstrategien für strategisches Lernen? In: Zeitschrift für Pädagogische Psychologie. Juli 2000 Vol. 14, (2/3), S. 72–84.

Bargel, Tino; Ramm, Michael; & Multrus, Frank (2012): Schwierigkeiten und Belastungen im Bachelorstudium – wie berechtigt sind die studentischen Klagen? In: Beiträge zur Hochschulforschung, 34(1), S. 26–41.

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

Blömeke, Sigrid & Zlatkin-Troitschanskaia, Olga (2013): Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor: Ziele, theoretischer Rahmen, Design und Herausforderungen des BMBF-Forschungsprogramms KoKoHs (KoKoHs Working Papers, 1). Berlin & Mainz: Humboldt-Universität & Johannes Gutenberg-Universität.

Campbell, Donald J. (1988): Task complexity: a review and analysis. In: *Academy of Management Review*, 13, 40–52.

Craik, Fergus; Lockhart, Robert S. (1972): Levels of Processing: A Framework for Memory Research. In: *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.

Hochschulrektorenkonferenz, Kultusministerkonferenz, Bundesministerium für Bildung und Forschung (2005): Qualifikationsrahmen für Deutsche Hochschulabschlüsse. Online verfügbar unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_04_21-Qualifikationsrahmen-HS-Abschluesse.pdf, letztes Abrufdatum: 19.3.2013.

Koepfen, Karoline; Hartig, Johannes; Klieme, Eckhard; Leutner, Detlev (2008): Current Issues in Competence Modeling and Assessment. In: *Zeitschrift für Psychologie/ Journal of Psychology*, 216(2), 61–73.

Kubinger, Klaus D.; Rasch, Dieter; Yanagida, Takuya (2011): Statistik in der Psychologie: Vom Einführungskurs bis zur Dissertation. Göttingen.

Marton, Ference; Säljö, Roger (1984): Approaches to learning. In Marton, Ference; Hounsell, D. and Entwistle, Noel (Hrsg.): *The Experience of Learning*. Edinburgh: Scottish Academic Press.

Prenzel, Manfred; Schindler, Christoph; Schulz, Florian (2013): Prüfungskultur an der Technischen Universität München. Erste Ergebnisse einer Befragung der Studierenden. Online verfügbar unter: http://www.lehren.tum.de/fileadmin/w00bmo/www/Downloads/Themen/Pruefungen/projektbericht_pruefungskultur_studierende.pdf, letztes Abrufdatum: 17.12.2013.

R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Wien: Österreich.

Renkl, Alexander (1994): Träges Wissen: Die „unerklärliche“ Kluft zwischen Wissen und Handeln München: Ludwig-Maximilians-Universität, Forschungsbericht des Psychologischen Instituts; 41.

Sambell, Kay; McDowell, Liz (1998): The construction of the hidden curriculum: messages and meanings in the assessment of student learning. In: *Assessment & Evaluation in Higher Education*, 23(4), S. 391–402.

Schaper, Niclas (2012): Fachgutachten zur Kompetenzorientierung in Studium und Lehre. Bonn: Hochschulrektorenkonferenz – nexus.

Seidel, T., & Hoppert, A. (2011): Merkmale von Lehre an der Hochschule. Ergebnisse zur Gestaltung von Hochschulseminaren mittels Videoanalysen. *Unterrichtswissenschaft*, 39(2), 154–172.

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

Schulmeister, Rolf; Metzger, Christiane (Hrsg.) (2011): Die Workload im Bachelor: Zeitbudget und Studierverhalten. Eine empirische Studie. Münster: Waxmann.

WR Wissenschaftsrat (2008): Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Köln: Geschäftsstelle des Wissenschaftsrates.

WR Wissenschaftsrat (2012): Empfehlungen zur Akkreditierung als Instrument der Qualitätssicherung. Köln: Geschäftsstelle des Wissenschaftsrates.

Manuskript eingegangen: 14.08.2013

Manuskript angenommen: 27.01.2014

Anschriften der Verfasser:

Florian Schulz

Fabian Zehner

Christoph Schindler

Professor Dr. Manfred Prenzel

TUM School of Education

Technische Universität München

Arcisstr. 21

80333 München

E-Mail:

f.schulz@tum.de

fabian.zehner@tum.de

christoph.schindler@tum.de

manfred.prenzel@tum.de

Florian Schulz, Fabian Zehner und Christoph Schindler sind wissenschaftliche Mitarbeiter am Lehrstuhl für Empirische Bildungsforschung der TUM School of Education. Manfred Prenzel ist Inhaber des Susanne Klatten-Stiftungslehrstuhls für Empirische Bildungsforschung an der Technischen Universität München und Dekan der TUM School of Education sowie Vorstandsvorsitzender des Zentrums für internationale Vergleichsstudien an der Technischen Universität München.