

Hochschulübergreifender Leitfaden für den kreativen Teil bei der Erstellung anwendungsorientierter Prüfungsfragen

Bernhardt-Melischnig, Dragosits, Trinko

Impressum

Hochschulübergreifender Leitfaden für den kreativen Teil bei der Erstellung
anwendungsorientierter Prüfungsfragen

herausgegeben vom Verein Forum Neue Medien in der Lehre Austria

Graz, 2013

Autor und Autorinnen

Johannes Bernhardt-Melischnig
Claudia Dragosits
Sandra Trinko

Layout

Johannes Bernhardt-Melischnig
Michael Raunig

Umschlaggestaltung

Michael Raunig

ISBN

9783732281329

Druck und Verlag

Books on Demand GmbH, Norderstedt

Vorwort

„Der Verein ‚Forum neue Medien in der Lehre Austria‘ bietet als etablierte Interessenvertretung ein lebendiges Netzwerk sowie die Entwicklung und Verbreitung von institutionsübergreifenden Maßnahmen und Modellen im Bereich der (technologiestützten) Bildungsangebote.“

Basierend auf diesem Mission Statement hat sich der Verein bereits 2010 dazu entschieden, seinen Mitgliedern Handreichungen und Praxisbeispiele zur Verfügung zu stellen, die sie bei der Realisierung und Verbesserung technologiegestützter Lehr- und Lernprozesse bestmöglich unterstützen. 2011 wurden zu diesem Zweck drei Arbeitsgruppen finanziert und 2012 wurde ein Forschungs- und Entwicklungscall ausgeschrieben, der den Einsatz neuer Medien in den Themenfeldern „Qualitativ hochwertige E-Learning in der Hochschullehre“, „Austausch hochschuldidaktischer Modelle“ und „Technologiegestützte Lehrinnovation“ fokussierte.

Die Ausschreibung war bewusst niederschwellig gehalten. Die Einreicher/innen wurden gebeten, ihr Projektvorhaben, dessen Zielsetzungen und die Maßnahmen zur Zielerreichung sowie die Projektkosten auf wenigen Seiten kurz und kompakt darzustellen. Ziel war es, möglichst viele zur Teilnahme an der Ausschreibung zu motivieren, ohne jedoch grundsätzliche Qualitätsstandards bei den Einreichungen außer Acht zu lassen. Teilnahmeberechtigt waren alle ordentlichen Mitglieder des Vereins, wobei im Sinne des Netzwerkgedankens Konsortien bevorzugt wurden.

Die Einreichungen wurden anhand eines Kriterienkataloges von einer Fachjury begutachtet. Die Jury empfahl die Unterstützung von insgesamt drei Projekten, die vom Verein mit einem Gesamtbetrag von 26.000 Euro gefördert wurden. Die nach der zwölfmonatigen Projektlaufzeit verfügbaren Ergebnisse stehen grundsätzlich nur Mitgliedern des Forum neue Medien in der Lehre Austria in uneingeschränkter Form zur Verfügung.

Der vorliegende „Hochschulübergreifende Leitfaden für den kreativen Teil bei der Erstellung anwendungsorientierter Prüfungsfragen“ ist eines dieser Ergebnisse. Er wurde in dem von der Medizinischen Universität Graz im Themenfeld „Austausch hochschuldidaktischer Modelle“ in Kooperation mit Lehrenden anderer österreichischer Hochschulen entwickelt. Zusätzlich förderte der Verein folgende zwei Projekte:

„Live-Streaming: Technologien, Systeme und Lösungen zur Übertragung von Lehrveranstaltungen“, eingereicht von der Universität Innsbruck im Bereich „Technologiegestützte Lehrinnovation“, sowie „eLearning-Qualitäts-Evaluationstool“, eingereicht von der Donau-Universität Krems im Bereich „Qualitativ hochwertiges E-Learning“.

Mit der Publikation der Ergebnisse aller drei geförderten Projekte erfüllt der Verein seine Zielsetzung, seinen Mitgliedern innovative Beispiele und Handreichungen für den Einsatz neuer Medien in der Hochschullehre zur Verfügung zu stellen. Soweit es seine finanzielle Situation zulässt, wird der Verein auch in Zukunft Projekte initiieren und unterstützen, um bestehende Netzwerkaktivitäten zu stärken und neue Kooperationen zu fördern.

In diesem Sinne wünschen wir Ihnen eine angenehme Lektüre und freuen uns auf Feedback unter praesidium@fnm-austria.at oder bei persönlichen Gesprächen.

Graz, im Sommer 2013

Martin Ebner und Stephan Waba
Präsidenten des Vereins
Forum neue Medien in der Lehre Austria

Vorwort

„Assessment drives learning“ – ein kurzes Statement, das die Bedeutung des Prüfens für das Lernergebnis verdeutlicht. Damit zahlt es sich aus, besonderes Augenmerk auf das Prüfungswesen im Hochschulraum zu legen, wobei heute neben der inhaltlichen Qualität besonders auch die formale Qualität, die Standardisierung und die Nachvollziehbarkeit im Vordergrund stehen.

Multiple-Choice-Fragen (kurz: MC-Fragen) bieten den Vorteil, dass sie besser als jedes andere Prüfungsverfahren standardisiert und statistisch evaluiert werden können. Um jedoch der Gefahr zu entgehen, bloße Fakten abzufragen, sich damit von den Lernzielen zu entfernen und bei einem „Lexikon des sinnlosen Wissens“ zu landen, müssen MC-Fragen hohe inhaltliche Anforderungen erfüllen. Das wirkliche Leben ist in jeder Wissenschaftsdisziplin täglich davon geprägt, dass man zwischen verschiedenen Optionen wählen muss – und genau das gilt es in MC-Fragen abzubilden. Derartige Fragen zu erstellen, ist allerdings ein kreativer Prozess, der persönliche Erfahrung einerseits, aber auch sprachliche Ausdrucksfähigkeit, Phantasie und Gestaltungskraft andererseits erfordert.

Der vorliegende Leitfaden trägt die Erfahrung aus verschiedenen Wissenschaftsdisziplinen gerade hinsichtlich dieser kreativen Gestaltungsnotwendigkeit zusammen. Er geht dabei einen wesentlichen Schritt über die üblichen Fragenleitfäden, die sich oft ausschließlich in Formalismen ergehen, hinaus und legt den Fokus auf das, worauf es ankommt: Realitätsnahe Situationen zu kreieren, in denen die Prüfungskandidatinnen und -Kandidaten komplexe Fähigkeiten unter Beweis stellen können.

Mein Dank gilt dem fnm-austria für die Unterstützung dieses Projekts und insbesondere den Projektverantwortlichen für die geduldige und konsequente Durchführung sowie allen Kolleginnen und Kollegen, die mit ihrer reichen Erfahrung zum Gelingen beigetragen haben.

Graz, im Sommer 2013

Josef Smolle
Rektor der Medizinischen Universität Graz

Zur Entstehung des Leitfadens

Dieser Leitfaden ist im Rahmen eines fnm-austria-Projektes (Forum neue Medien in der Lehre Austria) im Jahr 2012/2013 entstanden und von der Medizinischen Universität Graz zusammen mit der Wirtschaftsuniversität Wien federführend erstellt worden. Er ist vorrangig für alle fnm-austria-Mitgliedsinstitutionen gedacht.

Der im Rahmen dieses fnm-austria-Projektes entstandene Leitfaden zur Erstellung hochwertiger und anwendungsorientierter Multiple-Choice-Fragen ist für Fragenersteller/innen unterschiedlicher Fachrichtungen an tertiären Bildungseinrichtungen in Österreich gedacht. Der Leitfaden umfasst viele Aspekte des Themas „Prüfen mit Multiple-Choice-Fragen“, darunter: Grundlegendes zu schriftlichen Prüfungen mit Multiple-Choice-Fragen, wichtige Multiple-Choice-Fragen-Typen, deren Erstellung ebenso wie Qualitätsaspekte rund um Multiple-Choice-Fragen. Der Fokus liegt jedoch – anders als bei anderer Literatur zu diesem Thema – auf dem kreativ-inhaltlichen Erstellungsprozess. Anders formuliert, sollen Fragenersteller/innen beim kreativen Schreiben von Multiple-Choice-Fragen unterstützt werden, wobei formelle Kriterien vorausgesetzt oder nur kurz umrissen werden.

Zu den Recherchen und Interviews

Zwischen dem Spätsommer 2012 und dem Frühjahr 2013 wurden ca. 25 Interviews mit Lehrenden an österreichischen tertiären Bildungseinrichtungen geführt. Diese oft mehrstündigen Gespräche haben vorwiegend an den Institutionen der Lehrenden stattgefunden und wurden anhand eines strukturierten Interview-Fragebogens durchgeführt. Dieser Fragebogen hat aus 16 offenen Fragen bestanden, war jedoch oft nur der Initiator eines längeren und detaillierteren Gespräches oder einer intensiveren Diskussion.

Ziel dieser Interview-Reihe war das Sammeln von Informationen, Methoden, Tipps und Ideen von erfahrenen Lehrenden, wie man bei der Erstellung von geschlossenen Fragen vorgehen könnte. Diese gesammelten Informationen wurden anschließend in diesem Leitfaden zu Papier gebracht – Sie finden diese vorwiegend im Praxisteil, jedoch auch im Theorieteil dieses Leitfadens eingearbeitet.

Zur Verwendung des Leitfadens

Der Leitfaden ist in zwei große Kapitel sowie einen Anhang unterteilt:

Der erste Teil – der Theorieteil – bietet Fragenerstellenden einen systematischen und theoriefokussierten Einstieg in die Materie, welcher dennoch mit zahlreichen Beispielen versehen ist.

Der zweite Teil – der Praxisteil – bietet einerseits eine praxisorientierte Ergänzung; andererseits kann er von Fragenerstellenden, die noch weniger Erfahrung im Umgang mit MC-Fragen haben, auch **separat gelesen** werden, da der Theorieteil keine Voraussetzung darstellt.

Im Anhang sind weiterführende Literatur, ein Glossar, eine MC-Fragen-Checkliste und die Danksagung an alle Interview-Partner/innen zu finden.

In allen Abschnitten werden die erwähnten Themen anhand zahlreicher **Beispiele** anschaulich dargestellt, zudem finden Sie wichtige **Merkregeln**, aber auch **Praxistipps**, die gesondert gekennzeichnet wurden.

Beide Teile wurden getrennt voneinander verfasst. Zahlreiche Abschnitte weisen redundante Stellen auf, essentielle Passagen wurden jedoch bewusst wiederholt!

Inhalts- verzeichnis

Theorieteil	9
Einführung und Grundlagen	11
Einzelelemente einer Frage	33
Qualitätssicherung	49
Praxisteil	61
Wie sollte eine gute MC-Frage aufgebaut sein?	63
Wie vermeidet man versteckte Lösungshinweise?	77
Kreative Herangehensweisen	85
Anhang	91

Theorieteil

*„Item-writing is an art.
It requires an uncommon combination of
special abilities.“*

(M. C. Rodriguez, Michigan State University)

Autor: Johannes Bernhardt-Melischnig

Einführung und Grundlagen

Prüfen und schriftliches Prüfen

Der Leitfaden beginnt mit einer **Auswahl allgemeiner Aussagen zum Thema Prüfen**¹:

- „Prüfen dient dem Überprüfen des Erreichens a priori definierter Lernziele.“
- „Ergebnisse von Prüfungen sind nur dann von Nutzen, wenn daraus gültig und zuverlässig auf eine Leistung geschlossen werden kann, die über das Lösen der konkreten Prüfungsaufgabe hinausreicht.“
- „Ziel einer Prüfung ist generell das Erfassen und Beurteilen von Wissen, Fertigkeiten und Fähigkeiten, welche zuvor im Rahmen der Erstellung des Lehrplans als Lernziele für Lehrende und Studierende definiert worden sind.“
- „Einflüsse von sachfremden Faktoren und Zufälligkeiten führen zu unzuverlässigen Messungen, die eine sinnvolle Interpretation der Ergebnisse erschweren.“

Dabei sind zwei Punkte zu betonen: Zum einen ist der Begriff **des Lernziels** sehr wichtig, da ihm eine Schlüsselrolle in der Kommunikation über Lehr- und Lerninhalte zwischen Studierenden und Lehrenden zukommt, zum anderen sollte man die **Prüfung als Messinstrument** betrachten, welches je nach Handhabung brauchbare und gut interpretierbare Ergebnisse liefern kann. Dabei wird der Begriff des Lernziels wiederholt im Leitfaden aufscheinen, auf den zweiten Punkt wird vor allem im Abschnitt Qualitätssicherung eingegangen.

Es gibt zahlreiche *Prüfungsformate* mit vielen Vor- und Nachteilen. Die wichtigsten Formate sind *schriftliche Prüfungsformate* mit offenen und geschlossenen Fragestellungen, *mündliche Prüfungsformate*, die unterschiedlich strukturiert durchgeführt werden können, und sogenannte *beobachtende Prüfungsformate*, bei denen Studierende bei der Vorführung von Handlungen und Fer-

¹ Die aufgelisteten Aussagen stammen überwiegend aus Schulungsunterlagen des IML (Institut für Medizinische Lehre) der Universität Bern (Rene Krebs et al.).

tigkeiten beobachtet und bewertet werden. Alle genannten Formate ermöglichen *punktueller* Bewertungen von Studierenden, denen *longitudinale* Bewertungen gegenüberstehen (beispielsweise Portfolios). Abschließend sei betont, dass die Multiple-Choice-Prüfung (im Leitfaden kurz: MC-Prüfung) zu den geschlossenen schriftlichen Formaten gehört.

Für welches Prüfungsformat man sich entscheidet, hängt von mehreren Faktoren ab, wobei man zwei Faktoren hervorheben muss. Einerseits ist es das **Lernziel**, welches aufgrund des Inhaltes Einfluss auf das Format hat. Das Prüfen von Faktenwissen, das Prüfen auf Verständnis, aber auch das Anwenden von Wissen in neuen Situationen bestimmen schriftliche Prüfungsformate. Mündliche Prüfungsformate spielen dann ihren Vorteil aus, wenn es um das Erklären, Verknüpfen und allen voran Argumentieren geht. Beobachtende Formate sind bei der Beurteilung von Handlungen und Fertigkeiten (*skills*) zu bevorzugen.

Aufgrund des ersten genannten Faktors ist auch leicht ableitbar, dass eine sinnvolle Überprüfung eines Prüfungsfaches mehrere Prüfungsformate bedingt, wie in nachfolgendem Beispiel deutlich wird:

Beispiel 1 – Führerscheinprüfung²:

Wenn Sie an Ihre Führerscheinprüfung zurückdenken, haben Sie ein gutes Beispiel, in dem meist zwei, gelegentlich auch drei Formate kombiniert zum Einsatz gekommen sind. Sie haben eine schriftliche Prüfung über theoretische Themen zu Auto und Verkehr abgelegt, hatten eventuell auch ein mündliches Gespräch, am Ende hatten Sie eine Prüfungsfahrt, die beobachtet und beurteilt worden ist.

Andererseits sind es die Fragen nach den **vorhandenen Ressourcen** und der *zeitlichen Verfügbarkeit* dieser Ressourcen, die einen wichtigen Einfluss auf die Wahl des Formates haben. Eine kurze Übersicht über den Arbeitsaufwand bei unterschiedlichen Prüfungsformaten (neben der Objektivität und Reliabilität) ist in Abbildung 1 zu sehen.

² Auf Seite 47 wird auch die Validität anhand dieses Beispiels erklärt.

Prüfungsformate	Multiple-Choice	Short-Answer	Mündlich	Beobachtend
Schwerpunkt	Wissen, Verständnis, Anwendung	„aktive“ Formulierung	Mündlich- sprachlicher Ausdruck	Praktische Fertigkeiten
Objektivität	ooo	oo	o	o (o)
Reliabilität	oo	o	o	o
Aufwand				
Entwicklung	ooo	oo	oo	ooo (o)
Durchführung	o	o	ooo	oo (o)
Auswertung	o	oo	oo	o (o)

Abbildung 1: Test-Gütekriterien und Aufwandsabschätzungen verschiedener Prüfungsformate. Das MC-Format schneidet sehr gut bei Objektivität und Reliabilität ab, der Arbeitsaufwand konzentriert sich auf die Erstellung der Fragenitems. Näheres zur Objektivität und Reliabilität im Abschnitt Qualitätssicherung.

Der **Aspekt des Aufwandes** einer schriftlichen Prüfung mit geschlossenen Fragestellungen einerseits und offenen Fragestellungen andererseits sei hier kurz erläutert. Bei geschlossenen Fragen liegt die Hauptarbeit im Schreiben der Prüfungsfragen, die Durchführung und Auswertung ist dazu verglichen weniger aufwändig, gerade die Auswertung geschieht heute bereits automatisiert. Im Vergleich dazu sind offene Fragen etwas schneller zu erstellen, wengleich neben der Frage auch der Antwortschlüssel (eine systematische Aufzeichnung einer korrekten *Muster-Antwort*) erstellt werden muss. Neben einem vergleichbaren Aufwand bei der Durchführung liegt hier jedoch der Schwerpunkt des Arbeitsaufwandes im Korrigieren und Bewerten der gegebenen Antworten.

Schriftliche Prüfungsformate mit geschlossenen Fragen können sowohl bei *Paper-Pencil-Prüfungen* wie auch bei *Online- oder Computer-Prüfungen* eingesetzt werden. Darüber hinaus hat man jedoch den weiteren Vorteil, dass sie sich hervorragend für das *Self-Assessment* eignen. Beispielsweise kann man eine Muster- oder Probe-Prüfung, bestehend aus geschlossenen Fragen auf einer webbasierten Lernplattform, anbieten, damit Studierende ihre eigene Leistung bzw. ihre Vorbereitung auf die *eigentliche Prüfung* besser einschätzen können.

Die **Nachteile einer MC-Prüfung** liegen in der Einschränkung der überprüfbareren Wissensarten, einer möglichen Mehrdeutigkeit bei Formulierungen und möglicher Fehlbeurteilung von *Halbwissen*. Zusätzlich können Testfertigkeit (*test-wisenesness*, Näheres dazu auf Seite 46), ungewollte Lösungshinweise und einfaches Raten eine korrekte Deutung der Testergebnisse erheblich erschweren.

Prüfen mit Multiple-Choice-Fragen

Multiple-Choice-Fragen werden auch als *Mehrfachauswahl-* oder *Antwort-Wahl-Verfahren* bezeichnet. Ihr Hauptmerkmal ist das Vorhandensein mehrerer **vorformulierter Antwortalternativen**. Damit handelt es sich – wie bereits erwähnt – um einen geschlossenen Fragentyp, im Gegensatz zum offenen Fragentyp, bei dem Studierende eigene Antworten formulieren müssen.

Der Terminus *Multiple-Choice-Frage* (kurz: MC-Frage oder MC-Fragenitem) umfasst jedoch viele unterschiedliche Formate, die eben nur die Tatsache gemein haben, dass vorgefertigte Antwortalternativen zur Verfügung stehen. Angenommen, man bietet mehrere Antwortalternativen an, so können entweder eine oder mehrere Alternativen richtig sein. Zudem kann die Anzahl der richtigen Antwortalternativen entweder bekannt oder nicht bekannt sein.

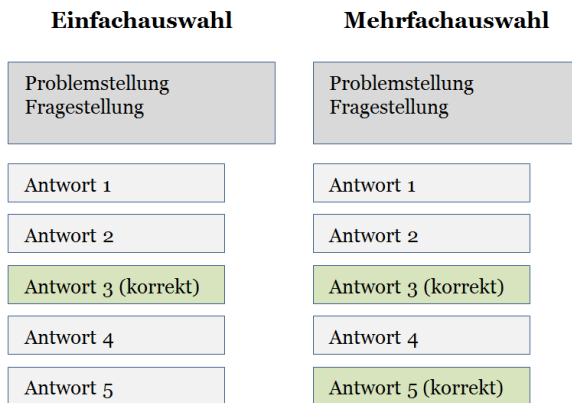


Abbildung 2: Verschiedene Varianten einer MC-Frage. Eine Antwortalternative ist korrekt (links) oder mehrere Antwortmöglichkeiten sind korrekt (rechts).

Die Variante einer richtigen Antwortalternative unter fünf gegebenen Antwortmöglichkeiten wird oft auch als Single-Choice-Frage bezeichnet. Der Terminus Multiple-Choice-Frage trifft dann zu, wenn mehrere richtige Antwortmöglichkeiten in einer Frage vorhanden sind. Korrekter wäre es jedoch, von Single- und Multiple-Response-Fragen zu sprechen, da immer mehrere Antwortmöglichkeiten zur Auswahl stehen (*Choices*). Die klassische Variante wird auch als „1-aus-5-“ oder „Best-of-five-Frage“ oder Frage des Typs A bezeichnet.

Beispiel 2 – eine Single-Response-Frage³:

Wachstumsbedürfnisse nach Maslow sind:

- (a) *Bedürfnisse nach Essen*
- (b) *Bedürfnis nach Selbstverwirklichung (korrekt)*
- (c) *Bedürfnisse nach Gesundheit*
- (d) *Bedürfnis nach Wertschätzung*

Beispiel 3 – eine weitere Single-Response-Frage⁴:

Wird ein Darlehen zur Gänze am Ende der Laufzeit zurückbezahlt, so spricht man von einer:

- (a) *Annuitätentilgung*
- (b) *Ratentilgung*
- (c) *Endfälligen Tilgung (korrekt)*
- (d) *Rententilgung*
- (e) *Keine der angeführten Antworten ist korrekt*

Beispiel 4 – eine Multiple-Response-Frage⁵:

Das Parlament ist zuständig für:

- (a) *den Vollzug des Gesetzes*
- (b) *die Genehmigung des Budgets (korrekt)*
- (c) *die Oberaufsicht über die Justiz (korrekt)*
- (d) *die Leitung der Verwaltung*

Inhaltlich werden MC-Fragen oft in **zwei Gruppen** unterteilt, in das **Richtig-Falsch-Format** und das **Beste-Antwort-Format**. Ersteres wird eingesetzt, wenn es um eine eindeutig zu beantwortende Frage geht, bei der eine Trennung

³ Zur Verfügung gestellt von Frau Gasteiger-Klicpera, Graz.

⁴ Zur Verfügung gestellt von Frau Sommersguter-Reichmann, Graz.

⁵ Beispiel aus den Schulungsunterlagen des IML der Universität Bern: „Multiple Choice Fragen? Ja, aber richtig“ (2008).

zwischen korrekt und inkorrekt indiskutabel im Sinne einer *Schwarz-Weiß-Entscheidung* möglich ist. Dies trifft auf viele Faktenfragen in Grundlagenbereichen zu. Das Zweite wird bei anwendungsorientierten und praxisnahen Fragestellungen eingesetzt, bei der mit Wahrscheinlichkeiten gearbeitet wird, die einer *Graustufenabwägung* entsprechen. Auf Seite 36 wird näher auf die Eindeutigkeit auf inhaltlicher Ebene eingegangen.

Beispiel 5 – eine Richtig-Falsch-Frage mit 2 Antwortalternativen:

Nach der kognitiven Sichtweise verändert sich Verhalten nicht, weil es von anderen verstärkt oder bestraft worden ist. Vielmehr wird es durch Zielanreize in Gang gesetzt und gesteuert.

- (a) Richtig
- (b) Falsch

Beispiel 6 – eine Richtig-Falsch-Frage mit 5 Antwortalternativen:

Welche Frequenz besitzt der Kammerton a'?

- (a) 220 Hertz
- (b) 420 Hertz
- (c) 440 Hertz (korrekt)
- (d) 450 Hertz
- (e) 880 Hertz

Beispiel 7 – eine Beste-Antwort-Frage:

Ein weibliches Frühgeborenes 35+5 SSW fällt am 3. Lebenstag wegen Hyperexzitabilität auf. Bei der klinischen Untersuchung fällt weiter eine verlängerte Rekapillarierungszeit und ein insgesamt fahl-blasses Hautkolorit auf. Im Blutbild ist eine deutliche Linksverschiebung mit einer IT-Ratio von 0,23 zu sehen, das CRP beträgt 36,5 mg/L.

Welche Diagnose trifft am ehesten zu?

- (a) IRDS
- (b) Wet lung
- (c) Kongenitale Zwerchfellhernie
- (d) Early-onset Sepsis (korrekt)
- (e) Vitium cordis

Um die Einteilung in das Richtig-Falsch- bzw. Beste-Antwort-Format etwas detaillierter auszuführen, werden hier kurz unterschiedliche Fragetypen vorgestellt. Zu den Beste-Antwort-Formaten gehört neben dem Typ A (Best-of-five) auch der Typ R, welcher bis zu 26 Antwortalternativen enthalten kann. Gerade beim computerunterstützten Prüfen etabliert sich zurzeit das sehr ähnlich aufgebaute *Long-Menu-Format*. Der Computer vergleicht die Studierendeneingaben mit der großen Liste an hinterlegten Antwortalternativen und bietet in Echtzeit die korrekte Schreibweise an. Der Typ PickN ist dem Typ A ähnlich, jedoch können mehr als eine Antwortalternative richtig sein.

In der Gruppe der Richtig-Falsch-Formate gibt es den Typ Kprim (der den Typ K ersetzen soll), der eine vierfache Richtig-Falsch-Entscheidung verlangt, sowie den Typ E, bei dem es gilt, zwei Aussagen und deren kausale Verknüpfung auf Korrektheit zu prüfen.

Der Typ K ist vielen Fragenerstellenden und Studierenden hinlänglich bekannt, soll jedoch weitgehend vermieden werden, da bei diesem Typ eine Beantwortung mit Teilwissen häufig zu einer korrekten Beantwortung führen kann.

Beispiel 8 – eine Typ-K-Frage (zu vermeiden!):

Welche der vier Aussagen zur Dateibenennung von Webpages ist/sind richtig?

- (1) keine Umlaute verwenden*
- (2) ausschließlich Großschreibung verwenden*
- (3) statt Leerzeichen den Unterstrich verwenden*
- (4) .html nicht als Dateiendung verwenden*
- (a) 1, 2 und 3 sind richtig*
- (b) 1 und 3 sind richtig (korrekt)*
- (c) 2 und 4 sind richtig*
- (d) nur 4 ist richtig*
- (e) alle sind richtig*

Beispiel 9 – eine Typ-Kprim-Frage (zu bevorzugen!):

Worauf ist bei der Dateibenennung von Webpages zu achten?

- (a) keine Umlaute verwenden (korrekt)*
- (b) ausschließlich Großschreibung verwenden*
- (c) statt Leerzeichen den Unterstrich verwenden (korrekt)*
- (d) .html nicht als Dateiendung verwenden*

Wie man am Beispiel 8 sehen kann, können Studierende mit Halbwissen – beispielsweise, dass Aussage (2) falsch ist – sehr schnell eine 50%-50%-Entscheidung herbeiführen, da sofort die Antworten (a), (c) und (e) eliminiert werden können. Da Aussage (4) leicht zu beantworten ist, müssen sich die Studierenden über Aussagen zu Umlauten und Leerzeichen nicht einmal Gedanken machen. Beim Kprim-Fragentyp werden alle vier Aussagen jedoch getrennt voneinander überprüft. Deshalb ist dieser Fragentyp bevorzugt einzusetzen.

Praxistipp:

Wenn Sie einen weiteren interessanten Fragentyp ausprobieren möchten, sehen Sie sich „Confidence-based Marking“ an⁶, welches Sie auch über die Lernplattform Moodle ausprobieren und anwenden können.

Aufbau einer guten MC-Frage

Wie bereits im ersten Teil erklärt, bildet die **Problemstellung** (oft auch als Vignette bezeichnet) zusammen mit der *Fragestellung* den **Fragenstamm**, darauf folgen die Antwortalternativen, wobei die falschen Antwortalternativen als **Distraktoren** bezeichnet werden. In der englischsprachigen Fachliteratur wird der Stamm als „stem“ und die Fragestellung als „lead-in“ bezeichnet, die Antwortalternativen als „options“, wobei die korrekte als „key“ und die inkorrekten als „distractors“ bezeichnet werden.

Dabei kann der Fragenstamm auch länger formuliert werden, sofern die Antwortalternativen so prägnant wie möglich formuliert sind (siehe Abbildung 3). Anders ausgedrückt:

⁶ <http://www.ucl.ac.uk/lapt/>

Merkregel:

Die zentrale Aussage gehört in die Frage, nicht in die Antwortalternativen!

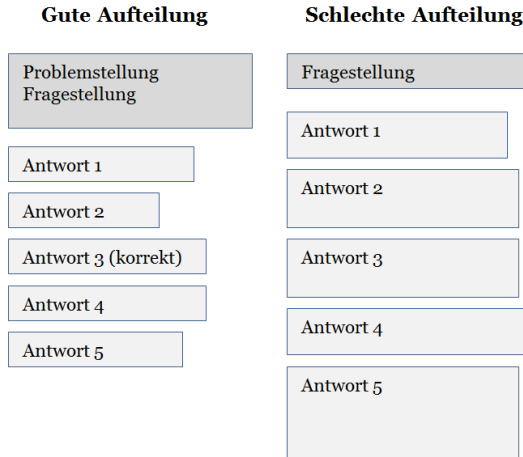


Abbildung 3: Der Vergleich einer guten und einer schlechten Aufteilung der Textmenge innerhalb einer MC-Frage.

Die Problemstellung soll realitätsnahe und positiv formuliert sein und alle für die Lösung notwendigen Informationen enthalten. Zusätzlich können alle AV-Medien (Audio-Visuelle-Medien) eingebaut werden, Näheres dazu auf Seite 35. Die Antwortalternativen sollen nach Möglichkeit homogen aufgebaut und ähnlich prägnant formuliert sein. Detaillierter werden die einzelnen Elemente ab Seite 33 beschrieben.

Allgemeines zur Qualität einer MC-Frage

Auch wenn viele Punkte in diesem Abschnitt banal erscheinen, so beeinflussen sie maßgeblich die Qualität einer Prüfungsfrage. Wichtig ist es, dass MC-Fragen eindeutig formuliert sind, keine Rechtschreibfehler und Tippfehler aufweisen sowie Abkürzungen und Fachbegriffe verwenden, die jeder Kandidatin bzw.

jedem Kandidaten bekannt sind. Sollten Abkürzungen und Fachbegriffe vorkommen, so sind diese so zu formulieren bzw. mit einer Erklärung zu ergänzen, dass es zu keinen Missverständnissen kommen kann.

Tipp- und Rechtschreibfehler können zum Teil ungeahnt schwerwiegende Auswirkungen haben, wie in Beispiel 10 zu sehen ist.

Beispiel 10 – eine Antwortalternative⁷:

„Eiweiß und Kaliummangel“ oder „Eiweiß- und Kaliummangel“?

Daneben ist es ebenso wichtig zu betonen, dass MC-Fragen stilistisch möglichst einfach formuliert sein sollen und Fangfragen ebenso wie kompliziert verschachtelte Satzkonstruktionen zu vermeiden sind.

Merkregel:

So kurz formuliert wie möglich, so lang wie für das Verständnis notwendig!

Meistens ist es nicht das erklärte Ziel, Textverständnis oder logisches Denken der Studierenden zu bewerten, auch wenn es das in Ausnahmefällen sein kann. In diesem Zusammenhang sollte auch auf Studierende hingewiesen werden, deren **Muttersprache** nicht die Sprache ist, in der die MC-Prüfung abgehalten wird. Dies soll aber nicht bedeuten, dass beispielsweise englischsprachige MC-Fragen bzw. -Prüfungen nicht sinnvoll oder gar unmöglich sind.

Rolf Porst hat in seiner Arbeit mit dem Titel *Question Wording*⁸ **zehn Gebote** formuliert, **wie Fragen formuliert werden sollen**. Kurz zusammengefasst lauten diese: einfache und unzweideutige Begriffe verwenden; lange und komplexe Fragen ebenso wie hypothetische Fragen vermeiden; doppelte Stimuli und Verneinungen sowie Unterstellungen und Suggestivfragen vermeiden; Fragen sollen einen eindeutigen zeitlichen Bezug haben; die Antworten sollen

⁷ Beispiel einer Antwort aus dem Fragenpool der Medizinischen Universität Graz.

⁸ Porst, R. (2009): Question Wording – Zur Formulierung von Fragebogen-Fragen. In: GESIS How-to-Reihe, Nr. 2, S. 11.

erschöpfend und disjunkt sein; der Fragenkontext darf sich nicht auf die Beantwortung auswirken und unklare Begriffe müssen definiert werden.

Sollte eine **Frage negativ formuliert** sein, also eine Negation beinhalten, so ist diese deutlich und nach Möglichkeit einheitlich hervorzuheben. Doppelte Verneinungen (beispielsweise in der Fragen- und Antwortformulierung zugleich) sind generell zu vermeiden. Laut vorherrschender Lehrmeinung⁹ sollen negativ formulierte Fragen nur dort zur Anwendung kommen, wo es auch darum geht, bestimmte Ausnahmen zu kennen.

Beispiel 11 – eine negativ formulierte MC-Frage¹⁰:

Sie wollen Ihre Präsentationsfolien möglichst wirksam gestalten. Welcher der folgenden Punkte dient diesem Ziel am wenigsten?

- (a) pro Folie nur ein Thema ansprechen*
- (b) pro Folie wenig Schlüsselwörter aufführen*
- (c) große Schrift verwenden*
- (d) mehrere Schriften einsetzen (korrekt)*
- (e) auf Farbkontraste achten*

Beispiel 12– eine weitere negativ formulierte MC-Frage¹¹:

Paul Ekman hat folgende kulturelle Darbietungsregeln beschrieben, außer:

- (a) Deintensivierung*
- (b) Maskierung*
- (c) Neutralisierung (korrekt)*
- (d) Supplikation*
- (e) Übertreibung*

Ebenso problematisch wie negative Fragen sind für Studierende mehrere Fragen in ein und demselben MC-Fragen-Item, d. h., wenn eine Fragestellung aus

⁹ Haladyna, T. M. & Downing, S. M. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334.

¹⁰ Beispiel aus den Schulungsunterlagen des IML der Universität Bern: „Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen“ (2002).

¹¹ Beispiel aus den Schulungsunterlagen des IML der Universität Bern: „Multiple Choice Fragen? Ja, aber richtig“ (2008).

zwei oder mehr Teilfragen besteht. Solche Fragenformulierungen sind dringend zu vermeiden.

Bedenken sollte man auch die **zeitliche Aktualität** einer Frage. Sollte sich diese auf eine kurze Zeitspanne beziehen oder auf ein Ereignis, welches sich in naher Zukunft ändern könnte, bedeutet dies auch, dass die Frage schneller zu überarbeiten sein wird. Wenn möglich sollte sich die/der Fragenerstellende diese Fragen notieren, um sie rechtzeitig überarbeiten zu können.

Es kommt zudem nicht selten vor, dass es **unterschiedliche Lehrmeinungen** zu ein und demselben Thema gibt. Hier kann nur geraten werden, entweder mögliche Verwirrungen von vornherein auszuschließen (Verweis auf Lehrveranstaltung, Lehrbuch oder Lehrmeinung selbst) oder alle Antwortalternativen und damit Meinungen als korrekt zu werten.

Wissens-, Verständnis- und Anwendungsfragen

Angelehnt an Benjamin Bloom und seine **Lernzieltaxonomie** sollte hier kurz der Unterschied zwischen diesen unterschiedlichen Fragearten erläutert werden. Seine Taxonomie besteht aus sechs Kategorien, die ersten drei lauten: Wiedergeben, Verstehen und Anwenden. So würde man beispielsweise bei einer Frage, die auf reines Wiedergeben abzielt, nach der Definition der statistischen Lagemaße *Mittelwert* und *Median* fragen. Eine Verständnisfrage würde darauf abzielen, sich eine Begründung geben zu lassen, warum in einem gegebenen Fall der Median und nicht der Mittelwert die richtige Kenngröße ist. Möchte man das Ganze als Anwendungsfrage adaptieren, könnte man die Daten (und deren Gewinnung) direkt präsentieren und abschließend nach dem sinnvollsten Lagemaß fragen. Nach Benjamin Bloom haben sich auch andere Wissenschaftler/innen mit solchen Kategorisierungen beschäftigt, die Bloom'schen abgeändert oder komplett andere Einteilungen beschrieben.

Beispiel 13 – eine Verständnisfrage¹²:

Eine Forschergruppe konnte experimentell zeigen, dass Männer bessere Leistungen in Raumvorstellungstests erbringen als Frauen. Wie wird das Geschlecht in diesem Untersuchungsplan bezeichnet?

- (a) Störvariable
- (b) unabhängige Variable (korrekt)
- (c) latente Variable
- (d) Moderatorvariable
- (e) Mediatorvariable

Beispiel 14 – eine kurze Anwendungsfrage¹³:

Wie lautet der Modalwert dieser Zahlenreihe: 1 1 2 3 8?

- (a) 1,0 (korrekt)
- (b) 1,5
- (c) 2,0
- (d) 2,5
- (e) 3,0

Josef Smolle¹⁴ beschreibt in seinem Buch *Klinische MC-Fragen rasch und einfach erstellen* (der Literaturliste auf Seite 93 zu entnehmen) die typischen Vorurteile gegenüber MC-Fragen – allen voran das Vorurteil, dass MC-Fragen nicht höhere kognitive Fähigkeiten prüfen könnten – und wie man diesen einfach und kreativ begegnet.

Herangehensweise und Hilfsmittel

Vor dem eigentlichen Erstellen neuer MC-Fragenitems sollte man sich Zeit für einige **Vorbereitungen** nehmen. Das bedeutet, alle fachspezifischen Unterlagen bereitzulegen, die man während der Erstellung benötigen könnte: Vorlesungsfolien, Lehrbücher, Lernziel- und Schlagwortkataloge sowie *Blueprints*. Beim *Blueprint* handelt es sich um eine „gewichtete Rasterung der Prüfungsin-

¹² Schulungsunterlagen des Psychologie-Institutes der Universität Wien.

¹³ Schulungsunterlagen des Psychologie-Institutes der Universität Wien.

¹⁴ Smolle, J. (2008). *Klinische MC-Fragen rasch und einfach erstellen*. Berlin, Boston: De Gruyter.

halte“ und diese wird im nachfolgenden Abschnitt „Strukturiertes Arbeiten“ näher erklärt.

Daneben benötigt man eventuell **Leitfäden** und **Ratgeber** zur MC-Fragenerstellung, wie beispielsweise den vorliegenden Leitfaden, sowie **Checklisten**, anhand derer man die erstellten Fragen überprüfen kann. Zusätzlich können auch Formulare und **Vorlagen** hilfreich sein, in die man die entsprechenden Fragen-Inhalte einträgt. Zudem sollte man sich als Autor/in mit einer bestehenden Fragen-Verwaltungssoftware oder bestehenden Peer-Review-Systemen vertraut machen. Gleiches gilt für den Fall, dass mit Lernplattformen wie beispielsweise Moodle¹⁵, ILIAS¹⁶ oder OLAT¹⁷ gearbeitet wird.

Es gibt zahlreiche Möglichkeiten, an die Multiple-Choice-Fragenerstellung heranzugehen – zahlreiche Aspekte, die variiert werden können: Wer arbeitet an der Fragenerstellung bzw. wer arbeitet mit? Wie arbeitet man – mit Papier und Stift, am Computer oder am Tablet-PC? Bringt das Erstellen neuer Fragen Vor- oder Nachteile im Vergleich zum Überarbeiten von Altfragen?

Praxistipp:

Behalten Sie bei Ihren Vorbereitungsarbeiten für eine Lehrveranstaltung oder beim Erstellen der Vorlesungsunterlagen das Prüfen der Inhalte im Hinterkopf!

STRUKTURIERTES ARBEITEN

Es gibt grundsätzlich zwei Möglichkeiten, wie man an die Erstellung neuer MC-Frageitems herangehen kann – eine **strukturierte** (diskursive) und eine **creative** (intuitive) Methode.¹⁸ Bedenken sollte man allerdings, dass diese beiden

¹⁵ <http://moodle.org>

¹⁶ <http://www.ilias.de>

¹⁷ <http://www.olat.org>

¹⁸ Hier wurden bewusst die Bezeichnungen „diskursiv“ und „intuitiv“ verwendet, wie sie auch im Abschnitt Kreativitätstechniken vorkommen.

Arten nicht miteinander in Konkurrenz stehen, sondern sich ergänzen, wobei man mit der strukturierten Vorgehensweise beginnen sollte.

Zu Beginn sollte man sich über den **Umfang, die Tiefe und die Gewichtung des prüfungsrelevanten Stoffes** Gedanken machen. Diese Strukturierung geschieht beispielsweise mit der Erstellung des oben bereits erwähnten Blueprints – dieser besitzt zumeist zwei oder mehr Dimensionen. Jede dieser Dimensionen wird dabei in mehrere Kategorien unterteilt.

Beispiel 15 – zweidimensionale Unterteilungen aus verschiedenen Fachbereichen:

Medizin: erste Dimension „Krankheiten der Atemwege“, zweite Dimension „Aspekte einer Erkrankung (Entstehung, Klinisches Bild, Diagnostik, Therapie, Prognose...)“

Kunst: erste Dimension „Kunstepochen“ (Romantik, Barock, Klassizismus...), zweite Dimension „Kunstrichtungen“ (Literatur, Musik, Bildnerische Kunst...)

Physik: erste Dimension „physikalische Gesetze“, zweite Dimension „Aspekte“ (Formale Beschreibung/Formel, Bedeutung, Anwendungsbeispiele...)

Anschließend überlegt man sich die Tiefe der einzelnen Themenblöcke: Bei welchen Themen ist mehr Detailwissen gefragt – bei welchen nicht? So sind in der Medizin naturwissenschaftliche Grundlagen von Bedeutung, allerdings müssen diese nicht detaillierter vermittelt werden als die medizinischen Inhalte selbst. Viele Randthemen sollen nicht „unter den Tisch fallen“, wie stark sie jedoch gewichtet werden, hängt von den jeweiligen Lehrenden ab.

Bei der Gewichtung werden schwerpunktmäßig bestimmte Themenbereiche hervorgehoben und betont, mögliche Gründe können sein:

- Häufig in der Praxis vorkommend
- Relevant für die spätere Berufsausübung
- Verbreitete Fehlmeinung
- Komplexes und schwieriges Thema
- Relevant für die weitere Ausbildung

Ebenso zu bedenken sind Betonungen von Themenbereichen aufgrund der ausführlichen Behandlung in der entsprechenden Lehrveranstaltung. Wurde in der Lehrveranstaltung auf Inhalte besonderen Wert gelegt oder wurden Inhalte bewusst weniger detailreich behandelt?

Ein wichtiger Aspekt diesbezüglich ist auch das Setzen unterschiedlicher Schwerpunkte bei Prüfungen und Prüfungsterminen, um den Studierenden ein **Taktieren zu erschweren**. Studierende spekulieren natürlich über mögliche Schwerpunkte und passen dementsprechend ihr Lernen an.

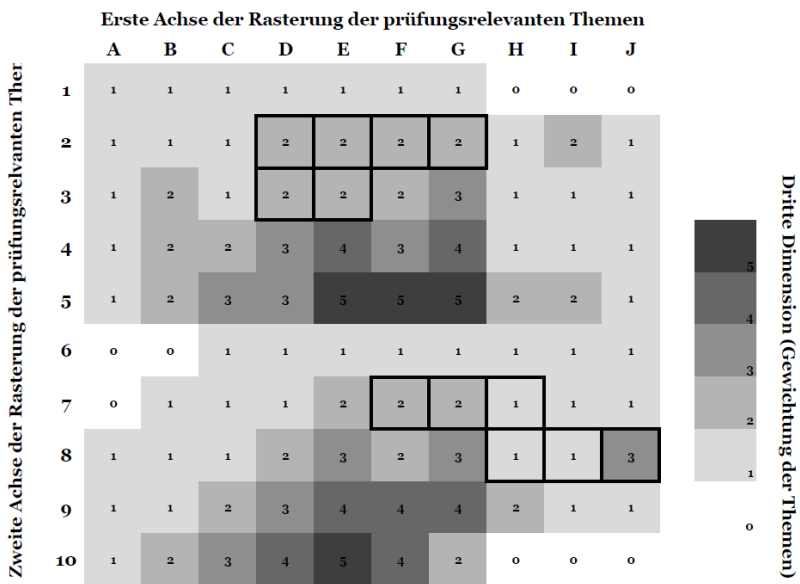


Abbildung 4: Eine schematische Darstellung einer Rasterung der prüfungsrelevanten Inhalte. Die Graustufen stehen für die Tiefe/für den Detailreichtum in einem Bereich, die Rahmen weisen auf sehr praxisnahe Inhalte hin (bspw. Erstellung anwendungsorientierter Fragenitems).

Die eben vorgestellte zweidimensionale und mit Schwerpunkten ergänzte Aufschlüsselung könnte um die Dimensionen „Art des Wissens“ und „Art des kog-

nitiven Prozesses“ ergänzt werden. Die erste Dimension zielt auf die verschiedenen **Formen von Wissen**¹⁹ ab: Fakten-, Konzept-, Prozedural- oder metakognitives Wissen. Die andere Dimension auf den **kognitiven Prozess**, wie er auch in der Lernziel-Taxonomie von Benjamin Bloom verwendet wird. Möchte man eine Frage stellen, die auf simple Wiedergabe oder auf Verständnis abzielt, oder möchte man, dass Studierende ihr Wissen in einer neuen Situation anwenden? Diese drei Kategorien – *Wiedergeben*, *Verständnis* und *Anwendung* – findet man unter anderem in der Bloom’schen Lernziel-Taxonomie, wie sie bereits zuvor auf Seite 22 beschrieben wurde.

Diese strukturierte Vorgehensweise ist jedoch nicht nur auf Fachebene, auf der das Thema grob strukturiert wurde, sinnvoll, sondern auch im Detail, wenn man einzelne inhaltliche Elemente ansieht. Hierfür zwei Beispiele aus unterschiedlichen Fachrichtungen:

Beispiel 16 aus Geschichte:

Kriegerische Auseinandersetzungen: Wann und wo hat die Auseinandersetzung stattgefunden? Zwischen welchen Konfliktparteien? Was waren allgemeine und spezielle Voraussetzungen dafür? Gab es einen Auslöser? Welche Phasen kann man unterscheiden bzw. welche besonderen Ereignisse sind hervorzuheben? Welche Konsequenzen hatte der Konflikt a) für beide Parteien, b) für Verbündete und c) für neutrale Länder?

Beispiel 17 aus dem Transportwesen:

Flugverkehr: Welche Arten von Flugverkehr gibt es? Wie und warum hat sich der Flugverkehr in den letzten 30 Jahren entwickelt? Was sind seine Vor- und Nachteile? Welche Konsequenzen ergeben sich daraus? Wie könnte man den Schadstoffaustausch und die Lärmbelastung eindämmen?

Man könnte eine solche strukturierte Vorgehens- oder Denkweise auch anders erklären – die nächsten beiden Absätze beschreiben zudem einen guten Übergang zu kreativeren Überlegungen:

In unserer (Wissens-)Welt gibt es „statische“ **Begrifflichkeiten und Handlungen** oder Prozeduren. Wir beschreiben reale Objekte, Personen und abs-

¹⁹ http://de.wikipedia.org/wiki/Wissen#Formen_des_Wissens

trakte Begriffe einerseits, **Fragewörter** (Interrogativpronomen) wie „wer“ und „was“ werden hier typischerweise eingesetzt. Auf der anderen Seite kann man nach Handlungen mit „wie“ (...geschieht etwas?) fragen. Wie sieht die Abfolge an Schritten für eine Prozedur aus? Was geschieht, wenn man die Reihenfolge der Schritte verändert? Wie weiß man, dass ein Schritt beendet ist und der nächste folgen kann? Neben weiteren Fragewörtern wie „wann“ oder „wo“ kann man jedoch auch mit „warum“ oder „weshalb“ fragen. Mit MC-Fragen, die nach dem Warum fragen, wird in vielen Fällen tieferes Verständnis abgefragt als mit einfachem Was. Beispielsweise könnte man statt nach der Bundeshauptstadt Deutschlands nach dem Grund für Berlin als Hauptstadt fragen. Studierende, die diese Gründe verstanden haben, werden auf alle Fälle auch die Hauptstadt richtig benennen können.

Zuletzt zwei weitere Anregungen, wie man andere interessante Fragestellungen finden könnte. Entweder man kehrt eine Fragestellung um: Statt beispielsweise zu fragen, welche Konsequenzen es hat, wenn ein bestimmter Bestandteil eines Ottomotors defekt ist, könnte man **umgekehrt fragen**, welcher Teil des Motors defekt sein könnte, nachdem man die „Symptome“ des nicht einwandfrei funktionierenden Motors beschrieben hat. Auf der anderen Seite könnte man bei bestimmten Fragen einfach einen Schritt weiterdenken. Statt nach einer bestimmten Antwort zu fragen, geben Sie eine richtige oder falsche Antwort vor und **fragen nach der Begründung** – ähnlich dem Berlin-Beispiel zuvor.

KREATIVITÄTSSTECHNIKEN

Es handelt sich dabei um sehr unterschiedliche Methoden zur Förderung der Kreativität. Auch wenn es sogenannte Geistesblitze gibt, so ist die Ideengenerierung ein ziemlich strukturierter Prozess, den man beeinflussen und steuern kann. Auch wenn damit keine Lösungen oder richtige Ergebnisse garantiert werden können, so sind es dennoch Verfahren, die sich in der Praxis als brauchbar und zielführend erwiesen haben.

Eine **kreative Person**, die offen und neugierig und zugleich konzentriert und fokussiert ist, sich jedoch weder geistig einschränken noch frustrieren lässt, ist ebenso notwendig wie ein **kreatives Umfeld**, welches je nach Vorlieben Ein-

drücke und Anregungen oder Ruhe bringt, um einen kreativen Prozess zu unterstützen.²⁰

Laut Csikszentmihalyi²¹ kann man den **kreativen Prozess** selbst in fünf Phasen unterteilen: Vorbereitung – Reifung – Einsicht – Bewertung – Ausarbeitung. Wichtig ist hier vor allem die intensive Auseinandersetzung mit dem Problem, denn diese ist Voraussetzung, damit sich während der Reifungsphase überhaupt kreative Ideen oder *Geistesblitze* einstellen können.

Viele der oben genannten Punkte sind sehr allgemein formuliert, kommen einem bekannt vor und geben jedoch nur bedingt eine gute Anleitung ab. Diese wäre aber notwendig, da wir oft mit geistigen Blockaden kämpfen, unabhängig davon, ob unsere Probleme kreativer oder systematischer Natur sind.

Kreative Blockaden können unterschiedliche Ursachen haben. Personen mit mangelndem Selbstbewusstsein und einer zu kritischen Grundhaltung müssen unter Zeitdruck kreative, brauchbare, bezahlbare Lösungen und Ideen hervorbringen. Eine solche zielorientierte Einstellung, zusammen mit gedanklichen Schranken und Scheuklappendenken – vielleicht kombiniert mit einer Tendenz zur Perfektion –, ist genauso wenig hilfreich wie fixes Festhalten an bereits Bekanntem.

Im Folgenden werden ausschnittsweise einige diskursive und intuitive Methoden vorgestellt: **Diskursive Methoden** hingegen legen ihren Schwerpunkt auf die Qualität und haben eine systematische Vorgehensweise. Beim *Morphologischen Kasten* beispielsweise werden alle Merkmale, Faktoren oder Dimensionen einer Fragestellung untereinanderbeschrieben und mit allen erdenklichen Ausprägungen ergänzt. Danach können die einzelnen Ausprägungen der Merkmale unterschiedlich kombiniert werden. Statt eines Tisches aus Holz mit vier Beinen kommt man auf einen Glastisch, der keine Beine hat – also möglicherweise von der Zimmerdecke hängt. Eine solche Methode profitiert von mehreren Personen und einer Moderation. Bei der Verwendung einer *Osborn-Checkliste* werden anfängliche Vorstellungen systematisch modifiziert. Dabei werden Fragen verwendet wie: Kann ich etwas hinzufügen oder vervielfachen? Was passiert über einen längeren Zeitraum oder wenn etwas häufiger auftritt? Kann ich reduzieren oder vereinfachen? usw. Eine sehr technische Variante ist das Erstellen eines *Ursache-Wirkungs-Diagrammes*, bei dem die Kausalitäten

²⁰ M. Rhodes: An Analysis of Creativity. In: Phi Delta Kappan. April 1961, S. 305–310.

²¹ M. Csikszentmihalyi: Kreativität. Klett-Cotta, Stuttgart 1997.

eines Problems herausgearbeitet und skizziert werden, um diese anschließend analysieren, diskutieren und gezielt verändern zu können.

Intuitive Methoden liefern oft viele Ideen, sie sollen gedankliche Assoziationen fördern, Unbewusstes bewusst machen und helfen, eigene Denkweisen zu verlassen. Eine sehr bekannte Methode ist das *Brainstorming*, das von Alex F. Osborn beschrieben wurde. Sie eignet sich gut für weniger komplexe Probleme auf sprachlicher Ebene, ist kostengünstig und einfach. In einer kleineren Gruppe mit 5–7 Personen werden freie Assoziationen gefördert, wobei Korrekturen und gegenseitiges Kritisieren nicht erwünscht sind. *Mind Mapping* bzw. das Erstellen von *Mind-Maps* ist eng verknüpft mit Tony Buzan und dreht sich um das Erschließen und Visualisieren eines bestimmten Themengebietes. Es handelt sich dabei um ein Baumdiagramm – ergänzt mit Anmerkungen zu den Inhalten –, das beim Planen, Strukturieren und Ideensammeln hilft. Die *Kopfstandtechnik* basiert auf einer Umkehrung der anfänglichen Aufgabenstellung und ist auch ein Bestandteil von anderen Kreativitätstechniken wie der Osborn-Checkliste oder der *Provokationstechnik* von Edward de Bono. Es können auch Wege gesucht werden, ein Problem zu verschärfen, wobei das Ergebnis am Ende wieder umgekehrt wird, um zu einer Lösung des anfänglichen Problems zu kommen.

Praxistipp:

Um gute Rahmenbedingungen zu schaffen, ist es oft besser, vom Computer Abstand zu nehmen und sich mit einem leeren Blatt Papier und einem Stift zu begnügen. Ebenso kann es von Vorteil sein, seinen gewohnten Arbeitsplatz zu verlassen und sich ins Freie oder in die Natur zu begeben.

Abschließend soll zusammengefasst werden, dass viele Wege zur Lösung eines Problems oder zum Finden einer neuen Idee beitragen können: Auf der einen Seite kann systematisch probiert und variiert werden. Man kann induktiv herangehen, sich viele Beispiele ansehen und anschließend verallgemeinern. Andererseits kann man verändern, vergrößern, verkleinern, umordnen oder allgemein Raum und Zeit modifizieren. Schlussendlich kann man an das Problem

herangehen, indem man es visualisiert oder modelliert, andere Perspektiven oder Kontexte sucht.

Praxistipp:

Besser eine Frage aufschreiben und später korrigieren, anstatt sie nie niederzuschreiben...

Einzelemente einer Frage

Die Einzelemente einer MC-Frage mit Fragenstamm, bestehend aus Problem- und Fragestellung sowie den Antwortalternativen, wurden bereits auf Seite 18 vorgestellt. Diese und wichtige Details dazu werden nun näher erläutert.

Die Problemstellung

Die Problemstellung ist die Darstellung eines realistischen oder realitätsnahen Problems, auf welches sich die eigentliche Fragestellung anschließend bezieht. Diese Problemstellung kommt vor allem bei anwendungsorientierten Fragen vor, also bei Verständnis- und Anwendungsfragen nach Benjamin Bloom. Solche Problemstellungen sollen authentisch sein und sich an reale Probleme anlehnen. Sie sollen sowohl Informationen enthalten, die für die Lösung notwendig sind, jedoch auch Kontextinformationen, um die Problemstellung authentischer werden zu lassen.

Josef Smolle nennt in seinem zuvor bereits erwähnten Buch, welches auch der Literaturliste zu entnehmen ist, neben **Authentizität** und Kontextbezogenheit auch noch Situiertheit und Immersivität. Das bedeutet, dass die Fragen räumlich und zeitlich realistisch eingebettet sein sollen und die Studierenden sich gut in die geschilderten Situationen hineinversetzen können müssen – sie sollen *in ihnen aufgehen*.

C. F. Herreid zitierend, ergänzt er folgende Punkte: Man soll eine **Geschichte erzählen** und dabei auf einen interessanten Aspekt fokussieren, der in der Gegenwart oder nahen Vergangenheit angesiedelt ist. Man soll Empathie zu den geschilderten Personen herstellen, die direkte Anrede verwenden – diese Beschreibung endet dann in einem Konflikt, der vom Studierenden eine **Entscheidung** verlangt.

Er beschreibt darin auch eine Methode, um den **Schwierigkeitsgrad** einer Frage **verändern** zu können, und nennt diese dabei *Typikalität*. Auch wenn es bei der Beschreibung von Patientinnen- und Patientenfällen als Problemstellung häufiger Anwendung findet, so ist das Beschreiben eines mehr oder weniger *typischen* Sachverhaltes in beinahe allen fachlichen Bereichen möglich. Das

trifft auf juristische Fallbeschreibung genauso zu wie auf betriebswirtschaftliche Überlegungen oder technische Probleme.

Bei der Formulierung des Fragenstamms müssen viele Aspekte berücksichtigt werden. Realitätsnahe und angewandte Beispiele sind gerne **Beste-Antwort-Fragen**. Statt der einen 100%ig korrekten Lösung kommen hier alle Antwortalternativen mehr oder weniger in Frage – für die Studierenden geht es dabei um das **Abwägen von Wahrscheinlichkeiten**. Auf der anderen Seite können anwendungsorientierte Fragen natürlich auch mit klaren Antworten arbeiten und als **Wahr-Falsch-Fragen** vorkommen. Die Problemstellungen solcher Fragen können auch umfangreicher sein und viele Einzelinformationen enthalten, wodurch jedoch das Abwägen der Wahrscheinlichkeiten erschwert wird. Auf der anderen Seite hat die/der Fragerstellende den Vorteil, dass viele Einzelinformationen sehr flexibel variiert und ausgetauscht werden können. Man kann somit sehr rasch die Anzahl an Einzelfragen ohne größeren Aufwand erhöhen.

Praxistipp:

Denken Sie an einen weiteren und für die Praxis sehr wesentlichen Vorteil anwendungsorientierter (MC-) Fragen: Sie können Ihre praktische und berufliche Erfahrung nutzen, wenn Sie derartige beispielhafte „Szenarien“ (Problemstellungen, Vignetten) beschreiben müssen.

Die Fragersteller/innen haben jedoch noch einen weiteren Vorteil, wenn sie anwendungsorientierte Fragen erstellen: Die Art, wie einzelne Informations-elemente in der Problemstellung präsentiert werden, ändert den Schwierigkeitsgrad einer Frage deutlich. Vorinterpretierte Informationen kommen den Studierenden entgegen, im Gegensatz dazu müssen **rohe Informationen** zuvor interpretiert werden, wie im folgenden Beispiel zu sehen ist:

Beispiel 18 – „roh“ oder „interpretiert“:

Rohe Präsentation von Information:

„Brille mit -1.5 Dioptrien“

Interpretierte Präsentation von Information:

„Brille zur Korrektur einer Kurzsichtigkeit“

Bildmaterialien, bei computergestützten Prüfungen auch generell **AV-Materialien** (Audio und Video), können jederzeit in Fragenitems – am besten in den Problemstellungen, aber auch in den Antwortalternativen – eingefügt werden. Entweder man verwendet Fotos, die zur Beurteilung des Sachverhaltes benötigt werden, oder es handelt sich um Abbildungen, Grafiken, Skizzen oder Diagramme, die zu interpretieren sind. Wichtig ist, dass diese Bildelemente nicht nur *lose* beigelegt werden, sondern Teil der Frage oder Problemstellung sind und für die Lösung der Frage entscheidend sind. Selbstverständlich muss garantiert sein, dass diese Elemente einwandfrei dargestellt werden und unmissverständlich zu deuten sind.

Bedenken Sie bei computergestützten Prüfungen, dass diese Bildelemente keinen Hinweis auf die korrekte Antwort der Frage liefern. Um ein Beispiel zu nennen, dürfen die Dateinamen der Bild- oder Mediendateien keinen Hinweis auf die korrekte Lösung geben.

Selbstverständlich sollten alle Materialien in ausreichender Qualität dargeboten werden. Auf korrekte Farbdarstellung, passende Auflösung, genaue Beschriftung und korrekte Größe ist genauso zu achten wie auf darstellbare Dateiformate bzw. Dateigrößen.

Praxistipp:

Gerade bei Rastergrafiken ist auf eine ausreichende Bildqualität zu achten!

Die Fragestellung

Die Fragestellung steht entweder alleine für sich oder wird anschließend an die Problemstellung präsentiert. Die Fragestellung soll essentielle Fragen in diesem Zusammenhang widerspiegeln, wobei diese direkt mit dem Problem verbunden sein muss. Entscheidend ist dabei, dass die Fragestellung eindeutig und unmissverständlich formuliert ist.

Beim Thema *Eindeutigkeit* sollen zwei Arten unterschieden werden: die *sprachliche* und die *inhaltliche Eindeutigkeit*.

Beginnend mit der inhaltlichen Eindeutigkeit: Fakten, die sehr konkret zu beschreiben und damit zu erfragen sind, kann man leicht als richtig oder falsch werten, wie im folgenden Beispiel 19 zu sehen ist. Auf der anderen Seite gibt es Fakten und damit Fragen, die angewandter und praxisorientierter sind. Diese lassen sich selten als 100%ig richtig oder 100%ig falsch bewerten, bei diesen geht es vorwiegend um Wahrscheinlichkeiten, abhängig von den genannten Rahmenbedingungen, unter denen etwas zutrifft.

Beispiel 19:

Das Prinzip „Die Auftriebskraft eines Körpers in einem Medium ist genauso groß wie die Gewichtskraft des vom Körper verdrängten Mediums“ ist nach welchem Gelehrten benannt?

- (a) Archimedes (korrekt)
- (b) Aristoteles
- (c) Newton
- (d) Galilei
- (e) Aristarchos

Diese Einteilung spiegelt sich sehr gut in den MC-Fragentypen wider, bei denen Richtig-Falsch-Fragen und Beste-Antwort-Fragen bereits erklärt wurden. Fragen nach Grundhaltungen, Meinungen und Argumentationsweisen sind dagegen mit geschlossenen Fragenformaten nur schwer zu formulieren. Hier wären schriftliche Formate mit offenen Fragen oder mündliche Formate besser geeignet, auf die Argumentationsweise in Hinblick auf eine Meinung zu fokussieren.

Bei der *sprachlichen Eindeutigkeit* ist es wichtig, mit präzisen Formulierungen zu arbeiten, die keinen Spielraum für Mehrdeutigkeiten zulassen. Bei den Antwortalternativen sind **absolute Häufigkeitsadverbien** wie immer, nie, absolut etc. zu vermeiden, wenn dadurch Studierende einen Lösungshinweis

erhalten. Auf der anderen Seite sind auch allzu vage Begriffe wie gewöhnlich, oft, gelegentlich oder in der Regel durch präzisere Beschreibungen zu ersetzen. Diese Präzisierung ist gut anhand der Beispiele 20 und 21 zu sehen.

Beispiel 20 – eine sehr vage Formulierung:

„Der junge Patient klagt häufig über Bauchschmerzen.“

Beispiel 21 – Umformulierung zu einer präziseren Formulierung:

„Der 8-jährige Niklas klagt seit ca. 2 Wochen über krampfartige Bauchschmerzen im rechten Oberbauch, die täglich in den Morgenstunden auftreten und 2–3 Stunden anhalten. Übelkeit oder andere Symptome sind nach seinen Aussagen noch nicht aufgetreten.“

Man versucht daher, die Problem- und die Fragestellung, jedoch auch die Antwortalternativen möglichst präzise zu formulieren, ohne von Absolut-Aussagen Gebrauch zu machen.

Kriterien für die Antwortalternativen

Alle Antwortalternativen (korrekte und inkorrekte Antwortalternativen) zusammen bzw. das „Set“ an Antwortalternativen sollen:

- homogen sein,
- so kurz wie möglich und ähnlich lang formuliert sein,
- eine Aussage pro Antwortalternative enthalten und
- keine ungewollten Lösungshinweise enthalten.

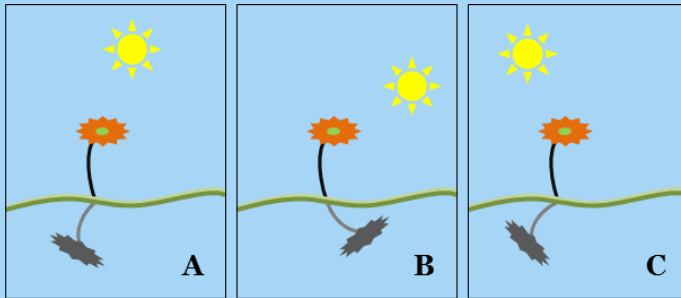
Kurz und prägnant zu formulieren ist essentiell, auch aus der Überlegung heraus, dass unnötig lange textliche Formulierungen die Wahrscheinlichkeit von Rechtschreib- und Tipp-Fehlern wie auch von mehrdeutigen Formulierungen erhöht.

Das folgende Beispiel 22 soll die **Homogenität** unter den Antwortalternativen zeigen: Auch wenn sich die drei Abbildungen im Detail unterscheiden, so sind sie dennoch völlig ident aufgebaut, umgelegt auf eine textliche Variante wäre genau diese Homogenität anzustreben.

Daneben zeigt dieses Beispiel, wie man Bildmaterial sinnvoll, in die Frage integriert einsetzen kann und, wie in diesem Fall, auch kreativ in den Antwortalternativen einsetzen kann.

Beispiel 22 – eine Abbildung homogener Varianten in den Antwortalternativen²²:

Welcher Schatten bei dieser Blume ist korrekt dargestellt?



Alle Antwortalternativen zusammen sollen die gegebenen Möglichkeiten möglichst erschöpfend und vor allem disjunkt (überschneidungsfrei) wiedergeben. Als Beispiel eine nicht überschneidungsfreie numerische Frage:

Beispiel 23 – eine numerische Frage mit überschneidenden Antwortalternativen (Medizin)²³:

Following a second episode of infection, what is the likelihood that a woman is infertile?

- (a) less than 20 %
- (b) 20 to 30 %
- (c) greater than 50 %
- (d) 90 %
- (e) 75 %

²² Die Anregung zu dieser Abbildung stammt von:
<https://sites.google.com/site/testwritingtutorial/multiple-choice-questions/writing-distracters>

²³ Beispiel aus: Case, S. M. & Swanson, D. B. (1998). Constructing written test questions for the basic and clinical sciences. National Board of Medical Examiners. Chapter 3.

Wie im zweiten Teil des Leitfadens ab Seite 77 detailliert beschrieben, sind **ungewollte Lösungshinweise**, die auch als Cues bezeichnet werden und deren Vorkommen als Cueing bezeichnet wird, zu vermeiden. Studierende erhalten durch diese formale oder sprachliche Hinweise, mit denen sie die richtige Antwort oder die falschen Antwortalternativen leichter identifizieren können. In diesem Zusammenhang sei auch auf die *test-wiseness* verwiesen, die auf Seite 46 näher beschrieben wird.

Noch ein paar Worte zur **Anzahl der Antwortalternativen**. Der klassische Fragentyp, der oben vorgestellt wurde, enthält meist fünf Antwortalternativen, von denen eine korrekt ist. Laut Literatur²⁴ können diese bis auf drei Alternativen reduziert werden, ohne deutlich schlechtere statistische Kennzahlen zur Konsequenz zu haben. Man sollte jedoch bedenken, dass damit auch die Ratewahrscheinlichkeit deutlich ansteigt. Je mehr Antwortalternativen bei einer MC-Frage angeboten werden, desto geringer ist die berechnete Ratewahrscheinlichkeit, vorausgesetzt, dass alle Antworten plausibel und sinnvoll sind.

Antwortraum und Antwortspreizung

Der künstliche Begriff des *Antwortraums* bezeichnet eine **Menge an Antwortalternativen** – aus dieser Menge können die Antwortalternativen einer Frage entnommen werden. Je größer dieser Raum ist, desto leichter ist es, geeignete Antwortalternativen zu finden. Bei einer Wahr-Falsch-Frage besteht dieser Antwortraum nur aus zwei Möglichkeiten (wahr und falsch), nimmt man beispielsweise die Himmelsrichtungen her, haben wir einen Antwortraum mit vier Möglichkeiten (N, S, E, W). Extrem große Antworträume sind jedoch ebenso zahlreich: europäische Städte, deutschsprachige Weltliteratur, Motoren-Typen und -Bauweisen, Krankheitserreger, Grammatikregeln etc.

Wichtig ist hierbei zu beachten, dass man sich nicht selbst – aufgrund der Fragestellung – zu sehr einschränkt, wie die beiden folgenden Beispiele zeigen sollen. Eine geänderte Fragenformulierung bewirkt oft eine deutliche Erleichterung bei der Suche nach Distraktoren.

²⁴ Rodriguez, M. C. (2005). Three Options Are Optimal for Multiple Choice Items: A Meta Analysis of 80 Years of Research. Educational Measurement: Issues and Practice.

Beispiel 24 – eine einschränkende Variante:

Hat die Bundeshauptstadt Wien mehr als eine Million Einwohner/innen?

- (a) Ja (korrekt)
- (b) Nein

(Hier stehen generell nur zwei Möglichkeiten zur Verfügung.)

Beispiel 25 – eine verbesserte Variante:

Welche der genannten österreichischen Städte hat mehr als eine Million Einwohner/innen?

- (a) Wien (korrekt)
- (b) Graz
- (c) Innsbruck
- (d) Linz
- (e) Salzburg

(Hier stehen Dutzende Städte (>10.000 Einwohner/innen) zur Verfügung.)

Unter dem Begriff *Antwortspreizung* versteht man die **Breite der Auffächerung** der Antwortalternativen. Damit lässt sich der Schwierigkeitsgrad der Frage gravierend beeinflussen, je breiter oder *gespreizter*, desto leichter wäre die Frage zu beantworten. Übertreibt man es in Bezug auf diesen Aspekt, würden extrem leichte und damit wenig trennscharfe Fragen zustande kommen.

Beispiel 26 – eine große Antwortspreizung²⁵:

Wer war der Verfasser der amerikanischen Unabhängigkeitserklärung?

- (a) Otto von Bismarck
- (b) Thomas Jefferson (korrekt)
- (c) Franklin Roosevelt
- (d) König Georg II
- (e) Katharina die Große

²⁵ Beispiel aus: Case, S. M. & Swanson, D. B. (1998). Constructing written test questions for the basic and clinical sciences. National Board of Medical Examiners.

Beispiel 27 – eine geringere Antwortpreizung:

Wer war der Verfasser der amerikanischen Unabhängigkeitserklärung?

- (a) George Washington*
- (b) Thomas Jefferson (korrekt)*
- (c) Alexander Hamilton*
- (d) Benjamin Franklin*
- (e) James Madison*

Eigenformulierungen, also selbst ausgedachte Antwortalternativen, die es im vorliegenden Fachbereich so nicht gibt, können im Spezialfall hilfreich sein, sind im Normalfall jedoch leicht zu durchschauen.

Kreative Beispiele zur Veranschaulichung:

- Otto von Habsberger*
- Albert Zweistein und seine Reaktivitätstheorie*
- Ultrarot (als Wortspiel zwischen Ultraviolett und Infrarot)*

Sollte es in der entsprechenden Situation passend sein, können bei Antwortalternativen auch **Kombinationen von Begriffen** und Formulierungen zielführend (siehe dazu untenstehendes Beispiel) sein, die oberste Regel lautet jedoch „eine Aussage pro Antwortalternative“.

Beispiel 28 – mit Kombinationen²⁶:

Der Inhaber der Long-Position in einem Zinsswap:

- (a) ist berechtigt, Zahlungen in variabler Höhe zu leisten.*
- (b) ist verpflichtet, Zahlungen in variabler Höhe zu leisten.*
- (c) ist verpflichtet, Zahlungen in fixer Höhe zu leisten. (korrekt)*
- (d) ist berechtigt, Zahlungen in fixer Höhe zu leisten.*

²⁶ Zur Verfügung gestellt von Frau Sommersguter-Reichmann, Graz.

Das Finden plausibler Distraktoren

Vielen Fragenersteller/innen geht es diesbezüglich ähnlich – es kann gelegentlich sehr schwierig sein, eine passende Anzahl sinnvoller Distraktoren zu finden. Viele der oben erklärten Hinweise können bei diesem Problem helfen. Allerdings gibt es auch eine andere Art, an dieses Problem heranzugehen – mit Hilfe der „*Bottom-up-Erstellung*“. Wie der Name bereits andeutet, wird hier **der umgekehrte Weg** beschritten. Normalerweise wird zuerst der Fragenstamm formuliert und anschließend die Antwortalternativen, bei der Bottom-up-Erstellung werden jedoch zuerst alle Antwortalternativen und anschließend der Fragenstamm formuliert.

Dieses Vorgehen hat **mehrere Vorteile**: Auf der einen Seite kann so leichter sichergestellt werden, dass die Antwortalternativen homogen sind. Man startet ohne Problem- oder Fragestellung, daher ist man nicht durch diese Elemente beeinflusst. Auf der anderen Seite fällt es im Fragenstamm oder – besser – in der Problemstellung leichter, die Formulierung so zu wählen, dass eben eine der zuvor genannten Antwortalternativen zur richtigen wird. Die Problemstellung enthält zahlreiche *Informationsbausteine* und diese können leichter abgeändert, ausgetauscht oder umformuliert werden, um die richtige Antwort von den falschen Antworten abzuheben.

Diese führt dann auch zu einer weiteren Methode, die Anzahl der Fragen oder deren Varianten zu erhöhen. Auch wenn es banal klingt: Man kann entweder die Frage- und Problemstellung belassen und die Antwortalternativen ganz oder teilweise austauschen; man kann jedoch auch die Antwortalternativen belassen und Frage- oder Problemstellungen abändern. Ebenso ist es möglich, die Problemstellung – bei gleichbleibender Fragestellung – abzuändern.

Mehrstufige Lösungswege

Jede MC-Frage soll zwar fokussiert sein, zwei oder mehrere Schritte zur Lösung und damit zur korrekten Antwortalternative sind jedoch prinzipiell erlaubt. Allgemein formuliert, könnte man eine (praxisnahe) Situation schildern und nach der (besten) Lösung fragen, auch wenn das den Zwischenschritt einer Problemdefinition notwendig macht. Als Beispiel könnte man in der Humanmedizin eine Patientin bzw. einen Patienten beschreiben und nach einer geeigneten Therapie fragen, auch wenn es das Finden einer Verdachts- oder Arbeitsdiagnose voraussetzt.

Überlegungen zur Cognitive Load Theory

Es gibt mehrere Gründe, die *Cognitive Load Theory* (CLT) von Chandler und Sweller zu erwähnen. Zum einen geht es um die Frage, wann man einen Fragenstamm verwenden soll, denn es ist nicht immer sinnvoll, eine Frage in einen realitätsnahen Rahmen zu betten. Eine reine Faktenfrage, die auf die Wiedergabe von Gelerntem abzielt, braucht keine „Geschichte rundherum“. Dieser Rahmen – diese Geschichte – ist nur bei anwendungsorientierten Fragen sinnvoll. Ansonsten würde sie von der eigentlichen Frage ablenken.

Hier noch ein paar Überlegungen zu ausführlichen Problemstellungen bei anwendungsorientierten Fragen: Viele Fragenersteller/innen fragen sich, ob das **Lesen und Verarbeiten** längerer Problemstellungen nicht zu viel Zeit während der Prüfung in Anspruch nimmt. Längere Textpassagen in einer Frage kosten sicher etwas mehr Lesezeit. Andererseits führen auch eine falsche Textverteilung in der Frage (siehe Abbildung 3), komplizierte Formulierungen und Satzkonstruktionen oder mehrdeutige Formulierungen zu einem erhöhten *Bearbeitungsaufwand*.

Ähnlich verhält es sich mit den Informationen und Informationsbausteinen im Fragenstamm. Die meisten Unterlagen zu diesem Thema verlangen, dass alle erforderlichen Informationen für das korrekte Beantworten der Frage vorhanden sein müssen, verlieren aber meistens kein Wort darüber, ob es auch mehr Informationen sein dürfen. So sind wir in unserem Berufsalltag ständig damit konfrontiert, aus einer Vielzahl an Informationen die für uns relevanten herauszufiltern. Bei anwendungsorientierten und praxisnahen Fragen könnte man diese Kompetenz jedoch auch zu einem (Neben-)**Ziel der Frage** machen.

Letztendlich stellt sich ebenso die Frage, inwieweit zusätzliche Medien – wie beispielsweise das oben erwähnte Bildmaterial – die Aufnahmefähigkeit der Studierenden beeinflussen.

Die *Cognitive Load Theory*²⁷ beschäftigt sich mit der kognitiven Belastung beim Lernen und wurde von Paul Chandler und John Sweller aufgestellt. Die Theorie beschreibt drei Arten von Belastungen:

- die *intrinsische kognitive Belastung (intrinsic cognitive load)*, die von der Schwierigkeit und Komplexität des *Stoffes* abhängt,
- die *extrinsische kognitive Belastung (extrinsic cognitive load)*, die von der Darstellung und Gestaltung der Materialien abhängt, und
- die *lernbezogene kognitive Belastung (germane cognitive load)*, die für den Aufwand steht, den Studierende haben, um den „Stoff“ zu verstehen.

Diese drei Arten der kognitiven Belastung summieren sich, daher sollte man versuchen, die Belastungen so zu verlagern, dass Kapazitäten für das Verstehen frei werden (lernbezogene Belastung). Da man die Schwierigkeit des Stoffes oder Themas so gut wie nicht beeinflussen kann, muss die Art der **Darbringung der Aufgabenstellung** (extrinsische Belastung) so reduziert wie möglich sein.

Praktisch würde das bedeuten, dass man einerseits alle überflüssigen Informationen und Elemente entfernt, die eine geistige Aufnahme und Verarbeitung der enthaltenen Elemente erschweren, andererseits muss man sprachlich einfach und eindeutig formulieren. Das bedeutet, dass – wie bereits an mehreren Stellen in diesem Leitfaden erwähnt – alle Formulierungen kurz und prägnant und ohne verschachtelte Sätze, doppelte Negationen und dergleichen formuliert sein müssen.

Rainer Baier von der Wirtschaftsuniversität Wien hat dies in seinen Schulungsunterlagen sehr gut dokumentiert: Bei einer Prüfungsfrage zu einem privaten Autoverkauf und der fraglichen Umsatzsteuerpflicht sind Details zum Auto oder gar Bilder des Autos nicht nötig – im Gegenteil, diese lenken von der Frage ab und kosten unnötig viel Zeit.

²⁷ Plass, J. L., Moreno, R. & Brünken, R. (Hrsg.) (2010). *Cognitive Load Theory*. Cambridge University Press.

Exkurs: die Studierendenperspektive

Hier sei kurz die Perspektive der Studierenden angesprochen. Wie könnten sie denken? Wie könnten sie taktieren?

Beginnen wir einmal mit den **studentischen Denkstrategien**: Wir würden uns wünschen, dass sich Studierende nach dem Lesen einer Frage aktiv eine eigene Antwort überlegen, diese anschließend mit den gegebenen Antwortmöglichkeiten vergleichen und schließlich die passende Antwort ankreuzen. Dies würde einer *Produktionsstrategie* entsprechen. In der Praxis greifen Studierende häufiger zur *Falsifikationsstrategie*: Sie gehen die Antworten der Reihe nach durch, schließen die aus ihrer Sicht nicht zutreffenden aus und versuchen dadurch, die Liste an Möglichkeiten zu reduzieren. Abschließend sei noch die *Trial-and-Error-Strategie* erwähnt. Hier widmen sich Studierende am Ende der Prüfung den nicht beantworteten Fragen, um diese versuchsweise zu beantworten und so durch Zufallstreffer zusätzliche Punkte zu bekommen. Dies natürlich nur, wenn dadurch keine Punkteabzüge drohen.

Zu bedenken ist jedoch auch die Ratewahrscheinlichkeit bei diversen Fragentypen: Bei einer Best-of-five-Frage (eine korrekte Antwort unter fünf gegebenen Antwortmöglichkeiten) ist die Ratewahrscheinlichkeit 20 % ($1/5$). Bei einer korrekten aus vier möglichen Antworten immerhin 25 % ($1/4$) und bei einer einfachen Richtig-Falsch-Beurteilung 50 % ($1/2$). Diese Überlegungen sind im Kontext der Bestehensgrenzen zu sehen.

Abschließend seien zwei weitere Begriffe erwähnt: „**Teilpunkte**“ und „**Punkteabzüge**“. Dies trifft nicht auf sogenannte Single-Choice-Fragen (oder korrekterweise: Single-Response-Fragen) zu, denn dort ist die Auswertung deutlich leichter – Punkte bei korrekter Beantwortung und keine Punkte bei falscher Beantwortung. Anders verhält es sich bei Fragen mit mehreren korrekten Antwortmöglichkeiten, so zum Beispiel 3 Richtigantworten unter 5 gegebenen sind richtig. Geben wir auch einen halben Punkt (also beliebig vereinbarte Teilpunkte), wenn jemand 2 der 3 korrekten Möglichkeiten richtig erkannt hat? Wenn es nur Punkte für korrekte Kreuze gibt, jedoch keine Abzüge für falsch gesetzte Kreuze, dann könnte jemand alle Kästchen ankreuzen? Diese Problematiken detailliert zu beschreiben und zu klären, würde jedoch den Rahmen sprengen.

Um die studentische Sichtweise noch etwas abzurunden, seien hier weitere Punkte erwähnt: Bei der *Augenscheinvalidität* (die auch auf Seite 51 erwähnt wird) geht es darum, den Studierenden den Eindruck zu vermitteln, dass die eingesetzten Fragen relevant und damit valide sind. Das Feedback zu einer

Prüfung ist ein ebenso wichtiger Aspekt. Es wird zwischen *formativen* und *summativen* Prüfungen unterschieden, dabei haben Erstere die Aufgabe, den Studierenden eine Rückmeldung zur eigenen Leistung zu geben. *Summative* Prüfungen haben eine andere primäre Aufgabe, sie sind Leistungsbeurteilungen, die mit mehr oder weniger harten Konsequenzen einhergehen. Wünschenswert wäre jedoch, dass jede Prüfung mit einer konstruktiven Rückmeldung an die Studierenden einhergeht.

Merkregel:

*„When the cook tastes the soup, that’s formative,
when the guests taste the soup, that’s summative.“*

(Robert E. Stake)

Die *Testfairness* ist ein Gütekriterium und bezieht sich auf „systematische Diskriminierung bestimmter Testpersonen aufgrund ihrer ethnischen, soziokulturellen oder geschlechterspezifischen Gruppenzugehörigkeit“²⁸.

Bedenken Sie andererseits auch, dass Studierende mit schriftlichen MC-Prüfungen bestens vertraut sind, beispielsweise ungewollte Lösungshinweise oft instinktiv aufnehmen, dass Sie einfach „**test-wiseness**“ mitbringen. Dies trifft vor allem auf die Studierenden zu, die in ihrer schulischen und studentischen Laufbahn bereits sehr viel mit MC-Prüfungen zu tun hatten. Millman, Bishop und Ebel haben diese 1965 folgendermaßen beschrieben: „subject’s capacity to utilize the characteristics and formats of the test and/or the test taking situation to receive a high score. Test wiseness is logically independent of the examinee’s knowledge of the subject matter for which the items are supposedly measures.“²⁹

²⁸ Testkuratorium (der Föderation deutscher Psychologenverbände) (1986). Mitteilung, *Diagnostica*, 32, 358–360.

²⁹ Millman, J., Bishop, C. H. & Ebel, R. (1965). Analysis of test wiseness in the cognitive domain. *Educational and Psychological Measurement*, 18, 787–790.

Praxistipp:

Bedenken Sie, dass Studierende generell einen wertvollen Beitrag beim Erstellen von Prüfungsfragen leisten können. Zwischenfragen während einer Vorlesung können Ideen für neue Fragen liefern. Viele Fragenersteller/innen binden Studierende auch beim Fragenerstellen direkt ein und schaffen dafür unterschiedliche Anreize. Ebenso können Studierende im Begutachtungsprozess hilfreich sein.

Qualitätssicherung

Gütekriterien und wie man sie messen kann

Jedes wissenschaftliche Messinstrument muss gewisse Anforderungen erfüllen, so auch das Messinstrument „schriftliche Multiple-Choice-Prüfung“, welches zur Beurteilung der Leistungen der Studierenden eingesetzt wird. Für psychodiagnostische Verfahren werden zahlreiche Haupt- und Neben-**Gütekriterien** aufgelistet. Ein erstes Gütekriterium ist die **Objektivität**, welche bei geschlossenen schriftlichen Prüfungsformaten eher gegeben ist als bei anderen Formaten. Dabei sind die Ergebnisse unabhängig von Einflüssen der Untersuchenden und der Untersuchungssituation zustande gekommen.

Die beiden anderen entscheidenden Gütekriterien sind die **Gültigkeit** (Validität) und die **Zuverlässigkeit** (Reliabilität). Eine erste Veranschaulichung dieser beiden Parameter ist in Abbildung 5 zu sehen:

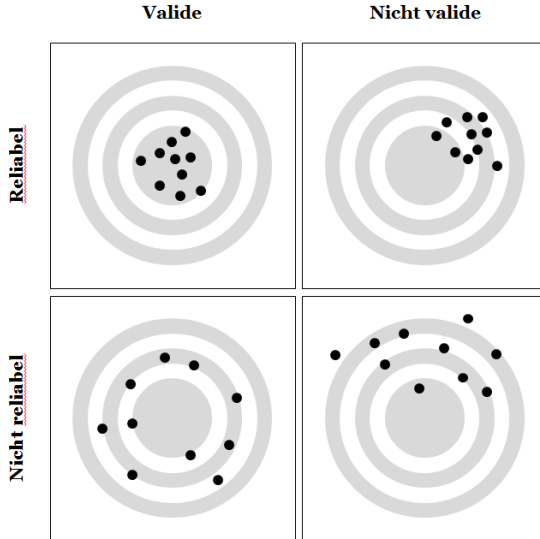


Abbildung 5: Veranschaulichung von Reliabilität und Validität.

Ein Prüfungsergebnis ist dann gültig oder **valide**, wenn es das misst, was zu messen beabsichtigt war. Um dieses Kriterium stark vereinfacht zu veranschaulichen: Mit einem Geschichtetest lässt sich kein Mathematikwissen beurteilen und ein Persönlichkeitstest lässt keine gültige Aussage über die Dauer einer Ehe zu. Um es in anderen Worten zu erklären, wird das anfangs erwähnte Beispiel 29 der Führerschein-Fahrprüfung wiederholt herangezogen.

Beispiel 29 – Validität anhand einer Führerschein-Prüfungsfahrt:

Nach einem schriftlichen oder mündlichen Wissenstest müssen die Kandidatinnen und Kandidaten ihre Fertigkeiten in einer Prüfungsfahrt unter Beweis stellen. Würde diese beispielsweise nur aus einer Beurteilung eines Einparkversuches bestehen, wäre die Validität sehr niedrig. Um diese zu erhöhen, müssen alle Kandidatinnen und Kandidaten sowohl in der Stadt wie auch auf Überlandstraßen und Autobahnen fahren. Einparken, Berganfahren mit der Handbremse sowie Rückwärtsfahren können ebenso Teil dieser Überprüfung sein. Dadurch erhöht sich die Validität deutlich.

Die Reliabilität könnte beispielsweise durch unterschiedliche Witterungsbedingungen, das Verkehrsaufkommen oder andere Zufälligkeiten beeinträchtigt sein, die Objektivität beispielsweise durch sehr unterschiedlich urteilende Fahrprüfer/innen.

Eine Frage trägt dann zur Validität einer Prüfung bei, wenn das Fragenthema **relevant in Hinblick auf zukünftige Anforderungen** ist, wenn das Anspruchsniveau, also der Schwierigkeitsgrad, stimmt, wenn es auf einen klar umschriebenen Inhalt fokussiert ist und wenn es eine eindeutig beste Lösung gibt. Relevant ist ein Fragenthema, wenn es im späteren Arbeitsalltag häufig vorkommt, wenn Fehler gravierende Folgen haben, wenn es sich um verbreitete Fehlmeinungen handelt oder wenn es für das Verständnis späterer Lehrinhalte entscheidend ist. Deshalb sind diese Punkte auch bei der Gewichtung von Prüfungsinhalten auf Seite 25 aufgelistet.

Es werden verschiedene Arten von Validität beschrieben: Die *externe Validität* versucht, objektive und nachprüfbare Außenkriterien anzugeben, um die Validität beurteilen zu können. Leider sind diese sehr selten, daher wird auf die *Kon-*

strukturaliditat zuruckgegriffen.³⁰ Bei einem *Konstrukt* handelt es sich um einen empirisch nicht erkennbaren Sachverhalt, der indirekt uber messbare Indikatoren erschlossen wird (beispielsweise das Konstrukt *Intelligenz* bei der Messung des Intelligenzquotienten). Bei der *Inhaltsvaliditat* wird der Prufungsinhalt von Expertinnen und Experten beurteilt, was ebenso subjektiv ist wie die *Augenscheinvaliditat* (*face-validity*). Diese ist gerade aus Studierendensicht interessant, da eine Frage oder Prufung, die Augenscheinvaliditat besitzt, von den Kandidatinnen und Kandidaten als valide und relevant angesehen wird. Denken Sie dabei an Aussagen wie: „Wofur brauchen wir das?“

Die Reliabilitat oder Zuverlassigkeit einer Prufung bezieht sich darauf, wie genau ein Merkmal gemessen wird. Sie ist dann **reliabel**, wenn ein Merkmal sehr genau gemessen wird oder, anders ausgedruckt, wenn bei einer Messwiederholung ein sehr ahnliches oder bestenfalls gleiches Ergebnis herauskommt.

Eine Frage tragt dann zur *Reliabilitat* einer Prufung bei, wenn der Schwierigkeitsgrad angemessen ist, wenn sie sprachlich einfach und klar formuliert ist und wenn sie keine ungewollten Losungshinweise (diese wurden auf Seite 38 und werden auf Seite 77 erlautert) enthalt.

Andreas Moltner hat dazu ein interessantes **Rechenbeispiel**³¹ publiziert: Ein Lehrender hat 250 Studierende und pruft deren Leistung mit einem Test, der aus 20 Fragen besteht. Da er erfahren mochte, wie stabil seine Ergebnisse sind, stellt er „20 plus 20“-Fragen und vergleicht die beiden Prufungsteile miteinander... Die Ergebnisse sind in Tabelle 1 zu finden. Man beachte, dass bei $r=0,5$ zwei Drittel der Studierenden unterschiedliche Noten bekommen, bei $r=0,8$ immer noch die Halfte, auch wenn es sich hauptsachlich nur mehr um eine Note Differenz handelt.

³⁰ Weitere Literatur zur Konstruktvaliditat:

Millman, J., Bishop, C. H. & Ebel, R. (1965). Analysis of test wiseness in the cognitive domain. *Educational and Psychological Measurement*, 18, 787–790.

Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Med Educ*. 37: 830–837.

³¹ Moltner, A., Schellberg, D. & Junger, J. (2006). Grundlegende quantitative Analysen medizinischer Prufungen – Basic quantitative analyses of medical examinations. *GMS Z Med Ausbild*, 23(3), Doc53.

Reliabilitätswert	r=0,5	r=0,65	r=0,8
Gleiche Note	84 (34 %)	99 (40 %)	123 (49 %)
Eine Note Differenz	114 (46 %)	116 (46 %)	112 (45 %)
Mehr Noten Differenz	52 (21 %)	35 (14 %)	15 (6 %)

Tabelle 1: In dieser Tabelle sieht man, wie viele der 250 Studierenden die gleiche Note auf die beiden Testteile zu je 20 Fragen in Abhängigkeit des Reliabilitätswertes bekommen.

Um die **Reliabilität** zu **steigern**, könnte man einerseits qualitativ hochwertigere Fragen verwenden, andererseits die Anzahl der Fragenitems erhöhen. Letztere kann nicht beliebig erhöht werden, da die Prüfungslänge ab einem bestimmten Punkt nicht mehr zu vertreten ist. Daher muss man sich zwangsläufig der Fragenqualität widmen, wobei zu erwähnen ist, dass auch andere Faktoren dieses Gütekriterium mitbestimmen: Sowohl in der Vorbereitung (Bekanntgabe der Prüfungsinhalte und -methoden) wie auch bei der Durchführung der Prüfung (geeignete Rahmenbedingungen) müssen alle Studierende gleich behandelt werden und die gleichen Voraussetzungen vorfinden. Wie bereits am Beginn des Leitfadens erwähnt, führen alle unerwünschten Einflüsse und eben auch Zufälligkeiten zu schlecht interpretierbaren – eben wenig zuverlässigen – Testergebnissen.

Wir möchten abschließend die beiden Begriffe Reliabilität und Validität nochmals anhand zweier Beispiele darstellen:

Beispiel 30 und 31 – Reliabilität und Validität:

Ein Lineal, welches (möglicherweise herstellungsbedingt) eine „verzerrte“ Skalierung aufweist, wird immer dieselben falschen Messergebnisse liefern, unabhängig davon, wer die Messungen durchführt. Diese jedoch wiederholt und damit zuverlässig. Diese Messungen sind durch eine hohe Reliabilität, aber eine niedrige Validität charakterisiert.

Wenn man unterschiedliche Personen nach der Uhrzeit fragt, ohne sie dabei auf eine Uhr sehen zu lassen, wird man, nachdem man einige Personen gefragt hat, eine recht gute Einschätzung der aktuellen Uhrzeit erhalten. Diese Messungen sind durch eine hohe Validität, jedoch eine niedrige Reliabilität charakterisiert.

Validität und Reliabilität stehen **nicht unabhängig** nebeneinander, vielmehr stehen sie in Beziehung zueinander. Ohne Objektivität kann man sich nicht der Reliabilität widmen, ohne dieser wiederum nicht der Validität. Überzeichnet formuliert: Wenn die Ergebnisse unzuverlässig sind, da sie sehr starken Schwankungen unterliegen, ist es sinnlos, sich Gedanken über die Gültigkeit oder deren Interpretation zu machen. Aber auch abseits dieser *Verkettung der Gütekriterien* gibt es – teils problematische – Zusammenhänge, einer ist beispielsweise das *Reliabilitäts-Validitäts-Dilemma*³².

Weitere Gütekriterien sind: Utilität (Nützlichkeit), Fairness, Testökonomie (Aufwand im Verhältnis zum Nutzen), Transparenz, Unverfälschbarkeit und Zumutbarkeit (zeitliche, psychische und körperliche Belastung im Rahmen der Prüfung). Letztere wurde bei der Verbesserung der Reliabilität durch das Erhöhen der Fragenanzahl einer Prüfung bereits erwähnt.

Der Begutachtungsprozess

Die **Begutachtung** (Korrekturlesen, (Peer-)Review) von erstellten Fragen vor deren erstem Einsatz in einer realen Prüfung ist eine der Möglichkeiten, die Fragenqualität zu verbessern. Bei der zweiten Möglichkeit werden die verwendeten Fragen nach der durchgeführten Prüfung einer statistischen Analyse unterzogen, dazu jedoch erst im nächsten Abschnitt.

Als erster Schritt wäre allen Fragenerstellenden zu empfehlen, ihre eigenen, neu erstellten Fragen nach einer Zeitspanne von ein paar Tagen nochmals gründlich durchzulesen und gegebenenfalls zu überarbeiten. Daneben empfiehlt es sich, die Fragen von zwei oder mehr Personen mit demselben fachlichen Hintergrund gegenlesen zu lassen. Bedenken sollte man auch die Möglichkeit einer Begutachtung in der Gruppe, beispielsweise bei einer Besprechung am Institut.

Doch auch das Gegenlesen durch fachfremde Personen wirkt sich positiv auf die Qualität aus. **Fachfremde** haben einen anderen – in gewisser Weise neutraleren – Blickwinkel auf die Frage, sehen leichter komplizierte Formulierungen oder schwierige Fachbegriffe, und wenn es um reine Tipp- und Rechtschreibfehler geht, ist sowieso kein Fachwissen notwendig. Auch wenn es um das Finden **ungewollter Lösungshinweise** geht, sind diese Personen oft im Vorteil.

³² Siehe <http://de.wikipedia.org/wiki/Reliabilitäts-Validitäts-Dilemma>.

Man sollte auch bei diesem Thema versuchen, kreativ zu denken, so ist das Gegenlesen durch Familienmitglieder genauso hilfreich wie das durch Studierende und Assistentinnen und Assistenten. Eine probeweise Durchführung einer Prüfung könnte ebenso Fehler bei den verwendeten Fragen aufdecken.

Abschließend ein Beispiel (von René Krebs) einer verbesserungswürdigen Frage, die anschließend umformuliert wurde:

Beispiel 32 – verbesserungswürdige Version der Frage³³:

Unsere tägliche Nahrung muss eine Reihe von Substanzen enthalten (teilweise in nur sehr geringen Mengen), die in unserem Körper nicht aus anderen Nahrungsbestandteilen synthetisiert werden können. Zu den Stoffen, die in unserer Nahrung nicht völlig fehlen dürfen, ohne dass Mangelerscheinungen auftreten, gehören folgende Stoffe, außer:

- (a) Vitamine (vor allem Vitamine A, B, C und D)*
- (b) Spurenelemente (vor allem Eisen, Magnesium, Kupfer)*
- (c) bestimmte Fette (essentielle Fettsäuren)*
- (d) Eiweiß (essentielle Aminosäuren)*
- (e) raffinierter Zucker*

Anmerkungen zu dieser Formulierung: Der erste Satz ist eine Belehrung, die auch weggelassen werden kann, mit „sehr geringen Mengen“ findet sich jedoch ein Hinweis auf Spurenelemente, der zweite Satz ist kompliziert formuliert und enthält eine doppelte Verneinung. Bei den Antwortmöglichkeiten finden sich Lösungshinweise, wie das Wort „essentiell“ in (c) und (d).

³³ Beispiel aus den Schulungsunterlagen des IML der Universität Bern: „Multiple Choice Fragen? Ja, aber richtig“ (2008).

Beispiel 33 – korrigierte Version des Beispiels 28:

Welcher der folgenden Stoffe kann in unserer Nahrung völlig fehlen, ohne dass Mangelercheinungen auftreten?

- (a) Vitamine D
- (b) Eisen
- (c) Fett
- (d) Eiweiß
- (e) Traubenzucker

Analyse der Prüfung und der Fragenitems

Die Prüfung wurde nun durchgeführt – welche Möglichkeiten bestehen im Nachhinein, sich die Ergebnisse näher anzusehen? Jede Optimierung führt zu einer besseren Messung bei der nächsten Prüfung. Man muss sich vor Augen halten, dass man die Prüfung auf drei Ebenen ansieht:

...auf der **Prüfungsebene** (Gesamtergebnis, Punkteverteilung, Reliabilität), auf der **Fragenitem-Ebene** (Itemschwierigkeit und -trennschärfe) und im Detail die **einzelnen Elemente** einer Frage (meist die Distraktoren).

Praxistipp:

„Debriefing“: Manche Lehrenden besprechen eine schriftliche Prüfung unmittelbar nach der Durchführung – dies bietet eine ideale Gelegenheit für unmittelbares Feedback, mit einer mündlichen Prüfungssituation vergleichbar.

Wir sehen uns als Erstes die Gesamtergebnisse an und können hierfür die **Häufigkeiten der Gesamtpunkte** der einzelnen Studierenden tabellarisch auflisten oder als Histogramm auftragen (und mit der Bestehensgrenze ergänzen, siehe beispielhaft Abbildung 6). Solche Histogramme sollen einen groben Überblick geben. Sie sind meist asymmetrisch und sollten bei Auffälligkeiten eine Überprüfung der Daten zur Folge haben. Ausreißer nach unten sind dabei oft Folge von Prüfungsabbrüchen, sie sind von Interesse, da sie weitere statistische Berechnungen beeinflussen können.

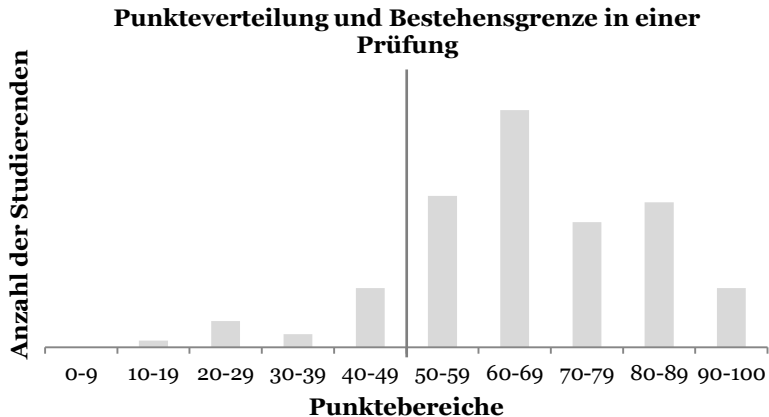


Abbildung 6: Ein Beispiel eines einfachen Histogramms mit Punkteverteilung und Bestehensgrenze.

Als Nächstes wendet man sich der Item-Ebene zu und sieht sich zuerst die **Item-Schwierigkeit** an. Sie bezeichnet das Verhältnis von Studierenden, die dieses Item richtig gelöst haben, zu allen Studierenden, bei denen dieses Item Teil der Prüfung war. Die Werte können somit zwischen 0 (keine/r hat das Item richtig gelöst) und 1 (alle haben das Item richtig gelöst) liegen. In der Literatur³⁴ wird ein **Bereich zwischen 0,4 und 0,8 empfohlen**. Die Verschiebung dieses Bereiches in Richtung erniedrigter Schwierigkeit ist damit begründet, dass zu viele schwere Items die Studierenden demotivieren. Die meisten Items sollten einen mittleren Schwierigkeitsgrad besitzen, aber auch leichte (um 0,8) und schwere (um 0,2) Items sind nötig, da wir ja auch die Leistungen im oberen und unteren Bereich *untersuchen* müssen.

Ein interessanter Punkt ist die mögliche Diskrepanz zwischen den erwarteten und den bestehenden Werten. Items können letzten Endes unerwartet schwer oder unerwartet leicht gewesen sein. Gründe können Formulierungsschwächen bei der Item-Erstellung oder inhaltliche Defizite in der Lehre sein.

³⁴ Möltner, A., Schellberg, D. & Jünger, J. (2006). Grundlegende quantitative Analysen medizinischer Prüfungen – Basic quantitative analyses of medical examinations. GMS Z Med Ausbild, 23(3), Doc53.

Der zweite Parameter, der im Zusammenhang mit einem einzelnen Item relevant ist, ist die **Item-Trennschärfe**. Diese wird durch den Trennschärfekoeffizienten angegeben und entspricht der Korrelation zwischen der Leistung beim einzelnen Item und der Gesamtleistung, daher liegt der Koeffizient zwischen -1 und +1. Es geht also um den Beitrag eines Items zum gesamten Test, dieser wird jedoch u. a. durch die Schwierigkeit, die Homogenität des Tests und die Item-Positionierung im Test beeinflusst. Anders ausgedrückt, handelt es sich um die Fähigkeit eines Items, zwischen guten und schlechten Studierenden zu unterscheiden.

Um diesen Parameter zu berechnen, gibt es mehrere Möglichkeiten, eine wäre der Korrelationskoeffizient nach Pearson-Bravais. Dabei geht es um die Korrelation der erreichten Punkte im Item mit den erreichten Punkten im gesamten Test. Die *korrigierte Trennschärfe* rechnet bei den Gesamtpunkten die Punkte dieser speziellen Aufgabe ab (part-whole-correction). Bei einer schiefen Verteilung kann auf andere Korrelationsmaße zurückgegriffen werden (Spearman und Kendall).

Trennschärfen über 0,3 sind anzustreben, Werte zwischen 0,2 und 0,3 sind akzeptabel und Werte zwischen 0,1 und 0,2 grenzwertig – anders ausgedrückt: Je näher an 1, desto besser.

Wenn man eine Ebene tiefer gehen will – um sich das Fragenitem noch genauer anzusehen –, muss man die Antwortalternativen analysieren und eine **Dis-traktorenanalyse** durchführen.

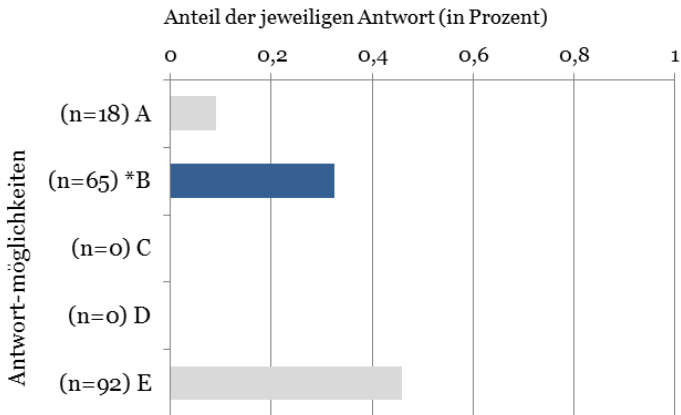
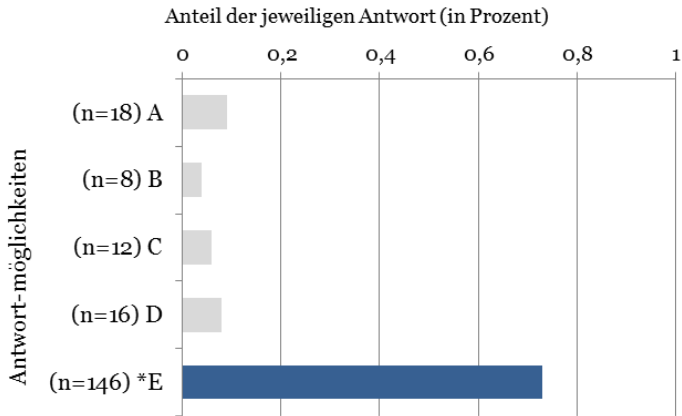


Abbildung 7: Grafische Darstellung einer passenden (linke Abbildung) und einer schlechten (rechte Abbildung) Distraktorenanalyse. Auf der rechten Seite fallen zwei Problem-bereiche auf, einerseits wurden die Antworten C und D von keinem Studierenden ge-wählt, andererseits wurde die Antwort E öfters gewählt als die korrekte Antwort B.

In beiden Grafiken der Abbildung 7 sieht man fünf Antwortmöglichkeiten (A bis E), von denen eine richtig ist (Fragentyp A oder „Best-of-five“), wobei „n“ die Anzahl der Studierenden ist, die diese Antwort gewählt haben; im horizontalen Balken daneben ist der Anteil der Studierenden in Prozent (die richtige Antwort ist mit einem Stern bzw. mit einem schwarzen Balken gekennzeichnet) angege-

ben. In der oberen Grafik sieht man eine gute Frage, alle Antwortmöglichkeiten wurden gewählt, die richtige am häufigsten.

Im Gegensatz dazu zeigt die untere Grafik zwei Problembereiche: Die Antwortmöglichkeiten C und D wurden gar nicht gewählt, was bedeutet, dass diese sehr leicht als falsch zu bewerten waren. Diese würde man vor der nächsten Verwendung des Items ändern müssen. Auf der anderen Seite zeigt sich eine interessante Konstellation bei den Antworten B und E, da die falsche Antwort E von mehr Studierenden gewählt wurde als die richtige Antwort B. Hier müsste man sich die Inhalte der beiden Antwortmöglichkeiten ansehen, ebenso die Vermittlung dieser Inhalte im Unterricht, um die Ursachen eines möglichen Missverständnisses aufzudecken.

Zu guter Letzt könnte man versuchen, die **Reliabilität der Prüfung** abzuschätzen, um zu sehen, wie zuverlässig und reproduzierbar die Prüfungsergebnisse sind. Wie wichtig dies ist, sieht man im Rechenbeispiel auf der Seite 51. Man kann jederzeit die Ergebnisse zweier Prüfungen vergleichen, wenn beispielsweise einerseits ein Test wiederholt wird (Test-Retest-Reliabilität) oder er andererseits parallel abgehalten wird (Paralleltest-Reliabilität). Man kann auch den Test in zwei Hälften teilen und diese miteinander vergleichen (Split-Half-Reliabilität). Wenn man diese letzte Überlegung fortführt und alle möglichen Test-Teilungen miteinander vergleicht, kommt man zu **Cronbachs α** ³⁵. Dieser Koeffizient steht für die interne Konsistenz und ist nicht gleichbedeutend mit der Reliabilität, α ist eine errechnete untere Grenze der Reliabilität, d. h., sie kann auch höher sein.

Abschließend sei noch betont, dass Cronbachs α ein möglicher Schätzer ist, dass jedoch auch andere statistische Verfahren existieren, um die Reliabilität zu schätzen, allerdings sind diese auch deutlich aufwändiger zu berechnen.

Auch wenn die in diesem Kapitel dargestellten statistischen Parameter entscheidende Hinweise zur Verbesserung eines Items bzw. einer Prüfung liefern können, so sollte man diese auch nicht über Gebühr strapazieren. Beispielsweise kann man auch ein Fragenitem mit schlechter Item-Schwierigkeit oder -Trennschärfe im Test belassen, wenn der Inhalt des Items essentiell ist und man sich sicher ist, dass es entsprechend unterrichtet wurde.

³⁵ Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

Der erste Teil des Leitfadens hat mit den Worten „Item-writing is an art“ begonnen und schließt passenderweise mit folgendem Zitat von Robert L. Ebel:

Schlusszitat:

„Item writing is essentially creative. Each item as it is being written presents new problems and new opportunities. Just as there can be no set formulas for producing a good story or a good painting, so there can be no set of rules that will guarantee the production of good test items. Principles can be established and suggestions offered, but it is the item writer's judgment in the application (and occasional disregard) of these principles and suggestions that determines whether good items or mediocre ones will be produced.“³⁶

³⁶ Ebel, R. L. (1951). Writing the test item. In E. F. Linquist (Ed.), Educational measurement (1st ed., pp. 185–249). Washington, DC: American Council on Education.

Praxisteil

*„Fragenerstellen ist wie das Niederschreiben
eines virtuellen Prüfungsgesprächs...“*

(Univ.-Prof. Dr. U. Frick)

Wie sollte eine gute MC-Frage aufgebaut sein?

Bei einer MC-Frage handelt es sich um eine geschlossene Aufgabenstellung, bei der den Studierenden mehrere Antwortalternativen vorgegeben werden. Beim Beantworten der Fragen müssen die Studierenden somit nicht selbständig eine Antwort formulieren und niederschreiben, sondern aus den vorgegebenen Antwortalternativen die richtigen auswählen und ankreuzen.

Hinweis:

In diesem Teil des Leitfadens werden unter MC-Fragen jene Fragen verstanden, die mindestens zwei richtige Antwortalternativen besitzen, im Gegensatz zu Single-Choice-Fragen, welche nur eine richtige Antwort besitzen.

Eine MC-Frage sollte immer aus einer Problemstellung, einer Fragestellung und den Antwortalternativen bestehen, wie in Abbildung 8 zu sehen ist. Während die Problemstellung und die Fragestellung den Stamm der MC-Fragen bilden, wird bei den Antwortalternativen zwischen richtigen und falschen Antworten unterschieden. Falsche Antworten werden auch Distraktoren genannt.

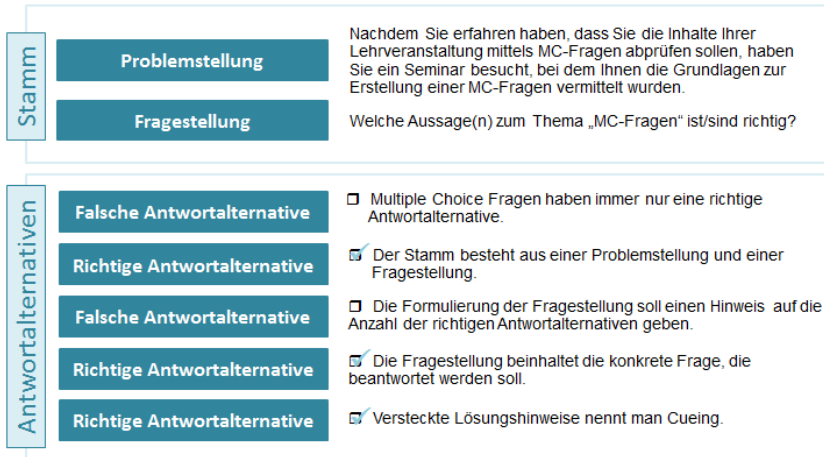


Abbildung 8: Die Elemente einer MC-Frage.

Die Problemstellung

Die Problemstellung dient dazu, den Studierenden alle Informationen, die Sie zur Beantwortung der Frage benötigen, bereitzustellen. In der Problemstellung wird somit der gesamte Problemaufriss bzw. die zu lösende Situation dargestellt.

Um die Frage richtig beantworten zu können, ist es daher essentiell, dass die Problemstellung **vollständig** ist und somit alle notwendigen Informationen enthält. Fehlen einzelne Informationen, kann die Frage mit den vorgegebenen Antwortalternativen womöglich nicht gelöst werden.

Praxistipp:

Wenn Sie das Feedback zu einer MC-Frage einholen, lassen Sie diese von Ihren Kolleginnen und Kollegen immer ohne Lösungshinweise lösen, um zu sehen, ob die Angabe vollständig ist oder wichtige Informationen fehlen.

Eine **kurze, klare und verständliche Formulierung** trägt dazu bei, einen guten Lesefluss und ein schnelles Verständnis des vorliegenden Problems zu gewährleisten. Solange das Lernziel der Frage nicht „lange Texte sinnerfassend lesen können“ heißt, sollten sich die Studierenden beim Lesen der Problemstellung auf die benötigten Informationen des Problemaufrisses und nicht auf die Komplexität des Textes konzentrieren können.

Eine weitere Möglichkeit, um Missverständnisse bei der Frage vorzubeugen, ist die Vermeidung von Fremdwörtern oder Abkürzungen. Dabei handelt es sich oft nicht nur um Fremdwörter im eigentlichen Sinne, sondern auch um Fachbegriffe einer bestimmten Branche oder eines Fachgebietes. Sobald ein Begriff für Verwirrung oder Missverständnisse beim Beantworten der Frage sorgen könnte, ist eine Erklärung des Begriffes für das Verständnis der Problemstellung von Vorteil:

Beispiel 34 – Begriffserklärung im Text:

„...erfahrungsgemäß geht der Stoff bei der ersten Wäsche um 0,25 cm pro Ballen ein („einlaufen“, „kleiner werden“). Für die Vorwäsche...“

„...trotzdem kommt es beim Zuschneiden zu einem 1/9 Verschnitt (=Stoffreste). Nur Stoff vom gleichen Zukauf...“

Der **Einsatz von Grafiken und Bildern** ist nur sinnvoll, wenn diese für das Verständnis und für die Beantwortung der Frage notwendig sind. Werden Bilder ohne Bezug zur MC-Frage verwendet, um beispielsweise eine lange Problemstellung aufzulockern, kann dies zu Missverständnissen führen, da Studierende in dem Bild eventuell Informationen suchen, die Sie zur Beantwortung der Frage benötigen.

Praxistipp:

Wird die MC-Frage bei einer Prüfung eingesetzt, so wird durch den Einsatz von Bildern das Schummeln erleichtert, da die Studierenden leicht erkennen können, welche Frage ihre Kolleginnen und Kollegen gerade behandeln.

Das Formulieren der Problemstellung kann bei komplexen Beispielen sehr aufwendig sein, da viele Informationen angegeben werden müssen, um die Frage beantworten zu können. Falls die komplexe Problemstellung es zulässt, dass mehrere Fragen aus den angegebenen Informationen gestellt werden, so kann es **nach einer komplexen Problemstellung auch mehrere Fragestellungen** inkl. Antwortalternativen geben.

Die Formulierung mehrerer Fragestellungen zu einer Problemstellung macht es einerseits möglich, die Informationen gleich für mehrere Fragen verwenden zu können, andererseits kommt dabei die Problematik der **Folgefehler** zum Tragen. In rechenintensiven Fächern wie Mathematik oder Buchhaltung kann bei MC-Fragen der Rechenweg der Studierenden nicht mehr nachvollzogen werden. Da nur das Endergebnis angekreuzt wird, können Folgefehler beim Korrigieren nicht berücksichtigt werden.

Praxistipp:

*Um die Problematik von Folgefehlern zu umgehen, können Sie Zwischenergebnisse angeben. Eine Angabe könnte beispielsweise so lauten:
„Nehmen Sie an, der Wert x ist 15,5. Berechnen Sie nun...“*

Diese Problematik kann umgangen werden, indem bei jeder Fragestellung eine kurze Angabe mit neuen Zwischenwerten angegeben wird. Die Studierenden rechnen immer mit den neu angegebenen Zwischenwerten weiter und erhalten so bei jeder Fragestellung die Möglichkeit, die Frage ohne Folgefehler beantworten zu können.

Die Fragestellung

Die Fragestellung sollte so formuliert sein, dass Sie auch als **offene Frage** gelöst werden kann. Alle Informationen, die zur Beantwortung der Frage benötigt werden, sind bereits in der Problemstellung enthalten, wodurch die Fragestellung nur aus der tatsächlichen Frage und der notwendigen Arbeitsanweisung wie „Kreuzen Sie die richtige(n) Aussage(n) an“ besteht.

Auch die Fragestellung sollte kurz und verständlich formuliert sein, um Missverständnisse zu vermeiden und den Studierenden die Möglichkeit zu geben, sich auf die Frage konzentrieren zu können.

Eine **optische Trennung** der Problemstellung und der Fragestellung erleichtert die Orientierung beim Lösen der Frage. Sind die Studierenden an das Format der optischen Trennung von Problemstellung und Fragestellung gewohnt, bietet dies mehr Flexibilität beim Beantworten der Frage. Studierende können so zuerst die Fragestellung lesen und daraufhin die Problemstellung bereits beim Lesen gezielt nach Informationen zur Beantwortung der Frage durchforschen.

Beispiel 35 – Trennung der Frage von der Problemstellung:

In der letzten Besprechung haben Sie von Ihrer/Ihrem Vorgesetzten erfahren, dass es eine Studienplanänderung gab. Aufgrund dieser Änderung werden sich die Studierendenzahlen in Ihrer Lehrveranstaltung stark erhöhen. Daher sollen Sie in Zukunft nur mehr MC-Prüfungen in dieser Lehrveranstaltung abhalten.

Welche Vorteile hat dieses Prüfungsformat für Sie als Prüfungsersteller/in? Kreuzen Sie die richtige(n) Aussage(n) an.

...

Um die MC-Frage schwieriger zu machen, werden oft Ausnahmen oder Verneinungen abgefragt, welche nicht die Frage an sich schwieriger machen, sondern den Denkweg der Studierenden beim Beantworten der Frage erschweren. Um das Lernziel der Frage und nicht das Denkvermögen der Studierenden abzuprüfen, sollten in der Fragestellung keine Verneinungen vorkommen. Sind diese nicht zu vermeiden, sollten sie hervorgehoben werden, damit diese nicht überlesen werden.

Beispiel 36 – Umgang mit Verneinungen:

Hervorgehobene Verneinung:

„Welche der folgenden Rechengrößen werden NICHT in den Datenpool der Kosten- und Leistungsrechnung übernommen?“

Verneinung vermeiden:

„Welche der folgenden Rechengrößen werden in den Datenpool der Kosten- und Leistungsrechnung übernommen?“

Das Lernziel dieser Frage war, dass die Studierenden wissen, welche Rechengrößen in den Datenpool der Kosten- und Leistungsrechnung übernommen werden, was eindeutig mit der zweiten Fragestellung abgeprüft wird. Mit der ersten Fragestellung müssen Studierende zuerst überlegen, welche Rechengrößen übernommen werden, und danach umdenken, welche der vorgegebenen Antwortalternativen somit nicht übernommen werden. Die Fragestellung wird durch die Verneinung inhaltlich jedoch nicht schwieriger, sondern wird von Studierenden häufig als „Falle“ wahrgenommen.

Wenn nicht bereits am Anfang der Prüfung ein Hinweis darauf gegeben wurde, ob bei den MC-Fragen immer nur eine oder mehrere Antwortalternative(n) richtig sind, kann durch die **grammatikalische Formulierung** der Arbeitsanweisung ein Hinweis auf die Anzahl der richtigen Antwortalternativen gegeben werden.

Praxistipp:

Um keinen Hinweis auf die Anzahl der richtigen Antwortmöglichkeiten zu geben, kann die Fragestellung beispielsweise so formuliert werden:

„Kreuzen Sie die richtige(n) Antwortalternative(n) an.“

Die Antwortalternativen

FORMALE KRITERIEN

Bei der Formulierung der Antwortalternativen sollte auf eine einfache und kurze Formulierung geachtet werden. Um die Korrektur und das Vergleichen der MC-Fragen zu erleichtern, sollte jede Antwortalternative mit einer Aufzählung versehen werden (z. B. a) b) c)).

Gleich wie bei der Formulierung einer Fragestellung sollten auch in den Antwortalternativen Verneinungen vermieden bzw. hervorgehoben werden. Wird sowohl in der Fragestellung als auch in der Antwortalternative eine Verneinung verwendet, handelt es sich um eine doppelte Verneinung, bei der das Lernziel der Frage und das Überprüfen der fachlichen Kompetenz in den Hintergrund rückt und sich das Abprüfen von logischem Verständnis in den Vordergrund drängt. Die Validität der Frage sinkt dabei maßgeblich:

Beispiel 37 – doppelte Verneinung:

Welche der Aussagen zum Thema Abschreibung trifft NICHT zu?

- (a) Wertpapiere sind kein abnutzbares bewegliches Anlagevermögen.*
- (b) Wertpapiere...*
- (c) ...*

Sind **alle Antwortalternativen gleich lang**, können keine Rückschlüsse darauf gezogen werden, welche Antwortalternative die richtige ist. Oft wird die längste Antwortalternative als richtig eingeschätzt, da davon ausgegangen wird, dass bei der Formulierung der richtigen Antwortalternative mehr Zeit aufgewendet wird bzw. bei dieser Antwortalternative besonders auf die Formulierung geachtet wird.

Die Antwortalternativen sollten immer untereinander und nicht nebeneinander angeordnet sein. Handelt es sich bei den Antwortalternativen um Zahlen, ist es bei einer auf- oder absteigenden Reihenfolge für die Studierenden einfacher, sich zu orientieren und die richtige Antwortalternative zu finden.

Variieren die Reihenfolge und Anzahl der richtigen bzw. falschen Antwortalternativen, ist für die Studierenden kein Schema erkennbar.

INHALTLICHE KRITERIEN

Die Antwortalternativen müssen so formuliert sein, dass sie von fachkundigen Personen als eindeutig richtig oder eindeutig falsch erkannt werden.

Bei MC-Fragen wird oft davon ausgegangen, dass diese durch die vorgegebenen Antwortalternativen leichter zu lösen sind als offene Fragen. Um die Fragen schwieriger zu machen, tendiert man leicht dazu, **Ausnahmen oder Randthemen** abzuprüfen. Um diese Tendenz zu vermeiden, kommt der Formulierung von Lernzielen eine große Bedeutung zu, um sich auf die wichtigsten Inhalte zu konzentrieren.

Merkregel:

Es wird die Regel geprüft und nicht die Ausnahme!

Gibt es falsche Antwortalternativen, die auch von Studierenden, die den Stoff nicht beherrschen, als eindeutig falsch erkannt werden, so können diese Antwortalternativen von den Studierenden sofort gestrichen werden. Die Wahrscheinlichkeit, die richtige Antwort zu erraten, wird durch das Streichen einzelner Antwortalternativen automatisch größer. Ein wichtiges Kriterium ist daher, dass sich die einzelnen Antwortalternativen nicht gegenseitig ausschließen.

Merkregel:

Falsche Antwortalternativen sollten so formuliert sein, dass sie „unwissenden“ Lernenden plausibel erscheinen.

HÄUFIGE FEHLER ODER FALSCHANTWORTEN VON STUDIERENDEN

Um eine gute und plausibel klingende falsche Antwortalternative zu erstellen, können häufige Fragen und Falschantworten aus der Lehrveranstaltung verwendet werden und in eine Antwortalternative umgewandelt werden. Stellen Sie in der Lehrveranstaltung häufig Fragen oder lassen Sie Studierende Rechnungen durchführen, um zu sehen, welche Fehler bei der Beantwortung der Frage gemacht werden. Arbeiten Sie mit Foren, so können Sie auch häufige Fragen daraus verwenden, um eine plausibel klingende falsche Antwortalternative zu erstellen. Werden im Forum auch die Rechenwege zur falsch errechneten Zahl angegeben, so können Sie genau diese Rechenwege mit den Angaben Ihrer Prüfung durchrechnen, um eine weitere falsche Antwortalternative zu erhalten.

EINSATZ VON FACHAUSDRÜCKEN

Durch den Einsatz von Fachausdrücken in den falschen Antwortalternativen werden diese Antwortalternativen von unwissenden Studierenden oft als richtig eingeschätzt. Für Studierende, die die Bedeutung der Fachausdrücke nicht kennen, hört sich der Fachausdruck bekannt an und erhöht den Erinnerungseffekt an den – vielleicht nur schnell durchgeblätterten – Stoff. Gleichzeitig wird das Gefühl vermittelt, dass in die Formulierung dieser Antwortalternative besonders viel Zeit investiert wurde. Dies wiederum lässt darauf schließen, dass diese Antwortalternative mit hoher Wahrscheinlichkeit richtig ist.

ANTWORTEN, DIE ZUM THEMA GEHÖREN, ABER KEINE ANTWORT AUF DIE FRAGE DARSTELLEN

Formulieren Sie Antwortalternativen, die zwar zum Thema gehören, aber keine Antwort auf die eigentliche Frage darstellen. Wenn Sie beispielsweise die Vorteile einer MC-Frage abprüfen, können Sie bei einer Antwortalternative die Definition einer MC-Frage verwenden. Diese ist generell richtig und wird somit von Studierenden, die die Frage nicht aufmerksam lesen, als richtig angekreuzt. Die Frage handelt jedoch von den VORTEILEN einer MC-Frage, wodurch die Definition einer MC-Frage nicht die richtige Antwort darstellt.

WORTWÖRTLICHE SÄTZE AUS DEN LERNMATERIALIEN

Eine ähnliche Funktion wie der Einsatz von Fachausdrücken haben auch wortwörtliche Sätze aus den Lehrmaterialien. Hier kann ein Satz wortwörtlich aus einem anderen Kapitel übernommen werden. Dieser stimmt zwar generell, ist aber im konkreten Fall keine Antwort auf die aktuelle Frage. Studierende, die den Prüfungsstoff nicht gelernt, sondern nur überflogen haben und dadurch auch die Zusammenhänge nicht verstanden haben, erkennen wortwörtliche Sätze im Zusammenhang mit dem Stoff wieder und schätzen diese als richtig ein, obwohl sie im Zusammenhang mit der konkreten Frage falsch sind.

WORTÄHNLICHKEITEN BZW. WORTSPIELEREIEN

Um eine falsche Antwortalternative für unwissende Studierende trotzdem plausibel klingen zu lassen, können Wörter mit ähnlichen Bedeutungen, leicht abgewandelte Wörter oder Wörter, die zwar in der Lehrveranstaltung, jedoch in einem anderen Zusammenhang verwendet wurden, eingesetzt werden. Ist beispielsweise das Wort „Fristenlösung“ die richtige Antwortalternative, so kann darüber nachgedacht werden, welche Bedeutung das Wort „Frist“ hat. Da es sich hier meist um eine zeitliche Begrenzung handelt, können die falschen Antwortalternativen dann ebenfalls Wörter sein, die mit einer zeitlichen Begrenzung im Zusammenhang stehen. Durch die Ähnlichkeit der Antwortalternativen wird das Beantworten der Frage für unwissende Studierende schwieriger.

FEHLER AUS OFFENEN PRÜFUNGEN

Werden die Inhalte in anderen Fächern oder an anderen Institutionen mittels offenen Prüfungen abgeprüft, so können die häufigsten Fehler der offenen Prüfung herangezogen werden, um plausibel wirkende falsche Antwortalternativen zu formulieren. Vor allem in Fächern, bei denen Rechenwege abgeprüft werden, können die falschen Rechenwege von Studierenden mit den Angaben der Prüfung nachgerechnet werden, um dadurch realistische Zahlen zu bekommen, die Studierende bei einem falschen Rechenweg als Lösung erhalten könnten.

ANLEGEN EINER DATENBANK

Werden die erstellten MC-Fragen als Prüfungsfragen verwendet, so hat es sich bewährt, eine Datenbank mit mehreren Antwortalternativen – sowohl richtigen als auch falschen – anzulegen, um die Frage immer wieder neu zusammenstellen zu können. Während die Problemstellung und die Fragestellung gleich bleiben, können immer wieder neue Antwortalternativen zu dem bereits bestehenden Stamm formuliert werden. Die Datenbank kann aus MC-Fragen bestehen, die aus zehn oder mehr möglichen Antwortalternativen bestehen. Beim Zusammenstellen der Prüfung können Sie die Antwortalternativen in beliebiger Reihenfolge kombinieren und so die Frage variieren.

Anwendungsorientierte Fragen

Ein oft genannter Nachteil von MC-Fragen ist die Erstellung anwendungsorientierter Fragen. Aus den Interviews sind mehrere Methoden hervorgegangen, wie auch MC-Fragen anwendungsorientiert gestaltet werden können, zusätzlich bietet die Bloom'sche Taxonomie – die auf Seite 22 beschrieben wird – gute Denkanstöße.

DATEN AUS ORIGINALEN HERAUSLESEN

Nicht immer müssen alle Angaben, die man für die Beantwortung der Frage benötigt, der Reihe nach in der Problemstellung angegeben werden. Durch die Verwendung von Originalen können Beispiele anwendungsorientierter gestaltet werden. Originale können im Rechnungswesen beispielsweise Rechnungen und Belege oder in der Medizin ein – natürlich anonymisierter – echter Blutbefund oder eine Röntgenaufnahme sein. Durch das Verwenden solcher Originale müssen die Studierenden einerseits wissen, welche Informationen Sie für die Beantwortung der Frage benötigen und wo Sie andererseits diese in dem angegebenen Original finden. Bei der Verbuchung einer Rechnung müssen weder der Nettopreis noch die Umsatzsteuer in der Problemstellung als Auflistung angegeben werden. Wird die Rechnung vom zu verbuchenden Verkauf in der Problemstellung angegeben, so müssen alle relevanten Zahlen aus dieser herausgelesen werden, wodurch die Frage nicht nur anwendungsorientierter wird, sondern auch zusätzliche Lernziele erreicht werden können.

EINE SITUATION BESCHREIBEN

Um MC-Fragen anwendungsorientierter und praxisnaher zu gestalten, können Situationen beschrieben werden, statt alle notwendigen Informationen aufzuzählen. Statt eine einfache Frage zu stellen, ob beispielsweise bestimmte rechtliche Aspekte in Österreich gestattet sind, kann ein Beratungsgespräch wiedergegeben werden, mit der Abschlussfrage, ob die in dem Gespräch genannten Aspekte für eine österreichische Staatsbürgerin bzw. einen österreichischen Staatsbürger zulässig bzw. richtig sind. Die Studierenden müssen nun selbst die benötigten Informationen aus dem abgedruckten Gespräch herausfiltern und diese auf die österreichische Rechtslage anwenden. Im medizinischen Kontext werden beispielsweise reale Patientinnen- und Patientengeschichten anonymisiert und anschließend in MC-Fragen zu diesen Musterpatientinnen und -Patienten umgewandelt – vom Erstgespräch bis hin zu einer bereits stattge-

fundenen Behandlung und danach auftretenden Problemen können hier viele Situationen beschrieben werden, in denen die Studierenden ihr Wissen anwenden müssen, statt auswendig Gelerntes einfach wiederzugeben.

MEHRERE KAPITEL MITEINANDER VERBINDEN

Durch das Abprüfen von kapitelübergreifenden Inhalten kann überprüft werden, ob die Studierenden nur einzelne Kapitel ohne Gesamtüberblick gelernt haben oder die Zusammenhänge der einzelnen Inhalte verstanden wurden. Durch die Verbindung von zwei oder mehreren Kapiteln in einer Frage wird die Aufgabe zwangsläufig anwendungsorientiert, da nicht mehr reines Faktenwissen ausreicht, um die Frage zu beantworten, sondern Zusammenhänge, Abhängigkeiten, Wirkungen etc. verstanden werden müssen. Werden in der Lehrveranstaltung beispielsweise einzelne Themenblöcke wie Sozialpolitik, Arbeitslosigkeit etc. durchgenommen, könnte als anwendungsorientierte Frage gefragt werden, welche Unterstützung eine arbeitssuchende Person vom Staat bekommen kann.

ALLGEMEINWISSEN MITEINBEZIEHEN

Um Fragen anwendungsorientiert zu gestalten, kann auch das Allgemeinwissen mit einbezogen werden. Bei der Prüfung zu europäischem Recht kann beispielsweise eine Frage einen Sachverhalt wiedergeben, dessen rechtliche Aspekte in der Lehrveranstaltung behandelt wurden. Bei der Frage findet die geschilderte Situation in der Türkei statt. Nun müssen Studierende wissen, dass die Türkei kein Beitrittsland der EU ist und somit das europäische Recht nicht angewandt werden kann. Durch die Einbeziehung des Allgemeinwissens wird somit auch das Lernziel abgedeckt, dass die Studierenden wissen müssen, welche Länder vom europäischen Recht betroffen sind.

Wie vermeidet man versteckte Lösungshinweise?

Der Begriff Cueing kommt aus dem Englischen und bedeutet, dass bei der Erstellung von MC-Fragen sehr oft unbewusste und ungewollte Lösungshinweise in der Fragestellung bzw. mehr noch in den Antwortalternativen gegeben werden.

Streng genommen stellt sogar schon die bloße Vorgabe der Antwortalternativen Cueing dar, da die/der Studierende bestimmte Formulierungen wiedererkennen kann und so der Erinnerungseffekt gesteigert wird. Als gutes Beispiel dient hier die bekannte Fernsehsendung „Die Millionenshow“. Viele der Fragen könnte man nur schwer aus dem Stegreif beantworten, wenn diese offen, sprich: ohne vorgegebene Antwortalternativen, gestellt würden. Mit Hilfe der (in diesem Fall vier) Antwortalternativen ist es sehr viel leichter, die Fragen zu beantworten.

Sind in MC-Fragen ungewollte Cues (Hinweise) enthalten, können dadurch diese oft schon mit Halbwissen oder durch Raten richtig gelöst werden.

Teilweise bietet die ÖH spezielle Kurse für Studierende an, in denen diese auf Cueing trainiert werden. Das Motto: Wie löse ich eine MC-Klausur, ohne dafür zu lernen?

Praxistipp:

Cues werden von Fragenerstellenden unbewusst meist in den Antwortalternativen hinzugefügt, da z. B. die richtige Antwort besonders schön formuliert sein soll. Daher sollte jede Frage auf Cues geprüft werden.

Zu den wichtigsten Cues, die unbewusst in Antwortalternativen erstellt werden, zählen:

- die Verwendung von absoluten bzw. weichen Aussagen,
- die Art der Ausformulierung bzw. die Verwendung der Sprache des Lehrbuchs,
- die Erstellung von unplausiblen Distraktoren (Kängurus),
- die Wortwiederholung und Verwendung von Fremdwörtern im Stamm der MC-Frage sowie in den richtigen Antwortalternativen,
- die grammatikalische Inhomogenität zwischen der Fragestellung und den Distraktoren sowie
- das sich gegenseitige Ausschließen von Antwortalternativen.

Diese werden in den folgenden Unterkapiteln näher erläutert.

Die Verwendung von absoluten bzw. weichen Aussagen

Bei diesem Cue geht es um die Art und Weise, wie Antwortalternativen formuliert werden.

Unbewusst ist man versucht, richtige Antwortalternativen eher mit weichen Aussagen bzw. Wörtern zu versehen, wie beispielsweise manchmal, oft, meistens, häufig oder selten. Das heißt, es werden keine absoluten Aussagen getroffen, sondern ein gewisser Spielraum für Ausnahmen bleibt gegeben. Absolute Wörter wie beispielsweise immer, nie, niemals, gar nicht, stets, muss, werden andererseits unbewusst als Cue in Distraktoren (falsche Antwortalternativen) verwendet.

Achtung: Bei Rechtsfächern kommt es aufgrund von gesetzlichen Bestimmungen oft zu einer Umkehr dieses Cues, da hier oft absolute Aussagen in den richtigen Antwortalternativen getätigt werden müssen.

In Cueing geschulte Studierende können somit einige Antwortalternativen ausschließen bzw. aufgrund solcher Formulierungen in die engere Auswahl nehmen.

Beispiel 38 – für diesen Cue:

Sie wurden auserwählt, im Rahmen einer Schulung im Bereich Erstellung von MC-Fragen eine Übungsfrage im MC-Format vorzubereiten. Da Sie sich gerade unsicher sind, wie man eine gute MC-Frage erstellt, suchen Sie nach Informationen darüber.

Kreuzen Sie die richtige(n) Aussage(n) in diesem Zusammenhang an!

(a) Sie können in Büchern und Fachzeitschriften nach Informationen zur Erstellung von guten MC-Fragen suchen. (korrekt)

(b) Sie können im Internet sich einen ersten Überblick zur Erstellung von guten MC-Fragen verschaffen, der Ihnen wahrscheinlich helfen wird. (korrekt)

(c) Sie müssen sich den Leitfaden zur Erstellung von guten MC-Fragen durchlesen, da Sie nur dort die richtige Herangehensweise lernen können. (korrekt)

Die Art der Ausformulierung bzw. die Verwendung der Sprache des Lehrbuchs

Richtige Antwortalternativen werden von Fragerstellenden meist besonders schön formuliert, da ihnen diese wichtig sind und daher mehr Zeit zur Gestaltung investiert wird.

Daher heben sich richtige Antwortalternativen durch eine besonders genaue Ausformulierung ab. Die verwendete Sprache, Grammatik und/oder auch die Länge der Aussage steht stark im Kontrast zu der Formulierung von Distraktoren. Häufig sind aufgrund dieser Formulierung die richtigen Antwortalternativen die längsten.

Außerdem sollte es vermieden werden, in den richtigen Antwortalternativen die Sprache des Lehrbuchs zu verwenden, da Studierende diese wiedererkennen und daher aufgrund eines Bauchgefühls eher ankreuzen würden.

Beispiel 39 – für diesen Cue:

Welche Merkmale von gut erstellten MC-Fragen kennen Sie?

Kreuzen Sie die richtige(n) Aussage(n) in diesem Zusammenhang an!

- (a) Die MC-Frage wurde inhaltlich richtig erstellt und würde von Fachexpertinnen/Fachexperten in gleicher Weise gelöst werden. (korrekt)*
- (b) Die Antwortalternativen sind so formuliert, dass sie von Fachexpertinnen/Fachexperten ebenfalls als eindeutig richtig bzw. eindeutig falsch erkannt werden. (korrekt)*
- (c) Es wurde darauf geachtet, dass Cues vorhanden sind.*
- (d) Es wurden viele Bilder verwendet.*

Wie man in diesem Beispiel sehen kann, sind die längsten und am genauesten ausformulierten Antwortalternativen die richtigen Antworten. Außerdem wurde hier die Sprache des Lehrbuchs (in diesem Fall des Guides) verwendet.

Erstellung von unplausiblen Distraktoren

Ist eine bestimmte Anzahl von Antwortalternativen gewünscht (z. B. es sollen pro Fragestellung immer fünf Antworten vorhanden sein), stößt man leicht auf das Problem, dass keine sinnvollen richtigen Antwortalternativen bzw. Distraktoren erstellt werden können.

Hier ist man sehr leicht versucht, unplausible bzw. abwegige Distraktoren zu verwenden. Diese Aussagen sind oft sehr weit hergeholt und werden zumeist auch von Unwissenden als eindeutig falsch erkannt. Solche Aussagen werden auch als Kängurus bezeichnet.

Praxistipp:

Lieber nur vier Antwortalternativen, bevor ein Distraktor ein Känguru ist!

Die Problematik bei solchen Kängurus liegt darin, dass sie einerseits überflüssig sind und dadurch anderen sinnvolleren Distraktoren den Platz wegnehmen und andererseits die Ratechancen der Studierenden massiv erhöhen. Da ja z. B. von fünf vorhandenen Antwortalternativen eine oder sogar mehrere eindeutig sofort als falsch erkannt werden, ist die Chance relativ hoch, dann aus den restlichen Antwortalternativen die richtigen zu erraten.

Beispiel 40 – für diesen Cue:

Von welchem/welchen Insekt(en) wird die afrikanische Schlafkrankheit übertragen?

Kreuzen Sie die richtige(n) Aussage(n) in diesem Zusammenhang an!

- (a) Moskito*
- (b) Termiten*
- (c) Tsetsefliege (korrekt)*
- (d) Känguru*

Wortwiederholung und Verwendung von Fremdwörtern im Stamm der MC-Frage sowie in den richtigen Antwortalternativen

Gerade bei der Erstellung von Theoriefragen kommt es oft zu der Problematik, dass nach bestimmten Fachausdrücken bzw. Theorien gefragt wird. Dabei benutzt man unbewusst in den richtigen Antwortalternativen diese Fachausdrücke wiederholt. Das heißt, sowohl im Stamm (Problemstellung, Fragestellung) als auch in den richtigen Antwortalternativen werden genau dieselben Fachausdrücke verwendet bzw. die Bezeichnungen von Theorien.

Beispiel 41 – für diesen Cue:

Welche Merkmale sollten Distraktoren aufweisen?

Kreuzen Sie die richtige(n) Aussage(n) in diesem Zusammenhang an!

- (a) Distraktoren sollten eindeutig falsch sein. (korrekt)*
- (b) Es sollte in Summe immer fünf Antwortalternativen geben.*
- (c) Distraktoren sollten plausibel klingen. (korrekt)*

Grammatikalische Inhomogenität zwischen Fragestellung und Distraktoren

Bei Vervollständigungssätzen kann es passieren, dass es zu grammatikalischen Inhomogenitäten zwischen der Fragestellung und den Distraktoren kommt. Bei den richtigen Antwortalternativen ist man meist noch sehr konzentriert und die Sätze lassen sich vervollständigen. Die falschen Antworten werden meist oberflächlicher erstellt und ihnen wird weniger Zeit bei der Erstellung gewidmet. Dadurch kann es zu grammatikalischen Fehlstellungen kommen.

Beispiel 42 – für diesen Cue:

Ist ein Distraktor völlig unplausibel, so spricht man von einem

...

- (a) ...Cue. (korrekt)*
- (b) ...falschen Antwort.*
- (c) ...Känguru. (korrekt)*

Antworten schließen sich gegenseitig aus

Werden die richtigen Antwortalternativen sowie die dazugehörigen Distraktoren so formuliert, dass sich diese gegenseitig ausschließen, kommt es zu einer erhöhten Rategefahr und dadurch zu einer erhöhten Wahrscheinlichkeit, die richtigen Antwortalternativen ohne „echtes“ Wissen anzukreuzen.

Beispiel 43 – für diesen Cue:

Kreuzen Sie die richtige(n) Aussage(n) an!

- (a) Um Cueing bei der Erstellung von Antwortalternativen zu vermeiden, sollte eine fachfremde Person diese darauf überprüfen. (korrekt)*
- (b) Um Cueing bei der Erstellung von Antwortalternativen zu vermeiden, sollte nicht eine fachfremde Person diese darauf überprüfen.*

Der bewusste Einsatz von Cueing bei der Erstellung von richtigen Antwortalternativen sowie Distraktoren

Studierende sind oft gut organisiert und daher auch gut auf Cueing trainiert. Deshalb ist es möglich, das nun erlangte Wissen rund um Cueing bewusst bei der Erstellung von richtigen Antwortalternativen sowie von Distraktoren einzusetzen.

So sollten in den richtigen Antwortalternativen alle genannten Cues vermieden werden und daher speziell darauf geprüft werden. Andererseits können vermeidbare Cues in den Distraktoren eingesetzt werden. Die falsche Antwortalternative könnte die längste Antwort sein, in der „weiche“ Ausdrücke verwendet werden. Oder eine richtige Antwortalternative wird so formuliert, dass diese für einen Nichtwissenden eindeutig wie ein Känguru aussieht.

Kreative Herangehensweisen

Aus den durchgeführten 25 Interviews mit Personen, die häufig mit der Erstellung von MC-Fragen beschäftigt sind, haben sich folgende Herangehensweisen für die Erstellung von MC-Fragen herauskristallisiert. Im zweiten Teil finden Sie ab Seite 23 ergänzend eine theoretische Abhandlung dieses Themas.

Lehrmaterialien (Skripten, Folien, Bücher)

„Bevor ich eine Prüfung zusammenstelle, setze ich mich noch einmal ganz intensiv mit dem Prüfungsstoff und mit der angegebenen Prüfungsliteratur auseinander. Manchmal sehe ich mir die letzten Prüfungen an, um zu sehen, welche Inhalte ich schon lange nicht mehr abgeprüft habe. Dadurch habe ich oft schon ein Grobkonzept, welche Inhalte ich diesmal abprüfen werde. Beim Durchblättern der Bücher und der Folien lese ich mir jene Inhalte, bei denen ich bereits weiß, dass ich sie abprüfen werde, genauer durch. Die restlichen Inhalte überfliege ich beim ersten Mal nur und notiere mir Stichworte, welche Inhalte ich abprüfen könnte. Danach sehe ich mir im Detail an, wie viele Stichworte ich habe, wie viele Fragen ich brauche und welche Fragen ich tatsächlich weiter ausformuliere, um den Stoff möglichst gut abzudecken.“

Zeitungen

„Ich verwende gerne aktuelle Tagesthemen oder Urteile, die gerade gefällt wurden, um dazu Prüfungsfragen zu gestalten. Durch das Lösen von Themen, die aktuell gerade in den Medien diskutiert werden, wird die Frage anwendungsorientierter, da die Studierenden sehen, dass die Inhalte der Lehrveranstaltung nicht nur reines Theoriewissen sind, sondern dass sie tatsächlich aktuelle Themen aus den Medien mit dem Wissen der Lehrveranstaltung lösen können. Mit solchen Fragen versuche ich gleichzeitig aber auch noch ein weiteres Lernziel zu erreichen: dass Studierende sich mit aktuellen Themen auseinandersetzen, diese hinterfragen und eine gewisse Zeitungskultur entwickeln.“

Wissenschaftliche Artikel

„Da ich auch in der Forschung tätig bin, lese ich sehr viele wissenschaftliche Artikel und nutze diese auch, um mir Ideen für meine nächsten Prüfungsfragen zu holen. Am Ende eines Artikels gibt es oft weiterführende Fragen, die ich dann in MC-Fragen umwandle. Dies ist jedoch ein laufender Prozess. Das heißt, ich setze mich nicht kurz vor einer Prüfung hin und sehe alle Artikel durch, sondern schreibe mir eigentlich das ganze Jahr über immer wieder passende Fragen in einem Dokument zusammen. Beim Zusammenstellen der Prüfung formuliere ich dann die konkrete Arbeitsanweisung und die Antwortalternativen.“

Bücher mit MC-Fragen

„In der Länderkunde gibt es viele Bücher, bei denen man bei den Zusammenfassungen in den einzelnen Kapiteln mehrere offene oder bereits formulierte MC-Fragen findet, um sein Wissen noch einmal überprüfen zu können. Ich sehe mir jedoch nicht nur die Bücher an, die ich in Österreich bekomme. Auch bei meinen Auslandsreisen bin ich immer wieder auf der Suche nach Büchern aus dem jeweiligen Land, aus denen ich mir Anregungen für meine MC-Fragen holen kann. Entweder nehme ich dann die Fragen aus den Wiederholungskapiteln als Idee für zukünftige MC-Fragen oder blättere das Buch durch und markiere mir stichwortartig, wozu ich mit den Inhalten meiner Lehrveranstaltung eine Frage formulieren könnte.“

Häufige Fragen

„Nach meiner Lehrveranstaltung notiere ich mir immer die Fragen, die mir Studierende gestellt habe. Bei manchen Übungen höre ich auch Fragen, die sich Studierende gegenseitig stellen und notiere diese, um sie als Anregung für neue MC-Fragen zu verwenden. Auch Fragen, die Studierende in diversesten Foren stellen, liefern mir oft Ideen für neue MC-Fragen. Wird beispielsweise ein Rechenweg im Forum diskutiert, hilft mir das auch gleich beim Formulieren der falschen Antwortalternativen.“

Anonymisierte Realbeispiele

„Die besten Ideen für MC-Fragen liefert mir oft die Praxis an sich. Statt mir immer wieder neue Beispiele zu überlegen, nehme ich gerne Realbeispiele, anonymisiere sie und wandle sie gegebenenfalls für meine Anforderungen ein bisschen ab. Die Beispiele sammle ich jedoch nicht nur kurz vor der Prüfung, sondern lege mir in einer Datenbank immer wieder sogenannte Musterpatienten an, sobald mir ein Fall interessant erscheint. Zu diesem Musterpatienten speichere ich alle Daten ab, die ich bekommen kann. Vom Erstgespräch bis hin zu Röntgenbildern und Befunden habe ich alles in der Datenbank gespeichert. So kann ich einen Musterpatienten auch für unterschiedliche Fächer und Prüfungen verwenden. Vor der Zusammenstellung der Prüfung sehe ich mir dann alle Musterpatienten an und wähle jene aus, die ich für diese Prüfung brauchen kann.“

Ich verarbeite mein Leben

„Bei der Erstellung von MC-Fragen verarbeite ich oft mein eigenes Leben. Urlaube, Ausflüge, Vorträge, an denen ich teilgenommen habe, Situationen, die ich in der Arbeit erlebt habe: Aus allem versuche ich, anwendungsorientierte Beispiele zu erstellen. Ich versuche aus meinen Erlebnissen fallbeispielähnliche Situationen zu beschreiben, anhand derer ich die verschiedensten fachlichen Richtungen abprüfen kann.“

So habe ich zum Beispiel in meinem Urlaub eine Straußenfarm besucht und machte darauf verschiedenste kostenrechnerische Beispiele rund um diesen Bauernhof. Gespickt mit vielen echten Infos, muss dann beispielsweise berechnet werden, wie viel der Bauer für ein Kilo Fleisch verlangen muss, damit all seine Kosten gedeckt sind.

Ein anderes Mal hatte ich Probleme mit einem neu gekauften Produkt. Daraus konnte ich, etwas ausgebaut, eine gute rechtliche Frage zu den Themen Gewährleistung, Mangel und Lieferverzug gestalten.“

Aktuelle Ereignisse bei der Erstellung von Fragen verwenden

„Ich suche mir immer ein aktuelles Thema, das gerade in den Medien besprochen wird oder meine Studierenden bewegt. Das kann ein Zeitungsartikel, ein

Bericht im Fernsehen, ein aktueller Börsenbericht, ein neues Bauvorhaben oder ein Thema, über das gerade gesprochen wird (wie eine gerade stattfindende Sportveranstaltung, ein Umweltthema, eine Katastrophe usw.) sein.

Bei der letzten Fußball-Europameisterschaft in Österreich und in der Schweiz konnte man sehr gut Beispiele zum Thema Marketing erstellen. Mit Hilfe des aktuellen Börsenberichts müssen Daten aus diesem genommen werden und können zu Finanzierungsbeispielen verarbeitet werden. Ein bekannter Lektor hat Notfall-Fragen im Bereich Medizin erstellt. Nach dem Motto: Sie sind Ersthelfer im Stadion, vor Ihnen liegt eine verletzte Person am Boden mit folgenden Symptomen usw.“

Superhelden- oder Märchen-Technik

„Nach einiger Zeit ist es mir und meinem Team immer schwerer gefallen, neue Fragen zu erstellen. Auch gemeinsame Ideenfindungs-Treffen haben uns nicht geholfen. Bei einem solchen Tiefpunkt meinte jemand: Was ist, wenn Superman bei einem Fenster vorbeifliegt, ein Blumentopf runterfällt und einen Fußgänger trifft? Wie sieht dann die rechtliche Situation aus? Was eigentlich als Scherz gemeint war, hat sich sofort in ein Lauffeuer an neuer Freude und Kreativität entwickelt. Jeder wollte ein noch „lustigeres“ Beispiel machen und der Tiefpunkt war überwunden.

Wenn wir nun wieder mal in einer Sackgasse stehen, wählen wir besonders absurde Rahmenbedingungen, wie zum Beispiel, dass in jeder MC-Frage zumindest ein Superheld vorkommen muss, ein Grimm-Märchen verarbeitet wird oder auch Comicfiguren von Entenhausen Probleme haben, die aufgearbeitet werden müssen. So wird das oft eintönige Erstellen von MC-Fragen richtig zum Spaß. Und natürlich wird auch immer die kreativste und lustigste Frage innerhalb des Teams gekürt.“

Ich gebe meinen Prüfungserstellenden ein zentrales Thema vor

„Gerade am Ende des Semesters ist oft die Luft aus unserem Team heraus und es will uns einfach nichts Kreatives mehr einfallen. In solchen Fällen hat es sich bewährt, ein zentrales allgemeines Thema vorzugeben. Das kann ein bestimmtes Unternehmen in einer bestimmten Branche, zu dem dann Beispiele erstellt werden sollen, sein. Oder aber: Es wird ein allgemeines Thema vorgegeben, wie

zum Beispiel Skifahren, da die Prüfung im Februar ist. Bei einem gemeinsamen Brainstorming werden dann diese Themen ausgebaut, verschiedenste Aspekte beleuchtet oder die fiktiven Unternehmen ausgebaut. Danach gehen die meisten mit vielen Ideen nach Hause und haben neue Ansätze.

So können zu einem fiktiven Unternehmen Beispiele im Bereich Mathematik, Recht, Personalführung, Organisation, Marketing, Buchhaltung, Kostenrechnung, Volkswirtschaftslehre, Sprachen usw. erstellt werden. Die Palette könnte weitergeführt werden. Hat man ein Bild vor Augen, schreiben sich die Fallbeispiele fast von selbst.“

Multiples Erstellen von Fragen mit Antwortalternativen

„Hat man einen guten komplexen Fragenstamm bzw. eine gute Problemstellung entwickelt, habe ich es in der Vergangenheit oft als Verschwendung gesehen, nur fünf Antwortalternativen, d. h. nur eine einzige Frage daraus zu machen. Daher bin ich irgendwann dazu übergegangen, eine beliebige Anzahl an Antwortalternativen und Distraktoren zu erstellen. Bei einer Prüfung habe ich mir dann fünf ausgesucht und konnte das gleiche Beispiel nur mit anderen Antwortalternativen zum nächsten Prüfungstermin nochmals geben. Mit der Zeit wurde das ganze immer größer und ich habe mir eine Fragen-Datenbank in Access erstellt. Dort vermerke ich nun sogar, welche Frage bzw. auch welche Antwortalternativen bei welchem Prüfungstermin gegeben wurden und kann so noch leichter eine neue Klausur erstellen.

Teilweise gehe ich dann auch von hinten vor und nehme die Antwortalternativen einer Frage und verändere die Problemstellung. Die Antwortalternativen müssen dann oft nur noch leicht angepasst werden. So kann ich in kürzester Zeit sehr effektiv einen großen Pool an Fragen erstellen. Habe ich mal zwischendurch Zeit, kann ich den Pool an Problemstellung bzw. an Antwortalternativen immer leicht erweitern.“

Anhang

Weiterführende Literatur

Bloom, B. S. (Hrsg.) (1976): *Taxonomie von Lernzielen im kognitiven Bereich*. 5. Auflage. Weinheim: Beltz Verlag.

Case, S. M. & Swanson, D. B. (1998). *Constructing written test questions for the basic and clinical sciences*. National Board of Medical Examiners.

Downing, S. M. (2003). *Validity: On the meaningful interpretation of assessment data*. *Med Educ.* 37:830–837.

Lienert G. A. & Raatz, A. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.

Millman, J., Bishop, C. H. & Ebel, R. (1965). *Analysis of test wiseness in the cognitive domain*. *Educational and Psychological Measurement*, 18, 787–790.

Möltner, A., Schellberg, D. & Jünger, J. (2006). *Grundlegende quantitative Analysen medizinischer Prüfungen*. *GMS Z Med Ausbild*, 23(3), Doc53.

Porst, R. (2009): *Question Wording – Zur Formulierung von Fragebogen-Fragen*. In: *GESIS How-to-Reihe*, Nr. 2, S. 11.

Smolle, J. (2008). *Klinische MC-Fragen rasch und einfach erstellen*. Berlin, Boston: De Gruyter.

Zahlreiche kurze Zusammenfassungen zu unterschiedlichen Prüfungsformaten finden Sie auch am/im:

„Institut für medizinische Lehre“ (IML) an der Universität Bern:
http://www.impl.unibe.ch/dienstleistung/assessment_pruefungen/ (Juli 2013)

Lehre-Bereich der Webseite der Medizinischen Universität Graz
<http://www.medunigraz.at/lehre> (Juli 2013)

„Department für medizinische Aus- und Weiterbildung“ (DEMAW) an der Medizinischen Universität Wien:
<http://www.meduniwien.ac.at/bemaw/bemaw/index.php> (Juli 2013)

Glossar

Review. (Peer-Review) Begutachtung, meist durch fachkundige, aber auch durch fachfremde Personen. Der Schwerpunkt dieser Begutachtung liegt dann entweder auf den formalen Kriterien oder auf den inhaltlichen Kriterien. Je aussagekräftiger eine Prüfung sein soll und je massiver die Konsequenzen für Studierende sind, desto mehr Personen sollten eine einzelne (MC-)Frage begutachten.

Antwortraum. Ein Begriff, der verdeutlichen soll, wie viele Antwortalternativen überhaupt möglich sind. Bei binären/dichotomen Antwortmustern wäre dieser sehr klein – er würde aus zwei möglichen Antwortalternativen bestehen. Alle Viren und Bakterien (Krankheitserreger) auf diesem Planeten würden einen sehr großen Antwortraum darstellen. Je größer dieser Antwortraum ist, desto leichter findet man geeignete Distraktoren für eine spezielle Frage.

Antwortspreizung. Antwortspreizung beschreibt das Auseinanderliegen der Antwortalternativen. Je größer diese Spreizung ist, desto weiter liegen die einzelnen Antwortalternativen auseinander.

Binär. (Binäre Antwortmuster) Binär bedeutet, dass es zwei mögliche Antwortalternativen gibt. Typisch dafür wären die Optionen *richtig* oder *falsch*. Siehe auch dichotom (dichotome Antwortmuster).

Blueprint. Es handelt sich dabei um ein mehrdimensionales Raster von Prüfungsinhalten, wobei auch eine Gewichtung der einzelnen Themenbereiche eingetragen werden kann. Dieser kann bzw. soll verwendet werden, um sich einen Überblick über Umfang und Tiefe der Prüfung zu verschaffen.

Bottom-up-Prozess. Dabei erfolgt die Fragenerstellung auf umgekehrte Weise. Es werden zuerst die Antwortalternativen erstellt und erst danach wird eine Problem- und Fragestellung gesucht, die diese Alternativen verwendet.

Cues. Ist die englische Bezeichnung für „ungewollte Lösungshinweise“, bei der die Korrektheit einer Antwortalternative durch die Formulierung oder durch den Inhalt in ungewollter Weise hervorgehoben wird. Im Sinne der Reliabilität muss dies unbedingt vermieden werden.

Dichotom. (Dichotome Antwortmuster) Siehe binär.

Distraktor(en). Es handelt sich um falsche Antwortalternativen. Sie sollen so formuliert und inhaltlich beschaffen sein, dass sie von Kandidatinnen und Kandidaten mit fehlenden Kenntnissen nicht als solche erkannt werden. Andererseits sollen Distraktoren die Kandidatinnen und Kandidaten nicht bewusst in eine „Falle“ locken.

Geschlossene Frage. Dabei sind Antwortalternativen bzw. Antwortmöglichkeiten vorgegeben. Die/der Studierende muss eine oder mehrere dieser Möglichkeiten wählen. Das Gegenteil wäre die offene Frage.

Item. (Fragenitem) Dieser Begriff kommt aus der psychologischen Testung und bezeichnet eine Frage bzw. eine Aufgabe, die im Rahmen einer Prüfung gestellt wird. Aufgrund der Tatsache, dass nicht alle Fragen mit einem Fragezeichen enden, wäre diese Bezeichnung „neutraler“.

Item-Analyse. Anhand dieser werden die Messeigenschaften eines einzelnen (Fragen-)items beurteilt. Neben der Schwierigkeit und der Trennschärfe einer Frage werden auch die einzelnen Antwortalternativen untersucht. Gerade schwache Studierende sollten durch die falschen Antwortalternativen (deshalb auch der Begriff „Distraktoren“) von der richtigen Antwort „abgelenkt“ werden – dies wird mit der Distraktoren-Analyse gemacht.

Item-Schwierigkeit. Bezeichnet das Verhältnis von Studierenden, die dieses Item richtig gelöst haben, zu allen Studierenden, bei denen dieses Item Teil der Prüfung war. Die Werte können somit zwischen 0 (keine(r) hat das Item richtig gelöst) und 1 (alle haben das Item richtig gelöst) liegen. Bei Multiple-Response-Formaten wären andere Lösungen notwendig, hier könnte man beispielsweise die erreichten Punkte durch die erreichbaren Punkte dividieren.

Item-Trennschärfe. Diese wird durch den Trennschärfe-Koeffizienten angegeben und entspricht der Korrelation zwischen der Leistung beim einzelnen Item und der Gesamtleistung, daher liegt der Koeffizient zwischen -1 und +1. Es geht also um den Beitrag eines Items zum gesamten Test, dieser wird jedoch u. a. durch die Schwierigkeit, die Homogenität des Tests und die Item-Positionierung im Test beeinflusst.

Lernziel. Beschreibt den angestrebten Lerngewinn einer/eines Lernenden/Studierenden – auf einen bestimmten Inhalt bezogen. Oft werden Lernziele in einem Lernziel-Katalog zusammengefasst.

Lernziel-Taxonomie (nach Benjamin Bloom) Siehe Taxonomie.

Multiple-Choice. (kurz: MC) Ist ein geschlossener Fragen-Typ in schriftlichen Prüfungen, bei dem mehrere Wahlmöglichkeiten („Choices“) zur Verfügung stehen. Oft wird in diesem Zusammenhang „Single-Choice“ so verstanden, dass nur eine der angegebenen Wahlmöglichkeiten zutrifft. Besser wäre es, von Single- und Multiple-Response-Formaten zu sprechen.

Objektivität. Steht für die Unabhängigkeit der Prüfungsergebnisse vom Untersuchenden. Es kann zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden werden.

Offene Frage. Dabei sind keine Antwortalternativen bzw. Antwortmöglichkeiten vorgegeben. Bei schriftlichen Prüfungsformaten würde die Antwort der Studierenden als „Freitext“ gegeben werden. Das Gegenteil wäre die geschlossene Frage.

Reliabilität. (Zuverlässigkeit) Ist ein wichtiges Gütekriterium und fragt danach, wie genau ein Merkmal gemessen wird. Die Reliabilität kann von folgenden Faktoren negativ beeinflusst werden: mangelnde Objektivität, Rateinflüsse, kleine und nicht repräsentative Fragenauswahl und andere Zufälligkeiten.

Single-Choice. (kurz: SC) Steht für Fragen, bei denen eine der angegebenen Wahlmöglichkeiten richtig ist, wobei es besser wäre, von Single-Response zu sprechen (siehe auch MC-Frage).

Taxonomie. Bedeutet Klassifikationsschema und dabei werden Objekte nach bestimmten Kriterien klassifiziert. Ein für das Fragerstellen wertvolles Schema ist die sechsstufige Lernziel-Taxonomie nach Benjamin Bloom, er klassifiziert u. a. die „kognitiven“ Lernziele in Wissen, Verstehen, Anwenden, Analyse, Synthese und Evaluation.

Validität. (Gültigkeit) Ist ein wichtiges Gütekriterium und fragt danach, ob die Frage bzw. die Prüfung wirklich das misst, was zu messen beabsichtigt ist. Es gibt folgende Aspekte der Validität: Inhalts-, Kriteriums-, Konstrukt- und Augenschein-Validität.

Beispiel einer Checkliste zur Kontrolle von MC-Fragen

Sind folgende allgemeinen Kriterien eingehalten worden?

- Prüft die Frage relevantes Fachwissen (häufig Vorkommendes, häufige Fehlmeinungen, für später Relevantes)?
- Ist der Schwierigkeitsgrad an die Zielgruppe angepasst?
- Ist die Frage klar und eindeutig formuliert?
- Haben Sie folgende Lösungshinweise vermieden?
 - Grammatikalische Hinweise?
 - Distraktoren unterschiedlich formuliert?
 - Verbale Assoziationen zwischen Stamm und Antworten?
 - Konvergenz-Cue?
 - Absolute Begriffe?
 - Gegenseitige Hinweise zwischen den einzelnen Fragen?
- Ist der passende Fragentyp gewählt worden?
- Sind Negationen hervorgehoben und doppelte Verneinungen vermieden worden?

Sind folgende Kriterien für eine gute Frage eingehalten worden:

- Ausführlicher Fragenstamm?
- Alle erforderlichen Informationen vorhanden?
- Wenig bis keine überflüssigen Informationen?
- Kann die Frage auch ohne Kenntnis der Antwortalternativen beantwortet werden?

Sind alle Kriterien für gute Antwortalternativen eingehalten worden:

- Alle gleiche Länge?
- Alle inhaltlich und sprachlich homogen?
- Nur eine Aussage enthaltend?
- Klare Gründe für alle Distraktoren?
- Ist die Frage eindeutig beantwortbar?

Danksagung

Die Autorinnen und der Autor möchten sich bei allen **Personen** für ihre Unterstützung und ihre Zeit **im Rahmen der Interviews** herzlichst bedanken (chronologisch geordnet nach den Interviewterminen):

Interviewgruppe Graz

Univ.-Prof. Dr. Stefan Schulz, Univ.-Prof.ⁱⁿ Dr.ⁱⁿ Angelika Hofer, FH-Prof. DI Dr. Harald Burgsteiner, Univ.-Prof. Dr. Karl Öttl, FH-Prof. DI Dr. Stefan Grünwald, Univ.-Prof.ⁱⁿ Dr.ⁱⁿ Judith Glück, FH-Prof.ⁱⁿ Dr.ⁱⁿ Eva Koban-Röß, MBA, FH-Prof. Dr. Ulrich Frick, Univ.-Prof. Dr. Thomas Griesbacher, Prof. Dr. Herbert Schwetz, Univ.-Prof. Dr. Andreas Zimmer, Univ.-Prof.ⁱⁿ Dr.ⁱⁿ Margit Sommersguter-Reichmann

Interviewgruppe Wien

Mag.^a Claudia Dragosits, Mag.^a Birgit Gatterer, Dr. Robert Hammerl, MMag.^a (FH) Irmgard Fallmann, Mag. Reinhard Schott, DIⁱⁿ Dr.ⁱⁿ Christine Prenner, Mag.^a Eva Trappl, Univ.-Prof.ⁱⁿ Dr.ⁱⁿ Gabriela Kornek, Dr. Günter Schopf, Mag.^a Carole Faux-Loewe

Zudem auch ein Dankeschön, dass uns alle Interviewpartnerinnen und Interviewpartner die Erlaubnis gegeben haben, ihre MC-Fragen exemplarisch im Leitfaden verwenden zu dürfen. Diese MC-Fragen wurden nur im Sinne eines Positiv-Beispiels verwendet und mit dem Namen der jeweiligen Person in einer Fußnote versehen.

Besonderer Dank gilt dem **Verein „Forum neue Medien in der Lehre Austria“** für die Ermöglichung und Unterstützung dieses Projektes sowie Frau **Univ.-Prof.ⁱⁿ DIⁱⁿ Dr.ⁱⁿ Andrea Berghold** und **Frau Dipl.-Kff. Katrin Althammer** für die Unterstützung und Koordinierung an der jeweiligen Institution. Die Autorinnen und der Autor möchten sich ebenso herzlich bei **Univ.-Prof. Dr. Stefan Schulz** für sein sorgfältiges Gegenlesen sowie bei **Univ.-Prof. Dr. Josef Smolle** für seine Inspiration und Unterstützung bedanken.