

Hochschuldidaktik

Hochschuldidaktik A - Z

Multiple-Choice-Prüfungen

Ausgangslage

Multiple-Choice-Prüfungen (MC-Prüfungen) sind in der Hochschullehre ebenso verbreitet wie umstritten. Sie gelten als ökonomisch, scheinen aber nach der Einschätzung mancher Dozierenden den Fokus zu stark auf Wissensabfragen zu richten.

Wie alle Formen von Leistungsnachweisen haben auch MC-Prüfungen eine grundlegende Funktion: Dozierende möchten, dass ihre Studierenden gewisse Lernziele erreichen. Wie die empirische Hochschulforschung nachweist, wird die Realität nicht nur bei MC-Prüfungen diesem Anspruch nicht immer gerecht. Oft klaffen Ziele, Lehre und Leistungsnachweise stärker auseinander, als dies den Dozierenden bewusst ist.

Mit Blick auf diese Ausgangslage ist kritisch zu fragen, wie MC-Prüfungen im Vergleich zu klassischen schriftlichen Prüfungsformen abschneiden.

MC-Prüfungen im Vergleich

Mögliches Kriterium (neben anderen) für einen Vergleich zwischen klassischen Prüfungsformen und MC-Prüfungen sind die testtheoretischen Gütekriterien, welche in der Prüfungspraxis angemessen zu berücksichtigen sind. Leistungsnachweise sind immer auch Leistungsmessungen. Als solche müssen sie objektiv, reliabel (zuverlässig), valide (gültig) und auch ökonomisch sein.

Wie MC-Prüfungen im Vergleich zu klassischen schriftlichen Prüfungen hinsichtlich dieser Gütekriterien abschneiden, lässt sich nicht exakt beziffern. In der Tendenz zeichnet sich aber etwa folgendes Stärken/Schwächen-Profil ab (siehe Abbildung 1):

Da die Korrektur von MC-Prüfungen weitgehend von den korrigierenden Personen unabhängig ist und oft automatisiert wird, ist die Objektivität deutlich höher als bei klassischen schriftlichen Prüfungen. Wenn MC-Prüfungen prüfungstechnisch sorgfältig gemacht sind (siehe dazu unten), kann auch eine hohe Reliabilität (Zuverlässigkeit) erreicht werden. Problematischer ist die Validität (Gültigkeit) von MC-Prüfungen. Es ist eine anspruchsvolle Herausforderung an die Konstruktion

von MC-Prüfungen, komplexe kognitive Prozesse in Auswahlfragen abzubilden. Nur wenn dies gelingt, messen MC-Prüfungen das, was sie zu messen vorgeben (und sind daher valide/gültig). MC-Prüfungen verlagern den Arbeitsaufwand in die Erstellungsphase, während bei der Korrektur viel an Zeit und Ressourcen

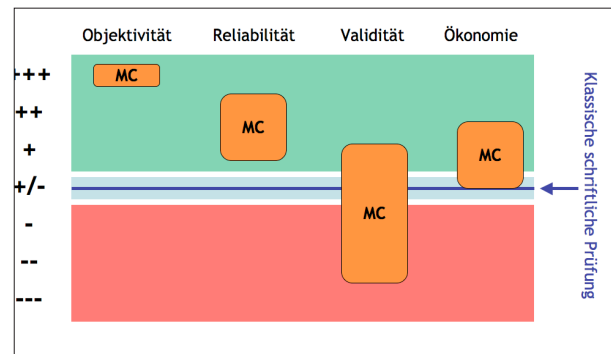


Abb. 1: Vergleich MC-Prüfungen und klassische schriftliche Prüfungen hinsichtlich der testtheoretischen Gütekriterien

gewonnen werden kann. Sie sind daher vor allem dann ökonomisch, wenn sie regelmässig und bei grossen Studierendenzahlen zum Einsatz kommen, so dass sich auch der Aufbau eines Fragen-Pools lohnt.

Die Validität von MC-Prüfungen

Die Validität von MC-Prüfungen ergibt sich nicht von selbst. Das grundsätzliche Problem besteht in der Diskrepanz, die zwischen den komplexen kognitiven Prozessen (die zum Lösen einer anspruchsvollen Aufgabe nötig sind) und der simplen äusseren Handlung des Ankreuzens einer Wahlantwort besteht. Wie können also differenzierte Analyseleistungen so in MC-Fragen übersetzt werden, dass auch wirklich das handlungsorientierte Durchdringen der Problem- und Sachlage und nicht bloss das Erkennen eines auswendig gelernten Faktums geprüft wird? Eine Frage, die vor allem in der Kompetenzdebatte im Rahmen von Bologna relevant wird.

Es sind vor allem zwei Kriterien, von welchen die Validität von MC-Prüfungen abhängt:

Hochschuldidaktik A - Z

Multiple-Choice-Prüfungen

2

1. Authentische Handlungssituation: Wissenschaftliche Ausbildungen zielen auf wissenschaftliche Handlungskompetenzen. In MC-Prüfungen kann vor allem jene Art von wissenschaftlicher Expertise gut gemessen werden, die das schnelle Beurteilen von unterschiedlichen Facetten eines vielschichtigen Themenkomplexes zum Ausdruck bringt. MC-Prüfungen können daneben auch wissenschaftliche Handlungssituationen simulieren, die das relativ schnelle Lösen von Anwendungen erfordern, wobei die Korrektheit des Handlungsergebnisses von zentraler Bedeutung ist.

2. Adäquate Aufgabenkomplexität: Die einzelnen Aufgaben/Fragen einer MC-Prüfung müssen gut auf die Lernziele und das didaktische Konzept der Lehrveranstaltung (beziehungsweise des Moduls) abgestimmt sein. Neben der inhaltlichen Übereinstimmung sollen auch die in der Lehrveranstaltung eingeübten Denkprozesse auf demselben Komplexitätsniveau liegen, wie sie in der Prüfung zu leisten sind. Um diesen Abgleich vorzunehmen, kann etwa auf eine Taxonomie verschiedener Komplexitätsstufen von Lernzielen und Prüfungsaufgaben zurückgegriffen werden. (Eine bekannte Taxonomie ist jene der Arbeitsgruppe um Benjamin Bloom. Darin werden sechs Stufen unterschieden: Wissensreproduktion, Verstehen, Anwenden, Analyse, Synthese und Beurteilung).

Vor dem Hintergrund einer solchen Systematik ist zu fragen, ob alle Komplexitätsstufen durch MC-Fragen geprüft werden können. Die Erfahrung mit MC-Tests in verschiedenen Disziplinen zeigt, dass Fragen auf den ersten vier Stufen (Wissensreproduktion, Verstehen, Anwenden, Analyse) gut ausgestaltet werden können. Synthese- und Beurteilungsfragen sind als MC-Fragen nur bedingt oder kaum zu realisieren.

Konstruktion einer MC-Prüfung - Fragetypen

Im Rahmen der Medizinalprüfungen in den USA werden MC-Prüfungen schon seit ca. 60 Jahren eingesetzt und intensiv erforscht. Dabei wurden auch viele verschiedene MC-Fragetypen entwickelt und getestet. Es sind mitunter zwei Erkenntnisse, die für die Konstruktion von MC-Prüfungen von besonderer Bedeutung sind:

- **Standardeinsatz:** Es empfiehlt sich, in einer MC-Prüfung nicht zu viele verschiedene Fragetypen einzusetzen und dabei testtheoretisch bewährte Typen in ihrer Standardversion zu bevorzugen. Auch die Bewertung (Punktevergabe für korrekte und falsche Antworten) sollte möglichst nach einem einheitlichen und einfachen Muster erfolgen.

- **Standardtypen:** Die drei testtheoretisch am besten ausgewiesenen Fragetypen sind (Beispiele siehe Anhang)

. *Typ A:* Frage mit 5 Wahlantworten, aus welcher die einzig richtige oder die beste Wahlantwort zu bestimmen ist.

. *Typ B:* Zu 5 vorgegebenen Wahlantworten folgen 2 bis 5 Fragen. Für jede dieser Fragen ist aus den 5 Wahlantworten die einzig richtige oder die beste Antwort zu bestimmen. Dies ergibt eine Serie von mehreren Fragen, die unabhängig voneinander bewertet werden (eine Wahlantwort kann mehrmals die richtige/beste sein).

. *Typ Kprim:* Zu einer Frage folgen 4 Aussagen. Bei jeder Aussage muss bestimmt werden, ob sie richtig oder falsch ist. Sind alle 4 Aussagen korrekt eingeschätzt, erhält der Kandidat/die Kandidatin 1 Punkt, sind 3 der 4 Aussagen korrekt eingeschätzt, werden 0.5 Punkte vergeben.

Literatur zur Vertiefung

Eine sehr hilfreiche Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen mit vielen Beispielen bietet die Zusammenstellung von René Krebs: http://www.iml.unibe.ch/dienstleistung/assessment_pruefungen/

Eine ausführliche Darstellung verschiedener MC-Fragetypen findet sich im Manual Constructing Written Test Questions For the Basic and Clinical Sciences des National Board of Medical Examiners (USA): <http://www.nbme.org/publications/item-writing-manual.html>

Eine kritische Darstellung und Weiterentwicklung der Taxonomie kognitiver Prozesse von Bloom leisten: Anderson, L. W. & Krathwohl, D. R. (Eds). (2001). A Taxonomy for Learning, Teaching and Assessing. New York: Longman.

Hochschuldidaktik A - Z

Multiple-Choice-Prüfungen

Anhang: Beispiele zu den Standardfragetypen

Typ A – Einfachauswahl

Klassische Genetik

(Markieren Sie, welche der Wahlantworten A – E die richtige oder beste Ergänzung des folgenden Satzanfanges ist.)

Die Hardy-Weinberg Regel besagt, dass...

- A die Kreuzung reinerbiger Individuen zu uniformen Nachkommen führt.
- B Kreuzungen der Nachkommen (F1) zweier reinrassiger Elternlinien untereinander zu Aufspaltung der Phänotypen nach bestimmten Zahlenverhältnissen führen.
- C Allelfrequenzen sich von Generation zu Generation nicht verändern, falls gewisse Bedingungen erfüllt sind.
- D die AB0-Blutgruppen des Menschen in den verschiedenen Populationen auf der Welt nicht gleich sind.
- E die natürliche Selektion und die Heritabilität von Merkmalen treibende Kräfte der Evolution sind.

3

Typ B – Zuordnungsfragen

Vergleichende Anatomie

(Auf die mit den Buchstaben A - E bezeichneten Wahlantworten folgen drei nummerierte Fragen. Markieren Sie zu jeder Frage die richtige oder beste Wahlantwort. Ein und dieselbe Wahlantwort kann mehr als einmal die richtige/beste sein.)

- A Der Wolff'sche Gang funktioniert im männlichen Geschlecht als Harnsamenleiter.
- B Das Pronephros ist die funktionale Adultniere.
- C Das Mesonephros wird zum Nebenhoden.
- D Der Müller'sche Gang ist im männlichen Geschlecht Harnleiter.
- E Es hat keine Veränderung stattgefunden.

Das Urogenitalsystem der Wirbeltiere hat im Laufe der Evolution verschiedene Veränderungen erfahren. Welche Veränderung (A - E) hat stattgefunden bei den

- | | | | | | |
|--------------|---|---|---|---|---|
| 1. Amphibien | A | B | C | D | E |
| 2. Reptilien | A | B | C | D | E |
| 3. Säugern | A | B | C | D | E |

Hochschuldidaktik A - Z

Multiple-Choice-Prüfungen

Typ Kprim – Vierfache Entscheidung „richtig/falsch“

Entwicklungsbiologie

(Beurteilen Sie bei jeder der Antworten 1- 4, ob sie in Bezug auf die folgende Aussage richtig oder falsch ist, und markieren sie die entsprechende Angabe. Es können 0 bis 4 Antworten richtig sein. Die korrekte Beurteilung aller 4 Antworten ergibt einen Punkt.

Für drei richtige Beurteilungen erhalten Sie einen halben Punkt.)

- 4 Eine wichtige Voraussetzung für entwicklungsbiologisches Denken war es, die Gleichwertigkeit der Blastomerenkerne (genomische Äquivalenz) nachzuweisen. Die genomische Äquivalenz ist durch die folgenden Experimente klar bewiesen:

1. Organisatortransplantationen durch Sperman und Mangold	richtig	falsch
2. Klonen von Dolly	richtig	falsch
3. Klonen von Amphibien durch Gurdon, Briggs und King	richtig	falsch
4. Herstellung transgener Organismen	richtig	falsch

Autor der Beispielfragen: Prof. Dr. Robert Stidwill, Universität Zürich