



Sie sind hier: ► [Universität Wien](http://www.univie.ac.at/) ► [Studienservice und Lehrwesen](http://studien-service-lehrwesen.univie.ac.at/) ► [Center for Teaching and Learning](http://ctl.univie.ac.at/) ► [Infopool besser lehren \(/\)](#) ► [Startseite \(/startseite/\)](#) ► [Prüfen & Beurteilen \(/startseite/pruefen-beurteilen/\)](#) ► [Beurteilen von Prüfungen](#)

Beurteilen von Prüfungen

Reinhard Schott

November 2018

Überblick

1. Bezugsnormen als Maßstab für die Beurteilung
 2. Beurteilen von schriftlichen und mündlichen Prüfungen
 - 2.1 Multiple Choice-Fragen und schriftliche Prüfungen mit einfachen offenen Fragen
 - 2.2 Schriftliche Prüfungen (Essay-Fragen) und mündliche Prüfungen
 - 2.3 Notenschlüssel
 3. Gütekriterien von Prüfungen
 4. Beurteilungsfehler
 - 4.1 Beurteilungsfehler vermeiden
-

Prüfungsnoten an Universitäten erfüllen verschiedene **Funktionen**: sie bewerten einen spezifischen Leistungsstand, geben den Studierenden eine Rückmeldung und informieren Dritte über das Leistungsniveau der AbsolventInnen, insbesondere andere Universitäten bei vertikaler Mobilität sowie Arbeitgeber^[1].

Diese Funktion können Noten dann erfüllen, wenn sie auf Basis von vorab definierten Zielen und Kriterien **zwischen unterschiedlichen Leistungsniveaus der Studierenden differenzieren**.

1. Bezugsnormen als Maßstab für die Beurteilung

Um eine Leistung fair beurteilen zu können, sind am besten die **vorab definierten Studienziele (die zu erwerbenden Kenntnisse und Kompetenzen der Studierenden) als Bezugsnorm** heranzuziehen.

Die Studienziele sollten auch den Studierenden bekannt gemacht werden, ebenso gilt es, ein **Kriterium** für „Ziel erreicht“ zu definieren (z.B.: „mehr als 50% der Punkte müssen für eine positive Note erreicht werden“ – siehe „Notenschlüssel“ weiter unten). Bei dieser **kriteriumsorientierten Bezugsnorm** gelten absolute **Leistungsstandards (in welchem Ausmaß werden die gesetzten Leistungsstandards bzw. Studienziele von der geprüften Person erfüllt?)**.^[2]

In der universitären **Praxis** kann der soziale Kontext in der Prüfungssituation eine Rolle spielen, besonders bei mündlichen Prüfungen. Allerdings ist eine **Orientierung an der sozialen Norm (die einzelne Leistung wird relativ zur Leistung aller anderen beurteilt) studienrechtlich problematisch** und den Studierenden gegenüber unfair. Bei Bewertungen nach der sozialen Norm geht man davon aus, dass die Leistungen der Studierenden einer Standardnormalverteilung (**Gauß'sche Glockenkurve**) folgen. Das führt im Extremfall dazu, dass die Besten eines Prüfungstermins – unabhängig vom Erreichen der Studienziele – ein „Sehr gut“ und die Schlechtesten ein „Nicht genügend“ erhalten. Ebenso könnten PrüfungsteilnehmerInnen mit der gleichen Leistung bei verschiedenen Prüfungsterminen unterschiedlich benotet werden. **Die Beurteilung nach der sozialen Norm ist mit den studienrechtlichen Vorgaben im Universitätsgesetz und der Satzung der Universität Wien nicht vereinbar.**

2. Beurteilen von schriftlichen und mündlichen Prüfungen

Allen Methoden der Leistungsüberprüfung gemeinsam ist, dass **Musterlösungen** und/oder **Kriterien für das Beurteilen der Antworten** eine möglichst standardisierte und faire Beurteilung der Leistungen ermöglichen. Der größte **Praxisvorteil bei Musterlösungen** ist, dass die einzelnen Prüfungen (von unterschiedlichen PrüferInnen) nach dem gleichen Schema und den gleichen Maßstäben, fair beurteilt werden können. Nicht zuletzt stellen Musterlösungen und Kriterien für das Beurteilen der Antworten eine nützliche Argumentationshilfe bei möglichen Prüfungseinsichten bzw. Einsprüchen dar (vgl. UG § 79 (5) und § 84).

Je nach Art der Prüfung ist dieser Schritt unterschiedlich aufwändig und sollte im Vorfeld erfolgen. **Generell gilt:** Je größer der Gestaltungsraum für die Studierenden beim Erbringen des Leistungsnachweises ist, desto anspruchsvoller ist das Festlegen und Aufschlüsseln von Kriterien für das Beurteilen der Antworten.

Studierende können sich ein gutes Bild über die Anforderungen einer Prüfung machen, wenn sie zuvor einige Beispielfragen mit Musterlösungen zur Verfügung gestellt bekommen.

2.1 Multiple Choice-Fragen und schriftliche Prüfungen mit einfachen offenen Fragen

Hier ist das Festlegen von **Kriterien für das Beurteilen der Antworten** meist einfach, indem als Kriterium „richtig/falsch angekreuzt“ bzw. „richtiges/falsches Wort eingefügt“ verwendet werden kann. Für jede richtige Antwort wird beispielsweise ein Punkt vergeben (oder auch ein halber Punkt für eine teilweise richtige Antwort) und ab einer bestimmten Anzahl von Punkten wird die Leistung positiv beurteilt.

	MC-Fragen	schriftliche Kurzantwortfragen
Beispiele für Kriterien zur Beurteilung der Antworten	angekreuzt nicht angekreuzt teilweise richtig angekreuzt	richtig falsch teilweise richtig

2.2 Schriftliche Prüfungen (Essay-Fragen) und mündliche Prüfungen

Gerade für die Korrektur/Beurteilung von offenen Fragen (schriftlichen und mündlichen) empfiehlt es sich, im **Vorfeld Musterlösungen und/oder Kriterien für das Beurteilen der Antworten** zu formulieren. In einer Musterlösung wird eine ideale Lösung formuliert, wobei alle wichtigen Aspekte und Begriffe genannt werden.

Musterlösungen eignen sich hier vor allem für Fragen, auf die es nur eine sinnvolle richtige Antwort gibt. Entsprechend der Klassifikation nach Bloom^[3] sind das Wissens-, Verständnis- und Anwendungsfragen.

Zusätzlich besteht die Möglichkeit, einzeln zu bewertende Kriterien für jede Frage zu erstellen, sodass **diese einzelnen Kriterien** am Ende ein **Beurteilungsschema** ergeben.

Offene schriftliche oder mündliche Fragen

Beispiele für Kriterien zur Beurteilung der Antworten	alle Teilaspekte wurden genannt; die geforderten Beispiele wurden angeführt; die zentralen Fachausdrücke wurden richtig verwendet; Zusammenhänge wurden erkannt und aufgezeigt; Lösungen wurden präsentiert und begründet; die Argumentation ist schlüssig; usw....
--	---

Beispiel für eine schriftliche Wissensfrage mit Musterlösung sowie ein Bündel von Kriterien (Beurteilungsschema)

Frage 1

- Was sind Nocizeptoren, wo kommen diese vor und worin unterscheiden sie sich von anderen Sinnen? (3 Punkte möglich)

Musterlösung

- Nocizeptoren sind Schmerzrezeptoren, die sich im gesamten Körper außer im Gehirn und in der Lunge befinden. Im Gegensatz zu anderen Sinnen tritt bei Wiederholung oder Andauern von Schmerzreizen keine Adaptation ein, sondern das Gegenteil, eine Sensibilisierung.

(Mögliche) Kriterien für das Beurteilen der Antworten

- Es wird erklärt, was Nocizeptoren sind (1 Punkt)
- Es wird darauf hingewiesen, wo Nocizeptoren vorkommen (1 Punkt)
- Der Unterschied zu anderen Sinnen wird erklärt (1 Punkt; wenn nur Adaptation oder nur Sensibilisierung genannt wird ½ Punkt)
- Weitere nicht direkt erfragte, aber relevante Anmerkungen (1 Punkt)
- Insgesamt aber nur 3 Punkte möglich

2.3 Notenschlüssel

Die Leistung der einzelnen Studierenden soll mit im Vorfeld definierten **Beurteilungskriterien** (Notenschlüssel) verglichen werden. Der Grad, in dem die auf Basis der Studienziele definierten Anforderungen erfüllt werden (kriteriumsorientierte Bezugsnorm), bestimmt die Note. Die Note der einzelnen Studierenden hängt also von der eigenen Leistung im Vergleich zu einem vorher festgelegten **Kriterium** und nicht von der Leistung der anderen Studierenden ab. **Im Normalfall wird eine Punkteanzahl bestimmt, für die die Studierenden gerade noch eine positive Note erhalten (Bestehensgrenze).** Die Bestehensgrenze und die Aufteilung der restlichen Noten werden in der Regel von den Lehrenden festgelegt. **In der Praxis werden beispielsweise oft „mehr als 50 %“ der Punkte für eine positive Note verlangt.** Zwischen der Bestehensgrenze und der maximalen Punktezahl können die restlichen Noten entweder linear bestimmt werden (der Abstand zwischen 4 und 3 ist gleich groß wie der Abstand zwischen 3 und 2 oder 2 und 1), oder man kann einer anderen Verteilung folgen (z.B.: der Abstand zwischen 4 und 3 ist größer als der Abstand zwischen 3 und 2, dieser Abstand ist wiederum größer als der Abstand zwischen 2 und 1).

3. Gütekriterien von Prüfungen

Prüfungen sollten idealerweise bestimmte **Gütekriterien** erfüllen.^[4] Um die Qualität einer Prüfung zu sichern oder zu heben, sind bei der Konstruktion und Konzeption einer Prüfung insbesondere folgende Fragen zu berücksichtigen:

- **Misst die Prüfung das, was sie messen soll, ist sie gültig (Validität)?** Deckt eine Prüfung den Umfang der Studienziele repräsentativ ab? Messen die Prüfungsaufgaben das in den Studenzielen definierte Wissen und Können?
Um diese Fragen in der **Praxis** zu klären, können sich **KollegInnen gegenseitig Feedback auf Prüfungsfragen** geben. Jede Prüfungsfrage wird dahingehend überprüft, ob sie „misst, was die Prüfung messen soll“, d.h. misst die Frage eines der Studienziele? Hierfür eignen sich KollegInnen, die mit dem Prüfungsstoff und den Studenzielen vertraut sind. Neben der Studienzielorientierung können die KollegInnen auch überprüfen, ob bei der Frage ungewollt Sprachverständnis oder Aufmerksamkeit mitgeprüft werden.
- **Wie genau misst die Prüfung (Reliabilität)?** Eine Prüfung ist dann reliabel, wenn das geprüfte Merkmal (Wissen und Können in einem bestimmten Fach) mit einem hohen Maß an Genauigkeit gemessen wird. **Eine höhere Anzahl mittelschwerer Prüfungsfragen führt in der Regel zu einer höheren Reliabilität.** Wenn die Umstände in der Praxis es zulassen, sollen in **Parallelgruppen dieselben Fragen** (vielleicht in unterschiedlicher Reihenfolge) vorgegeben werden.
- **Sind die Ergebnisse einer Prüfung unabhängig von der durchführenden Person (Objektivität)?** Sind die Ergebnisse einer Prüfung unabhängig davon, wer diese durchführt, auswertet oder interpretiert? Werden alle Studierenden nach den gleichen Kriterien beurteilt? In der Praxis ermöglichen im Vorfeld verfasste **Musterlösungen und/oder Kriterien** für die Beurteilung der Antworten, eine standardisierte und faire Beurteilung der Leistungen über alle Studierenden hinweg oder durch verschiedene PrüferInnen.

Daneben spielen natürlich u.a. die notwendigen Ressourcen (Zeit, Budget, Material etc.) gemessen am Informationsgewinn einer Prüfung (Gütekriterium der **Ökonomie**) und die zeitliche, psychische und körperliche Belastung für die Studierenden (**Zumutbarkeit**) eine Rolle.

4. Beurteilungsfehler

Um eine faire Korrektur von schriftlichen Prüfungen und eine fehler- und verzerrungsfreie Durchführung mündlicher Prüfungen zu gewährleisten, sollten mögliche Beurteilungstendenzen und Beurteilungsfehler bekannt sein, um - falls notwendig - gezielt gegensteuern zu können.^[5]

Erwartungseffekte: Positive oder negative Erwartungen von Lehrenden an die Studierenden können sich nach Art einer sich selbst erfüllenden Prophezeiung auf die Beurteilung der Leistungen auswirken. Erwartungen an die Notenverteilung können ebenfalls Auswirkungen haben. So widerstrebt es etwa vielen Lehrenden, fünf- oder zehnmal hintereinander ein „Sehr gut“ zu vergeben, da dies einer impliziten Annahme widerspricht, dass derartig viele „Sehr gut“ nicht hintereinander auftreten können.

Projektionsfehler und Halo-Effekt: Eigene Eigenschaften, Ansichten, Wünsche oder Fehler (zumeist unbewusst) werden auf die Studierenden übertragen und fließen in die Beurteilung ein. Von einem Merkmal wird auf ein anderes, vielleicht völlig unabhängiges Merkmal geschlossen. Die Wahrnehmung einer Person „überstrahlt“ die zu messende Leistung (Halo-Effekt), wenn sich etwa Kleidung, Sprache, Schriftbild oder Attraktivität von Studierenden auf die Beurteilung ihrer Leistung auswirken. Das kann z. B. dazu führen, dass die wahrgenommene „Begabung“ oder Persönlichkeitseigenschaften und nicht die Leistung beurteilt wird.

Reihenfolgeeffekt: Die Reihenfolge bei der Beurteilung von Prüfungen kann das Ergebnis beeinflussen. So besteht oft die Neigung, bei der Durchsicht die ersten Arbeiten strenger zu beurteilen als die letzten. Auch das Leistungsniveau vorangegangener Prüfungen kann sich auf die Beurteilung auswirken, etwa indem eine mäßige Arbeit besser beurteilt wird, wenn vorher einige sehr schlechte Arbeiten gelesen wurden bzw. umgekehrt. Bei mündlichen Prüfungen merkt man sich die Leistung zu Beginn und am Ende einer Prüfung am besten, diese haben so einen größeren Einfluss auf die Beurteilung.

Strenge- und Mildefehler: Schon „geringe Mängel“ fließen übermäßig stark in die Beurteilung ein und „gute“ Leistungen werden kaum gewertet (Strengefehler). „Gute“ Leistungen werden besonders stark gewichtet, „schlechte“ Leistungen fließen hingegen kaum in die Bewertung ein (Mildefehler). Die eigene Leistungserwartung (z.B. von „jugendstrenge“ oder „altersmilde“ BeurteilerInnen) sollte die Beurteilung nicht beeinflussen.

Tendenz zu Extremurteilen: Es wird vor allem zwischen „guten“ und „schlechten“ Leistungen unterschieden. Ist die Schwelle „gut“ erreicht, wird ungerechtfertigt die beste Beurteilung vergeben, bei Nichterreichen die schlechteste Beurteilung. Urteile in der Mitte des Notenspektrums werden vermieden.

Tendenz zur Mitte: Im Gegensatz zur Tendenz zu Extremurteilen kann auch die Tendenz bestehen, eindeutige (extreme) Urteile zu vermeiden. Die Tendenz zur Mitte tritt vor allem dann auf, wenn die Beurteilenden unsicher sind.

4.1 Beurteilungsfehler vermeiden

Lehrende haben verschiedenste Möglichkeiten, Beurteilungsfehler zu minimieren. Bei schriftlichen Prüfungen können Beurteilungsfehler wie etwa der Halo-Effekt, Reihenfolgeeffekte, Erwartungseffekte oder Strenge- und Mildefehler minimiert werden, indem **Prüfungen „quer“ beurteilt** werden. D.h. es wird jeweils eine Prüfungsaufgabe bei allen Prüfungsteilnehmenden kontrolliert, statt jeden Prüfungsbogen in einem Durchlauf zu bewerten. Der „Gesamteindruck“ über die Leistung sollte sich erst am Ende der Korrektur und nicht schon nach ein paar Fragen bilden. Ähnlich hilfreich sind **Musterlösungen und/oder Beurteilungsschemata** die bei mündlichen

Prüfungen beim Verringern der Beurteilungsfehler helfen können. Auch kollegiales Feedback, beispielsweise bei kommissionellen mündlichen Prüfungen, kann zu einer objektiven Beurteilung beitragen.^[6]

Bei längeren Korrekturarbeiten oder bei mündlichen Prüfungen mit vielen Studierenden sollten **Pausen** eingebaut werden. Schwankungen in der Strenge der Beurteilung, Kontrasteffekte zu vorherigen Prüfungen und eigene Ermüdungseffekte können so verringert werden.

Darüber hinaus kann es bereits helfen zu überlegen, inwieweit eine persönliche **Neigung zu bestimmten Beurteilungstendenzen** (Strenge oder Milde; Tendenz zu Extremurteilen etc.) besteht und diese bei der Beurteilung in Erinnerung zu behalten. Insbesondere bei mündlichen Prüfungen sollte versucht werden, hervorstechende Merkmale der Studierenden, die mit der eigentlich zu beurteilenden Leistung nichts zu tun haben, auszublenden. Möglicher Weise auftretender **Ärger** über die Leistungen sollte nicht über Gebühr in die Beurteilung einfließen.

Quellen

[1] Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht der Geschäftsstelle mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates. Hamburg 2012. www.wissenschaftsrat.de/download/archiv/2627-12.pdf.

[2] Metzger, Christoph, und Charlotte Nüesch. *Fair prüfen. Ein Qualitätsleitfaden für Prüfende an Hochschulen*. St. Gallen: Institut für Wirtschaftspädagogik, 2004.; Walzik, Sebastian. *Kompetenzorientiert prüfen. Leistungsbewertung an der Hochschule in Theorie und Praxis*. Opladen und Toronto: Verlag Barbara Budrich UTB, 2012; Zumbach, Jörg, und Hermann Astleitner. *Effektives Lehren an der Hochschule*. Stuttgart: Kohlhammer, 2016.

[3] Bloom, Benjamin S., Hrsg. *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim und Basel: Beltz, 1956/1972.

[4] Kubinger, Klaus. D. *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe, 2009.

[5] Zumbach und Astleitner. *Effektives Lehren an der Hochschule* [2]; Walzik. *Kompetenzorientiert prüfen* [2].

[6] Zumbach und Astleitner. *Effektives Lehren an der Hochschule* [2].

Empfohlene Zitierweise

Schott, Reinhard: *Beurteilen von Prüfungen*. Infopool *besser lehren*. Center for Teaching and Learning, Universität Wien, November 2018. [<https://infopool.univie.ac.at/startseite/pruefen-beurteilen/ beurteilen-von-pruefungen/>]

Dieser Text ist lizenziert unter Creative Commons

Namensnennung – Weitergabe unter gleichen Bedingungen 3.0 Österreich (CC BY-SA 3.0 AT)

Mehr Informationen unter <https://creativecommons.org/licenses/by-sa/3.0/at/>

(<https://creativecommons.org/licenses/by-sa/3.0/at/>)



(<https://creativecommons.org/licenses/by-sa/3.0/at/>)

Kontakt

T (Telefon): +43-1-4277-120 60
infopool@univie.ac.at

Universität Wien | Universitätsring 1 | 1010 Wien | T (Telefon) +43-1-4277-0

[Sitemap \(/sitemap/\)](/sitemap/) | [Impressum \(/impressum/\)](/impressum/) | [Barrierefreiheit \(https://www.univie.ac.at/ueber-uns/barrierefreiheit/\)](https://www.univie.ac.at/ueber-uns/barrierefreiheit/) | [Datenschutzerklärung \(https://dsba.univie.ac.at/datenschutzerklaerung\)](https://dsba.univie.ac.at/datenschutzerklaerung) | [Druckversion](#)