



ii.00

digital
kompetenzorientiert
prüfen

Fact Sheet für digitale kompetenzorientierte Prüfungen

Teil 5: Wie wird ausgewertet?

1. Bewertungskriterien	2
2. Bezugsnormen	3
3. Bewertungsraster	3
4. Bewertungsfehler	4
5. Bewertung schriftlicher Prüfungen (Klausuren)	6
6. Bewertung mündlicher Prüfungen und Präsentationen	6
7. Bewertung von Ausarbeitungen	7



Die Auswertung sollte bereits **vor** der Prüfung durchdacht und geplant werden. So können den Studierenden bereits vor der Prüfung die Bewertungskriterien genannt und gegebenenfalls Bewertungsfehler frühzeitig vermieden werden.

Bewertungskriterien

Bewertungskriterien reichen von der Festlegung, welche Antworten als richtig, teilweise richtig oder falsch gewertet werden, bis hin zu genauen Regeln der Punktevergabe. Die Objektivität, Reliabilität und Validität der Leistungseinschätzung kann durch gute und transparente Bewertungskriterien deutlich erhöht werden. (vgl. TU Darmstadt 2024)

Die Bewertung von Prüfungsaufgaben der unteren Taxonomiestufen ist einfacher, da hier die Antworten in der Regel nur korrekt und vollständig wiedergegeben werden müssen. Dies kann über die Identifikation von relevanten Aspekten und ein einfaches Richtig-Falsch-Schema geschehen. (vgl. Hagen et. al. 2015)

Prüfungsaufgaben, die sich auf eine höhere Taxonomiestufe beziehen, können nicht nach diesem System bewertet werden, da hier die Antwort auf die Prüfungsaufgabe über die Kriterien „richtig/ falsch“ oder „vollständig“ hinaus gehen muss. Die Bewertung sollte empirisch beobachtbar und für andere verständlich beschrieben werden, sodass nachvollzogen werden kann, wie die Bepunktung zustande gekommen ist. Auch hier gilt es, sich an die festgelegten Lernziele zu halten, da in den Lernzielen die zu erwerbenden Kompetenzen genau definiert sind. Die Prüfungsaufgabe selbst muss die Studierenden aber auch dazu auffordern, im Sinne des Lernzieles zu handeln und damit zu zeigen, auf welchem Niveau sie die dort beschriebenen Kompetenzen erworben haben. (vgl. Hagen et. al. 2015)



Bezugsnormen

Um eine erbrachte Leistung letztendlich als „gut“, „ausreichend“ oder „nicht ausreichend“ einstufen zu können, braucht es ein normiertes Referenzsystem, einen Standard. Der Standard wird als Bezugsnorm bezeichnet. Die Bezugsnorm gibt Aufschluss darüber, wie die erbrachte Leistung im Vergleich zu dem festgelegten Standard steht. Es können drei **Bezugsnormen** unterschieden werden:

Soziale Bezugsnorm

Bei der sozialen Bezugsnorm wird die erbrachte Leistung eines Individuums mit der Durchschnittsleistung der Gruppe in Relation gesetzt. Im Vergleich steht demnach die individuelle Leistung einer/ eines Studierenden mit den Leistungen der anderen Studierenden. Es wird erfasst, wie „gut“ oder „schlecht“ eine Person im Vergleich zu anderen Personen war.

Sachliche Bezugsnorm

Die Referenz der sachlichen Bezugsnorm ist ein vorab festgelegtes Lernziel. Es wird also überprüft, ob und in welchem Maß Studierende die mit dem Lernziel festgelegten Kompetenzen erworben haben. Wird die sachliche Bezugsnorm gewählt, können vorab objektive Bewertungskriterien festgelegt und wiederum in ein Bewertungsraster überführt werden.

Individuelle Bezugsnorm

Mithilfe der individuellen Bezugsnorm kann festgestellt werden, ob und wie stark sich die Studierenden im Laufe der Lehrveranstaltung/ des Studiums entwickelt haben. Dabei wird die Leistung einer Person über den zeitlichen Verlauf innerhalb eines bestimmten Zeitraums gemessen und mit den vorherigen Leistungen derselben Person in Relation gesetzt. Diese Bezugsnorm zeigt am deutlichsten den Lernfortschritt einzelner Personen auf und stärkt damit auch die Leistungsmotivation.

(vgl. Macke et. al. 2016; Rheinberg 2008; Schott 2018; Universität Zürich 2023)

Bewertungsraster

Hilfreich bei der Zuordnung der Antworten zu den jeweiligen Lernzielen und bei der Einordnung der Leistung im ersten Schritt in ein Punktesystem und im zweiten Schritt in ein Notensystem, ist ein sogenanntes **Bewertungsraster**. Dieses sollte im Sinne des Constructive Alignments bereits parallel zu und mit der Entwicklung der Lehrveranstaltung und der Lernziele erstellt werden. Laut Schindler (2015) „[enthält e]in vollständiges Bewertungsraster Bewertungskategorien, die die vorab definierten Kompetenzbeschreibungen abbilden“. Die Bewertungskategorien sind durch ein Regelwerk definiert, sodass eine eindeutige Zuordnung von Prüfungsantworten ermöglicht wird. Die Angabe von Ankerbeispielen sowie Angaben zur Punktevergabe tragen zu einer weiteren Erhöhung der Objektivität, Reliabilität und Validität der Korrektur und Auswertung bei. Mit einem Bewertungsraster kann die Wahrscheinlichkeit erhöht werden, dass in den Prüfungsantworten die für das zu überprüfende Konstrukt relevanten Aspekte identifiziert und bewertet werden. Ein Bewertungsraster beinhaltet typischerweise Bewertungskategorien wie z.B. richtig oder falsch und eine Beschreibung bzw. Charakterisierung dieser Kategorien in Form von Regelwerken sowie Ankerbeispielen. Darüber hinaus muss angegeben werden, welche (Leistungs-)Punkte für jede Kategorie vergeben werden. Die Festlegung von Bewertungskategorien ergibt sich aus den Lernzielen und der Art der gestellten Anforderungen. (vgl. Schindler 2015)



Bewertungsfehler

Die Bewertung einer Leistung fußt in der Regel auf der (subjektiven) Wahrnehmung einer/ eines Lehrenden. Diese Subjektivität kann z.B. durch Einhaltung der Gütekriterien und der Orientierung an einem objektiven Bewertungsraster reduziert werden. Dennoch bleibt das Risiko von Bewertungsfehlern bestehen. Ist man sich dieser möglichen Bewertungsfehler jedoch bewusst, können sie durch Selbstreflexion aktiv vermieden werden. Es gibt Bewertungsfehler, die aufgrund der Persönlichkeit der Beurteilenden entstehen, im Folgenden **persönlichkeitsbedingte Bewertungsfehler** genannt. Es gibt zudem Bewertungsfehler, die durch Wahrnehmungsverzerrung bedingt sind, im Folgenden **wahrnehmungsverzerrte Bewertungsfehler** genannt, zudem gibt es **Tendenzfehler**.

Persönlichkeitsbedingte Bewertungsfehler

Sympathie/ Antipathie

Studierende, die den Prüfenden sympathisch sind, werden milder beurteilt und als kompetenter sowie motivierter eingeschätzt. So besteht das Risiko, dass die Kompetenzen von Studierenden, die als unsympathisch wahrgenommen werden, nicht „korrekt“ beurteilt werden.

Milde-Streng-Fehler

Der Milde-Streng-Fehler beschreibt die starke Über- oder Unterbewertung bestimmter Aspekte. Ein „kleiner Fehler“ wird dabei übermäßig streng geahndet (Streng-Fehler), korrekte Antworten werden übermäßig stark positiv gewichtet (Milde-Fehler).

Erwartungs-/ Bestätigungseffekt

Studierende, die in der Lehrveranstaltung durch ihre Leistung positiv aufgefallen sind, unterliegen der Erwartung, dass ihre Leistung auch in der Prüfung hoch sein muss, werden aber tendenziell auch milder beurteilt. Die Erwartung einer höheren Leistung einer/ eines Studierenden wird außerdem auch durch nonverbale Kommunikation vermittelt und verstärkt. Umgekehrt kann eine negative Erwartungshaltung zu schlechteren Prüfungsergebnissen führen. Faktoren, die die Erwartung von Prüfenden beeinflussen können, sind: Geschlecht und Alter, Auftreten, Aussehen und Ausstrahlung, die rhetorische Gewandtheit, Sprache und Dialekt usw.

(vgl. Macke et. al. 2016; Schott 2018)



Bewertungsfehler

Wahrnehmungsverzerrte Bewertungsfehler

Halo-Effekt

Beim Halo-Effekt überstrahlt ein einzelnes, besonders auffälliges Merkmal oder eine Kompetenz, andere davon völlig unabhängige. Auch das Gegenteil ist möglich, der Gesamteindruck einer Person kann sich auf die Wahrnehmung von einzelnen Merkmalen auswirken.

Reihungseffekt

Die Reihenfolge von Prüfungen bzw. Prüfungsergebnissen kann Einfluss auf deren Beurteilung haben. So kann eine unmittelbar vorangegangene Prüfungsleistung Einfluss auf die Beurteilung der nachfolgenden Prüfungsleistung haben.

Primacy- & Recency-Effekt

- Beim Primacy-Effekt prägen die **zuerst** wahrgenommenen Leistungen die weitere Bewertung der Prüfenden, nachfolgende Leistungen werden dadurch überbewertet.
- Beim Recency-Effekt hingegen überstrahlen oder überschatten die **zuletzt** wahrgenommenen Leistungen vorherige.

(vgl. Macke et. al. 2016; Schott 2018)

Tendenzen

Tendenz zur Mitte

Bei einer Tendenz zur Mitte werden Prüfungsleistungen oft nur mittelmäßig bewertet, obwohl eigentlich ein stärkeres Urteil positiv wie negativ, möglich oder nötig wäre. Vor allem Prüfende, die sich (noch) unsicher in ihrer Leistungsbeurteilung sind, tendieren zur Mitte.

Tendenz zur Extreme

Hier tendieren die Prüfenden stärker zu (ungerechtfertigten) extremen Beurteilungen. Gute Leistungen werden dann als sehr gut oder als die beste(n) Leistung(en) eingestuft und schlechtere Leistungen als besonders schlecht. Beurteilungen zur Mitte werden vermieden.

(vgl. Macke et. al. 2016; Schott 2018)



Tipps zur Vermeidung von Bewertungsfehlern

- Um persönlichkeitsbedingte Bewertungsfehler zu minimieren, empfiehlt es sich, Prüfungen anonym zu korrigieren, sofern möglich.
- Um wahrnehmungsverzerrte Beurteilungsfehler zu minimieren, empfiehlt es sich, Kriterien festzulegen, die dafür sorgen, dass alles beurteilt wird.
- Es sollte reflektiert werden, ob die Leistung des vorhergehenden Studierenden die Bewertung beeinflusst.



Bewertung schriftlicher Prüfungen (Klausuren)

Für die Beurteilung schriftlicher Prüfungen (Klausuren) ist es ratsam, vorab Musterlösungen und/ oder Bewertungskriterien zu erstellen und festzulegen. Eine Musterlösung bildet die optimale Lösung einschließlich aller relevanten Aspekte und Begriffe ab. Musterlösungen sind gemäß der Taxonomie nach Anderson und Krathwohl insbesondere für Wissens-, Verständnis- und Anwendungsfragen geeignet. Es empfiehlt sich außerdem, für jede einzelne Frage spezifische Bewertungskriterien festzulegen. Diese einzelnen Kriterien bilden schließlich ein Bewertungsschema. (vgl. Schott 2018)

Bewertung mündlicher Prüfungen und Präsentationen

Bei mündlichen Prüfungen und Präsentationen steht der Inhalt des Gesagten im Vordergrund, aber dennoch spielt neben den inhaltlichen Kriterien auch die Performance (bewusst oder unbewusst) eine Rolle bei der Bewertung. „Performance“ bedeutet hier, wie sich die/ der Studierende während der Prüfung präsentiert, einschließlich ihrer/ seiner Körpersprache, Sprechgeschwindigkeit usw. (vgl. Limburg 2022)

Folgende **Kriterien** sollten nach Limburg (2022) bei mündlichen Prüfungen und Präsentationen beachtet werden:

Inhaltliche Darstellung

- Nennung zentraler Aspekte eines Themas
- Nachvollziehbare und schlüssige Argumentation
- Logisch strukturierte Darstellung

Sprachliche Darstellung

- Verwendung von Fachsprache
- Vermeidung von Umgangssprache
- Definition von zentralen Begriffen (auf Nachfrage)



Bewertung von Ausarbeitungen

In Ausarbeitungen werden zusätzlich zu den inhaltlichen, also den lernzielbezogenen Kriterien, auch die gestalterischen Kriterien herangezogen. Die gestalterischen Kriterien beinhalten die Einhaltung von Konventionen für das jeweilige Format (z.B. Unterschiede in der Gestaltung einer Seminararbeit und einer Projektarbeit) sowie die weiteren Kriterien wie „sprachliche Eindeutigkeit“, „Grammatik“, „Rechtschreibung“ und „Formatierung“. (vgl. Limburg 2022)

Nach Limburg (2022) sollten folgende **Kriterien** bei der Bewertung der Ausarbeitung beachtet werden:

Bearbeitung des Themas

- Eingrenzung des Themas, sodass es in dem vorgegebenen Rahmen detailliert bearbeitet werden kann.
- Behandlung zentraler Aspekte des Themas.
- Es werden nur Aspekte dargestellt, die zur Beantwortung der Fragestellung dienen.
- Nachvollziehbare und schlüssige Argumentation.

Umgang mit Quellen

- Verwendung notwendiger sowie einschlägiger Quellen zur Beantwortung der Fragestellung.
- Belegung aller Aussagen (außer allgemeingültigen).
- Begründung von Positionierung/Kritik.
- Unverfälschte Wiedergabe von Quellen
- Quellen wurden eindeutig zitiert.

Struktur der Ausarbeitung

- Es liegt eine klar erkennbare Fragestellung vor.
- Die Einleitung beinhaltet alle wichtigen Aspekte.
- Das Fazit enthält alle wichtigen Aspekte.
- Alle Teile der Arbeit sind vorhanden.
- Die Kapitel der Arbeit tragen zur Beantwortung der Fragestellung bei.
- Sinnvolle Gliederung der Absätze.

Sprachliche Gestaltung

- Verwendung von Fachsprache.
- Definition zentraler Begriffe.
- Eindeutige Formulierungen.
- Eindeutige Bezüge.

Korrektheit

- Einhaltung von Rechtschreibung und Grammatik.
- Einheitliche Formatierung, Quellenangaben, Zitierweise und Literaturangaben.



Literatur:

Hagen, Michael; Szczyrba, Birgit; Wunderlich, Antonia (2015): Prüfungen auswerten mit dem Niveaustufenmodell. Abrufbar unter: https://www.th-koeln.de/mam/downloads/deutsch/hochschule/profil/lehre/steckbrief_niveaustufen.pdf; Stand: 21.02.2022

Limburg, Anika (2015): Leistungsbeurteilung. Abrufbar unter: <https://dbs-lin.ruhr-uni-bochum.de/lehre-laden/planung-durchfuehrung-kompetenzorientierter-lehre/leistungsbeurteilung/begriffliche-differenzierung/?pdf=1054>; Stand: 27.02.2023

Macke, Gerd; Hanke, Ulrike; Viehmann-Schweizer, Pauline; Raether, Wulf (2016): Kompetenzorientierte Hochschuldidaktik: lehren – vortragen – prüfen – beraten. Weinheim / Basel: Beltz.

Rheinberg, Falko. (2008): Bezugsnormen und die Beurteilung von Lernleistungen. In: Scheider, Wolfgang & Hasselhorn, Marcus (Hrsg.): Handbuch der Pädagogischen Psychologie, Bd. 10, 1. Auflage, S. 178 – 186. Göttingen: Hogrefe.

Schindler, Christoph Josef (2015): Herausforderung Prüfen: Eine fallbasierte Untersuchung der Prüfungspraxis von Hochschullehrenden im Rahmen eines Qualitätsentwicklungsprogramm. München. Abrufbar unter: <https://d-nb.info/1079001778/34>; Stand: 27.02.2023

Schott, Reinhard (2018): Beurteilen von Prüfungen. Abrufbar unter: <https://infopool.univie.ac.at/startseite/pruefen-beurteilen/beurteilen-von-pruefungen/>; Stand: 27.02.2023

TU Darmstadt: Festlegen von Bewertungskriterien und Erstellung eines Bewertungsrasters. Abrufbar unter: https://www.einfachlehren.tu-darmstadt.de/themensammlung/details_9536.de.jsp; Stand: 21.02.2022

Universität Zürich: An welcher Norm orientiere ich mich bei der Benotung meiner Prüfungen? Abrufbar unter: <https://teachingtools.uzh.ch/de/tools/benoten-von-pruefungen>; Stand: 21.02.2023.