

Leistungsbeurteilung

Thesen

- Leistungsbeurteilung meint die Überprüfung eines Lernzuwachses im Bereich Wissen und/oder Fähigkeiten.
 - Lernende, aber auch Lehrende erhalten eine Rückmeldung über den Lernstand.
 - Bei Leistungsbeurteilung zählen Objektivität, Validität und Reliabilität.
 - *Bewertung* und *Leistungsbeurteilung* meinen unterschiedliche Aspekte.
 - Eine bewusste Auseinandersetzung mit Leistungsbeurteilung ist sehr hilfreich.
-

Begriffliche Differenzierung

Unter Leistungsbeurteilung wird die Überprüfung eines Lernzuwachses im Bereich Wissen und/oder Fähigkeiten verstanden. Dieser Lernzuwachs wird häufig nicht individuell, sondern in Bezug auf eine vergleichbare Leistung anderer oder gemessen an einem Ideal beurteilt und mit einer Note beziffert. Aus didaktischer Perspektive gerät die Leistungsbeurteilung idealerweise erst dann in den Blick, wenn **Prüfungsformat** und -gestaltung, **Lernziele** und didaktische Planung einer Veranstaltung sinnvoll aufeinander ausgerichtet sind (s. **constructive alignment**). Insbesondere die lernzielbezogene Planung der Veranstaltung und respektive der Prüfung erleichtert die Beurteilung, weil dadurch implizit bereits Kriterien existieren, auf die die Lehre ausgerichtet wurde.

Ein Ziel der Leistungsbeurteilung ist es, sowohl Lernenden als auch Lehrenden eine Rückmeldung über den Lernstand zu geben. Lernende erhalten einen Eindruck, wie Lehrende ihr Wissen und ihre Fähigkeiten einschätzen. Lehrende hingegen erhalten eine Rückmeldung darüber, in welchem Ausmaß es ihnen gelungen ist, Wissen und Fähigkeiten zu vermitteln. Deshalb sollte eine Leistungsbeurteilung möglichst intersubjektiv nachvollziehbar sein – oft wird von objektiv, valide und reliabel gesprochen.

Bei der Beschäftigung mit dem Thema Leistungsbeurteilung ist eine begriffliche Differenzierung hilfreich: Unter *Bewertung* verstehen wir im Folgenden einen spontanen, meist unbewussten Vorgang, bei dem Menschen mit einem Werturteil auf etwas reagieren, hier auf eine Prüfungsleistung oder einen Aspekt von ihr (beispielsweise einen Tippfehler auf dem Deckblatt einer Seminararbeit). Diese Bewertung geschieht auch auf der Grundlage von mehr oder weniger unbewussten Bewertungsmaßstäben und -vorgängen, so dass sie zunächst einmal nicht völlig beeinflussbar ist. Die *Leistungsbeurteilung* erfolgt hingegen in einem reflektierten Prozess, bei dem eine Leistung an Kriterien gemessen wird. Daher ist es für diesen Prozess nützlich, zuvor Kriterien aufzustellen und sich die eigenen impliziten Bewertungsmaßstäbe bewusst zu machen.

Im Einzelnen stellen sich im Zuge der Leistungsbeurteilung die Fragen, welche Kriterien an eine Prüfungsleistung angelegt werden können, wie eine Leistung an diesen Kriterien gemessen werden kann und wie Lehrende – ganz konkret – an das Produkt der Prüfung herangehen können (Rotstift?, Korrekturzeichen?, etc.). Der Beurteilungsprozess und die Beurteilung selbst sollten für die Prüflinge möglichst nachvollziehbar und begründet dargestellt werden, so dass sie – im Idealfall – auch daraus lernen können. Antworten auf diese Fragen finden Sie in den folgenden Abschnitten. Zuvor möchten wir jedoch darlegen, inwiefern die Leistungsbeurteilung ein wichtiges Thema der Hochschuldidaktik darstellt und eine bewusste Auseinandersetzung mit ihr sehr hilfreich ist. An dieser Stelle noch ein kurzer Hinweis zum Nachteilsausgleich: Informationen dazu finden Sie [hier im Lehre Laden](#).

Warum es schwierig ist, Prüfungsleistungen zu benoten

Lehrende sind bei einer Leistungsbeurteilung in Notenform – so zeigen uns etliche Studien – mit mindestens drei Schwierigkeiten konfrontiert:

1. erstens sind Notenerurteile
 - nur bedingt *valide*, weil in die Beurteilung auch Faktoren einfließen, die nicht zur Prüfungsleistung zählen,
 - nicht *objektiv*, weil sie bei unterschiedlichen Lehrenden unterschiedlich ausfallen und
 - nicht *reliabel*, weil sie bei Wiederholungen nicht zum selben Ergebnis führen.

2. zweitens erfordert es Aufwand und im Hochschulkontext wenig verbreitete Methoden, die Leistungsbeurteilung so zu vollziehen, dass Studierende durch die Rückmeldung auf ihre Prüfungsleistungen auch lernen können und
3. drittens ist der Betreuungsaufwand gerade in Massenfächern groß, was eine solche Orientierung auf das Lernen erschwert.

Zu 1) Studien aus dem schulischen und dem universitären Bereich belegen, dass Noten nicht in dem gewünschten Maß valide, reliabel und objektiv sind. Das ist zum Beispiel darauf zurückzuführen, dass zur Beurteilung nur eine subjektive Perspektive eingenommen werden kann, die von unterschiedlichen Anforderungen an und Vorstellungen von einer perfekten Prüfung geprägt sind. Auch fließen Aspekte der Bewertung (s.o.) in die Beurteilung: sei es das äußere Erscheinungsbild eines Prüflings bei einer mündlichen Prüfung oder der Tippfehler auf dem Deckblatt einer Hausarbeit (Jachmann 2003, S. 46-52). Nicht zuletzt beeinflusst die Bezugsnorm die Notenfindung: Rein der *sachorientierten Bezugsnorm* folgend würde eine Prüfungsleistung ausschließlich an einem vordefinierten Leistungsstand gemessen, dem meist eine Idealvorstellung zugrunde liegt. Bei der *sozialen Bezugsnorm* erhielten für eine konkrete Prüfung diejenigen mit der besten Leistung die beste Note, diejenigen mit der schlechtesten Leistung die schlechteste Note. Bei einer *individuellen Bezugsnorm* hingegen wird ausschließlich der Lernzuwachs der Einzelperson beurteilt. Diese Bezugsnorm fördert zwar am ehesten die Lernbereitschaft und die Lernfähigkeit (Meyer 1984; Bohl 2004), ist aber im Universitätsbetrieb meist nicht umsetzbar und wird von vielen als ungerecht empfunden. In der Praxis wird daher meist eine Mischform zwischen sozialer und sachlicher Bezugsnorm gewählt, bei der nur im Ausnahmefall individuelle Aspekte miteinbezogen werden.

Angesichts der hier beschriebenen Ausgangssituation können Lehrende also gar nicht erreichen, dass ihre Notengebung im statistischen Sinne Kriterien wie Validität, Reliabilität und Objektivität genügt, die häufig mit einer gerechten Beurteilung assoziiert werden. Für eine möglichst gerechte Benotung ist es deshalb notwendig, anderen Beurteilungsprinzipien zu folgen; wir empfehlen Ihnen, einem Prinzip der ‚transparenten Subjektivität‘ zu folgen, bei dem Sie Ihre Beurteilungskriterien bereits vor der Prüfung so offenlegen, dass sie für Studierende nachvollziehbar sind. (Siehe dazu weiter unten.)

Zu 2) Prüfungen haben u.a. die Funktion, den aktuellen Stand der fachspezifischen Denk-, Handlungs- und Schreibweisen zu ermitteln. Dabei geht es sowohl um Wissensbestände, also im engeren Sinne Fachwissen und Kenntnis fachspezifischer Darstellungs- und Argumentationskonventionen, als auch um Fähigkeiten wie beispielsweise Tiefenlern-, Lese- oder Schreibstrategien, die für ein erfolgreiches Studium unabdingbar sind. Solche Wissensbestände und Fähigkeiten können Studierende nur durch das Studium und die damit verbundenen Prüfungen *erlernen* (s. z.B. Pohl 2007), müssen jedoch in diesen Prüfungen

gleichzeitig *beweisen, dass sie darüber verfügen.*

Zwar wird im Bereich des Fachwissens meist – dem didaktischen Prinzip des ***constructive alignment*** entsprechend – nur das geprüft, was zuvor vermittelt wurde. Im Bereich der fachspezifischen Lese-, Schreib- und Lernstrategien sowie der Prinzipien wissenschaftlich angemessener Darstellung hingegen besteht das Problem, dass Studierende sie sich oftmals ohne Instruktion aneignen müssen. Vor diesem Hintergrund ist nachvollziehbar, dass viele Studierende in den ersten Semestern insbesondere mit schriftlichen Prüfungen überfordert sind, die an deutschen Hochschulen immerhin ca. 80% aller Prüfungen ausmachen (vgl. Kerres/Schmidt 2011).

In der Lehr- und Prüfungspraxis bleibt der Aspekt des ‚aus der und durch die Prüfung Lernens‘ oft sowohl auf Seiten der Studierenden als auch auf der der Lehrenden unbeachtet. Nur wenn Lehrende die Prüfung auch selbst als Lernanlass begreifen und kommunizieren, kann sich dieser Gedanke auch bei Studierenden besser durchsetzen. Solange dies nicht geschieht, verlaufen Lernprozesse langsamer und schlechter, als es möglich wäre. Dass dies zum Nachteil für die Studierenden ist, liegt auf der Hand. Doch auch für Lehrende bringt dies Aufwand und Ärger mit sich, vor allem über schlechte Texte und hohen Betreuungsaufwand.

Zu 3) Vor dem Hintergrund der beschriebenen Problematik, dass Noten nicht im statistischen Sinne objektiv sind, Lehrende ihre Beurteilungskriterien nur selten vor einer Prüfung für alle Studierenden verständlich offenlegen, und Studierende die für Prüfungen erforderlichen Wissensbestände und Strategien erst sukzessiv erlernen – ist es nachvollziehbar, dass Studierende viele Fragen haben. Für viele Lehrende ist der Betreuungsaufwand im Rahmen von Prüfungsvorbereitungen (gerade bei Examensprüfungen) daher oft sehr groß und insbesondere in Massenfächern nicht oder nicht gut leistbar. Daraus ergibt sich eine für beide Seiten frustrierende Situation: Lehrende sind während der Sprechstunde immer häufiger mit langen Schlangen vor ihrer Tür konfrontiert und Studierende trauen sich u.a. deswegen oft nicht, ihre legitimen Fragen zu stellen (vgl. Boettcher/Meer 2000). Als Konsequenz daraus entspricht auch das Prüfungsergebnis oft nicht dem, was sich beide Seiten vorstellen.

Um trotz hoher Lehrbelastung Prüflinge gut zu betreuen, können Sie gezielt unterschiedliche Informationsmedien einsetzen. Seien es schriftlich ausformulierte Leitfäden und Handouts, mit denen Sie Konventionen wissenschaftlicher Texte oder Ihre Anforderungen vermitteln, oder der Verweis auf Ratgeberliteratur, die Sie für gut befinden. Egal welche Materialien Sie nutzen: Wir möchten Ihnen empfehlen, nicht nur die formalen Aspekte (wie Zitation und Formatierung) damit abzudecken, sondern beispielsweise auch auf den Zweck der jeweiligen Prüfungsleistung und dafür sinnvolle Lese-, Schreib- oder Lernstrategien einzugehen. Um Zeit zu sparen ist es auch sinnvoll, Materialien um Antworten auf Fragen zu ergänzen, die Studierende in Ihren Sprechstunden oft stellen. Das **Schreibzentrum** berät und unterstützt Sie, wenn Sie Materialien

entwickeln oder ausarbeiten möchten.

Beurteilungskriterien festlegen

Unterschiedliche Studien legen nahe, dass es sinnvoll ist, Prüfungsleistungen anhand einer überschaubaren Zahl von Kriterien zu beurteilen. So konnten Hartog und Rhodes bereits 1936 nachweisen, dass eine Beurteilung anhand von 700 Einzelkriterien genauso wenig intersubjektiv vergleichbare Ergebnisse liefert (Hartog/Rhodes 1935) wie eine anhand eines reduzierten Kriterienkatalogs von drei Einzelkriterien (Grzesik/Fischer 1985, S. 241). Arbeiten Sie also mit ca. zehn bis 15 Kriterien. Bei Prüfungsformen, bei deren Bearbeitung Studierende einen größeren Freiraum haben (z. B. Klausuren in Textform, Hausarbeiten, Essays) können Sie Ihren Beurteilungsspielraum erweitern, indem Sie eine Rubrik für „Unvorhergesehenes“ aufnehmen.

Kriterien ergeben sich zum einen aus den fachlichen Inhalten, deren Beherrschung geprüft wird (lernzielbezogene Kriterien), zum anderen aus der Beherrschung der jeweilig angemessenen Darstellungsform:

Lernzielbezogene Kriterien

Viele der Kriterien für die Beurteilung einer Prüfungsleistung ergeben sich unter Berücksichtigung der Lernziele der entsprechenden Veranstaltung bzw. des Moduls oder – bei Examensarbeiten – des jeweiligen Studiengangs. Kriterien zu formulieren fällt umso leichter, je präziser Lernziele formuliert wurden.

Beispiel:

Mit einem Seminar wurde das Lernziel verfolgt, Probleme betrieblicher Praxis vor dem Hintergrund theoretischer Lerninhalte zu analysieren und auf dieser Grundlage Lösungsansätze zu entwickeln. Geprüft wurden diese Lernziele durch eine Textklausur, in deren Rahmen ein Problem der betrieblichen Praxis mithilfe einer selbst gewählten Theorie analysiert und auf dieser Grundlage eine Handlungsempfehlung formuliert werden sollte. Sinnvolle Beurteilungskriterien wären demnach:

- Wurde aus den im Seminar vermittelten Theorien eine Theorie ausgewählt, mit der das ausgewählte Problem sinnvoll analysiert werden kann?
- Wurde die Theorie korrekt dargestellt?
- Wurde die Theorie sinnvoll auf das Problem angewendet?
- Ist der Bezug zwischen den Ergebnissen der Analyse und der Handlungsempfehlung nachvollziehbar?
- Ist die Handlungsempfehlung aus praktischer und theoretischer Sicht sinnvoll, d.h. beispielsweise zielführend und praktikabel?

Weil lernzielbezogene Kriterien auf inhaltliche Aspekte abzielen, gelten sie immer nur für eine konkrete Prüfung. Sie sollten Sie daher immer dann neu formulieren, wenn Sie mit einer Prüfung andere Lernziele verfolgen.

Tipp: Nutzen Sie die **Lernzieltaxonomie** von Bloom (1976), um sich die verschiedenen Ebenen potentieller Lernziele zu vergegenwärtigen und um die Lernziele auszuformulieren.

Allgemeine Kriterien bei schriftlichen Arbeiten

Bei schriftlichen Arbeiten kommen zu den lernzielbezogenen Beurteilungskriterien textsortenbezogene Kriterien hinzu. Sie betreffen zum einen die Einhaltung der Konventionen, die für die jeweilige Textsorte gelten (z.B.: Ist bei einer Hausarbeit an jeder Stelle der Ursprung einer Information erkennbar?). Zum anderen geht es um typische textbezogene Kriterien wie die der sprachlichen Eindeutigkeit, der korrekten Grammatik, Rechtschreibung und Zeichensetzung oder auch der angemessenen Formatierung. Insbesondere die Einhaltung solcher Kriterien wird während des Studiums nicht vermittelt und Studierende benötigen daher hier vielfältige Lernanlässe.

Übersicht allgemeiner Kriterien bei schriftlichen Arbeiten:

- Zur Bearbeitung des Themas
 - Das Thema wurde so eingegrenzt, dass es in der gegebenen Seitenzahl detailliert bearbeitet werden kann.
 - Die zentralen Aspekte des Themas wurden behandelt.
 - Es wurden keine Aspekte dargestellt, die nicht zur Beantwortung der Fragestellung beitragen.

- Die Argumentation ist nachvollziehbar.
- Die Argumentation ist schlüssig.
- Zur Struktur der Arbeit
 - Der Arbeit liegt eine klar erkennbare Fragestellung zugrunde.
 - In der Einleitung wurden alle relevanten Aspekte bearbeitet (Relevanz des Themas, Ziel der Arbeit/Fragestellung, Themenein- und -abgrenzung, ggf. besondere Untersuchungsmethodik, Material, Aufbau der Arbeit samt Begründung).
 - Im Fazit wurden alle relevanten Aspekte bearbeitet (Beantwortung der Fragestellung, Zusammenfassung der zentralen Ergebnisse, Ausblick).
 - Alle Teile der Arbeit sind vorhanden (Deckblatt, Inhaltsverzeichnis, Textteil, Literaturverzeichnis, ggf. Anhang).
 - Die einzelnen Kapitel leisten einen sichtbaren Beitrag zur Beantwortung der Fragestellung/zur Bearbeitung des Themas.
 - In den Kapiteln sind Absätze sinnvoll gebildet (d.h. > als ein Satz, < als ein Gedankengang, = ein Hauptgedanke).
- Zur sprachlichen Gestaltung der Arbeit
 - Es wurde wo nötig Fachsprache benutzt.
 - Es wurde keine Umgangssprache benutzt.
 - Die zentralen Begriffe sind definiert.
 - Der Text ist präzise und eindeutig formuliert.
 - Bezüge sind an jeder Stelle eindeutig.
- Zum Umgang mit Quellen
 - Es wurden die notwendigen und hinreichenden Quellen verwendet, um die Frage zu beantworten.
 - Es wurden einschlägige Quellen verwendet.
 - Alle Aussagen (außer allgemeingültige) sind belegt.
 - Jede Positionierung/Kritik ist begründet.
 - Die Quellen wurden unverfälscht wiedergegeben.

- Die Quellen/Positionen wurden zueinander in Bezug gesetzt.
- An jeder Stelle ist deutlich, ob eine Quelle referiert wurde oder eigene Gedanken dargestellt wurden.
- Korrektheit
 - Der Text entspricht den Regeln der Rechtschreibung.
 - Die Zeichensetzung folgt den Regeln.
 - Die Grammatik entspricht den Regeln.
 - Die Formatierung, die Quellenangaben, die Zitierweise und die Literaturangaben sind einheitlich.

Allgemeine Kriterien bei mündlichen Prüfungen/Präsentationen und Praxisprüfungen

Ob bei mündlichen Prüfungen, Präsentationen und Praxisprüfungen zu den lernzielbezogenen Kriterien weitere hinzukommen, handhaben Lehrende sehr unterschiedlich. Dabei ist zu berücksichtigen, dass Gesprochenes flüchtig ist und Inhalte nicht gut von der Art ihrer Darbietung getrennt wahrgenommen werden können, so dass zu den inhaltlichen Kriterien immer auch – ob bewusst oder unbewusst – Kriterien der Performanz hinzutreten. Dazu gehört beispielsweise, ob Körpersprache, Sprechtempo, Wortwahl etc. den Prüfling souverän und kompetent erscheinen lassen, und ob seine Art der Darbietung suggeriert, er könne aus dem Vollen schöpfen. Deshalb ist es sinnvoll, Studierende darüber zu informieren, dass in diesem Punkt mündliche Prüfungen und Präsentationen durchaus mit Bewerbungsgesprächen vergleichbar sind: Es gilt zu zeigen, was man kann, und nicht, was man nicht kann.

Darüber hinaus können Sie folgende Kriterien an mündliche Prüfungen und Präsentationen anlegen:

- Zur inhaltlichen Darstellung
 - Die zentralen Aspekte des Themas werden benannt, unwichtige Aspekte werden nicht benannt.
 - Die Argumentation ist nachvollziehbar.
 - Die Darstellung ist logisch strukturiert.
 - Die Argumentation ist schlüssig.

- Zur sprachlichen Darstellung
 - Die einschlägige Fachsprache wird verwendet.
 - Es wird keine Umgangssprache verwendet.
 - Zentrale Begriffe können (auf Nachfrage) definiert werden.
- Zum Umgang mit Forschungsliteratur
 - Es wird auf sinnvolle Forschungsliteratur Bezug genommen.
 - Der Ursprung zentraler Gedanken kann (auf Nachfrage) benannt werden.
 - Es wird eine Position zur Forschungsliteratur eingenommen.
 - Die Informationen/Gedanken aus der Forschungsliteratur werden unverfälscht wiedergegeben.
 - Positionen aus der Forschungsliteratur werden zueinander in Bezug gesetzt.

Kriterien hierarchisieren

Auch wenn Beurteilungskriterien feststehen, ergibt sich die Note bei der Beurteilung nicht von selbst: Kriterien können fachbezogen, hinsichtlich der Lernziele und auch hinsichtlich einzelner Lernender unterschiedlich relevant sein und auch ihr Zusammenwirken kann unterschiedlich beurteilt werden kann. So stellt sich die Frage, wie eine inhaltlich schwache aber eloquent geschriebene, formal korrekte Arbeit im Gegensatz zu einer inhaltlich starken, aber schlecht geschriebenen und formal fehlerhaften Arbeit beurteilt werden sollte. Dies handhaben nicht nur einzelne Lehrende unterschiedlich, sondern es hängt auch davon ab, wie wichtig im entsprechenden Fach und für die konkrete Prüfung die lernzielunabhängigen Beurteilungskriterien sind.

Um Kriterien zu hierarchisieren, können Sie die Aspekte detaillierter ausdifferenzieren, die Ihnen wichtiger sind, und die Ihnen unwichtigeren Aspekte grob zusammenfassen. So könnte es in einer Bachelorarbeit im Maschinenbau sinnvoll sein, die lernzielbezogenen Kriterien detaillierter ausdifferenzieren, sprachliche und formale Kriterien aber zu einem einzigen Beurteilungskriterium zusammenzufassen. Denn zum einen konnten die Studierenden zuvor kaum Schreiberfahrungen sammeln, zum anderen gehört das (mit lernzielbezogenen Kriterien erfasste) fachspezifische Handlungswissen zur Kernkompetenz von Maschinenbauern; gute Texte zu schreiben ist eine weniger starke Anforderung an sie. Für eine philologische Arbeit hingegen wäre möglicherweise eine Ausdifferenzierung der text(sorten)bezogenen Kriterien

sinnvoll, weil deren Berücksichtigung eine Kernkompetenz in philologischen Fächern ist.

Bei einer Hierarchisierung von Kriterien können Sie auch den Lernstand der Studierenden einbeziehen. So zeigte sich, dass sich Studienanfänger vielfach schwer tun, diskursive und argumentative Aufgaben angemessen zu bewältigen. Deshalb empfehlen wir Ihnen, solche Kriterien für sie schwächer zu gewichten. Ähnlich können Sie mit der Gewichtung auch auf die individuelle Situation einzelner Studierender reagieren, indem Sie z. B. zur Bewertung schriftlicher Arbeiten von Studierenden, die in einer Fremdsprache schreiben, Fehler im Bereich sprachliche Korrektheit weniger stark gewichten, als bei Studierenden, die in ihrer Erstsprache schreiben. Zudem ist es wichtig, auch jenseits der festgelegten Kriterien offen zu sein für überraschende gute Aspekte einer Leistung, d.h. die Festlegung von Kriterien soll ein Hilfsmittel zur Leistungsbeurteilung sein, das Sie flexibel einsetzen können.

Prüfungsleistungen an Kriterien messen

Anders als beispielsweise an Schulen bestehen an Universitäten große Spielräume hinsichtlich der Möglichkeiten, zu einem Notenurteil zu gelangen. Neben dem verbreiteten eher intuitiven Vorgehen, bei dem die Prüfungsleistung aus sich heraus beurteilt wird, ohne explizit zuvor formulierte Kriterien anzulegen, werden auch im universitären Kontext oft Kriterienkataloge zur Leistungsbeurteilung angesetzt. Sie bieten den Vorteil, dass die Beurteilung in einem stärker standardisierten Prozess erfolgen kann, und dass sichergestellt wird, dass jede Prüfungsleistung an denselben Kriterien gemessen wird. Kriterienkataloge zu nutzen führt daher sowohl bei Lehrenden als auch bei Studierenden eher dazu, dass eine Beurteilung als gerecht empfunden wird.

Kriterienkataloge können ganz unterschiedlich gestaltet sein, wobei meist eine Form der Skalierung verwendet wird. Verbreitet ist, Kriterien in sechs Abstufungen zu skalieren, so dass die Einschätzung mit der Nutzung der herkömmlichen Ziffernnoten assoziiert wird. Hiermit wird suggeriert, dass die endgültige Note im arithmetischen Mittel der Einzelnoten liegt. Da damit aber kaum Spielräume zur individuellen Gewichtung bestehen, empfehlen wir Ihnen, eine geringere Menge an Abstufungen zu verwenden. Eine ungerade Zahl – sinnvollerweise also drei oder fünf Kriterien – begünstigen, dass die Mitte gewählt wird, wenn eine Leistung in Bezug auf ein Kriterium nicht eindeutig gut oder schlecht umgesetzt wurde. Um Leistungen schnell zu beurteilen, ist dies sinnvoll, es ist aber nicht unbedingt im Sinne einer differenzierten Beurteilung. Wir empfehlen Ihnen, ruhig mehrere Skalierungsweisen zu erproben, bevor Sie sich

endgültig für eine entscheiden.

Wenn Sie bei der Beurteilung einer Leistung anhand der Skalierung oftmals unsicher sind, wo genau Sie sie einordnen möchten, ist es hilfreich, ein sogenanntes **Rubric** für diese konkrete Prüfung zu erstellen. Dabei wird für jedes einzelne Kriterium und jede Stufe der Skalierung ausformuliert, welche Kennzeichen eine Leistung hat, die genau dieser Skalierungsstufe zugeordnet wird.

Einen solchen Kriterienkatalog zu nutzen geht in der Regel mit der Entscheidung einher, die Beurteilung zumindest schwerpunktmäßig an der sachlichen Bezugsnorm auszurichten. Die soziale Bezugsnorm dient dann als Form des Korrektivs, so wenn die Noten aller Prüflinge um eine Notenstufe angehoben werden, falls der Notendurchschnitt zu niedrig war. Hier kann aus den verhältnismäßig schlechten Ergebnissen der Schluss gezogen werden, dass die Lernziele der Veranstaltungen nicht ideal erreicht werden konnten. Auch dient die soziale Bezugsnorm als Entscheidungshilfe bei der Beurteilung qualitativ ähnlicher Prüfungsleistungen. Die individuelle Bezugsnorm nutzen viele Lehrende bei der Beurteilung von schriftlichen Arbeiten, die von internationalen Studierenden geschrieben wurden – dass deren Arbeit auf sprachlicher Ebene nicht der einer Muttersprachlerin entspricht, kann und sollte nicht stark ins Gewicht fallen (es sei denn vielleicht bei einem Germanistikstudium). Umgekehrt könnte die Anwendung einer individuellen Bezugsnorm auch bedeuten, dass an das russische Essay von Studierenden in der Slavistik, deren Muttersprache Russisch ist, strengere Maßstäbe angelegt werden als an die Arbeiten von deutschsprachigen Studierenden.

Um sich nicht dem eigenen und fremden Vorwurf auszusetzen, ungerecht zu benoten, können Sie sich bewusst für die Anwendung der Bezugsnorm entscheiden und Ihre Entscheidung, wenn nötig, offen legen. Um diese Entscheidung treffen zu können, kann es hilfreich sein, zu überlegen, welche Funktion die Benotung in der konkreten Situation primär erfüllen soll: z. B. einen Leistungsvergleich in der Gruppe herzustellen, eine lernorientierte Rückmeldung zu geben oder einen bestimmten Leistungsstand zu bescheinigen, der für das weitere Studium oder den Beruf vorausgesetzt wird.

Kriterienraster erstellen

Sie haben allgemeine Kriterien für z.B. die schriftliche Prüfung Ihrer Studierenden erstellt, die die Bearbeitung des Themas, die Struktur der Arbeit, die sprachliche Gestaltung, den Umgang

mit Quellen und die Korrektheit umfassen, und weitere **Lernziel**-bezogene Kriterien definiert:

- Wurde aus den im Seminar vermittelten Theorien eine Theorie ausgewählt, mit der das ausgewählte Problem sinnvoll analysiert werden kann?
- Wurde die Theorie korrekt dargestellt?
- Wurde die Theorie sinnvoll auf das Problem angewendet?
- Ist der Bezug zwischen den Ergebnissen der Analyse und der Handlungsempfehlung nachvollziehbar?
- Ist die Handlungsempfehlung aus praktischer und theoretischer Sicht sinnvoll, d.h. beispielsweise zielführend und praktikabel?

In einem weiteren Schritt haben Sie die Kriterien in eine Hierarchie gebracht, um die Aspekte detaillierter auszdifferenzieren, die Ihnen wichtiger sind, und die Ihnen unwichtigeren Aspekte grob zusammenfassen. Wie können Sie die studentischen Ausarbeitungen möglichst objektiv beurteilen?

Analytisches Raster

Dafür bietet sich, wenn Sie sich an der sachlichen (kriterialen) Bezugsnorm orientieren, ein analytisches Kriterienraster (Kreuzer 2018, S. 3f.) an, das in der englischsprachigen Literatur oft als „Rubric“ bezeichnet wird. In einem analytischen Kriterienraster notieren Sie für jedes Einzelkriterium in Matrix-Form drei oder vier Erwartungsniveaus. Die Einteilung funktioniert nach dem verbreiteten Prinzip von „+, o, -“ oder „++, +, o, -“. Für jedes Kriterium erstellen Sie einen Deskriptor. Damit beschreiben Sie, was genau Sie auf das jeweilige Kriterium bezogen mit der Beurteilung von z.B. „+“ meinen. Diese Beschreibung sollte leicht verständlich und eindeutig formuliert sein, so dass alle Anwender*innen des Kriterienrasters auf dieselbe Weise das Geschriebene interpretieren. Dafür ist es u.a. wichtig, auf Formulierungen wie „gut“, „angemessen“ oder „sinnvoll“ zu verzichten.

Beispiel

Sie haben als Learning Outcome Ihrer Lehrveranstaltung definiert: „Die Studierenden können Probleme betrieblicher Praxis vor dem Hintergrund theoretischer Lerninhalte analysieren und auf dieser Grundlage Lösungsansätze entwickeln und Handlungsempfehlungen formulieren.“

und eine Klausur mit offenen Fragen als Prüfungsform und Aufgabentyp gewählt. Fünf beispielhafte Kriterien mit je 3 Erwartungsniveaus und zugehörigen Deskriptoren finden sich in Ihrem Kriterienraster:

| | Erwartungsniveau + | Erwartungsniveau 0 | Erwartungsniveau - |
|--|--|---|--|
| Kriterium 1: Theorieauswahl | Es wurde eine Theorie ausgewählt, mit der das Problem zielführend analysiert werden kann. | Es wurde eine Theorie ausgewählt, mit der das Problem analysiert werden kann. | Es wurde keine Theorie ausgewählt, mit der das Problem analysiert werden kann. |
| Kriterium 2: korrekte Darstellung der Theorie | Die Theorie wurde in allen Bestandteilen korrekt dargestellt. | Die Theorie wurde in einigen Bestandteilen korrekt dargestellt. | Die Theorie wurde nicht korrekt dargestellt. |
| Kriterium 3: Anwendung der Theorie | Die Theorie wurde in allen Aspekten sachlogisch auf das Problem angewendet. | Die Theorie wurde größtenteils sachlogisch auf das Problem angewendet. | Die Theorie wurde nicht sachlogisch auf das Problem angewendet. |
| Kriterium 4: Bezug zwischen Analyse und Handlungsempfehlungen | Der Bezug zwischen den Ergebnissen der Analyse und der Handlungsempfehlung ist in allen erwähnten Punkten schlüssig und nachvollziehbar. | Der Bezug zwischen den Ergebnissen der Analyse und der Handlungsempfehlung ist in einigen Punkten schlüssig und/oder nachvollziehbar. | Der Bezug zwischen den Ergebnissen der Analyse und der Handlungsempfehlung ist weder schlüssig noch nachvollziehbar. |
| Kriterium 5: Logik der Handlungsempfehlung | Die Handlungsempfehlung ist aus praktischer und theoretischer Sicht zielführend und praktikabel. | Die Handlungsempfehlung ist aus praktischer und/oder theoretischer Sicht zielführend oder praktikabel. | Die Handlungsempfehlung ist aus praktischer oder theoretischer Sicht weder zielführend noch praktikabel. |

Tabelle: Fünf Kriterien mit je drei Erwartungsniveaus. Quelle: eigene Darstellung

Dieser Matrix, die ein beispielhafter Auszug ist, können Sie nun Ihre Hierarchie der Kriterien in Form einer Gewichtung hinzufügen. Sie können dabei Punkte vergeben oder die Kriterien mit Prozentanteilen versehen. Wenn Ihnen z.B. die Theorieauswahl und die korrekte Darstellung der Theorie weniger wichtig sind als ihr Lehr-Schwerpunkt, die Anwendung der Theorie, und weniger

wichtig als die beiden Kriterien zu den Handlungsempfehlungen, dann könnten Sie z.B. für die ersten beiden Kriterien je 15 Prozent der Gesamtnote vergeben, für Kriterium drei 30 Prozent und für die beiden letzten Kriterien je 20 Prozent.

Ein solches Raster zu erstellen ist, insbesondere beim ersten Mal, aufwendig. Und gleichzeitig hat es bei wiederkehrenden Lehrveranstaltungen und **Aufgabenschemata** Vorteile. Denn Sie müssen sich nicht jedes Mal neue Kriterien überlegen, sondern können auf die Raster aus anderen Lehrveranstaltungen zurückgreifen. Selbiges gilt für die Deskriptoren, mit denen Sie Ihre Erwartungen an die studentischen Leistungen beschreiben. Zudem liegt ein Vorteil in der Korrekturphase: Mithilfe dieses Kriterienrasters können Sie sich die **Notenbegründung**, die Sie sonst zusätzlich anfertigen müssen, sparen. Eine Prüfungsbeurteilung, die nur aus Symbolen wie „+“ und „o“ besteht, und die weder intersubjektiv nachvollziehbar ist noch die Garantie bietet, dass Sie nach einiger Zeit noch zum selben Schluss kommen, hat vor Gericht keine Chance. Deshalb benötigen Sie für jede Note im Prüfungskontext an der Hochschule eine nachvollziehbare Begründung. Diese kann aus dem Kriterienraster bereits abgelesen werden. Was letztlich auch erweiterte Möglichkeiten des **Feedbacks** an Studierende ermöglicht. Denn wenn Sie den Studierenden anhand solcher Kriterien und Beschreibungen verständlich machen, wo deren Stärken und Schwächen in der Prüfung lagen, unterstützen Sie den individuellen Lernprozess.

Literatur

Kreuzer, Pia (2018): **Kriterienraster**. Handreichung der Prüfungswerkstatt. Mainz: Johannes Gutenberg-Universität.

Studierende über Kriterien informieren

Mindestens vier Gründe sprechen dafür, Studierende möglichst präzise über Beurteilungskriterien zu informieren: Erstens können sie erst während des Studiums lernen, was in Prüfungen von ihnen erwartet wird. Zweitens haben Lehrende – wie oben gezeigt – unterschiedliche Anforderungen. Drittens ist es eine beruflich relevante Kompetenz, z.B. beim Schreiben unterschiedliche Vorgaben berücksichtigen zu können und viertens erleichtert Ihnen die Offenlegung von Kriterien die Betreuung *und* die Beurteilung, so dass Sie dadurch Zeit

einsparen.

Um Kriterien offenzulegen, können Sie vielfältige und unterschiedliche Settings und Materialien nutzen: Von Sprechstundengesprächen und Seminarsitzungen über Leitfäden, ‚Fahrpläne‘ und Style Sheets bis hin zu Handouts und Kriterienkatalogen (s.u.). Wenn Sie das Ziel verfolgen, den Betreuungsaufwand zu reduzieren, können Sie Studierende schriftlich über Ihre Kriterien informieren. Weil aber ja unter einem Kriterium wie „der Text ist verständlich“ verschiedene Lehrende Unterschiedliches verstehen, ist es wichtig, dass Sie die Kriterien nicht nur benennen, sondern möglichst anhand von Beispielen illustrieren, was genau Sie damit erfassen und wie Sie es auf eine Prüfungsleistung anwenden. Solche Beispiele können Sie auch im Rahmen einer Seminarsitzung oder im Kolloquium gemeinsam mit den Studierenden diskutieren. Bestimmt freuen Sie sich, wenn Prüfungen gut verlaufen – warum also nicht Studierende auch einmal explizit darüber informieren, was sie für eine sehr gute Note tun können?

Beispiel aus einem Betreuungsleitfaden für eine mündliche Prüfung:

- Ihr Ziel für die Prüfung sollte sein, mir zu beweisen, dass Sie das Thema beherrschen. Deshalb dürfen Sie mich unterbrechen, wenn ich etwas sage, das Sie selber wissen!
- Nutzen Sie die knappe Zeit in der Prüfung geschickt: Bilden Sie beim Lernen eine Wissenshierarchie! Würden Sie sich z.B. über Goethe prüfen lassen, sollten Sie also nicht seinen Lebenslauf darstellen (ich möchte nicht prüfen, wie gut Sie auswendig lernen), sondern lieber kontroverses Detailwissen einbringen, das nicht in jedem Buch steht.

Wir empfehlen Ihnen, auch Ihre Beurteilungskriterien als Teil eines solchen informativen Betreuungsleitfadens offenzulegen, mit dem Sie Prüflinge zudem darüber informieren können, wie Sie die Betreuung gestalten, also wann und wie Sie bereit sind zu unterstützen, was Sie als Eigenleistung erwarten, wie viel Zeit Sie für die Betreuung zur Verfügung stellen, ob die Möglichkeit für Zwischenrückmeldungen besteht etc.

Notenbegründung und lernförderliche Rückmeldung

Damit Studierende aus einer Prüfung lernen können, sind sie auf Ihre Rückmeldungen angewiesen!

Aus der am häufigsten gewählten Rückmeldeform bei schriftlichen Arbeiten, nämlich der Nutzung von Korrekturzeichen, lernen Studierende in der Regel nicht (Lee 2007). Dies gilt auch für den kurzen Kommentar oder das Notengutachten. Sie haben eher die Funktion die Note zu legitimieren, auch um Rechtssicherheit herzustellen. Korrekturen, Kommentare und Notengutachten folgen damit anderen Regeln als eine lernförderliche Rückmeldung, weil mit ihnen das Ziel verfolgt wird, möglichst alle benotungsrelevanten (Stärken und) Schwächen sachlich zu beschreiben. Formulierungen, die Sie hier wählen, müssen wegen der Rechtssicherheit beispielsweise immer eine Entsprechung zur gewählten Note haben. Nützliche Formulierungshilfen dafür finden Sie bei den Hinweisen zur Notenbegründung der Kultusministerinnenkonferenz für den **schulischen Bereich**. Jeder Note werden hier einzelne Formulierungen zugeordnet, die den Grad der Erfüllung eines Kriteriums beschreiben.

Damit Studierende aus der Prüfung selbst besser lernen können, sollten daher Korrektur, Kommentar und Gutachten um eine lernförderliche Rückmeldung ergänzt werden. Dafür sind folgende Prinzipien wichtig:

- Sprechen Sie nur das Wichtigste an:
Rückmeldungen können sich auf Higher und Later Order Concerns beziehen. Zu Higher Order Concerns in wissenschaftlichen Texten zählt beispielsweise, ob eine Fragestellung entwickelt und konsequent verfolgt wird, ob der Text sachlich korrekt und für die Adressaten verständlich ist, ob die Relevanz des Themas deutlich wird etc. Later Order Concerns hingegen betreffen Syntax, Wortwahl, sprachliche oder formale Korrektheit. Bei einer lernförderlichen Rückmeldung sollten Sie sich auf die Beobachtungen im Bereich der Higher Order Concerns beschränken, die aus Ihrer Sicht am wichtigsten für die Qualität des Textes sind. Mehr als drei oder vier sollten es nicht sein, damit Studierende einen Fokus für ihre Gestaltung des weiteren Lernprozesses haben.
- Benennen Sie Stärken des Texts:
Positive Rückmeldung ist nicht nur motivierender und kann leichter akzeptiert werden als defizitorientierte (vgl. z.B. Ilgen et al. 1979), sondern führt auch infolge zu besseren Texten. Durch rein defizitorientiertes Feedback hingegen büßen die Texte nach einer Überarbeitung an Qualität ein (Hillocks 1986). Nehmen Sie sich daher Zeit dafür, die Stärken eines Textes zu identifizieren und sprechen Sie auch hier diejenigen an, die am wichtigsten für die Qualität

des Textes sind.

- Begründen Sie Ihre Rückmeldung:
Eine Rückmeldung basiert auf Vorstellungen davon, welchen Konventionen und Darstellungsmustern ein Text entsprechen sollte. Weil Studierende Konventionen kennen und anwenden lernen müssen, ist es besonders wichtig, dass Sie Ihre Rückmeldungen auf solche Prinzipien beziehen (z.B. mit Formulierungen wie „Weil es in wissenschaftlichen Texten üblich ist...“; „Weil mir bei wissenschaftlichen Texten wichtig ist...“) (vgl. Limburg 2016: S. 149–150).

Literaturtipps

1. Boettcher, W./Meer, D. (2000): „Ich hab nur ne ganz kurze Frage“ – Umgang mit knappen Ressourcen. Sprechstundenkommunikation an der Hochschule. Neuwied: Luchterhand Verlag.
2. Bohl, T. (2004): Prüfen und Bewerten im offenen Unterricht. Neuwied: Kriftel.
3. Grzesik, J./Fischer, M. (1985): Was leisten Kriterien für die Leistungsbeurteilung? Theoretische, empirische und praktische Aspekte des Gebrauchs von Kriterien und der Mehrfachbeurteilung nach globalem Ersteindruck. Opladen: Westdeutscher Verlag.
4. Hartog, P. J./Rhodes, E.C. (1935): An Examination of Examinations. London: Macmillan.
5. Jachmann, M. (2003): Noten oder Berichte?: Die schulische Beurteilungspraxis aus der Sicht von Schülern, Lehrern und Eltern. Wiesbaden: VS Verlag für Sozialwissenschaften.
6. Kerres, M./ Schmidt, A. (2011): Zur Anatomie von Bologna-Studiengängen. Eine empirische Analyse von Modulhandbüchern. In: Die Hochschule. Journal für Wissenschaft und Bildung 2, S. 173–191.
7. Lee, I. (2007): Ten mismatches between Teachers' beliefs and Written Feedback Practice. ELT Journal 63 (1).
8. Limburg, Anika (2016): Betreuung von studentischen Arbeiten. In: Wissen, was zählt, 2. Auflage, S. 149–152.
9. Meyer, W. (1984): Das Konzept von der eigenen Begabung. Stuttgart: Huber.
10. Pohl, T. (2007): Studien zur Ontogenese wissenschaftlichen Schreibens. Niemeyer: Tübingen.

Autor*in

Dr. Anika Limburg, Mitarbeiterin im Schreibzentrum des Zentrums für Wissenschaftsdidaktik der Ruhr-Universität Bochum, anikalimburg