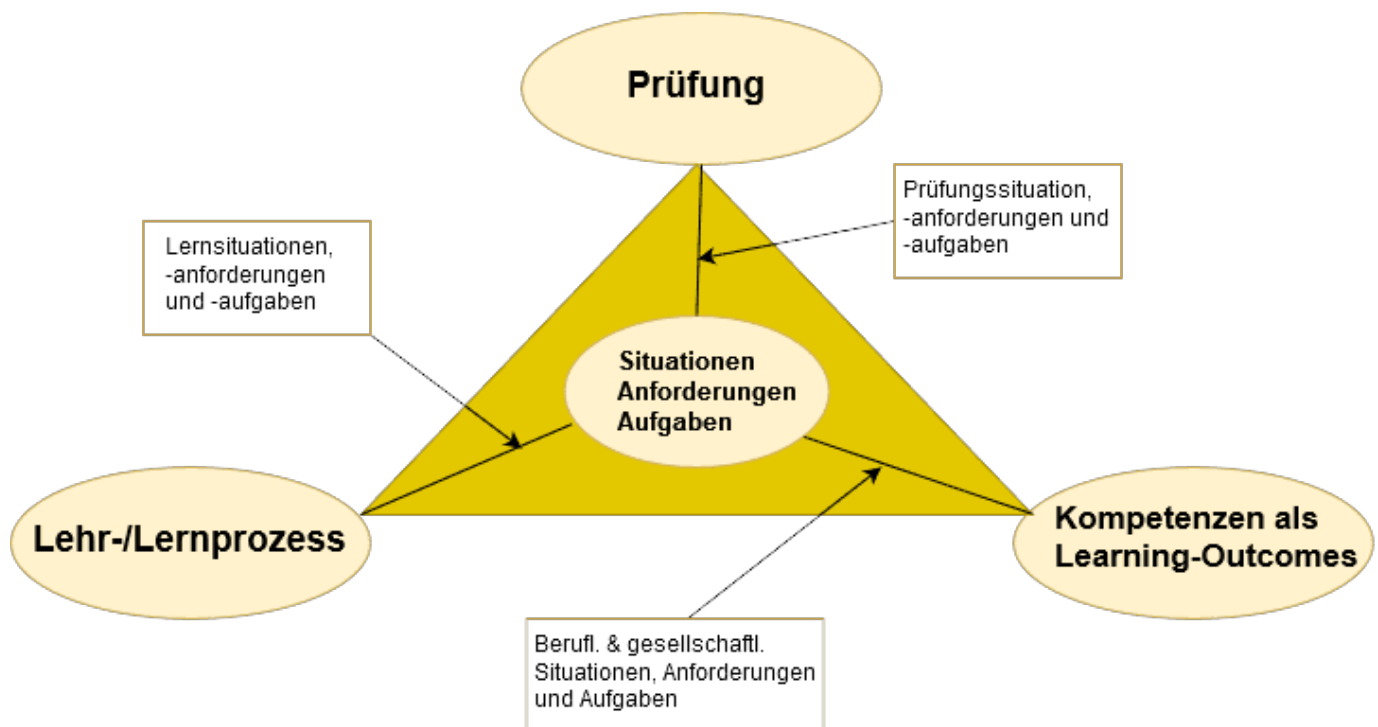


# Selbstlernkurs: Kompetenzorientiert Prüfen

## Kompetenzen

Der Prozess der Prüfungserstellung beginnt nicht am Ende der Lehrveranstaltung. Bereits bei der Lehrplankonzeption müssen Lehre, Lernzielen und Prüfungen gemeinsam gedacht und geplant werden. Bildlich dargestellt ist es hier im Constructive Alignment nach Johannes und Beatrix Wildt:



Es zeigt, dass alle drei Bereiche miteinander zusammenhängen. Zunächst Lehrende stellen Learning-Outcomes auf, also Kompetenzen und Lernziele, die im Modul erreicht werden sollen. Abhängig davon wird der Lehr-/Lernprozess gestaltet und Aufgaben konzipiert, die mit der Anwendung von erlernten Kompetenzen zu lösen sind.

## ► Kompetenzorientiert Prüfen I

# Taxonomien

Die Lernaktivität der Studierenden hängt stark davon ab, was am Ende der Lehrveranstaltung geprüft wird. Wird Faktenwissen verlangt, werden die Studierenden sehr wahrscheinlich auch nur Faktenwissen lernen. Daher erfordert eine kompetenzorientierte Prüfung auch eine kompetenzorientierte Lehre. Eine gute Orientierung bietet die Lernzieltaxonomie nach Anderson und Krathwohl:

### Taxonomie nach Anderson & Krathwohl



Quelle: Universität Zürich (2022): Lerntaxonomie nach Anderson und Krathwohl; erweitert und modifiziert OTH Regensburg;

Die angestrebten Kompetenzen sind mit Lernzielen verknüpft. Nach oben zur Spitze der Pyramide hin steigt die Komplexität und der Schwierigkeitsgrad. Idealerweise sollte eine Prüfung aus mehreren Taxonomien zusammengestellt werden, um eine gewisse Qualität der Prüfung zu erreichen und auch die Leistungen der Studierenden besser voneinander differenzieren zu können.

## Lernziele

Zusammen mit den Veranstaltungsinhalten sollten auch die Lernziele formuliert werden, die dann nützlich bei der Prüfungserstellung sind. Eine Orientierung bieten folgende Leitfragen:

- Was sind die Studierenden am Ende in der Lage zu tun?
- Womit können Studierende das erlernen
- Wozu sollen Studierende das erlernen?

# Formulierungshilfen für Fragen

Damit Sie die gewünschten Taxonomien und Kompetenzen ansprechen können, sind auf dieser Seite verschiedene Verbtabelle dargestellt (deutsch und englische Verben). Ebenso ist es unbedingt wichtig, auch den Studierenden in der Hinsicht zu sensibilisieren, damit ihnen genau bewusst ist, was von ihnen verlangt wird. Teilweise muss auch dieses Wissen erst vermittelt werden, welches sich über Übungsaufgaben im Laufe der Veranstaltungen gut umsetzen lässt.

# Prüfungskonzepte

Das Prüfen kann mehr sein als Wissen und Verständnis zu überprüfen. Besonders für das formative Prüfen ist es eine Chance, dass man während der Prüfung etwas mitnimmt, auch wenn man nicht besteht. Je nach Zulässigkeit durch die Prüfungsordnungen sollte die Vielfalt an diversen Fragetypen genutzt werden. Ebenfalls sind alternative Prüfungsformen auch ein gutes Mittel. Auf dieser Seite befindet sich bereits eine Sammlung mit vielen Prüfungsformen.

# Blueprint

Ein Blueprint ist ein Schema, das vorab für eine Prüfung anlegt wird, um die Inhalte zu gewichten. Damit soll sichergestellt werden, dass die Anteile von Themenbereichen, Taxonomiestufen und der Bepunktung einzelner Fragen ausgeglichen ist. Die Verwendung eines Blueprint erhöht zudem die Reliabilität und Validität der Prüfung (werden im folgenden Abschnitt erklärt).

# BLUEPRINT PHYSIOLOGIE

## z. b. Medizin

	Definitionen	Zusammenh.	Pathophys.	gesamt
Zellphysiol.				5%
Nerv				5%
Muskel				10%
Blut				10%
Atmung				10%
Säure/Basen				5%
Niere				10%
Herz/Kreislauf				10%
Verdauung				10%
...				25%
Umfang	20%	50%	30%	100%

Quelle: <https://lehreladen.rub.de/planung->

durchfuehrung-kompetenzorientierter-lehre/

kompetenz-pruefen/multiple-choice-

pruefungen/

## Gütekriterien für Fragen

Das Gütekriterium **Objektivität** bezieht sich auf die Unabhängigkeit der Prüfung vom Prüfenden und von der Prüfungssituation. Das Prüfungsergebnis soll unabhängig von der Person sein, die die Prüfung durchführt und bewertet. Um die Durchführungsobjektivität sicherzustellen, müssen alle Studierenden die gleichen Informationen über den Prüfungsablauf, die Bearbeitungszeit, die erlaubten Hilfsmittel und Hinweise zur Bearbeitung und Beantwortung der Fragen erhalten. Um auch bei der Bewertung der Prüfung Objektivität (Auswertungsobjektivität) gewährleisten zu können, müssen alle Prüfungen nach den gleichen Bewertungskriterien korrigiert werden. Deshalb sollte vor der Prüfung eine Musterlösung erstellt werden. Die Interpretationsobjektivität zielt darauf ab, dass unterschiedliche Bewerter:innen die gleichen Schlüsse aus dem Testergebnis ziehen, die gleiche Anzahl von Punkten vergeben und so zu den gleichen Noten kommen. Aus diesem Grund sollten die Punktevergabe im Bewertungsraster und die Noten in einem Notenschlüssel festgelegt werden.

Das Gütekriterium **Reliabilität** ist erfüllt, wenn durch die Prüfungsfragen/-aufgaben zuverlässig der wahre Leistungsstand eines bzw. einer Studierenden in einer Prüfung gemessen wird. Die Reliabilität kann in der Praxis durch folgende Faktoren erhöht werden:

Die Prüfungsdauer sollte so gewählt sein, dass mindestens 80% der Studierenden alle Prüfungsfragen/-aufgaben beantworten können.

Generell sollten die Fragen/Aufgaben nicht zu schwer und nicht zu leicht sein, das heißt, es sollten überwiegend Fragen/Aufgaben mit mittlerem Schwierigkeitsgrad gestellt werden. Eine gute

Reliabilität kann durch eine hohe Anzahl an Prüfungsfragen/-aufgaben, bzw. eine umfassende Prüfung, erreicht werden. Je mehr Fragen von den Studierenden beantwortet werden, desto besser kann der Leistungsstand und die Erreichung der Lernziele gemessen werden.

Wenn eine Prüfung als zuverlässig gilt, bleibt die Frage, ob sie das misst, was sie messen soll und somit valide ist. Der inhaltliche Rahmen einer Prüfung ist durch die Modulbeschreibung der Lehrveranstaltung gegeben. Der Schwierigkeitsgrad wird durch die Lernziele und den damit adressierten Taxonomiestufen beschrieben.

Die **Validität** einer Prüfung erfordert deshalb eine starke Orientierung an der Modulbeschreibung einer Lehrveranstaltung und den ggf. im Laufe der Lehrveranstaltungsdurchführung erfolgten begründeten Anpassungen der Lehrinhalte und Schwierigkeitsgrade.

# Optimierung von Fragetypen

## Geschlossene Fragetypen erweitern

Bei geschlossenen Fragetypen können nur Antworten aus einer Vorgabe ausgewählt werden. Mit Abstand werden am häufigsten Multiple-Choice-Fragen zur reinen Wissensabfrage eingesetzt. Dadurch kommen aber nur die Lernzieltaxonomien "Erinnern" und "Verstehen" zur Anwendung. Gute Fragen sollen aber auch höhere Stufen adressieren. Die beiden nächsthöheren Stufen können mittels einer Optimierung des Fragenstammes erreicht werden. Die Anreicherung mit Medien ist eine gute Abwechslung und bietet eine Grundlage, auf die sich die Frage beziehen kann. Das kann eine kurze Audio-Datei sein, wo in der darauffolgenden Frage nach einem Inhalt aus dem Gehörten gefragt wird oder auch ein Diagramm oder eine Fallbeschreibung.

Für ein langfristiges Arbeiten mit elektronischen Prüfungen und zum einfachen Erweitern der Fragensammlung kann ein Aufgabenschema erstellt werden: "Erläutern Sie, welche Bedeutung X für Y hat." Mit dem simplen Austausch der Komponenten im Fragestamm kann schnell eine neue Frage oder Alternativversion erstellt werden. Selbstverständlich darf man nicht vergessen, die Antwortmöglichkeiten anzupassen.

Im Grunde gewinnt eine Frage an Qualität, wenn sie eine höhere Kompetenzstufe erreicht und ein Verstehen statt einem Wissen abgefragt wird.

Hier ist ein nicht gut gelungenes Beispiel dargestellt. Eine MC-Frage besteht aus einem Aufgabenstamm und den Antwortoptionen. Dort ist die Aufgabe oder die Frage dargestellt. Diese kann in einem kurzen Kontext eingebettet sein. Zusätzlich sind weitere Bestandteile ergänzend möglich: eine Fallbeschreibung, Datensätze oder Analyse-Materialien wie Diagramme. In diesem Beispiel soll die korrekte Konjugationsform eines Verbes anhand eines sprachverwandten Verbes bestimmt werden. Dabei werden die Taxonomiestufen 3 und 4 adressiert. Es geht bei der Aufgabe eher um das Verstehen und Transferieren statt um das Abrufen von Wissen. Nur ist der Aufgabenstamm zu lang und mit zu vielen unnötigen Aussagen vollgepackt. Ebenso sind die Distraktoren nicht gut gewählt.



<b>Aufgabenstamm mit Frage-/Aufgabentext</b>	{	Die Bildung von Verben ist essentiell für das Erlernen einer Sprache. Es wird konjugiert wie ein regelmäßiges Verb mit der Endung -ar. Wählen Sie die 3. P. Sg. vom spanischen Verb <i>cenar</i> . Übersetzt heißt es „essen“. Daneben steht das italienische Vorbild.
<b>weitere Bestandteile (langer Aufgabenstamm)</b>		<div><b>cenare</b> ceno ceni cena ceniamo cenate cenano</div> <div><div>A: cena B: cenáis C: conas D: cenamos</div><div>}</div><div>Antwortoptionen</div></div>

Nachfolgend ist ein besseres Beispiel abgebildet. Hier ist der Fragenstamm kurz und knapp gehalten, es gibt keine unnötigen Angaben oder versteckte Hinweise.



<b>Aufgabenstamm mit Frage-/Aufgabentext</b>	{	Vervollständigen Sie den Satz. Orientieren Sie sich bei der Bildung des Verbes an der italienischen Version. Sébastien _____ una manzana.
<b>weitere Bestandteile (langer Aufgabenstamm)</b>		<div><b>mandare</b> mando mandi manda mandiamo mandate mandano</div> <div><div>A: mandas B: mandes C: manda D: mande</div><div>}</div><div>Antwortoptionen</div></div>

## ► Gute Prüfungsfragen: MC-Fragen

# Offene Fragen

Eine gute Frage sollte ohne Nachfragen seitens des Prüflings auskommen. Die Aufgabe sollte also mehrfach durchgelesen und kontrolliert werden, ob alle enthaltenen Fachbegriffe erklärt und die Fragestellung klar definiert ist. Es ist auch besonders peinlich, wenn Fachbegriffe falsch verwendet werden. Dazu gehört ebenfalls die Angabe über den Umfang der zu erwartenden Antwort und ein transparentes Bewertungsraster.

Eine offene Frage kann schnell mit zu vielen Aufgaben überfüllt wirken. Wandeln Sie daher eine umfangreiche Frage in mehrere kleinere Fragen um oder arbeiten Sie mit Teilaufgaben. Hier eignet sich auch die Kombination mit einer geschlossenen, bspw. einer MC-Frage. Die Erläuterung zur Wahl der ausgewählten Antwort findet in der offenen Frage statt.

Für die Konzeption von Aufgaben zu höheren Taxonomiestufen, wie dem Analysieren oder Beurteilen, kann weiteres Material hinzugefügt werden: Diagramme, Karten, Texte oder Tabellen mit Zahlenwerten. Diese Analyse kann mittels erlaubtem Zugriff auf Forschungsdatenbanken und -bibliotheken erfolgen, womit gleichzeitig weitere Kompetenzen geprüft werden. Schließlich ist das schnelle und zuverlässige Auffinden relevanter Informationen und das kompetente Einarbeiten in neue Fachthemen eine Schlüsselkompetenz in zahlreichen Disziplinen.

Offene Fragen stehen aber nicht nur für die Beantwortung mit Texten. Auch Aufgaben zum Schreiben von Programmcodes sind möglich. Der spezielle Fragetyp CodeRunner in Moodle stellt eine Umgebung bereit, in dem Codes von verschiedenen Programmiersprachen geschrieben und ausgeführt werden. Wichtig ist dabei immer an das **Constructive Alignment** zu denken und die Fragen an den Lernzielen im Kurs auszurichten.

Bei diesem guten Beispiel aus der germanistischen Sprachwissenschaft sollen Monophthongierungen identifiziert werden. Bei dem zweiten Teil der Frage soll die Wahl begründet werden. Dabei erfolgt die Anwendung der Stufe 3 Anwenden (gelerntes Wissen anwenden) und der Stufe 5 Beurteilen (Begründung der Wahl) von der Lernzieltaxonomie von Anderson und Krathwohl. Durch die Kombination aus zwei Teilaufgaben kann der Prüfling sich quasi selbst überprüfen, ob er sein Wissen angewendet. Und durch das Fordern einer Begründung wird der Versuch eines Ratens auf gut Glück zumindest etwas eingeschränkt.

**a) Markieren Sie im folgenden Text Monophthongierung:**

Wan ob er danne daz **getuot**,  
sô **muoz** er sagen wol vür wâr  
daz die tugende **guot** sint gar,  
wan niht anders zer werlde ist  
daz gar sî **guot** zaller vrist.  
Sô mac ouch niht gereichen baz  
zem obersten **guote**, wizzet daz.  
Daz swære daz **muoz ziehen** nider  
und kumt von sîner kraft niht wider.

**b) Begründen Sie kurz Ihre Auswahl unter Angabe der Diphthonge!**

► Gute Prüfungsfragen: Offene Fragen

# Überprüfung

Nachdem die erste Version einer Prüfungsfrage erstellt wurde, beginnt nun der Prozess der Überarbeitung. Mit Hilfe einer Suchmaschine oder auch ChatGPT kann kontrolliert werden, ob sich die Lösung der Frage nicht dort leicht finden lässt. Geprüft werden alle Elemente, der Aufgabenstamm, die Distraktoren und die zusätzlichen Materialien. Man sollte sich auch selbst nach dem „Warum und Wie“ fragen statt dem „Wer oder Was“.

## Aufgabenstamm

Die Aufgabe sollte nicht nach den falschen Antworten fragen. Diese könnten sich im Gedächtnis als richtig festsetzen. Eine gute Frage ist selbsterklärend und sollte ohne Nachfragen seitens des Prüflings auskommen. Die Aufgabe sollte also mehrfach durchgelesen und kontrolliert werden, ob alle enthaltenen Fachbegriffe erklärt und die Fragestellung klar definiert ist. Eine fehlerfreie Grammatik und Ausdruck ist dabei selbstverständlich. Die Inhalte sollten kurz und prägnant gehalten, keine doppelten oder überhaupt Verneinungen verwenden werden.

## Distraktoren

Die Distraktoren, also Antwortoptionen, sollten klar und verständlich formuliert sein. Dafür ist es wichtig, dass jeder der Distraktoren nur eine Antwort enthält. Indem sich alle Distraktoren in Länge und Struktur ähneln, kann ein Erraten der korrekten Option, meist die längste Antwort, ausgeschlossen werden. Selbstverständlich sollte auch die grammatische Struktur zum Aufgabenstamm bzw. zur gestellten Frage passen. Die Verständlichkeit kann optimiert werden, indem eine Negation in der Aufgabe und bei den Antworten vermieden wird. Statistiken haben gezeigt, dass ab vier Distraktoren die Ratewahrscheinlichkeit deutlich sinkt. Es sollten aber auch nicht zu viele davon erstellt werden, um die Verständlichkeit und Übersicht zu bewahren. Einfach mal die Position der korrekten Antwort variieren, jedoch sollten sie bei entsprechender Fragestellung stets logisch angeordnet werden und sich auch aufeinander beziehen. und die erstellten Distraktoren auf versteckte Hinweise, die die korrekte Antwort verraten könnten.

## Review-Prozess

Für die Entwicklung von qualitativ hochwertigen Fragen ist die Etablierung eines Review-Prozesses empfehlenswert. Dabei gibt eine Kollegin oder ein Kollege aus dem gleichen oder einem ähnlichen Fachbereich Feedback zu den Fragen. Bei diesem Korrekturlesen kann diese Checkliste hilfreich sein, die die zu begutachteten Kriterien auflistet:

- Grammatik



- Begriffsverwendung
- Logik
- Distraktoren
- Übereinstimmung mit Lehrziel/Blueprint/Taxonomie
- Punktzahl

# Bewertungen

Für formatives wie auch für summatives Prüfen, also mit einer einzigen Abschlussprüfung, ist es empfehlenswert, sich Bewertungskriterien zu erstellen. Sie bauen auf den **Lernzielen** und **Taxonomien** auf. Sie erleichtern den Lehrenden eine gerechte Bewertung und den Studierenden eine Orientierung, was in der Prüfung gefordert wird. Das gegebene Feedback gibt den Studierenden einen detaillierten Einblick in ihren Leistungsstand. Es zeigt auch deren Potenziale und Entwicklungsbedarfe an. Zum Entwickeln eines eigenen Bewertungsrasters werden zunächst die Minimalanforderungen und danach die Maximalanforderungen festgelegt. Ergänzend definiert man Zwischenstufen und gewichtet die einzelnen Kriterien und vergibt die zu erreichenden Punkte. Auf dieser Seite sind **Vorlagen zu Bewertungsmatrix** gespeichert. Bei mehreren Aufgaben in einem Test empfiehlt sich die gleichen Aufgaben nacheinander zu korrigieren. Sind mehrere Prüfende beteiligt, sollten statt einzelner Klausuren die einzelnen Aufgaben zur Bewertung zugeteilt werden. Weiterhin empfiehlt sich besonders bei erstmals verwendeten Fragen eine Sichtung mehrerer Antworten, um eine Orientierung für die Lösungsbewertung zu erhalten.

## Bewertungsfehler

Die Bewertung von Prüfungsleistungen basiert nicht selten auf einer subjektiven Wahrnehmung der Lehrenden. Zur Reduzierung und möglichst vollständiger Vermeidung sollten sich Lehrende an festgelegte Bewertungsraster und Gütekriterien halten. Es gibt zudem Bewertungsfehler, die durch Wahrnehmungsverzerrung bedingt sind oder auch Tendenzfehler. Ein weiterer Tipp zur Vermeidung von Bewertungsfehlern ist die anonyme Korrektur, sodass keine Verbindungen zum Prüfling hergestellt werden und die Bewertung positiv oder negativ beeinflussen.

## 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.

## 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.

## 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.