

RUHR-UNIVERSITÄT BOCHUM

„KI-ROBUSTE“ PRÜFUNGEN: EIN SCHRITT ZURÜCK FÜR EINEN WEG NACH VORN

NADINE LORDICK & JULIA BECKER

Agenda

- Einführung
- Konzept
- Beispiele & Rückmeldungen
- Reflexion
- Schritt für Schritt-Anleitung

Was ist „Künstliche Intelligenz“?

- Sammelbegriff für eine Vielzahl von Methoden, Anwendungen und Modellen, die die Automatisierung von bisher menschlichen Tätigkeiten zum Ziel haben
- Problem: Begriff ist missverständlich
- Wovon wir heute sprechen: generative Große Sprachmodelle (auch: Large Language Models [LLMs]), die häufig auch als Chatbots fungieren (z. B. ChatGPT, Gemini, Lumo, ...)



Workshop zu KI & Prüfen

- 3–4-stündiger Workshop
- Prämissen & Konzept:
 - kein „one-size fits all“
 - Ziel: Reflexion & Erarbeitung statt fertiger Lösungen
 - schrittweises Vorgehen mit Hilfe eines Reflexionsblattes



Zentrum für
Wissenschafts-
didaktik



Reflexionsblatt zu Prüfungen

Welche Prüfungsform habe ich? *(Bitte eine auswählen)*
☐ Klausur ☐ Hausarbeit ☐ Mündliche Prüfung ☐ Sonstiges: _____

Welche Lernziele will ich mit der Prüfung abprüfen?

Was bewerte ich?

☐ Ergebnis ➡ ☐ Wie viel Toleranz für Fehler ist in Ordnung? _____ %

☐ Lernzuwachs ➡ ☐ Sozialer Vergleich ☐ Individueller Vergleich

☐ Prozess ➡ ☐ Ergebnis muss stimmen ☐ Lösungsweg muss richtig sein

Wie gewährleiste ich Vergleichbarkeit?

☐ Kriterienraster ☐ Musterlösung ☐ Kennzeichnung von LLM-Nutzung

☐ Sonstiges: _____

Welche Bewertungskriterien lege ich an?

Passen meine Bewertungskriterien zu den Lernzielen?

☐ Ja, komplett ☐ Ja, in Teilen ☐ Nein, noch nicht

Was will ich, dass die Studierenden wirklich selbst zeigen?

Wo sind Hilfsmittel egal?

Wie können Hilfsmittel ggf. gekennzeichnet werden?

Wie sehr will ich die Studierenden kontrollieren, wie sehr ihnen vertrauen?

Kontrolle ← _____ → Vertrauen

Welche alternativen Prüfungsformen passen zu meinen Lernzielen (und den rechtlichen Vorgaben)?

Eine Beispieldiskussion: Abstracts schreiben

Person A:

Ich versuche den Studierenden deutlich zu machen, was sie davon haben, Abstracts selber zu schreiben, um sie zu motivieren, auf Sprachmodelle zu verzichten.

Person B:

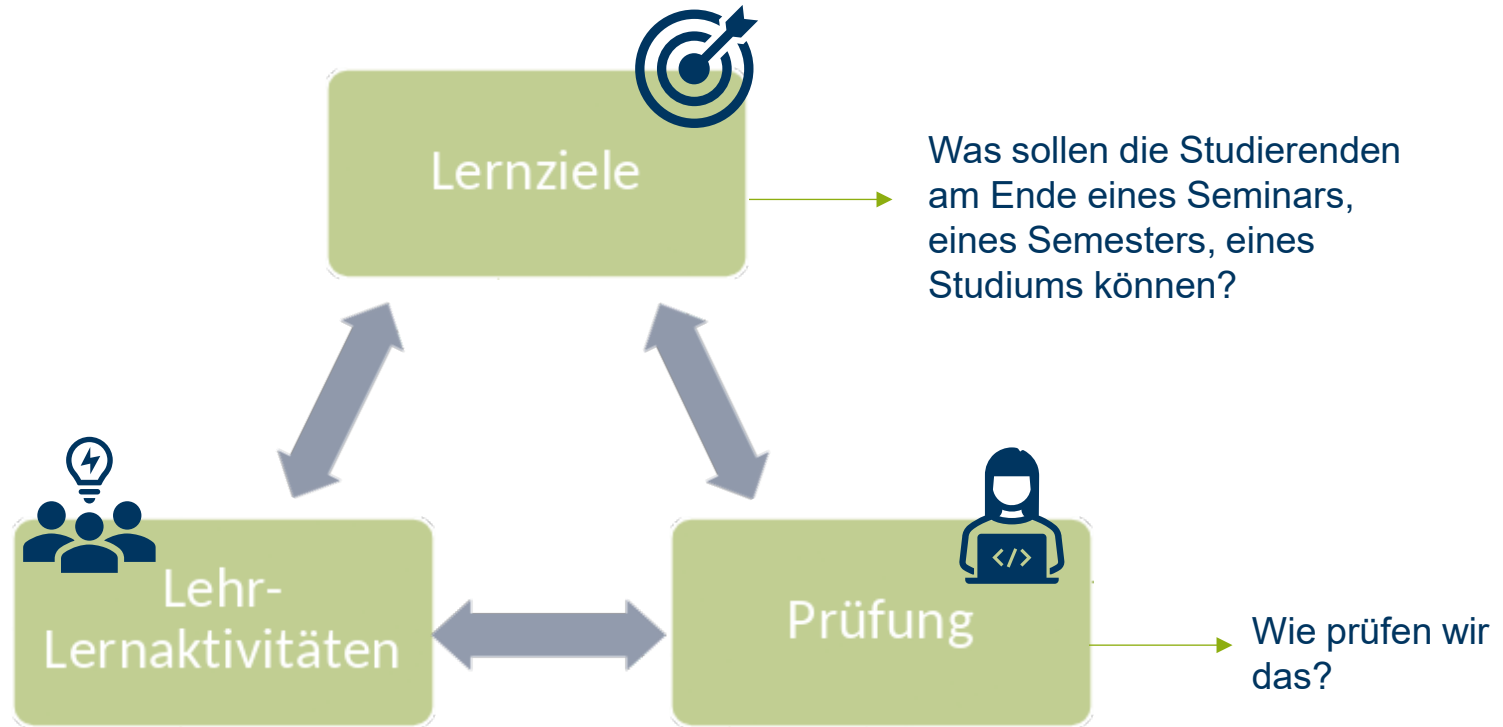
Aber müssen Studierende denn überhaupt lernen, Abstracts zu schreiben, wenn das in Zukunft sowieso von Sprachmodellen übernommen wird?

Person C:

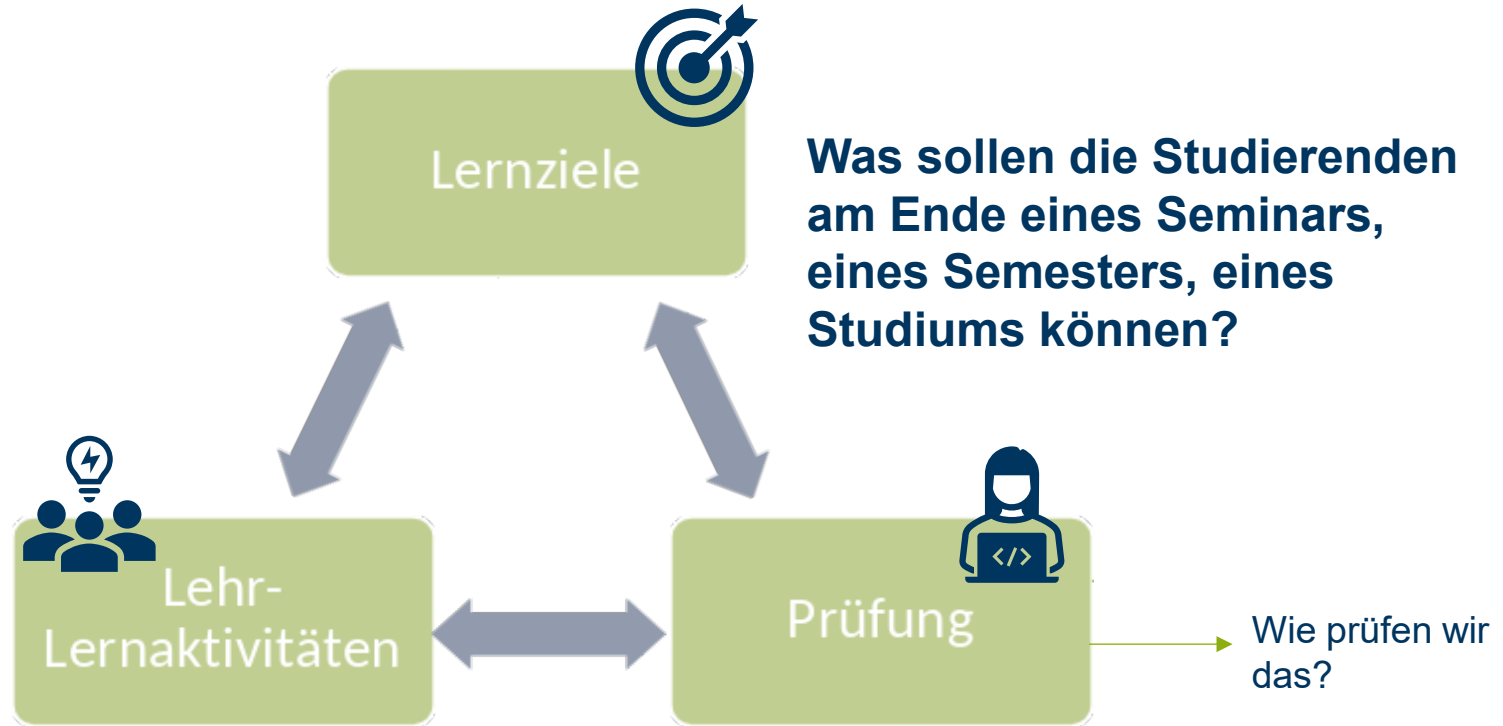
Beim Abstract schreiben geht es ja um mehr als nur ‚Abstract schreiben‘. Ich lerne dabei, Wichtiges von Unwichtigem zu trennen – und das ist eine wichtige Fähigkeit...

Was müssen Studierende (noch) lernen?

Back to the Basics: Constructive Alignment



Back to the Basics: Constructive Alignment



Beispiel: Was tun wir beim wissenschaftlichen Schreiben?

- Fragestellung entwickeln
- Fachbegriffe definieren
- Forschungsliteratur zusammenfassen, zitieren & paraphrasieren
- Argumentieren
- Methoden darlegen
- Hypothesen aufstellen
- Überarbeiten
- Formatieren
- ...



Beispiel Paraphrasieren: mit und ohne LLM



Ohne LLM

- interpretieren
- Wichtiges von Unwichtigem trennen
- l
- n
- lernen, dass es unterschiedliche Wege gibt, eine Idee auszudrücken
- ...

Mit LLM

- etwas über das Tool lernen
- kritisch prüfen, ob der Inhalt richtig
- se das
- leistet, was sie leisten soll
- lernen, dass es unterschiedliche Wege gibt, eine Idee auszudrücken
- ...

Man lernt immer etwas – die Frage ist nur, was.

Beispiel Paraphrasieren: Mögliche Lernziele

Die Studierenden sollen am Ende der Veranstaltung in der Lage sein,

... mithilfe generativer KI Paraphrasen zu erstellen, die wissenschaftlichen Kriterien standhalten.

... zu erläutern, was eine gute Paraphrase ausmacht.

... eine Idee auf unterschiedliche Weise auszudrücken.

... durch Paraphrasen zu überprüfen, ob sie einen Inhalt so verstanden haben, dass sie ihn wiedergeben können.

...

Schritt 1 & 2: Prüfungsform & Lernziele



Zentrum für
Wissenschafts-
didaktik



Reflexionsblatt zu Prüfungen

Welche Prüfungsform habe ich? *(Bitte eine auswählen)*

☐ Klausur

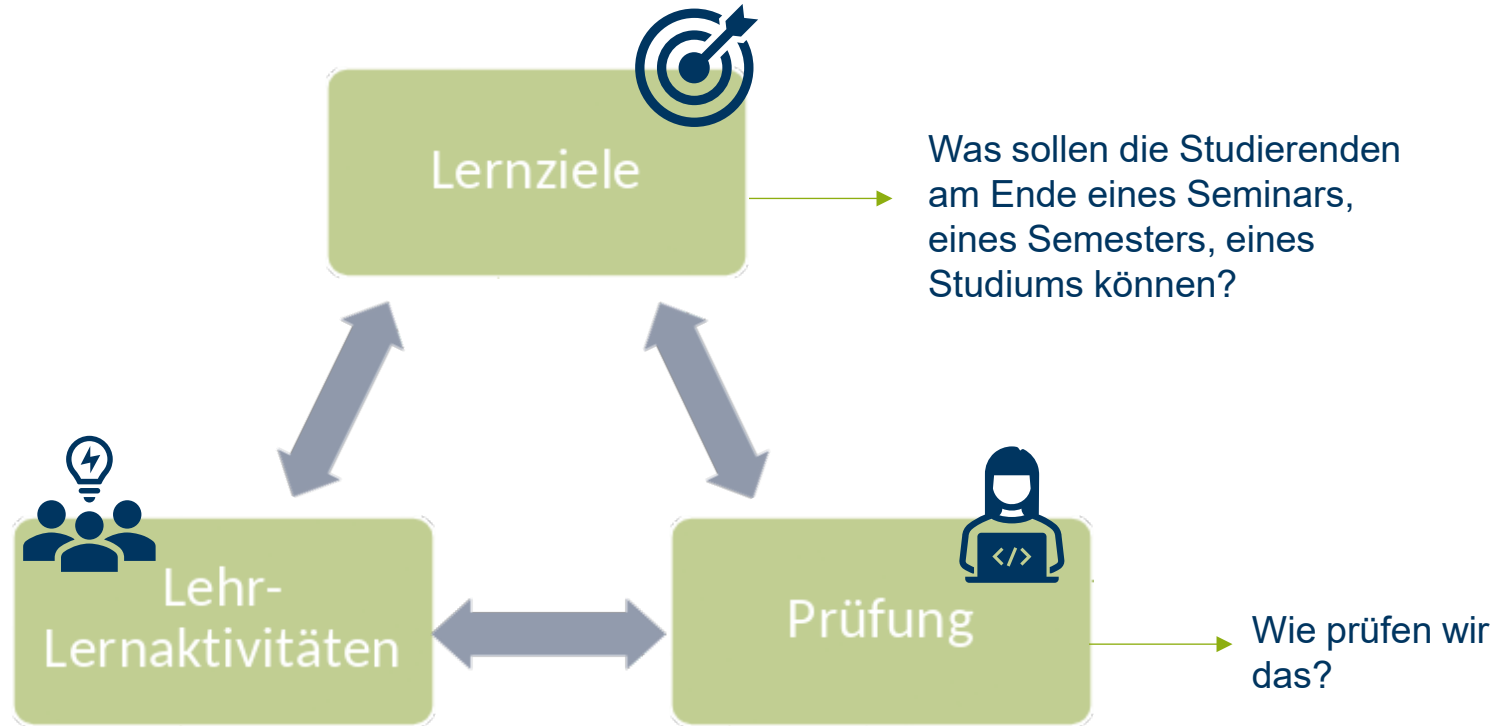
☐ Hausarbeit

☐ Mündliche Prüfung

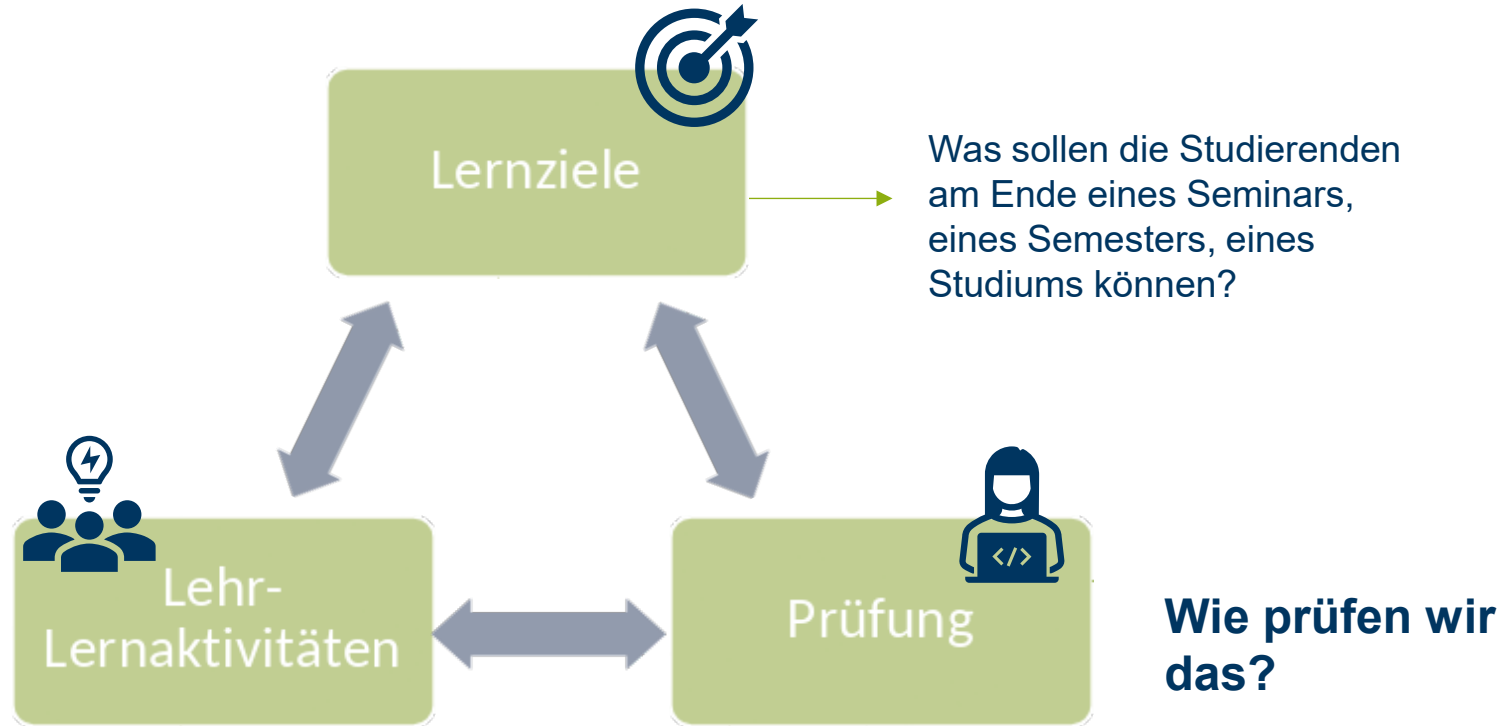
☐ Sonstiges: _____

Welche Lernziele will ich mit der Prüfung abprüfen?

Back to the Basics: Constructive Alignment



Back to the Basics: Constructive Alignment



In Prüfungen bewerten wir die Fähigkeit der Studierenden, wissenschaftlich zu handeln.

In Prüfungen bewerten wir die Fähigkeit der Studierenden, wissenschaftlich zu handeln.

Was wird bewertet?	Das Ergebnis?	Wie viel Toleranz für Fehler ist in Ordnung?
	Der Lernzuwachs?	Mache ich einen sozialen oder einen individuellen Vergleich?
	Der Prozess?	Muss das Ergebnis stimmen, oder reicht der richtige Weg?
Wie gewährleiste ich die Vergleichbarkeit?		
Welche Bewertungskriterien lege ich an?		

Schritt 3 & 4: Bewertung und Vergleichbarkeit

Was bewerte ich?

- ☐ Ergebnis → ☐ Wie viel Toleranz für Fehler ist in Ordnung? _____%
- ☐ Lernzuwachs → ☐ Sozialer Vergleich ☐ Individueller Vergleich
- ☐ Prozess → ☐ Ergebnis muss stimmen ☐ Lösungsweg muss richtig sein

Wie gewährleiste ich Vergleichbarkeit?

- ☐ Kriterienraster ☐ Musterlösung ☐ Kennzeichnung von LLM-Nutzung
- ☐ Sonstiges: _____

Beispiel Paraphrasieren: Mögliche Kriterien

■ Ohne LLM (Beispiel):

	Erwartungsniveau +	Erwartungsniveau 0	Erwartungsniveau -
Kriterium: Paraphrasieren	Zitate wurden vollständig mit passenden Worten (inkl. Synonymen) paraphrasiert und somit eindeutig verständlich wiedergegeben.	Zitate wurden überwiegend mit passenden Worten (inkl. Synonymen) paraphrasiert und somit zumeist verständlich wiedergegeben.	Zitate wurden nicht mit passenden Worten paraphrasiert und somit nicht adäquat wiedergegeben.

■ Mit LLM (Beispiel):

	Erwartungsniveau +	Erwartungsniveau 0	Erwartungsniveau -
Kriterium: Prompt formulieren für Paraphrase	Es wurden mehrere Eingaben probiert und die passende Formulierung begründet verwendet.	Es wurden mehrere Eingaben probiert ohne Begründung der Auswahl.	Es wurden keine alternativen Eingaben ausprobiert.
Kriterium: Paraphrase kritisch betrachten	Alle durch das KI-Tool generierten Phrasen wurden in ihrer Passung zum Original eingeordnet und bewertet.	Einige der durch das KI-Tool generierten Phrasen wurden auf ihre Passung hin betrachtet.	Keine der durch das KI-Tool generierten Phrasen wurde kontextualisiert und hinterfragt.

Schritt 5, 6 & 7: Kriterien, Passung & Leistung

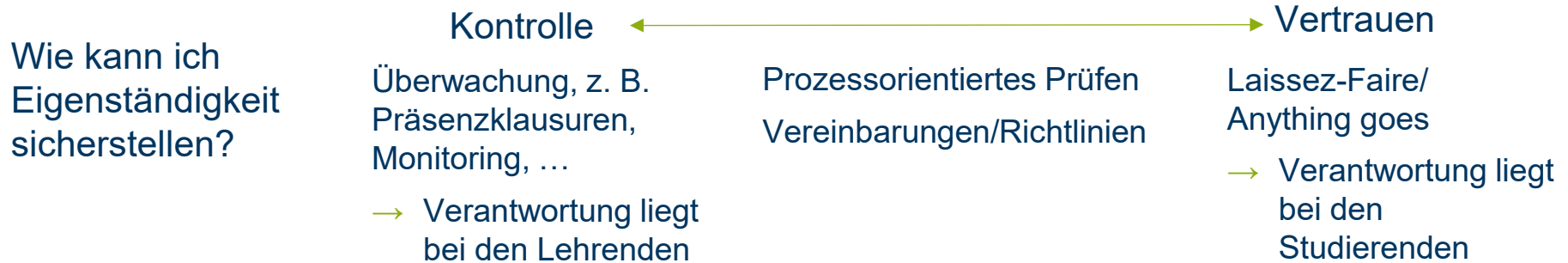
Welche Bewertungskriterien lege ich an?

Passen meine Bewertungskriterien zu den Lernzielen?

☐ Ja, komplett ☐ Je in, in Teilen ☐ Nein, noch nicht

Was will ich, dass die Studierenden wirklich selbst zeigen?

In Prüfungen bewerten wir die Fähigkeit der Studierenden, wissenschaftlich zu handeln.



Schritt 8 & 9: Hilfsmittel, Kennzeichnung, Vertrauen vs. Kontrolle

Wo sind Hilfsmittel egal?

Wie können Hilfsmittel ggf. gekennzeichnet werden?

Wie sehr will ich die Studierenden kontrollieren, wie sehr ihnen vertrauen?

Kontrolle ←————→ Vertrauen

Alternative Prüfungsformate

Ebene I: Kategorisierung von Prüfungsformaten (mit generischen Bezeichnungen z.B. für Prüfungsordnungen)						
Prüfungsmodus	mündlich		schriftlich		(forschungs-)praktisch	
	Mündlich ist die Prüfung aus der Studierenden-Perspektive Der Prüfling ist in einer zuhörenden Rolle „Sprechwerk“ als Prüfungsgegenstand		Schriftlich ist die Prüfung aus der Studierenden-Perspektive Der Prüfling ist in einer lesenden Rolle „Schriftwerk“ als Prüfungsgegenstand		Praktisch ist die Prüfung aus der Studierenden-Perspektive Der Prüfling ist in einer (m)echanischen Rolle „Handwerk“ als Prüfungsgegenstand	
Prüfungsformat	dialogisch	monologisch	in Präsenz	in Absenz	situativ	materialisiert
	Prüfender und Studierender sprechen miteinander Frage-Antwort-Sequenzen	Studierender spricht und Prüfender hört zu Vortragsituation	Prüfung festgelegt auf Zeitpunkt und Ort synchron, zentralisiert	Prüfung festgelegt auf Zeitraum und Form asynchron, dezentral	Studierender demonstriert handelnd Wissen/Können Handlungsprozess	Studierender produziert im Handeln Wissensartefakte Handlungsergebnis
Bezeichnung	Gespräch	Vortrag	Klausur	Hausarbeit	Demonstration	Produktion

Ebene II: Ausgestaltung von Prüfungsformen (in kreativer Vielfalt mit Erläuterung im Modul- und Veranstaltungskontext)						
	Gespräch	Vortrag	Klausur	Hausarbeit	Demonstration	Produktion
Gestaltungsdimensionen						
Sozialform	individuell, im Zweier-team, in Gruppen	individuell, im Zweier-team, in Gruppen	in der Regel individuell	individuell, im Zweier-team, in Gruppen	individuell, im Zweier-team, in Gruppen	individuell, im Zweier-team, in Gruppen
Ressourcen	keine, ein, mehrere Anker wie Portfolio, Poster etc.	ohne Visualisierung, mit Visualisierung	open book, closed book	Literatur frei wählbar, (teilweise) vorgegeben etc.	unter Feldbedingungen, unter Laborbedingungen	unter Feldbedingungen, unter Laborbedingungen
Medium	in der Realisation, in der Videokonferenz	in der Realisation, in der Audio-, Videokonferenz	Paper-Pencil-Klausur, E-Klausur	als Text, Text-Bild-Kombination, Hypertext	in der Realisation, videografiert	als Text, Bild, Audio, Video, Multimedia
Zweck	Wissenswiedergabe, -anwendung etc.	Wissenskombination, -anwendung, -schaffung etc.	Wissenserkenntnis, -wiedergabe, -anwendung etc.	Wissenskombination, -anwendung, -schaffung etc.	(Handlungs-)Wissenswiedergabe, -anwendung etc.	(Handlungs-)Wissensanwendung, -schaffung etc.
Kontext	auf Literatur basierend, in ein Projekt eingebettet etc.	in der Lehrveranstaltung, lehr-extern im Feld	in der Regel auf Literatur basierend	in der Lehrveranstaltung, für lehr-externes Feld	in der Lehrveranstaltung, lehr-extern im Feld	in der Lehrveranstaltung, für lehr-externes Feld
[...]						
Beispiele	Einzelgespräch auf der Basis von Thesen Einzelgespräch auf der Basis eines Portfolios Gruppengespräch zu einem Team-Projekt Einzelgespräch zu einer Fallgeschichte Videokonferenz-Gespräch zu Auslandsseminar	Kolleg auf der Basis von Literatur Gruppenpräsentation zu einem eigenen Projekt Videografierte Vortrag auf einer Konferenz Einzelpräsentation eines wissenschaftlichen Posters Präsentation in einer Audio-/Videokonferenz	Multiple Choice-Klausur Open Book Klausur mit offenen Fragen Fall-Klausur ohne Ressourcen E-Klausur mit multimedialen Anwendungsaufgaben Themenorientierung mit Auswahlmöglichkeit	Einzelarbeit zu einem selbst gewählten Thema Sammlung und Reflexion eigener Texte (Portfolio) Gruppendokumentation eines Projekts Hausarbeit in Form eines Wikis in Teamarbeit Artikel für eine Fachzeitschrift	Szenische Aufführung Rollenspiel Simulation Ausführung einer Aufgabe unter authentischen Bedingungen Videografierte Ausführung einer Aufgabe	Erstellung eines Audioprodukts Erstellung eines Videoprodukts Gestaltung eines Business Plans Konzeption von Instrumenten für Forschungsprojekte Sammlung mehrerer Artefakte (Portfolio)

Schritt 10: Prüfungsformat

Welche alternativen Prüfungsformen passen zu meinen Lernzielen (und den rechtlichen Vorgaben)?

Zwei Beispiele



Zentrum für
Wissenschafts-
didaktik



Reflexionsblatt

Welche Prüfungsförmung bevorzuge ich?

☒ Klausur

Welche Lernaktivitäten bevorzuge ich?

Die Studierenden sollen in der Lage sein, die Lernaktivitäten in Beschreibungen zu ordnen.

Prüfungsförmung

Was bewerte ich?

☒ Ergebnis

o Wie viel Toleranz für Fehler ist in Ordnung? 0 % (Klausur mit geschlossenen Fragen)

o Lernzuwachs

o Sozialer Vergleich

o Individueller Vergleich

o Prozess

o Ergebnis muss stimmen

o Lösungsweg muss richtig sein

Wie gewährleiste ich Vergleichbarkeit?

☐ Kriterienraster

☒ Musterlösung

☐ Kenn...

o Sonstiges:

Welche Bewertungskriterien lege ich an?

Klausurthemen sind inhaltsvalid, d.h. sie decken

Jede Klausuraufgabe erfasst eindeutig, ob ein theoretisches

von anderen, verwandten, oder leicht verwandtbaren Kon-

zum Beispiel: Eigenschaften komplexer Probleme: Untersche-

Passen meine Bewertungskriterien zu den Lernzie-

o Ja, komplett

☒ Ja, in Teilen

o Nein, no

Was will ich, dass die Studierenden wirklich selbst

Zuordnung von Fachbegriffen anhand von praxis-

Gedächtnis, d.h. ohne externe Wissensspeicher.

Wie sind Hilfsmittel egal?

Übersetzung der Fragen und Antwortoptionen in

Wie können Hilfsmittel ggf. gekennzeichnet werden?

Die Prüfungsumgebung (Moodle-Exam) erlaubt k

Wie sehr will ich die Studierenden kontrollieren, v

Kontrolle

Welche alternativen Prüfungsformen passen zu me

Vorgaben?

Studierende dürfen laut Prüfungsordnung zwischen

mündliche Prüfung, Hausarbeit. In der Regel wird ein

ist selten; Hausarbeit wird nie gewählt, da diese am au

stellt (teilweise Studie planen und durchführen).

E-Klausur

- Ideal nicht erreichbar (z. B. aufgrund externer Vorgaben, anderer Prioritäten, ...)
- Lernzielpassung: „Jain“
- Aber:
- Anlass zum Überdenken von Prüfungen
- Vergewisserung über Ziele
- Identifizieren von Problemen



Zentrum für
Wissenschafts-
didaktik



Reflexionsblatt zu Prüfungen

Welche Prüfungsförmung bevorzuge ich?

☐ Klausur

Welche Lernaktivitäten bevorzuge ich?

(1) Studie

(2) Studie

(3) Studie

(4) Studie

Was bewerte ich?

☒ Ergebnis

o Wie viel Toleranz für Fehler ist in Ordnung? 0 %

o Lernzuwachs

o Sozialer Vergleich

o Individueller Vergleich

o Prozess

o Ergebnis muss stimmen

o Lösungsweg muss richtig sein

Wie gewährleiste ich Vergleichbarkeit?

☐ Kriterienraster

☒ Musterlösung

☐ Kenn...

o Sonstiges:

Welche Bewertungskriterien lege ich an?

Klausurthemen sind inhaltsvalid, d.h. sie decken

Jede Klausuraufgabe erfasst eindeutig, ob ein theoretisches

von anderen, verwandten, oder leicht verwandtbaren Kon-

zum Beispiel: Eigenschaften komplexer Probleme: Untersche-

Passen meine Bewertungskriterien zu den Lernzie-

o Ja, komplett

☒ Ja, in Teilen

o Nein, no

Was will ich, dass die Studierenden wirklich selbst

Zuordnung von Fachbegriffen anhand von praxis-

Gedächtnis, d.h. ohne externe Wissensspeicher.

Wie sind Hilfsmittel egal?

Übersetzung der Fragen und Antwortoptionen in

Wie können Hilfsmittel ggf. gekennzeichnet werden?

Die Prüfungsumgebung (Moodle-Exam) erlaubt k

Wie sehr will ich die Studierenden kontrollieren, v

Kontrolle

Welche alternativen Prüfungsformen passen zu me

Vorgaben?

Studierende dürfen laut Prüfungsordnung zwischen

mündliche Prüfung, Hausarbeit. In der Regel wird ein

Hausarbeit (Studienbesprechung)

es gibt viele Punkte, die kritisch diskutiert werden können. Nicht alle können erschöpfend beschrieben werden.

Was bewerte ich?

☒ Ergebnis

o Wie viel Toleranz für Fehler ist in Ordnung? 0 %

o Lernzuwachs

o Sozialer Vergleich

o Individueller Vergleich

o Prozess

o Ergebnis muss stimmen

o Lösungsweg muss richtig sein

Wie gewährleiste ich Vergleichbarkeit?

☐ Kriterienraster

☒ Musterlösung

☐ Kenn...

o Sonstiges:

Welche Bewertungskriterien lege ich an?

Klausurthemen sind inhaltsvalid, d.h. sie decken

Jede Klausuraufgabe erfasst eindeutig, ob ein theoretisches

von anderen, verwandten, oder leicht verwandtbaren Kon-

zum Beispiel: Eigenschaften komplexer Probleme: Untersche-

Passen meine Bewertungskriterien zu den Lernzie-

o Ja, komplett

☒ Ja, in Teilen

o Nein, no

Was will ich, dass die Studierenden wirklich selbst

Zuordnung von Fachbegriffen anhand von praxis-

Gedächtnis, d.h. ohne externe Wissensspeicher.

Wie sind Hilfsmittel egal?

Übersetzung der Fragen und Antwortoptionen in

Wie können Hilfsmittel ggf. gekennzeichnet werden?

Die Prüfungsumgebung (Moodle-Exam) erlaubt k

Wie sehr will ich die Studierenden kontrollieren, v

Kontrolle

Welche alternativen Prüfungsformen passen zu me

Vorgaben?

Studierende dürfen laut Prüfungsordnung zwischen

mündliche Prüfung, Hausarbeit. In der Regel wird ein

mündliche Prüfung; eingeschränkt, da es ein Nebenziel der Hausarbeit ist, Argumente sprachlich/textuell präzise und nachvollziehbar darzustellen.

Rückmeldungen

- „Ich hab bisher unterschätzt, wie wichtig Lernziele sind.“
- „Danke für die strukturierte Herangehensweise, es macht so viel Sinn, sich über das alles klar zu werden.“
- „ganz ehrlich, es verändert sich doch eigentlich nichts, auch nicht mit KI“
- „Für mich ist das gerade frustrierend, weil ich die Prüfungsform nicht verändern darf. Als wir noch keine Klausur in dem Fach hatten, passten die Lernziele besser.“

Schlussfolgerungen

- keine einfachen Fragen, altbekannte Fragen
- strukturierte Reflexion, aber auch Entscheidungen notwendig
- Lernziele explizieren
- Probleme sind wie immer Rahmenbedingungen (PO etc.)
- offen: Wie kann man bei E-Klausuren mit automatischer Auswertung höhere Taxonomiestufen abprüfen?
- etablierte Prüfungsformen können und sollten hinterfragt werden
- Fachkulturen sind wichtig

Schritt für Schritt-Anleitung

1. Prüfungsform (Klausur, Hausarbeit, mündliche Prüfung)
2. Lernziele
3. Bewertung (Ergebnis, Lernzuwachs, Prozess; Fehlertoleranz)
4. Vergleichbarkeit (Kriterienraster, Musterlösung, Kennzeichnung LLM, Sonstiges)
5. Bewertungskriterien
6. Passung Bewertungskriterien und Lernziele (ja, jein, nein)
7. Leistung Studierender
8. Hilfsmittel & Kennzeichnung von Hilfsmitteln
9. Kontrolle bis Vertrauen
10. alternative Prüfungsformen für Lernziele

 Zentrum für
Wissenschafts-
didaktik



Reflexionsblatt zu Prüfungen

Welche Prüfungsform habe ich? *(Bitte eine auswählen)*
☐ Klausur ☐ Hausarbeit ☐ Mündliche Prüfung ☐ Sonstiges: _____

Welche Lernziele will ich mit der Prüfung abprüfen?

Was bewerte ich?
☐ Ergebnis ➡ ☐ Wie viel Toleranz für Fehler ist in Ordnung? _____ %
☐ Lernzuwachs ➡ ☐ Sozialer Vergleich ☐ Individueller Vergleich
☐ Prozess ➡ ☐ Ergebnis muss stimmen ☐ Lösungsweg muss richtig sein

Wie gewährleiste ich Vergleichbarkeit?
☐ Kriterienraster ☐ Musterlösung ☐ Kennzeichnung von LLM-Nutzung
☐ Sonstiges: _____

Welche Bewertungskriterien lege ich an?

Passen meine Bewertungskriterien zu den Lernzielen?
☐ Ja, komplett ☐ Jein, in Teilen ☐ Nein, noch nicht

Was will ich, dass die Studierenden wirklich selbst zeigen?

Wo sind Hilfsmittel egal?

Wie können Hilfsmittel ggf. gekennzeichnet werden?

Wie sehr will ich die Studierenden kontrollieren, wie sehr ihnen vertrauen?
Kontrolle ←————→ Vertrauen

Welche alternativen Prüfungsformen passen zu meinen Lernzielen (und den rechtlichen Vorgaben)?

Hilfreiche Links

Becker, J. (2019): Lernzieltaxonomien im Vergleich. <https://doi.org/10.60806/2c73-4081>

Becker, J. (2022): Schemata für Prüfungsaufgaben. <https://doi.org/10.60806/0nf4-9h34>

im Brahm, G. (2012): Operationalisierung von Lernzielen.
<https://doi.org/10.60806/z70c-az71>

Lordick, N. (2023). Textgenerierende Technologien in der Lehre.
<https://doi.org/10.60806/9vcn-1d75>

Limburg, A. (2015). Leistungsbeurteilung. <https://doi.org/10.60806/3x7e-rj02>

Reinmann, G. (2014): Prüfungen – einfach in der Kategorisierung – komplex in der Ausgestaltung. https://gabi-reinmann.de/wp-content/uploads/2014/05/Pr%C3%BCfungsformen_Mai2014.pdf

**Vielen Dank für die
Aufmerksamkeit!**

Nadine Lordick & Julia Becker

Ruhr-Universität Bochum
Zentrum für Wissenschaftsdidaktik (ZfW)

nadine.lordick@rub.de
j.becker@rub.de

**ZfW-Blog
abonnieren**

