

Kompetenzorientiertes Prüfen im Projekt DigiTeLL

*Handreichung der Beratungsstruktur zum
Querschnittsthema „Kompetenzorientiertes Prüfen/E-
Prüfungen“*

Prof. Dr. Andreas Frey

Cosima Schenk

Arbeitseinheit

Pädagogische Psychologie
mit dem Schwerpunkt Beratung,
Diagnostik und Evaluation

05. September 2023

Was genau versteht man unter „kompetenzorientiertem Prüfen“?

Seit Inkrafttreten der Bologna-Reform besteht die Forderung an die Hochschullehre, den Fokus auf die Vermittlung von Kompetenzen zu legen. Studierende sollen also im Rahmen ihres Studiums nicht nur fachspezifisches Wissen erwerben, sondern auch dazu in der Lage sein, fachspezifische Anforderungen auf Basis des erworbenen Wissens zu lösen. Unter Kompetenzen werden dabei kognitive (das Denken betreffende) Leistungsdispositionen verstanden.

Das Prinzip des Constructive Alignments (Biggs, 1996) spielt im Hinblick auf Prüfungen eine wichtige Rolle. Hierbei handelt es sich um ein didaktisches Konzept, das vorsieht, dass Lernziele, Lernaktivitäten und Prüfungen aufeinander abgestimmt sein sollen (siehe Handreichung zu Constructive Alignment auf der DigiTeLL Website). Bezogen auf Prüfungen stellt sich daher die Frage, wie diese gestaltet sein müssen, damit die in den Lernzielen beschriebenen und durch die Lernaktivitäten gefestigten Kompetenzen tatsächlich abgefragt werden.

Die vorliegende Handreichung gibt einen Überblick über die einzelnen Schritte, die für die Gestaltung kompetenzorientierter Prüfungen relevant sind. Die Ausführungen basieren auf dem Buch von Spoden und Frey (2021), in dem die Thematik umfassend besprochen wird. Die folgenden Schritte werden im Folgenden erläutert:

1. Spezifikation des Messgegenstandes
2. Entwicklung von Klausuraufgaben
3. Setzen von Notenstufen und Bestehensgrenzen
4. Auswertung von Klausuren

Anschließend wird ein kurzer Ausblick auf die Methode des adaptiven Testens vorgenommen; dabei wird die Auswahl der Items an bisherige Antworten der Studierenden angepasst, sodass diejenigen Items ausgewählt werden, die das höchste Ausmaß diagnostischer Information bieten. Diese hoch effiziente Art des Testens kann optional bei kompetenzorientierten Prüfungen eingesetzt werden. Ein geeignetes Hilfsmittel hierfür ist die KAT-HS-App (Fink et al., 2021), welche im Rahmen des Projekts „Kriteriumsorientiertes adaptives Testen in der Hochschule“ (KAT-HS) entwickelt wurde und kostenlos auf der Website <https://kat-hs.uni-frankfurt.de/> heruntergeladen werden kann.

Das Dokument enthält zuletzt auch ein Begriffsglossar, in dem die im Zusammenhang mit kompetenzorientierten Prüfungen wichtigsten Begriffe kurz erläutert werden.

1. Spezifikation des Messgegenstandes

Worum geht es?

In einem ersten Schritt ist es wichtig, sich konkrete Lernziele zu überlegen und diese in Kompetenzanforderungen zu überführen. Hierbei werden Kompetenzen als kognitive Leistungsdispositionen verstanden (Klieme & Leutner, 2006); das bedeutet, dass kompetenzorientierte Hochschullehre fachspezifische Fähigkeiten vermitteln sollte, mit denen konkrete Handlungsanforderungen erfolgreich bewältigt werden können. Wichtig ist hierbei zu beachten, dass die erwähnten Handlungsanforderungen sich nicht nur auf das konkrete Wissen beziehen, das Studierende erwerben sollen, sondern auch höhere kognitive Prozesse umfassen, die dann als Grundlage von konkretem Handeln dienen.

Tipps zur Umsetzung

Bei der Formulierung konkreter Kompetenzanforderungen ist es nützlich, die Gesamtkompetenz zu unterteilen. Die Gesamtkompetenz ist also die Summe der Teile. Für diese Teile können nun Teillernziele formuliert werden. Hierbei sollten sowohl die inhaltlichen Themenbereiche (z.B. Kapitel bei einer Vorlesung) abgedeckt werden als auch die erwarteten kognitiven Prozesse. Bei der Differenzierung der kognitiven Prozesse kann man sich an der Taxonomie kognitiver Lernziele nach Bloom (1984) orientieren. Diese umfasst die kognitiven Prozesse Wissen, Verstehen, Anwendung, Analyse, Synthese und Evaluation. Eine detaillierte Beschreibung der Taxonomie sowie weiterführender Versionen der Taxonomie ist bei Armstrong (2010) zu finden.

Für jeden Inhaltsbereich kann überlegt werden, welche kognitiven Prozesse man von erfolgreichen Studierenden am Ende der Lehrveranstaltung erwartet. Die gemeinsame Darstellung von Inhaltsbereichen und kognitiven Prozessen kann übersichtlich mit einer Messgegenstandsmatrix dargestellt werden. Da man typischerweise nicht für jeden Inhaltsbereich jeden kognitiven Prozess erwartet (z.B. reicht es bei Grundlagen eines Themenbereichs meistens schon aus, über Wissen zu verfügen und die Inhalte verstanden zu haben), sind in der Regel auch nicht alle Zellen einer Messgegenstandsmatrix besetzt. Das Kompetenzziel der Lehrveranstaltung ist die Summe der besetzten Zellen. Durch die Angabe von Prozentwerten können bei Bedarf noch Gewichtungen der einzelnen Aspekte vorgenommen werden. Für alle gefüllten Zellen der Messgegenstandsmatrix sind Operationalisierungen in Form von Prüfungsaufgaben zu formulieren. Ein Beispiel für eine Messgegenstandsmatrix für eine Vorlesung im Bereich Forschungsmethoden ist in der folgenden Abbildung 1 dargestellt.

Inhaltsbereich	Kognitiver Prozess						Summe
	Wissen (40 %)	Verstehen (40%)	Anwendung (20%)	Analyse (0%)	Synthese (0%)	Evaluation (0%)	
Geschichte der Psychologie (10 %)	5	5	-	-	-	-	10
Wissenschaftstheorien (20 %)	10	10	-	-	-	-	20
Psychologische Methoden (30 %)	15	15	-	-	-	-	30
Experiment (20 %)	5	5	10	-	-	-	20
Grundlagen des Messens (20%)	5	5	10	-	-	-	20
Summe	40	40	20	0	0	0	100

Abbildung 1. Beispiel für eine Messgegenstandsmatrix für eine Vorlesung im Bereich Forschungsmethoden

2. Entwicklung von Klausuraufgaben

Worum geht es?

Steht fest, welche kognitiven Prozesse in welchen Inhaltsbereichen geprüft werden sollen, können entsprechende Klausuraufgaben entwickelt werden. Diese sollen Rückschlüsse darüber erlauben, ob bzw. in welchem Ausmaß das Teillernziel der betreffenden Zelle erreicht wurde. Eine Aufgabe bezieht sich immer auf genau eine Zelle der Messgegenstandsmatrix. In dieser Handreichung soll insbesondere die Entwicklung von Multiple Choice (MC) Fragen im Vordergrund stehen. Diese bieten den Vorteil, dass sie von Studierenden i.d.R. sehr schnell bearbeitet werden können, weshalb innerhalb der vorgegebenen Zeit der Prüfung mehr Inhaltsbereiche geprüft werden können. Wichtig dabei ist, dass die öfters formulierte Aussage, dass mit MC-Fragen nur Wissen abgeprüft werden kann, falsch ist. Durch zielgerichtete Aufgabenkonstruktion lassen sich auch höhere kognitive Prozesse mit MC-Aufgaben prüfen (siehe auch das Dokument „Empfehlungen zur Konstruktion von Multiple-Choice-Testaufgaben für höhere kognitive Prozesse“ der Beratungsstruktur „Kompetenzorientiertes Prüfen/E-Prüfungen“ auf der DigiTeLL Website). MC-Items bestehen aus einem Stimulus (einem kurzen Beschreibungstext zu Beginn, über den die Studierenden erfahren, wonach genau gefragt wird), einer richtigen Antwortmöglichkeit sowie mehreren Distraktoren (falschen Antwortmöglichkeiten). Letztere sollten zwar hinsichtlich der Formulierung (also bzgl. der Grammatik, Länge und des Stils) der richtigen Antwortmöglichkeit ähneln, aber inhaltlich eindeutig als falsch identifizierbar sein. Der Name Multiple Choice resultiert aus dem Sachverhalt, dass aus mehreren Antwortalternativen ausgewählt werden kann. Der teilweise vorzufindende Begriff „Single Choice“ wird im vorliegenden Text nicht verwendet.

Tipps zur Umsetzung

Ausgehend von der bereits vorgestellten Messgegenstandsmatrix sollten Items so formuliert werden, dass sie sich zum einen auf eine bestimmte Inhaltsdimension, zum anderen auf einen bestimmten kognitiven Prozess beziehen. Wichtig ist hierbei, dass – falls sich ein Item auf einen höheren kognitiven Prozess bezieht – eine voraussichtlich für die zu testende Person unbekannte Aufgabenstellung beschrieben wird. Sonst wäre lediglich ein Erinnern und somit ein einfacher Wissensabruf zum Lösen nötig. Abbildung 2 zeigt beispielhaft, wie ein Item, das sich auf einen höheren kognitiven Prozess bezieht (in diesem Fall den Prozess Anwendung, bezogen auf Inhalte der deskriptiven Statistik), aussehen kann.

Sie haben von fünf Personen das Alter erhoben und erhalten folgende Werte:

<i>Person A</i>	<i>Person B</i>	<i>Person C</i>	<i>Person D</i>	<i>Person E</i>
<i>21 Jahre</i>	<i>21 Jahre</i>	<i>19 Jahre</i>	<i>18 Jahre</i>	<i>21 Jahre</i>

Berechnen Sie das durchschnittliche Alter der Personen!

- a. 18 Jahre*
- b. 19 Jahre*
- c. 20 Jahre*
- d. 21 Jahre*

Abbildung 2. Beispiel für ein Item, das höhere kognitive Prozesse abdeckt

Für die Entwicklung der konkreten Aufgabenstellung empfiehlt es sich, sich an gängigen Standards zu orientieren, wie sie z.B. in Fink et al. (2021) beschrieben werden. In diesen werden die wichtigsten Punkte genannt, die im Hinblick auf den Inhalt der Items, das Format und den Stil zu beachten sind – beispielsweise, dass keine trivialen Inhalte abgefragt und dass bei der Aufgabenformulierung Verneinungen vermieden werden sollten. Eine vollständige Übersicht über diese Richtlinien ist im Anhang dieser Handreichung zu finden.

3. Setzen von Notenstufen und Bestehensgrenzen

Worum geht es?

Die Aufgabe bei der Bewertung von Klausuren besteht darin, ausgehend von den Antworten auf die vorgelegten Items zu ermitteln, inwieweit eine Studentin oder ein Student über die für die betreffende Lehrveranstaltung geforderten Kompetenzen verfügt. Das Ergebnis muss somit am Prüfungsinhalt (repräsentiert durch die Messgegenstandsmatrix) und nicht etwa an der Leistung der anderen Studierenden orientiert sein. Auf diese Weise ist es möglich, ausgehend von dem Testergebnis darauf zu schließen, in welchem Ausmaß die kompetenzorientierten Lernziele tatsächlich erreicht wurden. Bei einem solchen kriteriumsorientierten Vorgehen sind lediglich Grenzwerte festzulegen, die beispielsweise „Bestehen“ und „Nicht Bestehen“ oder verschiedene Noten voneinander trennen. Das kann beispielsweise bedeuten, dass bei einer Prüfung, bei der lediglich zwischen Bestehen und Nicht-Bestehen unterschieden wird, nur ein Grenzwert festgelegt wird, der angibt, welche Fähigkeiten Studierende innehaben, wenn sie die Prüfung bestehen sollten. Sollten bei der Prüfung verschiedene Notenstufen erreicht werden können, ergeben sich entsprechend auch mehrere inhaltliche Kategorien und mehrere Grenzwerte, die zuvor bestimmt werden müssen.

Tipps zur Umsetzung

Das Bestimmen der Grenzwerte kann über ein sogenanntes Standard-Setting erfolgen. An diesem sollten mindestens zwei fachkundige Personen teilnehmen, von denen eine die Prüferin bzw. einer der Prüfer ist. Für die Items sollten vorab auf Basis der Antworten der Studierenden bei der Klausur die Itemschwierigkeiten berechnet worden sein. Dies kann bei Verwendung der bereits genannten KAT-HS-App auf Basis der Item-Response-Theory erfolgen (was im Hinblick auf kriteriumsorientierte Testwertinterpretationen optimal ist) oder als prozentualer Anteil korrekter Antworten berechnet werden. Ausdrücke der Items werden dann nach Schwierigkeit aufsteigend sortiert. Die Aufgabe der fachkundigen Personen besteht nun darin, beginnend vom einfachsten Item jenes zu identifizieren,

das die Studierenden mindestens mit einer Wahrscheinlichkeit von 50% lösen können müssen, um die Klausur zu bestehen. Dieses Item markiert die Grenze zwischen bestanden und nicht bestanden. Analog kann für andere Grenzen zwischen Notenstufen vorgegangen werden. Bei der Verwendung der Item-Response-Theory können diese inhaltlichen Grenzen auch bei Folgeklausuren wieder genutzt werden. Hierbei können auch andere Klausuritems zum Einsatz kommen, solange einige Items der vorherigen Klausur zur Verbindung der Jahrgänge eingesetzt werden. Das Vorgehen wird durch die KAT-HS-App unterstützt. Wenn die Klausur nicht auf Basis der Item Response Theory ausgewertet wird, muss das Setzen der Grenzwerte bei jeder Klausur aufs Neue erfolgen. Eine differenzierte Beschreibung des Vorgehens basierend auf der Item Response Theory findet sich bei Frey (2021).

4. Auswertung von Klausuren

Worum geht es?

Den letzten Schritt stellt schließlich die Auswertung von Klausuraufgaben dar. Dieser Schritt bestand herkömmlicherweise in der Auszählung korrekter Antworten. Dieses Vorgehen ist bei Hochschulklausuren allerdings nicht angemessen, da nur aufgrund der Summe korrekter Antworten keine inhaltlichen Aussagen darüber abgeleitet werden können, in welchem Ausmaß eine Studentin oder ein Student über die geforderte Kompetenz verfügt. Hierfür ist das oben beschriebene kriteriumsorientierte Vorgehen zu nutzen. Dieses Vorgehen erlaubt Rückschlüsse darüber, was Studierende tatsächlich können. Dass solche Rückschlüsse auf Basis der Summe oder des prozentualen Anteils korrekter Antworten nicht möglich sind, wird direkt klar, wenn man sich das extreme Beispiel einer sehr einfachen Klausur vorstellt, bei der fast alle Personen 100% der Aufgaben richtig beantworten. Der hohe Prozentwert heißt nicht, dass alle Personen über die geforderten Kompetenzen verfügen. Dies *könnte* der Fall sein, es könnte aber auch an den sehr leichten Aufgaben liegen. Mit mathematischen Modellen der Item Response Theory ist es möglich, die Personenfähigkeit von der Aufgabenschwierigkeit zu trennen. Dies ermöglicht einen Rückschluss von Itemantworten auf die gemessene Kompetenz, unabhängig von den eingesetzten Items.

Tipps zur Umsetzung

Die Auswertung von Klausuren auf Basis der Item Response Theory kann mit der speziell für diesen Zweck an der Goethe-Universität entwickelten KAT-HS-App erfolgen, welche kostenfrei über die Website des KAT-HS-Projekts (<https://kat-hs.uni-frankfurt.de/>) heruntergeladen werden kann. Die App verfügt über eine intuitive Benutzeroberfläche und kann nach einer Schulung auch zur selbständigen Analyse von Klausurdaten genutzt werden.

Ausblick: Adaptives Testen

Die Nutzung der KAT-HS-App zur Konstruktion, Durchführung und Auswertung von Hochschulklausuren erleichtert nicht nur die Umsetzung von kriteriumsorientierten Klausuren, sie bietet auch eine Funktion zum adaptiven Testen.

Unter adaptivem Testen versteht man ein Testverfahren, bei dem sich die Itemauswahl an dem bisherigen Antwortverhalten einer Person orientiert (Frey, 2020). Anders als bei herkömmlichen Klausuren, bei denen alle Personen eines Testdurchgangs üblicherweise dieselben Aufgaben bearbeiten, liegt bei adaptivem Testen also eine individuelle Auswahl an Klausuraufgaben vor, die sich an den Fähigkeiten einer jeweiligen Person orientiert. Personen, die eine besonders hohe Kompetenz aufweisen, bekommen vergleichsweise schwierige Aufgaben vorgelegt, während Personen, die eine eher niedrige Kompetenz aufzeigen, leichtere Aufgaben zu beantworten haben. Die Klausur endet, wenn ein vorab definiertes Abbruchkriterium erfüllt ist, beispielsweise wenn eine festgelegte Anzahl an Aufgaben bearbeitet wurde oder ein Zeitlimit erreicht wurde. Wichtig ist auch hier, zu beachten, dass die vorläufige Schätzung der individuellen Kompetenz und die Auswahl von Items mit geeignetem Schwierigkeitsgrad die Verwendung von Messmodellen der Item Response Theorie voraussetzt. Zwei Vorteile von adaptivem Testen sind, dass das Verfahren generell sehr präzise misst und eine hohe Messpräzision auch für Personen mit besonders hoher oder niedriger Kompetenz erreicht wird. Während herkömmliche Klausuren i.d.R. nur im mittleren Bereich des Kompetenzspektrums genau messen, kann bei adaptiven Klausuren sichergestellt werden, dass auch in den Randbereichen des Kompetenzspektrums präzise gemessen und zwischen unterschiedlichen Fähigkeitsausprägungen gut differenziert werden kann, ohne eine große Menge an Items darbieten zu müssen.

Noch Fragen?

Bei Unklarheiten oder weiterführenden Fragen zu kompetenzorientierten Prüfungen, adaptivem Testen oder der Umsetzung von Klausuren mithilfe der KAT-HS-App wenden Sie sich gerne an das [Beratungsangebot zum Thema kompetenzorientiertes Prüfen](#).

Begriffsglossar

<i>Item</i>	Frage bzw. Aufgabe in einem Test/einer Prüfung
<i>Item Response Theory</i>	Testtheorie, bei der die Schwierigkeit eines Items und die Fähigkeit einer Person in direkten Bezug zueinander gebracht werden können, Grundlage für kompetenzbezogene Klausurinterpretationen
<i>Itemschwierigkeit</i>	der Grad der Schwierigkeit, den eine einzelne Testaufgabe für die Teilnehmenden des Tests darstellt
<i>Kompetenz</i>	kontextspezifische kognitive Leistungsdisposition, die sich auf konkrete Situationen und Anforderungen in einem bestimmten Anwendungsbereich bezieht
<i>Messgegenstand</i>	Konzept oder Einheit, welche in einem bestimmten Kontext gemessen wird
<i>Personenparameter</i>	die Fähigkeit der getesteten Person, welche mithilfe des jeweiligen Tests erfasst werden soll
<i>Standard-Setting</i>	Verfahren zur Festlegung spezifischer Kriterien, die verschiedene Kompetenzniveaus für eine Beurteilung in einem Test beschreiben

Literatur

- Armstrong, P. (2010). Bloom's taxonomy. *Vanderbilt University Center for Teaching*. Available online: <https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/> (accessed on 5 September 2023)
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education*, 32(3), 347-364. <https://doi.org/10.1007/BF00138871>
- Bloom, B.S. (1984). *Taxonomy of Educational Objectives. The Classification of Educational Goals*. Longman.
- Fink, A., Born, S., Spoden, C., & Frey, A. (2021). Operationalisierung von Lernzielen bei Hochschulklausuren. In C. Spoden & A. Frey (Hrsg.), *Psychometrisch fundierte E-Klausuren für die Hochschule* (S.17-26). Pabst Science Publishers.
- Fink, A., Spoden, C., Frey, A., & Naumann, P. (2021). Kriteriumsorientiertes adaptives Testen mit der KAT-HS-App. *Diagnostica*, 27(2), 110-114. <https://doi.org/10.1026/0012-1924/a000268>
- Frey, A. (2020). Computerisiertes adaptives Testen. *Testtheorie und Fragebogenkonstruktion*, 501-525. https://doi.org/10.1007/978-3-662-61532-4_20
- Frey, A. (2021). Hochschulklausuren als kriteriumsorientierte Testverfahren. In C. Spoden & A. Frey (Hrsg.), *Psychometrisch fundierte E-Klausuren für die Hochschule* (S.39-48). Pabst Science Publishers.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876-903. <https://doi.org/10.25656/01:4493>
- Spoden, C., & Frey, A. (Hrsg.). (2021). *Psychometrisch fundierte E-Klausuren für die Hochschule*. Pabst Science Publishers.

Anhang

Aus: Fink, A., Born, S., Spoden, C., & Frey, A. (2021). Operationalisierung von Lernzielen bei Hochschulklausuren. In C. Spoden & A. Frey (Hrsg.), *Psychometrisch fundierte E-Klausuren für die Hochschule* (S.17-26). Pabst Science Publishers.

Richtlinien zum Verfassen von Aufgaben im Antwort-Wahl-Verfahren

Inhalt

1. Jede Aufgabe sowohl auf eine bestimmte Inhaltsdimension als auch auf einen bestimmten kognitiven Prozess beziehen (Messgegenstandsmatrix)
2. Unbekanntes Material verwenden, um auf komplexere kognitive Prozesse (z.B. Verstehen) zu prüfen
3. Wahrscheinlichkeit, eine Aufgabe zu lösen, unabhängig von der Bearbeitung anderer Aufgaben halten
4. Aufgaben sollen auf wichtigem Lerninhalt basieren; keine trivialen Inhalte prüfen
5. Meinungsbasierte Aufgaben vermeiden
6. Trick-Aufgaben („auf eine falsche Fährte locken“) vermeiden

Stil und Format

7. Aufgaben vertikal anordnen
8. Jede Aufgabe überprüfen und ggf. ändern (Inhalt, Rechtschreibung, Grammatik und Format)
9. Grad an linguistischer Komplexität an Zielgruppe anpassen
10. Lesezeit der Aufgaben minimieren

Aufgabenstellung

11. Zentrale Idee klar und deutlich in Aufgabenstellung darlegen, nicht in Antwortoptionen
12. Aufgabenstellung positiv formulieren

Antwortoptionen

13. Nur plausible und für Aufgabenstellung relevante Antwortoptionen nutzen
 14. Nur eine Antwortoption sollte die richtige Antwort sein
 15. Position der richtigen Antwort in Bezug auf die Anzahl der Antwortmöglichkeiten variieren, jedoch nicht nach einem bestimmten Muster
 16. Antwortoptionen in eine logische oder numerische Reihenfolge bringen
 17. Antwortoptionen unabhängig voneinander halten und Überschneidungen vermeiden
 18. Optionen „Keine der genannten“, „Alle der genannten“ oder „Ich weiß nicht“ vermeiden
 19. Antwortoptionen positiv formulieren, negative Wörter wie „nicht“ vermeiden
 20. Antwortmöglichkeiten, die einen Hinweis auf die Lösung geben, vermeiden
 - a. Antwortoptionen sollten ähnlich lang sein
 - b. Spezifische Determinanten, wie z.B. immer, niemals, komplett oder absolut vermeiden
 - c. Lautassoziationen, also Antwortoptionen, die nicht aufgrund logischer Überlegungen, sondern aufgrund ihrer klanglichen Ähnlichkeit gewählt werden, und Optionen, die identisch bzw. ähnlich zu Worten in der Aufgabenstellung sind, vermeiden
 - d. Paare oder Triplets von Antwortoptionen, die dem Prüfling einen Hinweis auf die korrekte Lösung geben, vermeiden
 - e. Offensichtlich absurde oder lächerliche Antworten vermeiden
 21. Distraktoren (falsche Antwortmöglichkeiten) so plausibel wie möglich formulieren; für Distraktoren typische Fehler der Studierenden verwenden (z.B. Antworten auf offenen Fragen)
 22. Humorvolle Formulierungen in Klausuren vermeiden
-