

Handreichungen zum kompetenzorientierten Prüfen

Inhalt

Überblick	2
Schritt 1: Angestrebte Lernergebnisse als Grundlage für die Prüfungserstellung	3
Schritt 2: Die Erstellung von Prüfungsaufgaben	9
Schritt 3: Abgleich von Prüfungsaufgaben und angestrebten Lernergebnissen	20
Schritt 4: Die Festlegung von Bewertungskriterien	29
Schritt 5: Die Auswertung von Prüfungsergebnissen	36

Überblick

Die Reihe „Handreichungen zum Kompetenzorientierten Prüfen“ beinhaltet ausführliche Hilfestellungen und Anregungen zur Planung, Erstellung, Bewertung und Reflexion von Kompetenzorientierten Prüfungen. Die Handreichungen sind im Rahmen des durch den Stifterverband geförderten Kooperationsprojekt „Herausforderung Prüfen“ entstanden. Die Reihe ist in fünf nacheinander folgende Schritte gegliedert.

Schritt 1 „Angestrebte Lernergebnisse als Grundlage für die Prüfungserstellung“:

Der erste Schritt umfasst die Formulierung angestrebter Lernergebnisse als Ausgangsbasis zur Erstellung kompetenzorientierter Prüfungsfragen. Zu Beginn wird das Konzept „angestrebte Lernergebnisse“ beschrieben sowie deren Bedeutung für die Lehre. Danach fokussiert die Handreichung auf die konkrete Formulierung von angestrebten Lernergebnissen und bietet konkrete Kriterien für gut formulierte Lernergebnisse an.

Schritt 2 „Die Erstellung von Prüfungsaufgaben“:

Im zweiten Schritt steht das Erstellen von Prüfungsaufgaben im Vordergrund. Es wird kurz auf unterschiedliche Aufgabentypen und Antwortformate als Basis eingegangen, wichtige Prinzipien bei der Aufgabenerstellung werden dargestellt, sowie verschiedene Techniken vorgestellt, um offene Aufgaben (d.h. Aufgaben die mehr das Konzeptwissen als das reine Faktenwissen der Studierenden abprüfen) zu konstruieren.

Schritt 3 „Abgleich von Prüfungsaufgaben und angestrebten Lernergebnissen“:

Der dritte Schritt ist das Anwenden eines Verfahrens, mit dem sich überprüfen lässt, inwiefern die Prüfungsaufgaben die zuvor formulierten angestrebten Lernergebnisse repräsentativ abbilden. Dieses Verfahren (logische Aufgabenanalyse) umfasst insgesamt sieben Schritte, die in der Handreichung ausführlich beschrieben werden.

Schritt 4 „Die Festlegung von Bewertungskriterien“:

Neben der Erstellung von kompetenzorientierten Prüfungsaufgaben spielt die Erstellung von Bewertungskriterien eine zentrale Rolle, um objektives, reliables und valides Prüfen überhaupt erst zu erreichen. Dieser vierte Schritt sollte immer parallel mit der Aufgabenerstellung (Schritt 2) erfolgen, wird hier jedoch in einem gesonderten Schritt ausführlich behandelt. Es wird skizziert, wie sich ein umfassendes Bewertungsraster in wenigen Schritten erstellen lässt. Zwei ergänzende Beispiele veranschaulichen diese Schritte und die Bestandteile des Bewertungsrasters.

Schritt 5 „Die Auswertung von Prüfungsergebnissen“:

Der fünfte Schritt besteht aus einer umfassenden Analyse der Prüfungsergebnisse. Deshalb geht dieser Schritt auf wichtige Voraussetzungen ein, die gegeben sein müssen, um solch eine Analyse durchzuführen. Anschließend werden zentrale statistische Kennwerte erläutert und wichtige Hinweise für die Analyse von Prüfungsergebnissen gegeben. Es wird zudem kurz erläutert, welche Rückschlüsse man auf Basis der Ergebnisse auf die Prüfung, einzelne Aufgaben aber auch auf die zugrundeliegende Lehrveranstaltung ziehen kann. Im letzten Abschnitt wird die Erstellung einer Aufgabendatenbank auf Basis von Aufgabencharakteristika beschrieben.

Die aktuelle Fassung der kompletten Sammlung (inklusive Deckblatt und Inhaltsverzeichnis) finden Sie stets unter <http://www.lehren.tum.de/themen/pruefungen/herausforderung-pruefen/>.

Schritt 1: Angestrebte Lernergebnisse als Grundlage für die Prüfungserstellung

Der erste Schritt umfasst die Formulierung angestrebter Lernergebnisse als Ausgangsbasis zur Erstellung kompetenzorientierter Prüfungsfragen. Zu Beginn wird das Konzept „angestrebte Lernergebnisse“ beschrieben sowie deren Bedeutung für die Lehre. Danach fokussiert die Handreichung auf die konkrete Formulierung von angestrebten Lernergebnissen und bietet konkrete Kriterien für gut formulierte Lernergebnisse an.

Was sind eigentlich angestrebte Lernergebnisse?

Unter einem angestrebten Lernergebnis oder angestrebten Lernergebnissen (*intended learning outcomes*) werden im Allgemeinen Aussagen darüber verstanden, was Studierende nach dem Absolvieren eines Studiengangs, eines Moduls oder einer Veranstaltung können sollen. Themen und Inhalte werden dabei ausschließlich in ihrem Anwendungszusammenhang benannt, so dass direkt ersichtlich wird, für welche Kompetenzen und Fähigkeiten die vermittelten Inhalte relevant sind.

Die Bedeutung von angestrebten Lernergebnissen

- Sie zeigen Studierenden die konkreten Ziele eines Studiengangs, eines Moduls oder einer Veranstaltung auf und verdeutlichen deren praktische Relevanz.
- Sie geben den Studierenden eine Orientierung, in welcher Form bestimmte Inhalte am Ende der Veranstaltung überprüft werden.
- Sie dienen dem oder der Lehrenden als Ausgangspunkt für die Erstellung von Prüfungsaufgaben.
- Sie können zu einer höheren Validität der Prüfung führen, da sie zu überprüfende Fähigkeiten und Kompetenzen konkret benennen und damit operationalisierbar machen.
- Auf Basis angestrebter Lernergebnisse können klare Bewertungskriterien für Aufgaben festgelegt werden. Prüfungsergebnisse werden damit reliabler (=zuverlässiger).

Die Formulierung von angestrebten Lernergebnissen

Wie definiert man, was gute Lernergebnisse sind?

In der Hochschullehre sind grundsätzlich die Dozierenden für die curriculare Gestaltung ihrer Veranstaltungen und damit auch die Formulierung von angestrebten Lernergebnissen zuständig. Vorgegeben wird aber mittlerweile, dass angestrebte Lernergebnisse in Form von Kompetenzen formuliert sein sollen, also beschreiben sollen, was Studierende nach Absolvieren der Veranstaltung idealerweise tun können. Diese Vorgabe ist als Hilfestellung für die Explikation von Zielen der eigenen Lehre zu verstehen.

Angestrebte Lernergebnisse sollten demnach auf relevante Anforderungen Bezug nehmen, die nach Abschluss der Veranstaltung im Zusammenhang mit den vermittelten Inhalten auf die Studierenden zukommen können. Für die Identifikation solcher Anforderungen eignen sich zweierlei Herangehensweisen: Entweder können die Anforderungen selbst empirisch eruiert (Primäranalyse) oder bereits bestehende Veröffentlichungen zum jeweiligen Fachbereich gesichtet werden (Sekundäranalyse).

Möglichkeit 1: Primäranalyse von Anforderungen

Inhalt von Angestrebten Lernergebnissen können zum Beispiel (1) *Ausgangsvoraussetzungen von weiterführenden Veranstaltungen im Studium sein.*

Entsprechend können für die Identifikation von relevanten Lernergebnissen Profile des Studiengangs analysiert werden. Ebenso bietet es sich an, sich mit Kolleginnen und Kollegen über die Voraussetzungen von weiterführenden Veranstaltungen auszutauschen. Vor allem wenn Veranstaltungen in Modulen integriert sind (was sie im Idealfall sein sollten), sollten die vermittelten Kenntnisse und Fähigkeiten systematisch aufeinander aufbauen. Existieren für Fähigkeiten und Kompetenzen mehr oder weniger elaborierte Entwicklungsmodelle für einzelne Kompetenzen, so wird unter anderem wirksam verhindert, dass Lerninhalte doppelt vermittelt werden.

Angestrebte Lernergebnisse können – je nach Anspruchsniveau – außerdem an (2) *beruflichen Anforderungen* ausgerichtet sein.

Berufliche Einsatzfelder sind im akademischen Bereich nicht immer ganz eindeutig zu identifizieren, weshalb es nicht ganz einfach ist, relevante Ziele für alle Studierenden zu finden. Dennoch gibt es für viele Inhalte typische Anwendungsbereiche, die auch bei der Formulierung von Lernergebnissen bedacht werden sollten, vor allem dann, wenn sich das Anforderungsniveau zum Beispiel auf die Anwendung von Inhalten bezieht.

Weiterhin können auch Kontakte außerhalb der Universität – etwa an der Schnittstelle von Industrie und Forschung – wichtige Informationen über aktuelle Trends und Anforderungen liefern. Der Vorteil ist, dass durch die Aufnahme aktueller Themen und Anforderungen in den Lernergebnissen nicht nur ein externes Kriterium für die Validität zur Verfügung steht, sondern durch das Relevanz erleben auch die Motivation der Studierenden gesteigert wird.

Möglichkeit 2: Sekundäranalyse von Anforderungen

Angestrebte Lernergebnisse können (1) auf *bestehenden Kompetenzmodellen für einzelne Fachbereiche* aufbauen.

Die Forschung entwickelt und überprüft laufend Modelle zur Struktur und zum Erwerb domänenspezifischer Kompetenzen. Eine Möglichkeit ist daher, zu recherchieren, ob zum jeweiligen Fachbereich nicht bereits geprüfte Modelle vorliegen, an denen man sich für die Definition von angestrebten Lernergebnissen orientieren kann.

Außerdem liefern (2) *Lehrbücher in der Regel Fachsystematiken, die als Grundlage für die Überlegungen zu angestrebten Lernergebnissen dienen können.*

Dabei ist zu beachten, dass die herangezogenen Lehrbücher möglichst aktuell sind und dass die jeweilige Systematik – wenn nicht bereits durch die Autoren geschehen – an den Anspruch einer Kompetenzorientierung angepasst wird, dass also Inhalte nicht nur thematisch, sondern anhand von Anwendungsbereichen etc. systematisiert sind.

Ebenso stellen (3) *übergeordnete Institutionen* immer wieder *Orientierungshilfen für die Formulierung von Kompetenzen* zur Verfügung.

Zum Beispiel liegt mit dem Qualifikationsrahmen für deutsche Hochschulabschlüsse der Hochschulrektorenkonferenz ein domänenübergreifendes Kompetenzprofil vor, auf dessen Grundlage angestrebte Lernergebnisse formuliert werden können. Dabei ist zu beachten, dass es sich hierbei um domänenunabhängige, das heißt, sehr allgemein gehaltene Kompetenzen handelt, die nach Möglichkeiten für die jeweiligen Fachbereiche noch einmal spezifiziert werden sollten.

Die konkrete Formulierung von angestrebten Lernergebnissen mit Hilfe von Leitfragen

Bei der konkreten Formulierung von angestrebten Lernergebnissen kann es hilfreich sein, sich immer wieder an Leitfragen zu orientieren. Diese helfen dabei, Lernergebnisse einerseits inhaltlich zu differenzieren und sich andererseits Gedanken über verschiedene Niveaus innerhalb der einzelnen Dimensionen zu machen.

1. *Was sind die Studierenden nach Durchlaufen des Moduls in der Lage zu tun?*

Mit dieser Frage soll sichergestellt werden, dass die Lernergebnisse auch tatsächlich das unmittelbare „Können“ der Studierenden abbilden. Das Lernergebnis kann dann direkt als Antwort auf diese Frage formuliert werden, etwa: „Die Studierenden können ein aktuelles Fahrzeugkonzept bewerten“.

2. *Können mit den angestrebten Lernergebnissen verschiedene Anforderungsdimensionen (z.B. verschiedene Kompetenzen) eindeutig unterschieden werden?*

Durch die Benennung von Anforderungsdimensionen können angestrebte Lernergebnisse weiter konkretisiert werden. Zum Beispiel: „Die Studierenden können ein aktuelles Fahrzeugkonzept in mindestens drei relevanten Dimensionen bewerten“.

3. *Sind die einzelnen angestrebten Lernergebnisse konkret genug formuliert, um überprüfbar zu sein?*

Mit dieser Leitfrage soll sichergestellt werden, dass auf Basis eines angestrebten Lernergebnisses relativ unmittelbar eine entsprechende Prüfungsaufgabe entworfen werden kann, die dieses Lernergebnis abbildet. Zum Beispiel: „Die Studierenden können ein aktuelles Fahrzeugkonzept in mindestens drei relevanten Dimensionen bewerten, indem sie anhand von zentralen Kriterien dessen Güte beschreiben“.

Was sind häufig gemachte Fehler bei der Formulierung von angestrebten Lernergebnissen?

- Ein angestrebtes Lernergebnis ist *zu allgemein* gehalten, wie zum Beispiel: „Die Studierenden können ingenieurwissenschaftlich relevante Konstruktionsaufgaben lösen.“
- Ein angestrebtes Lernergebnis ist *zu spezifisch*, wie zum Beispiel: „Die Studierenden können mittels der Software x und der Konstruktionsmethode y die Aufgabe der Konstruktion eines Kühlkreislaufes unter den Bedingungen a, b und c innerhalb einer Zeitspanne von n Minuten lösen.“
- Statt eines angestrebten Lernergebnisses wird ein *Lernziel* beschrieben, wie zum Beispiel: „Die Studierenden sollen die Konstrukte der Programmiersprache Java erlernen.“
- Statt eines angestrebten Lernergebnisses wird ein *Lerninhalt* bezeichnet: „In diesem Modul werden Kenntnisse und Fähigkeiten in den Bereichen Approximationsalgorithmen, Backtracking und Branchand-Bound vermittelt.“

Konkrete Kriterien für gut formulierte angestrebte Lernergebnisse

Die folgende Liste enthält einzelne Aspekte, auf die hin sich die angestrebten Lernergebnisse überprüfen lassen. Diese Aspekte können – neben den Leitfragen – auch bereits während des Definitionsprozesses als Hilfestellung dienen.

- ✓ *Die angestrebten Lernergebnisse enthalten eine Inhalts- und eine Verhaltenskomponente:* Die Verhaltenskomponente gibt an, was man an einem bestimmten Inhalt (=Inhaltskomponente) tun bzw. können soll.
- ✓ *Inhalts- und Verhaltenskomponente sind ausreichend spezifisch,* d.h. sie sind so formuliert, dass deren Bedeutung eindeutig ist (und damit messbar). Dadurch kann zweifelsfrei entschieden werden, welche Anforderungen in entsprechenden Aufgaben abgebildet sein müssen.
- ✓ *Die angestrebten Lernergebnisse enthalten einen Beherrschungsgrad* (z.B. in welcher Qualität oder Zeit man die Anforderungen ausführen soll).
- ✓ *Die angestrebten Lernergebnisse sind übersichtlich,* d.h. es sollte nur eine begrenzte Anzahl formuliert werden (max. 8-10). Falls zu viele angestrebte Lernergebnisse formuliert sind, so ist gegebenenfalls eine Abstraktion vorzunehmen (siehe häufig gemachte Fehler).
- ✓ *Die angestrebten Lernergebnisse beinhalten ein konkretes, zweifelsfreies (Aktions-)Verb.* Damit soll sichergestellt sein, dass der oder die Studierende anhand des angestrebten Lernergebnisses erkennen kann, was von ihm oder ihr erwartet wird (statt „wissen xy“ z.B. „können xy nennen“). Zudem kann dadurch aus dem Lernergebnis direkt eine Aufgabenstellung erfolgen.
- ✓ *Die angestrebten Lernergebnisse sind realistisch* (im Blick auf den gegebenen Zeitrahmen und die vorhandenen Ressourcen), das heißt, es gilt zu überlegen, welche Anforderungsstufe realistischer Weise vermittelbar und auch überprüfbar ist. Einige Anwendungsfälle können beispielsweise zu komplex für den Rahmen einer Veranstaltung sein.
- ✓ *Die angestrebten Lernergebnisse sind mit den Inhalten und Zielen des Moduls abgestimmt.* Das heißt, bereits bei der Identifikation von angestrebten Lernergebnissen orientiert man sich zum Beispiel an Folgeveranstaltungen oder vorangegangenen Veranstaltungen.

Allgemeine Prinzipien

- ✓ *Reflexion über Vorkenntnisse und Fähigkeiten.* Damit soll sichergestellt sein, dass angestrebte Lernergebnisse tatsächlich veranstaltungsspezifische Fähigkeiten benennen.
- ✓ *Begründung der Relevanz der angestrebten Lernergebnisse* (siehe Abschnitt zur Analyse von Anforderungen).
- ✓ *Fokus auf Resultate.* Stellt sicher, dass tatsächlich *Ziele* gesetzt und überprüft werden.
- ✓ *Formulierung weder zu allgemein noch zu konkret.* Angestrebte Lernergebnisse stellen einen Mittelweg zwischen Abstraktion auf Ebene zu allgemeiner Fähigkeiten und einer zu konkreten (beispielhaften) Beschreibung einzelner Fähigkeitsaspekte dar.
- ✓ *Formulierung von angestrebten Lernergebnissen sollte nicht mit der Beschreibung von Lerninhalten verwechselt werden.* Angestrebte Lernergebnisse bestehen aus Inhalts- und Verhaltenskomponente.

Hilfsmittel: Die Taxonomie nach Bloom

Taxonomien sind hilfreiche Instrumente in Form von Schemata, um Anforderungen zu klassifizieren. Eine der bekanntesten Taxonomien ist die Taxonomie nach Bloom, Englhart, Furst, Hill & Krathwohl (1956) bzw. – in erweiterter Fassung - von Anderson & Krathwohl (2001; siehe Abb. 1). Sie erlaubt die Klassifizierung von Fähigkeiten nach Erkenntnisstufen, beginnend beim „Erinnern“ bis hin zum „Entwickeln“, wobei jede Stufe die vorausgegangene voraussetzt. Zu den jeweiligen Erkenntnisstufen sind Verben angegeben, die sich eignen, um die jeweilige Erkenntnisstufe in den angestrebten Lernergebnissen möglichst „verhaltensnah“ zu benennen.

Erinnern

abrufen, anführen, angeben, auflisten, aufsagen, aufzählen, ausführen, benennen, bezeichnen, darlegen, darstellen, definieren, erkennen, identifizieren, kennzeichnen, nennen, reproduzieren, sagen, skizzieren, wiederholen, wiedergeben, zeichnen, zitieren

Verstehen

anordnen, anpassen, argumentieren, beschreiben, darstellen, diskutieren, einordnen, erklären, erläutern, generalisieren, illustrieren, identifizieren, klären, klassifizieren, übersetzen, veranschaulichen, vergleichen, zuordnen

Anwenden

ausführen, benutzen, berechnen, bilden, demonstrieren, durchführen, erfassen, erstellen, handhaben, implementieren, lösen, modifizieren, umsetzen, umwandeln, verwenden, voraussagen, zusammenstellen

Analysieren

ableiten, auswerten, aufdecken, auswählen, Bedeutung ermitteln, Beziehungen darstellen, charakterisieren, differenzieren, entdecken, Fehler überprüfen, gliedern, kennzeichnen, organisieren, strukturieren, unterscheiden, zergliedern, verallgemeinern

Bewerten

abstimmen, abwägen, auswerten, beurteilen, diskutieren, einschätzen, entscheiden, ermitteln, evaluieren, gewichten, kommentieren, kritisieren, Stellung nehmen, testen, überprüfen, überwachen, werten

Entwickeln

entwerfen, entwickeln, erfinden, folgern, gestalten, konstruieren, konzipieren, kreieren, planen, produzieren

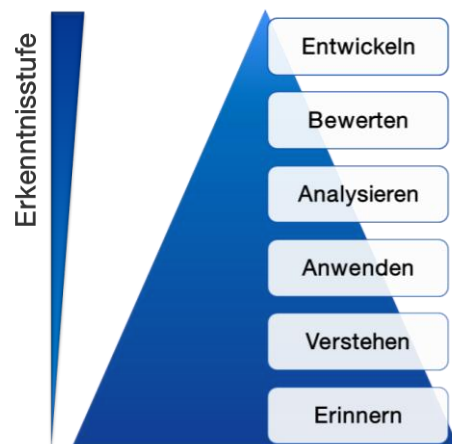


Abbildung 1: Taxonomie nach Bloom, Englhart, Furst, Hill & Krathwohl (1956) adaptiert durch Anderson & Krathwohl (2001)

Literatur

- Anderson, L.W. & Krathwohl, D.R., Hrsg. (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York, Addison-Wesley.
- Bloom, B. S., Enghart, M., Furst, E., Hill, W., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. New York: David McKay.
- Declan, K. (2008). Lernergebnisse (Learning Outcomes) in der Praxis. Ein Leitfaden, Bonn, DAAD.
- Leitfaden zur Formulierung von Lernergebnissen in Modulbeschreibungen (2008). TUM Stabsstelle QM/RW, Fassung vom 17.09.2008.
- Schaper, N. (unter Mitwirkung von Reis, O., Wildt, J., Horvath, E., Bender, E.) (2012). „Fachgutachten zur Kompetenzorientierung in Studium und Lehre“. https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/fachgutachten_kompetenzorientierung.pdf (abgerufen: 25.7.2016)
- Wegweiser zur Erstellung von Modulbeschreibungen (2016). TUM Hochschulreferat Studium und Lehre. https://www.lehren.tum.de/fileadmin/w00bmo/www/Downloads/Themen/Studiengaenge_gestalten/Dokumente/Wegweiser_Modulbeschreibungen_Version3_Stand_Maerz_16.pdf (abgerufen: 14.07.2016)

Schritt 2: Die Erstellung von Prüfungsaufgaben

Nachdem im ersten Schritt beschrieben wurde, wie Lehrende gute Lernergebnisse als Basis für ihre Prüfung formulieren, geht es im zweiten Schritt um die konkrete Erstellung von Prüfungsaufgaben. Es wird kurz auf unterschiedliche Aufgabentypen und Antwortformate als Basis eingegangen, wichtige Prinzipien bei der Aufgabenerstellung werden dargestellt, sowie verschiedene Techniken vorgestellt, um offene Aufgaben (d.h. Aufgaben die mehr das Konzeptwissen als das reine Faktenwissen der Studierenden abprüfen) zu konstruieren.

Funktion und Struktur von Aufgaben

Mit Hochschulprüfungen soll festgestellt werden, welchen Kompetenz- oder Wissenstand Studierende aufweisen bzw. ob und inwieweit sie die angestrebten Lernergebnisse erreicht haben. Die zu vermittelnden Kompetenzen sind dabei meist nicht direkt beobachtbar - man sieht es einem Studierenden also nicht an, ob er eine bestimmte Kompetenz besitzt oder nicht. Daher ist es notwendig, aus dem gezeigten Verhalten eines Studierenden in einer Anforderungssituation Rückschlüsse zu ziehen. Prüfungsaufgaben stellen dabei standardisierte Anforderungssituationen dar, die es Studierenden ermöglichen und sie dazu anregen, das erwartete Verhalten, hinter dem eine Kompetenz vermutet wird, zu zeigen. Mit Prüfungsaufgaben werden also gewissermaßen Verhaltensstichproben gezogen, die möglichst verlässliche Rückschlüsse auf die zugrunde liegende Kompetenz erlauben sollen.

Jede Aufgabe besteht aus zwei Komponenten: dem **Aufgabenstamm** und dem **Antwortformat**. Der Aufgabenstamm stellt ganz allgemein die Anforderungssituation dar, zum Beispiel in Form einer Frage, einer Aussage, eines Bildes, einer Geschichte, Zeichnung oder Rechenaufgabe. Die Anforderungssituationen müssen dabei so gestaltet sein, dass das fragliche Verhalten herbeigeführt wird, Studierende also zum Zeigen des relevanten Verhaltens angeregt und angeleitet werden.

Die Auswahl von Inhalten und Wissensstufen, kognitivem Niveau, Aufgabenoffenheit, Bearbeitungsumfang, und Komplexität sollte dabei stets zielgerichtet auf das angestrebte Lernergebnis erfolgen. Gleichzeitig soll beurteilt werden, in welchem Maße ein bestimmtes Lernergebnis erreicht worden ist. Aufgaben sollten deshalb verschiedene Fähigkeitsniveaus abbilden, um eine differenzierte Bestimmung der erzielten Fähigkeiten zu ermöglichen.

Das Antwortformat dient dagegen zur Registrierung des Testverhaltens. Grundlegend wird bei Tests zwischen offenen (oder freien) und geschlossenen (gebundenen) Antwortformaten unterschieden.

Offenes Antwortformat

Bei einem offenen Antwortformat formuliert die getestete Person selbst eine Antwort in einem allgemein verständlichen Zeichensystem. Die dabei registrierten Antworten müssen auf dieser Basis in Bewertungskategorien einzuordnen sein. Das offene Antwortformat wird vorzugsweise zur Messung von konzeptionellem Verständnis, der Fähigkeit Sachverhalte in einer Domäne schriftlich oder grafisch darzustellen und zu erläutern, sowie für Berechnungs-, Bewertungs-, Analyse- und Problemlöseaufgaben eingesetzt.

Innerhalb der offenen Antwortformate kann zwischen zwei Typen unterschieden werden. Bei sogenannten Kurzwortantwortaufgaben müssen die Studierenden selbstständig eine (kurze) Antwort formulieren. Die Antworten beschränken sich auf das Aufzählen von Begriffen, Aussagen, Definitionen oder Fakten oder erfordern das Lösen von Berechnungsaufgaben. Kurzwortantwortaufgaben müssen so formuliert sein, dass es nur eine richtige, kurze Antwort gibt. Man erwartet eine bestimmte Antwort. Der Studierende muss erkennen welche zuvor gelernte Regel angewendet oder welcher zuvor gelernte Fakt hingeschrieben oder welche zuvor gelernte Lehrbuchdefinition wiedergegeben werden muss.

Bei erweiterten offenen Antwortformaten sind die erwarteten Antwort länger und erfordern mehrere verschiedene Konzepte, Lösungs- oder Konstruktionsschritte, Wissenseinheiten, Argumentationen oder die Bewertung mehrerer Konzepte.

Geschlossenes Antwortformat

Bei geschlossenen Antwortformaten wird demgegenüber eine Auswahl von Antworten (Wahlantwortformat) angeboten, so dass von der Testperson eigens keine Antwort formuliert werden, sondern lediglich aus den gewählten Alternativen gewählt werden muss. Jede Wahlantwort besteht aus einem Aufgabenstamm, einer richtigen Antwort sowie mehreren (idealerweise 2-3) falschen Antworten (sogenannten Distraktoren). Diese Distraktoren sollten dabei so gewählt sein, dass sie attraktive Antwortalternativen darstellen und nicht einfach durch Plausibilitätsannahmen ausgeschlossen werden können.

Wahl eines Aufgabentypus und Antwortformats

In erster Linie sollte ein Aufgabentypus gewählt werden, der zur gestellten Anforderung (also dem dazugehörigen angestrebten Lernergebnis) passt. Es gilt deshalb – insbesondere auch bei der Auswahl eines passenden Antwortformats - darauf zu achten, dass das fragliche Verhalten tatsächlich gezeigt werden kann. Geht es zum Beispiel um die Fähigkeit, sicher mit einem Auto am Straßenverkehr teilzunehmen, kann das dazu notwendige Verhalten nicht in einem schriftlichen Test gezeigt werden. Soll jedoch getestet werden, ob eine Person das benötigte Wissen über Verkehrsregeln besitzt, um sicher am Straßenverkehr teilnehmen zu können, kann dieses Wissen sehr wohl in einem schriftlichen Test abgefragt werden, indem die Testpersonen die Antwort entweder selber formulieren oder aus vorgefertigten Wahlantworten auswählt.

Folglich eignet sich nicht jeder Frage- bzw. Aufgabentyp gleich gut, um verschiedene Fähigkeiten zu überprüfen. Abbildung 2 und die darauffolgende Tabelle 1 geben eine (grobe) Orientierung zu Einsatzgebieten von Fragetypen sowie Vor- und Nachteilen des jeweiligen Formats.

Im darauffolgenden Abschnitt sollen noch einmal die wichtigsten Grundprinzipien bei der Aufgabengestaltung zusammengefasst werden.

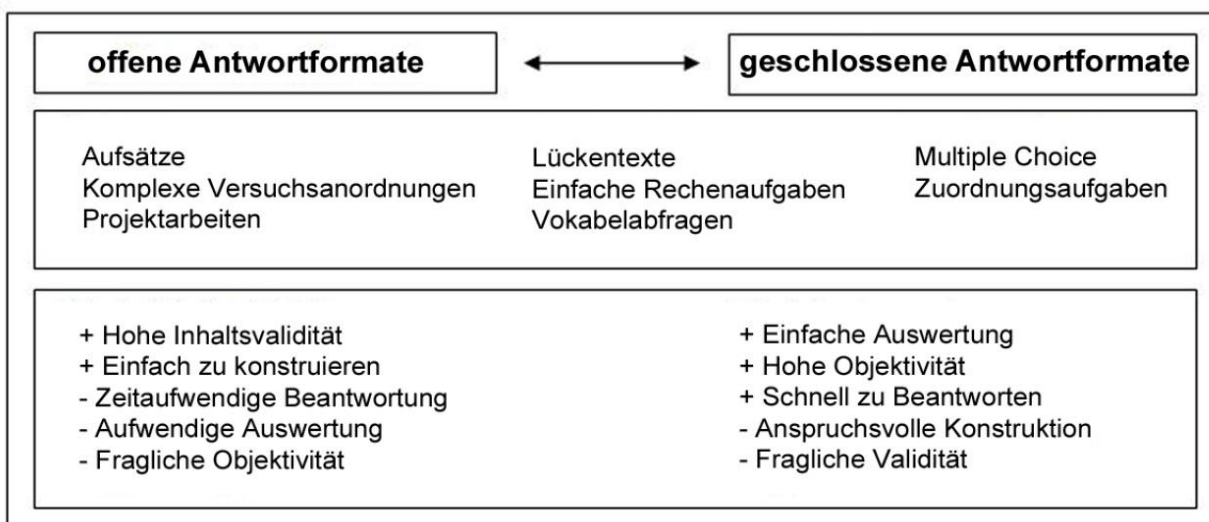


Abbildung 2: Kontinuum offener und geschlossener Antwortformate für Kompetenztests mit Beispielen sowie Vor- und Nachteilen (Hartig & Jude 2007, S. 30).

Tabelle 1: Mögliche Einsatzgebiete verschiedener Fragetypen

Fragetyp	Angestrebte Lernergebnisse
Multiple Choice	<ul style="list-style-type: none"> ➤ Begriffe wiedererkennen ➤ Konzepte verstehen ➤ Zusammenhänge verstehen
Zuordnungsaufgabe	<ul style="list-style-type: none"> ➤ Begriffe in Beziehung setzen ➤ Konzepte einander zuordnen ➤ Hierarchien erkennen
Sortierungsaufgabe	<ul style="list-style-type: none"> ➤ Prozesse analysieren ➤ historische Entwicklungen reproduzieren
Kurzantwortaufgabe	<ul style="list-style-type: none"> ➤ Begriffe und Definition reproduzieren ➤ Berechnungen durchführen ➤ kurze Bewertungen ➤ Entscheidungen treffen
Erweiterte offene Antwortaufgabe	<ul style="list-style-type: none"> ➤ Problemlöseprozess durchführen und beschreiben ➤ Analysieren ➤ Standpunkt argumentieren ➤ Einschätzungen darlegen und begründen

Die wichtigsten Prinzipien bei der Aufgabenerstellung

Im Folgenden werden noch einmal die wichtigsten Prinzipien für die Aufgabenstellung vorgestellt. Wie bereits im einleitenden Abschnitt formuliert, sollten Aufgaben differenziert ein bestimmtes Lernergebnis erfassen und dabei die Bewertungen von Niveauunterschieden innerhalb der Lernergebnisse zulassen. Darüber hinaus sollten Aufgaben kompetenzorientiert gestaltet werden.

Zwei Grundanforderungen „guter“ Aufgaben

- ✓ **Jede Aufgabe bildet lediglich eine Anforderungsdimension (in der Regel = angestrebtes Lernergebnis) ab**

Zum Beispiel darf bei einer Aufgabe, die Rechenfähigkeit erfassen soll, auch nur diese Fähigkeit für die Lösung der Aufgabe eine Rolle spielen.

- ✓ **Aufgaben berücksichtigen Aufgabenniveaus**

Mit mehreren Aufgaben innerhalb einer Anforderungsdimension sollte das Leistungsniveau über den Schwierigkeitsgrad der Aufgaben differenzierbar sein.

Wichtigstes Kriterium: Lernergebnisbezug

Zentral für die Aufgabenerstellung ist, dass Aufgaben immer **mit Bezug auf (ausschließlich) ein bestimmtes angestrebtes Lernergebnis** erstellt werden. Es gilt also zu überlegen, *welches Lernergebnis* bzw. auch *welches Bearbeitungsniveau dieses Lernergebnisses* mit der Aufgabe abgedeckt werden soll.

Im nächsten Schritt sollte überlegt werden, *mit welchem Aufgabentyp das Lernergebnis auf dem entsprechenden Niveau überprüft werden kann*. Dabei kann es durchaus sinnvoll sein, für bestimmte angestrebte Lernergebnisse jeweils einen passenden Aufgabentyp zu entwickeln.

Zielführend kann es sein, eher kurze und bündige Aufgaben zu formulieren bzw. große Aufgaben in Teilaufgaben zu zerlegen, da Anforderungen so präziser zu identifizieren und zu beschreiben sind.

Je nachdem, ob für die Definition von angestrebten Lernergebnissen eine Taxonomie zum Einsatz kam, lässt sich diese auch für die Erstellung von Aufgaben nutzen. Am Beispiel der erweiterten Lernzieltaxonomie nach Bloom (1956) von Anderson und Krathwohl (2002) (vorgestellt in der Handreichung zur Definition von angestrebten Lernergebnissen) lassen sich zu den einzelnen Erkenntnisstufen anhand der dazu aufgeführten Verben Ansätze für Aufgaben überlegen. So kann beispielsweise zur Überprüfung eines angestrebten Lernergebnisses, das die Analyse von Daten vorsieht, eine Teilaufgabe gestellt werden, bei der Ergebnisse vorgegeben und dann verallgemeinert oder auf Fehler überprüft werden sollen.

Ansonsten kann bereits bei der Konzeption von Aufgaben nach Prinzipien der rationalen Aufgabenanalyse, die sich in erster Linie für die Überprüfung bestehender Aufgaben eignet, vorgegangen werden (siehe weiter unten).

Darüber hinaus lassen sich bei der Umsetzung bestimmte allgemeine Kriterien beachten, die die Passung zwischen angestrebten Lernergebnissen und Aufgaben erleichtern.

Ein Lernergebnisbezug ist zusätzlich abgesichert, wenn die Aufgabe...

- ✓ ...ausschließlich Informationen enthält, die zur Lösung des Problems relevant sind.
- ✓ ...keine Lösungshinweise für andere Aufgaben in derselben Prüfung enthält.
- ✓ ...damit eine von anderen Aufgaben unabhängige Lösung erlaubt.
- ✓ ...eine klare Problemstellung und Handlungsaufforderung beinhaltet.
- ✓ ...in klarer, eindeutiger Sprache formuliert ist.
- ✓ ...nur in notwendigem Maß auf Fachsprache zurückgreift.
- ✓ ...einen an Kriterien beurteilbaren Schwierigkeitsgrad aufweist.

Weiteres Kriterium: Kompetenzorientierung

In der Lehre ist nach aktuellen Standards auf Kompetenzorientierung zu achten. Sogenanntes träges Wissen soll reduziert und stattdessen Handlungsfähigkeit in komplexen Situationen ins Zentrum gerückt werden. In schriftlichen Prüfungen kann Kompetenzorientierung umgesetzt werden, indem Aufgaben *die kognitiven Voraussetzungen für kompetentes Handeln erfassen*. Kompetenzorientierung in Aufgaben kann darüber hinaus über folgende Kriterien gesteuert werden:

- Über eine stärkere Fokussierung auf konzeptionelles und strategisches Wissen (in Übereinstimmung mit dem jeweiligen Lehrziel/angestrebten Lernergebnis)
- Über die Aufgabenoffenheit, also inwiefern die Aufgabe verschiedene Lösungswege zulässt
- Über die sprachlogische Komplexität der Aufgabe
- Über Kontext- und Situationsbezüge

Dabei ist zu beachten, dass die Gestaltung von Aufgaben nach Merkmalen einer Kompetenzorientierung natürlich in Einklang mit den angestrebten Lernergebnissen erfolgen sollte. Nicht jede Aufgabe muss zwangsläufig kontextualisiert sein, um kompetenzorientiert zu sein. So kann auch bereits die Begründung ausreichen, dass zum Beispiel ein abstrakter Rechenschritt eine Voraussetzung für eine bestimmte Kompetenz darstellt.

Techniken zur Aufgabenkonstruktion

Vorab definierte Lernergebnisse bilden den Ausgangspunkt für die Aufgabengestaltung. Jedes angestrebte Lernergebnis beschreibt und definiert eine Aufgabenmenge. In Aufgaben werden standardisierte Anforderungssituationen abgebildet, in denen die Lernenden das zur Bewältigung einer Aufgabenmenge erforderliche Verhalten an den entsprechenden Inhalten ausführen können. Die in den Aufgaben dargestellten Anforderungssituationen müssen den in den angestrebten Lernergebnissen festgelegten Situationen und Inhalten entsprechen, also möglichst originalgetreu in das entsprechende Format umgesetzt werden. Das Antwortformat muss die Ausführung des angestrebten Verhaltens zulassen. Die Übersetzung von Zielen in Anforderungen stellt eine große Herausforderung dar und erfordert neben fundierten fachlichen Kenntnissen zum Prüfungsgegenstand auch eine Portion Kreativität. Die im Folgenden dargestellten Techniken sollen dabei helfen, kreativ neue Aufgaben zu entwickeln. Die Techniken können außerdem sehr gut auf bereits bestehende Aufgaben angewendet werden.

Technik 1: Dosierte Öffnen von Aufgaben durch Zielumkehrung geschlossener Aufgaben

Aufgaben lassen sich dosiert öffnen, indem eine Umkehrung geschlossener Aufgaben erfolgt. Bei der Umkehrung von Aufgaben geht man immer vom Endzustand aus. Statt Ausgangswerte vorzugeben und nach dem Endergebnis zu fragen, geben Sie das Endergebnis einer Aufgabe vor und fragen nach den Ausgangswerten. Studierende sind somit aufgefordert, beispielsweise die Ausgangswerte einer bestimmten Rechnung mit Hilfe des Endergebnisses zu ermitteln, anstatt einfach nur die vorgegebenen Zahlen in eine Formel einzusetzen und den Rechenweg auszuführen.

Das dosierte Öffnen von Aufgaben soll an folgendem Beispiel (Aufgabenbeispiel aus SINUS Transfer) aufgezeigt werden:

Statt

„Zeichne ein Parallelogramm, dessen Grundseite 8 cm lang ist und dessen Höhe 3,5 cm lang ist und berechne den Flächeninhalt dieses Parallelogramms.“

Zielumkehrung

„Zeichne ein Parallelogramm mit einem Flächeninhalt von 24cm^2 .“

Dieses Beispiel zeigt, dass bei der Zielumkehrung das Konzept - Flächeninhalte von Parallelogrammen berechnen sowie Parallelogramme zeichnen - verstanden sein muss. Es reicht nicht aus, vorgegebene Werte in eine Formel einzusetzen und den Flächeninhalt zu berechnen. Vielmehr muss das Prinzip der Berechnung verstanden sein, um vom Flächeninhalt auf die einzelnen Ausgangswerte zu gelangen. Demzufolge lassen sich durch die Umkehrung geschlossener Aufgaben reine Produktionsaufgaben – ein reines Einsetzen von Werten in eine Formel – in verstehensorientierte Aufgaben umwandeln, in denen mehr das Konzeptwissen fokussiert wird – also das: „Was hat der Lernende hinsichtlich des Berechnungsprinzips und des Zeichnens verstanden?“.

Weitere Beispiele für „Dosiertes Öffnen“ durch Aufgabenumkehrung:

Diese Beispiele zeigen ein dosiertes Öffnen der Aufgabenstellung, insofern das Endergebnis bereits angegeben wird und von dort ausgehend gearbeitet werden soll.

Aufgabenbeispiel aus PISA 2003 (<http://pisa.ipn.uni-kiel.de>)

FRAGE 2: TAGESLICHT

In der Abbildung wird gezeigt, wie Lichtstrahlen von der Sonne auf die Erde scheinen.

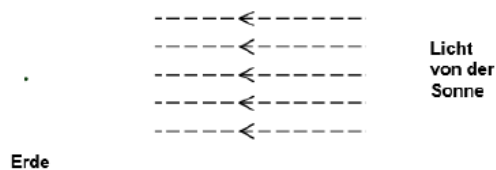


Abbildung: Lichtstrahlen von der Sonne

Nimm an, es wäre der kürzeste Tag in Melbourne.

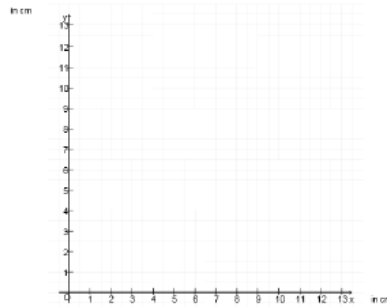
Zeichne die Erdachse, die nördliche Hemisphäre, die südliche Hemisphäre und den Äquator in die Abbildung ein. Beschrifte alle Teile deiner Antwort.

Aufgabenbeispiel aus den Vergleichsarbeiten Mathematik (<http://bildungsserver.berlin-brandenburg.de/>)

Aufgabe 1

Zeichne in das Koordinatensystem:

- a) eine Fläche mit einem Flächeninhalt von $A = 12 \text{ cm}^2$
- b) eine weitere Fläche mit einem Umfang von $u = 16 \text{ cm}$ ein



Bezeichne die Eckpunkte jeder Fläche und gib ihre Koordinaten $(x; y)$ an.
Wie heißen die entstandenen Flächen?

Technik 2: Initiierung von Begründungen oder Einschätzungen

Das Initiieren von Begründungen oder Einschätzungen wird am folgenden Beispiel (Aufgabenbeispiel aus SINUS Transfer) aufgezeigt:

α , β und γ bezeichnen die Innenwinkel in einem Dreieck.

Beurteile jeweils, ob die folgenden Aussagen zutreffen können und begründe deine Einschätzung:

- a) ‚Die Summe $\sin \alpha + \sin \beta + \sin \gamma$ kann größer als 3 sein.‘
- b) ‚Die Summe $\sin \alpha + \sin \beta + \sin \gamma$ kann gleich 0 sein.‘
- c) ‚Das Produkt $\sin \alpha * \sin \beta * \sin \gamma$ ist immer größer als 0.‘

In diesem Beispiel muss neben der Einschätzung, welche der Aussagen zutreffend sind, noch begründet werden, warum diese falsch oder richtig sind. Es reicht also nicht aus, den richtigen Lösungsfall zu raten oder auswendig gelernt aus dem Skript wiederzuerkennen. Die Lösung muss auf Grundlage des in der Vorlesung erworbenen Wissens erklärt werden.

Aufgaben, die Begründungen oder Einschätzungen verlangen, weisen ebenso den Vorteil auf, Konzeptwissen anstelle von reiner Reproduktion von Wissen abzufragen. Um dies zu erreichen, geben Sie bei dieser Technik Statements vor, die von Seiten der Lernenden eingeschätzt werden sollen. Dies ermöglicht es Ihnen, näher an das Konzeptwissen des Lernenden zu gelangen. So lassen sich aus reinen Reproduktionsaufgaben, in denen die Lernenden nur ihr Wissen reproduzieren, ohne dies notwendigerweise verstanden zu haben, Aufgaben generieren, die das konkrete Verstehen eines bestimmten Sachverhalt aufzeigen.

Aufgabenbeispiel aus PISA 2006

Frage 5: SONNENSCHUTZ

Das lichtempfindliche Papier ist dunkelgrau und verändert seine Farbe zu hellgrau bei schwacher Sonneneinstrahlung und zu weiß bei starker Sonneneinstrahlung.

Welche der folgenden Abbildungen zeigt ein Muster, das auftreten könnte? Erkläre, warum du sie ausgewählt hast.

<p>A</p> <table style="width: 100%; text-align: center;"> <tr> <td>○ M</td> <td>● S1</td> <td>● S2</td> </tr> <tr> <td>● ZnO</td> <td>● S3</td> <td>● S4</td> </tr> </table>	○ M	● S1	● S2	● ZnO	● S3	● S4	<p>C</p> <table style="width: 100%; text-align: center;"> <tr> <td>● M</td> <td>● S1</td> <td>● S2</td> </tr> <tr> <td>● ZnO</td> <td>● S3</td> <td>○ S4</td> </tr> </table>	● M	● S1	● S2	● ZnO	● S3	○ S4
○ M	● S1	● S2											
● ZnO	● S3	● S4											
● M	● S1	● S2											
● ZnO	● S3	○ S4											
<p>B</p> <table style="width: 100%; text-align: center;"> <tr> <td>● M</td> <td>● S1</td> <td>● S2</td> </tr> <tr> <td>○ ZnO</td> <td>● S3</td> <td>● S4</td> </tr> </table>	● M	● S1	● S2	○ ZnO	● S3	● S4	<p>D</p> <table style="width: 100%; text-align: center;"> <tr> <td>○ M</td> <td>● S1</td> <td>● S2</td> </tr> <tr> <td>○ ZnO</td> <td>○ S3</td> <td>● S4</td> </tr> </table>	○ M	● S1	● S2	○ ZnO	○ S3	● S4
● M	● S1	● S2											
○ ZnO	● S3	● S4											
○ M	● S1	● S2											
○ ZnO	○ S3	● S4											

Antwort:

Erklärung:

Weitere Beispiele für die Initiierung von Begründungen oder Einschätzungen:

Diese Beispiele zeigen, wie Begründungen und Einschätzungen initiiert werden können. Dabei werden mögliche Lösungen vorgegeben, die richtigen Lösungswege sollen begründet werden. Dies erfordert von Seiten der Studierenden ebenfalls mehr als das reine Wiedergeben von Faktenwissen.

Aufgabenbeispiel aus den Vergleichsarbeiten Mathematik

Aufgabe 2

Welche Aussage ist wahr? Kreuze an. Begründe.

- Jedes Rechteck ist auch ein Quadrat.
- Jedes Quadrat ist auch ein Rechteck.
- Jedes Parallelogramm ist auch eine Raute/ein Rhombus.
- Jedes Parallelogramm ist auch ein Rechteck.

Technik 3: Erfragen von Anwendungsbeispielen oder Grenzen

Das Erfragen von Anwendungsbeispielen oder Grenzen ermöglicht es auch, näher an das Konzeptwissen der Studierenden zu gelangen. Mit dieser Technik entwickeln Sie Aufgaben, in denen Sie die Lernenden beispielsweise auffordern, Beispiele zu finden, oder Sie lassen die Lernenden Grenzen aufzeigen oder beurteilen.

Beispiel:

Gib jeweils einen Wachstumsprozess an, der sich besonders gut durch eine

- a) Exponentialfunktion
- b) quadratische Funktion
- c) lineare Funktion

beschreiben lässt.

Um diese Aufgabe lösen zu können, muss das Konzept des Wachstumsprozesses verstanden sein. Auch ein mathematisches Grundverständnis der angegebenen Beispiele ist Grundvoraussetzung, um das geeignete Beispiel finden zu können. Diese Aufgabenstellung geht damit über das reine Auswendiglernen von Faktenwissen hinaus und setzt Konzeptwissen voraus.

Das folgende Beispiel stellt noch einmal expliziter auf die Grenzen in der Anwendung bestimmter Funktionen ab, indem die Schülerinnen und Schüler eine Beobachtung nennen sollen, die mit diesen Funktionen nicht modelliert werden kann.

Aufgabenbeispiel aus SINUS Transfer

Gib einen Wachstumsprozess an, der sich weder durch eine Exponentialfunktion, noch durch eine quadratische Funktion, noch durch eine lineare Funktion angemessen beschreiben lässt.

Technik 4: Von Begriffen ausgehen

Ein wesentliches Ziel im Studium ist es, dass Studierende zentrale Begriffe einer Domäne kennen und verstehen lernen. Viele Prüfungsaufgaben fragen daher nach Begriffsdefinitionen, Einsatzgebieten oder Vor- und Nachteilen, etc. von Begriffen. Aufgaben, die von Begriffen ausgehen bzw. auf diese abzielen, können in ihrer Offenheit variiert werden. Zwei Typen werden im Folgenden an Beispielen dargestellt, die jeweils auf den Begriff „Mittelsenkrechte“ führen:

Typ 1: Die Aufgabenbearbeitung führt unmittelbar auf den Begriff.

Folgendes Beispiel verdeutlicht eine Aufgabenstellung, die zu dem gesuchten Begriff führt:

„Zeichne eine Strecke AB. Schlage dann einen Kreis um A und einen Kreis mit demselben Radius um B, so dass sich beide Kreise schneiden. Verbinde die Schnittpunkte. Welche Eigenschaft weißt die entstehende Gerade auf?“

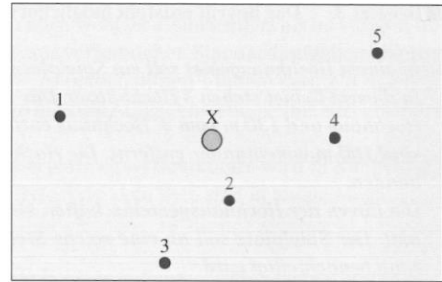
Die Lösung der Aufgabe führt direkt zum Mittelsenkrechtenbegriff. Zum Problemlösen führt sie nicht, denn es gibt nur einen vorgezeichneten Weg und nur eine vorherbestimmte Entdeckung. Dieser Typ von Aufgaben ist in seiner Offenheit eingeschränkt. Die Problemstellung ist sehr stark vordefiniert, so

dass lediglich das Abarbeiten oder Erinnern von Lösungsschritten bzw. der richtigen Lösung (häufig in Form einer Wiedergabe richtiger Begriffe, Vor- und Nachteilen, Einsatzgebieten oder Definitionen) notwendig ist.

Typ 2: Der Begriff entsteht notwendig bei der Lösung des Problems.

Anstatt den Begriff direkt vorzugeben, kann man als Aufgabenersteller/in von allen Aspekten und Anwendungen des Begriffs „rückwärts“ denken und ein Problem konstruieren, das – vorwärts bearbeitet – notwendigerweise auf diesen Begriff hinführt.

*„Die Karte zeigt ein Stück Land.
Es gibt fünf Brunnen in diesem Gebiet.
Stelle dir vor, du stehst bei X mit einer Herde von Schafen, die Durst haben. Zu welchem Brunnen gehst du? Die Wahl war natürlich nicht schwierig. Du gehst zum nächstgelegenen Brunnen.“*



Entwickle nun eine Einteilung des Landes in fünf Gebiete, so dass zu jedem Ort in einem Gebiet der Brunnen in diesem Gebiet der nächstgelegene ist.“

In diesem Beispiel ist der Mittelsenkrechtenbegriff nicht Endprodukt der Aufgabe, sondern Teil des Lösungsweges. Das Problem führt zwar sehr zielstrebig auf den vom Lehrenden ins Auge gefassten Begriff, lässt aber viele individuelle Vorgehensweisen zu. Der spezifische Reiz liegt jedoch darin, dass ein Begriff nicht als fertiges Produkt dargeboten wird, sondern wie selbstverständlich als Lösung für ein Problem durch die Lernenden erfunden bzw. entdeckt werden muss.

Literatur

- Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Addison Wesley Longman, Inc.
- Bloom, B. S., Enghart, M., Furst, E., Hill, W., & Krathwohl, D. R. (Eds.) (1956). *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. New York: David McKay.
- Büchter, A. (2006). *Verstehensorientierte Aufgaben als Kern einer neuen Kultur der Leistungsüberprüfung*. SINUS-Transfer: Erläuterung zu Modul 10. Online-Publikation. http://sinus-transfer.uni-bayreuth.de/fileadmin/MaterialienDB/427/Buechter_Modul_10.pdf (abgerufen: 14.07.2016)
- Büchter, A. & Leuders, T. (2011). *Mathematikaufgaben selbst entwickeln. Lernen fördern – Leistung überprüfen*. 5. Auflage. Berlin: Cornelsen Scriptor.
- Hartig, J. & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.): *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17-36). Bonn, Berlin: Bundesministerium für Bildung und Forschung.

Literatur Aufgabenbeispiele

- Büchter, A. (2006). *Verstehensorientierte Aufgaben als Kern einer neuen Kultur der Leistungsüberprüfung*. SINUS-Transfer: Erläuterung zu Modul 10. Online-Publikation. http://sinus-transfer.uni-bayreuth.de/fileadmin/MaterialienDB/427/Buechter_Modul_10.pdf (abgerufen: 14.07.2016)
- OECD Programme for International Student Assessment (2006): PISA 2006: Beispielaufgaben aus dem Naturwissenschaftstest. http://archiv.ipn.uni-kiel.de/PISA/PISA06_Science_Beispielaufgaben.pdf (abgerufen: 14.07.2016)
- Vergleichsarbeiten (VERA). Bildungsserver Berlin-Brandenburg. <https://www.iqb.hu-berlin.de/vera/aufgaben/ma1> (abgerufen: 14.07.2016)

Schritt 3: Abgleich von Prüfungsaufgaben und angestrebten Lernergebnissen

Nachdem in den ersten beiden Schritten der Fokus auf die gute Formulierung von Lernergebnissen als Ausgangsbasis für die Erstellung kompetenzorientierter Prüfungsfragen und auf die konkrete Erstellung von Prüfungsaufgaben lag, geht es im dritten Schritt um die Überprüfung, inwiefern die Prüfungsaufgaben die zuvor formulierten angestrebten Lernergebnisse repräsentativ abbilden. Dieses Verfahren (logische Aufgabenanalyse) umfasst insgesamt sieben Schritte, die im Folgenden ausführlich beschrieben werden.

Misst die Prüfung tatsächlich die Fähigkeiten oder Kompetenzen bei den Studierenden, die sie messen soll? Diese Frage nach der sogenannten Validität ist ein zentrales Qualitätskriterium einer jeden Prüfung. Die Inhaltsvalidität einer Prüfung wird daran festgemacht, ob die Prüfungsaufgaben die formulierten angestrebten Lernergebnisse repräsentativ abdecken. Ein Verfahren, um die Passung zwischen Prüfungsaufgaben und angestrebten Lernergebnissen zu überprüfen, ist die logische Aufgabenanalyse.

Ziel ist es, zu entscheiden, ob die in den Aufgaben dargestellten Anforderungssituationen und das in der Aufgabe erwartete Verhalten auf die Erreichung des angestrebten Lernergebnisses rückschließen lässt. Jede Prüfungsaufgabe soll genau die Inhalte enthalten und das Verhalten hervorrufen, das ersichtlich macht, ob bzw. inwieweit ein angestrebtes Lernergebnis erreicht wurde.

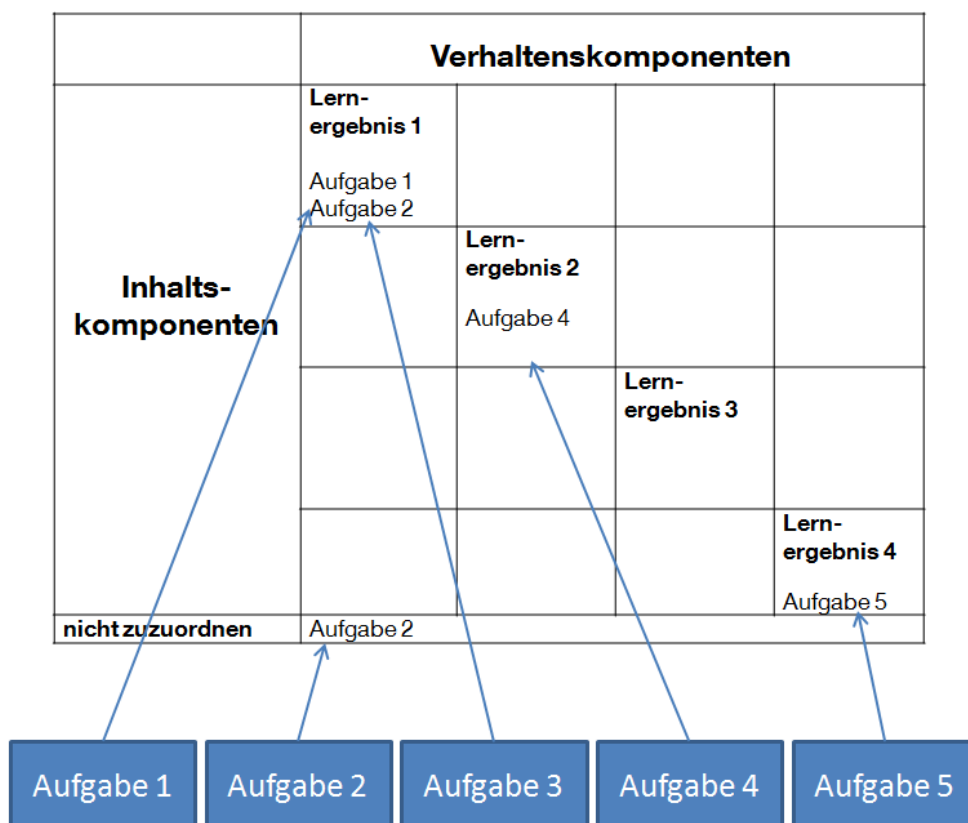


Abbildung 3: Zuordnung von Aufgaben zu angestrebten Lernergebnissen

Die logische Aufgabenanalyse

Schritt 1: Formulieren der angestrebten Lernergebnisse

Die angestrebten Lernergebnisse sollten schriftlich vorliegen (siehe Handreichung „angestrebte Lernergebnisse als Grundlage für die Prüfungserstellung“), bevor mit der Aufgabenanalyse begonnen wird. Zunächst ist es sinnvoll, die Inhalts- von der Verhaltenskomponente in den jeweiligen Lernergebnissen zu trennen, das heißt, kenntlich zu machen, welches Verhalten die Studierenden an welchem Inhalt zeigen sollen.

Beispiele für die Aufteilung in Inhalts- und Verhaltenskomponente:

Im Folgenden sind beispielhaft vier Lernergebnisse aufgeführt. Markiert ist darin jeweils die Inhaltskomponente (=gelb) und die Verhaltenskomponente (=blau).

Die Studierenden sind in der Lage:

1. Fragestellungen, theoretische Hintergründe und zentrale Ergebnisse aktueller Studien aus der Berufsbildungsforschung zu beschreiben.
2. Forschungsmethodische Vorgehensweisen und Befunde aktueller empirischer Studien aus der Berufsbildungsforschung zu erklären.
3. Befunde aus aktuellen Studien hinsichtlich ihrer wissenschaftlichen Geltungskraft und berufspraktischen Relevanz einzuschätzen und zu bewerten.
4. Theoretische Annahmen und empirische Befunde aktueller Studien auf den Schulkontext anzuwenden.

Schritt 2: Eintragen der angestrebten Lernergebnisse in eine Inhalts-Verhaltensmatrix

Die Verhaltenskomponenten werden anschließend in die Spaltenbeschriftung einer Matrix übertragen, die Inhaltskomponenten in die Zeilenbeschriftung. In den jeweiligen Zellen steht dann jeweils das angestrebte Lernergebnis als Kombination einer Verhaltens- mit einer Inhaltskomponente (siehe Tabelle 2).

Anschließend können die einzelnen Prüfungsaufgaben den Zellen und somit den angestrebten Lernergebnissen zugeordnet werden (siehe Schritt 3).

Anmerkung: Die Erstellung einer solchen Matrix ist nicht unbedingt notwendig, um die Passung zwischen Aufgaben und Lernergebnissen zu überprüfen. Sie eignet sich aber vor allem dann, wenn nicht ganz klar ist, inwieweit angestrebte Lernergebnisse systematisch durch Aufgaben abgedeckt sind bzw. ob bestimmte angestrebte Lernergebnisse möglicherweise noch ergänzt werden müssen.

Tabelle 2: Inhalts-Verhaltensmatrix

Inhaltskomponenten	Verhaltenskomponenten			
	beschreiben	erklären	einschätzen und bewerten	auf den Schulkontext anwenden
Fragestellungen, theoretische Hintergründe und zentrale Ergebnisse aktueller Studien aus der Berufsbildungsforschung	Angestrebtes Lernergebnis 1			
Forschungsmethodische Vorgehensweisen und Befunde aktueller empirischer Studien aus der Berufsbildungsforschung		Angestrebtes Lernergebnis 2		
Befunde aus aktuellen Studien hinsichtlich ihrer wissenschaftlichen Geltungskraft und berufspraktischen Relevanz			Angestrebtes Lernergebnis 3	
Theoretische Annahmen und empirische Befunde aktueller Studien				Angestrebtes Lernergebnis 4
nicht zuzuordnen				

Schritt 3: Zuordnen der Aufgaben zu den angestrebten Lernergebnissen

Im dritten Schritt wird nun versucht, die einzelnen Aufgaben der Prüfung den Zellen der Matrix und damit den angestrebten Lernergebnissen zuzuordnen (vgl. Abbildung 3). Dabei sollte die Frage gestellt werden, wie die Studierenden die Aufgabe aller Wahrscheinlichkeit lösen werden. Welches Wissen, welche Fähigkeiten oder Kompetenzen werden zur Lösung benötigt?

In einem ersten Schritt können diese Überlegungen relativ frei erfolgen bzw. anhand von fachlichen Überlegungen angestellt werden. Hilfreich kann es auch sein, die Aufgaben von Kolleginnen und Kollegen probelösen zu lassen und dabei über mögliche Lösungswege zu reflektieren. Diese Herangehensweise entspricht einem bekannten Verfahren, dem sogenannten „Lauten Denken“, das in der Forschung zur Validierung von Aufgaben Anwendung findet.

Zur besseren Einordnung der Aufgaben können aber auch zusätzlich folgende Aspekte herangezogen werden:

Inhaltliche Übereinstimmung

Stimmen Inhaltsaspekte im angestrebten Lernergebnis und der Aufgabe überein?

Beispiel: Das angestrebte Lernergebnis hat als Inhalt mechanische Fragen zur Stereo-Statik. Überprüfen Sie, ob die fragliche Aufgabe ebenfalls eine Problemstellung mit einer mechanischen Fragestellung zur Stereostatik beinhaltet.

Angesprochene Wissensart

Wurde bei der Formulierung von Lernergebnissen auf die Unterscheidung nach Wissensarten (Faktenwissen, Prozedurales Wissen, Konzeptwissen und Strategiewissen; siehe im Anhang dieses Dokuments) zurückgegriffen, so kann überlegt werden, welches Wissen in einer Aufgabe benötigt wird.

Beispiel: Nach dem angestrebten Lernergebnis sollen mathematische Formeln wiedergegeben werden (Faktenwissen). Eine Prüfungsaufgabe, die diesem Lernergebnis entspricht, sollte nun auch auf die Wiedergabe von Formeln abzielen und nicht zum Beispiel den Verwendungszusammenhang der Formel (Konzeptwissen) oder deren Anwendung auf ein Problem (Prozedurales Wissen) verlangen.

Plausibilität der Verhaltenskomponenten

Stimmen Verhaltenskomponenten im angestrebten Lernergebnis mit denen der Aufgabe überein? Das heißt, werden die Studierenden tatsächlich mit der Aufgabe dazu angehalten, das besagte Verhalten zu zeigen?

Beispiel: Das angestrebte Lernergebnis bezeichnet das „Bewerten von Sachverhalten“. Entsprechend sollte auch die Aufgabe eine entsprechende Aufforderung enthalten. Verlangt die Aufgabe stattdessen die Anwendung eines Algorithmus zur Berechnung einer Lösung zielt die Aufgabe eher darauf ab, Prozeduren durchzuführen

Aspekte im Antwortformat können ebenfalls herangezogen werden. Verlangt ein angestrebtes Lernergebnis eigenständiges Formulieren oder Skizzieren, so muss das Antwortformat eine eigenständige Formulierung zulassen, also offen angelegt sein.

Weiterhin kann differenziert werden, in welcher Art und Weise das jeweilige Wissen eingesetzt werden muss. Unterschieden werden kann dabei zwischen Reproduktion, nahem und weitem Transfer von Wissen sowie kreativem Problemlösen (siehe im Anhang dieses Dokuments). Es ist dann abzugleichen, ob das jeweilige angestrebte Lernergebnis und die entsprechende Aufgabe auf den gleichen kognitiven Prozess abzielen.

Aufgabenoffenheit

Stimmen die Offenheit des angestrebten Lernergebnisses und der Aufgabe überein?

Durch Variation der Engführung bzw. Offenheit der Problemstellung (gut vs. schlecht definierte Problemstellung; siehe im Anhang dieses Dokuments) können gezielt verschiedenartige angestrebte Lernergebnisse abgefragt werden. Umfasst die Aufgabe zum Beispiel einen eindeutigen Arbeitsauftrag bzw. eine klar identifizierbare Fragestellung, zu deren Bearbeitung alle benötigten Informationen vorliegen, und ist nur eine Lösung gesucht bzw. richtig, eignet sich die Aufgabe wahrscheinlich eher für die Reproduktion von Wissen oder das Durchführen bestimmter Prozeduren.

Beispiel wenig offenes, angestrebtes Lernergebnis:

„Die Studierenden sind in der Lage, xy zu berechnen.“

Bei diesem angestrebten Lernergebnis liegt der Fokus auf der prozeduralen Fähigkeit „Berechnen“. Die Offenheit der Aufgabe kann deshalb eingeschränkt sein, indem alle benötigten Informationen bereitgestellt sind, so dass zur Lösung der Aufgabe lediglich die Berechnung durchgeführt werden muss.

Beispiel offenes, angestrebtes Lernergebnis:

„Die Studierenden sind dazu in der Lage, einen Lösungsansatz für ein komplexes Problem im Bereich xy aufzustellen.“

Hier sollte die Problemstellung in einer Aufgabe möglichst offen vorgegeben sein und z.B. die Auswahl relevanter Informationen, die Auswahl eines geeigneten Verfahrens oder einer

Problemlöseheuristik verlangen. Weniger offene Aufgaben erhöhen die Wahrscheinlichkeit einer Lösung durch vorab eingeübte Lösungsalgorithmen, die nicht dem angestrebten Lernergebnis entsprechen.

Aufgabenkomplexität

Stimmt die Komplexität des Lernergebnisses mit der der Aufgabenstellung überein?

Die Komplexität der Aufgabe kann über verschiedene Stellschrauben verändert werden, wie zum Beispiel über die sprachliche Komplexität (Syntax, Textmenge, Terminologien, Einfachheit der Sprache usw.), die Menge und Darbietung der Informationen in der Aufgabe, die Verwendung mehrere Darstellungsformen (Zeichnungen, Graphen, Text usw.) sowie die Angabe mehrerer Bedingungen.

Schritt 4: Begründen der Zuordnung

Die Zuordnung von Einzelaufgaben zu angestrebten Lernergebnissen sollte intersubjektiv nachvollziehbar sein, das heißt, andere Personen sollten nachvollziehen können, warum eine Aufgabe einem angestrebten Lernergebnis zugeordnet wurde. Sichergestellt werden kann dies über eine explizierte Begründung für die Zuordnung. Die Begründung muss nicht ausschließlich „frei“ bzw. fachdidaktisch erfolgen, sondern kann auch auf Basis der generischen Aspekte aus dem vorangegangenen Arbeitsschritt erfolgen.

Idealerweise wird die Begründung im Sinne einer „didaktischen Überlegung“ für jede Aufgabe schriftlich festgehalten. Dies erleichtert die Kategorisierung von Aufgaben, zum Beispiel zur Erstellung einer Aufgabendatenbank.

Schritt 5: Überprüfung der Inhalts-Verhaltensmatrix

Nachdem alle Aufgaben zugeordnet worden sind, sollten folgende Aspekte überprüft werden:

- Werden alle angestrebten Lernergebnisse durch Aufgaben abgedeckt?
- Gibt es Aufgaben, die ein angestrebtes Lernergebnis erfassen, das gar nicht beabsichtigt ist?
- Gibt es Aufgaben, die nicht eingeordnet werden konnten? Was überprüfen diese Aufgaben? Wie können diese Aufgaben abgeändert werden, um ein angestrebtes Lernergebnis zu erfassen? Muss ggf. ein weiteres angestrebtes Lernergebnis formuliert werden?
- Gibt es angestrebte Lernergebnisse, die durch überproportional viele Aufgaben abgedeckt werden? Gibt es angestrebte Lernergebnisse, zu denen nur eine oder zwei Aufgaben existieren (zur Sicherstellung der Reliabilität der Messung sollten mindestens drei bis vier Aufgaben pro angestrebtes Lernergebnis eingesetzt werden – je nach dessen Komplexität bzw. Abstraktionsgrad)?

Schritt 6: Optimierungsbedarf feststellen

Im nächsten Schritt kann nun expliziert werden, welche Veränderungen in den Aufgaben (oder auch in den angestrebten Lernergebnissen) anzustreben sind, um die Passung zwischen Aufgaben und angestrebten Lernergebnissen - und damit auch die inhaltliche Validität - zu erhöhen.

Beispiele:

- Ein angestrebtes Lernergebnis wird nicht durch Prüfungsaufgaben abgedeckt: Es sollten Aufgaben formuliert werden, die genau dieses angestrebte Lernergebnis überprüfen und in die Prüfung aufnehmen.
- Eine Aufgaben erfasst etwas, das gar nicht als angestrebtes Lernergebnis definiert wurde: Entweder kann das Lernergebnis (falls sinnvoll) mit aufgenommen werden oder die Aufgabe sollte abgeändert oder aus der Prüfung entfernt werden.

- Eine Aufgabe erfasst nur Teilaspekte des angestrebten Lernergebnisses, aber nicht das Lernergebnis in seiner ganzen Komplexität. So kann z.B. in der Aufgabe gefordert sein, dass Studierende die theoretischen Annahmen und empirischen Befunde aktueller Studien beschreiben, aber nicht - wie in unserem Beispiel (Tabelle 2) verlangt - diese auf den Schulkontext anwenden müssen. In diesem Fall müsste die Aufgabe um den fehlenden Aspekt erweitert werden.

Schritt 7: Qualitätssicherung

Um sicherzustellen, dass die Zuordnung nicht ausschließlich auf der Einschätzung einer Einzelperson beruht, sollten Sie die Zuordnung der Aufgaben zu den angestrebten Lernergebnissen zusätzlich von einer Kollegin oder einem Kollegen durchführen lassen, die/der ein vergleichbares Wissen über den abzuprüfenden Themenbereich besitzt. Im Anschluss kann analysiert werden, wie gut die Übereinstimmung der Zuordnungen ist. Über Aufgaben, die unterschiedlich zugeordnet wurden, sollte gesprochen werden. Worin liegen die Gründe für die unterschiedliche Einordnung? Ist es möglich, zu einer gemeinsamen Einschätzung zu kommen? Sollten Anpassungen an der Aufgabe vorgenommen werden, um eine eindeutigere Zuordnung zu ermöglichen?

Bei neu erstellten Aufgaben, die gezielt für ein angestrebtes Lernergebnis erstellt worden sind, kann auch eine weitere Methode angewandt werden. Die erstellten Aufgaben können –nicht systematisiert - an Kolleginnen und Kollegen weitergegeben werden, mit der Bitte, diese den angestrebten Lernergebnissen zuzuordnen. Insofern die Zuordnung der eigenen Zuordnung entspricht, ist von einer guten Passung auszugehen.

Eine weitere Möglichkeit zur Überprüfung der Inhaltsvalidität ist die Durchführung einer kognitiven Validierung mit Studierenden (Cognitive labs). Die Studierenden lösen Beispielaufgaben und legen dabei ihre Denkprozesse offen (Think-Aloud-Technik). Die verbalisierten Denkprozesse werden protokolliert und analysiert.

Anhang

Klassifizierung von Wissen

Faktenwissen („Wissen, dass“)

Faktenwissen beinhaltet Definitionen und Fakten (meistens in der Form von Begriffen, Aufstellungen, Aussagen, Regeln oder Beschreibungen) und ist explizit verbalisierbar.

Beispiel: Schmelzpunkte verschiedener Materialien bei verschiedenen Temperaturen (Fakt).

Prozedurales Wissen („Wissen, wie“)

Prozedurales Wissen ist Wissen darüber, wie etwas aus- oder durchgeführt werden kann. Es enthält „Wenn-Dann“-Produktionsregeln oder eine Sequenz an Schritten, die ausgeführt werden müssen. Durch Praxis und zielgerichtetes Üben kann prozedurales Wissen in automatisierte Abläufe überführt werden.

Beispiele: Vorgehen zur Messung der Dichte eines Objekts; Durchführung der ersten Ableitung, Ausführen einer Differentialgleichung, Berechnung der Geschwindigkeit eines Autos, Zeichnen eines Graphen, der die Geschwindigkeit eines Autos abbildet; Anwenden von Algorithmen zum Aufstellen chemischer Reaktionsgleichungen.

Konzeptwissen („Wissen, warum“)

Konzeptwissen ist ein Wissen über systematische Zusammenhänge. Im weiteren Sinn wird Konzept dabei verstanden als...

- Sammlung von Leitsätzen oder Prinzipien oder
- Skizzen eines Vorhabens.

Konzeptwissen beinhaltet typischerweise die Anwendung wissenschaftlicher Prinzipien, Erklärungsmodelle, Taxonomien, Schemata oder Begriffsnetze oder das Argumentieren mit Modellen oder Konzepten. Die Studierenden nutzen zur Lösung entsprechender Aufgaben ihr Wissen über Theorien. Die Aufgabe verlangt von Studierenden, Voraussagen zu treffen oder unter Rückgriff auf Theorien oder Modelle zu argumentieren.

Konzeptwissen kann verwendet werden, um Probleme zu interpretieren, Fehlersuchen durchzuführen, zu erklären, zu modellieren und um Voraussagen zu treffen, wie Veränderungen in einem Konzept andere Konzepte beeinflussen.

Beispiel: Erklärung, warum Tag und Nacht auf der Erde auftreten; Erklärung, wie ein bestimmter Virus wirkt; Erklärung, welcher Zusammenhang zwischen Aufgaben-Format und kognitiven Prozessen besteht.

Strategiewissen (Wissen wann, wo und wie)

Strategiewissen umfasst Wissen darüber, wann, wo und wie Wissen in neuen Situationen eingesetzt werden kann. Strategiewissen beinhaltet domänenspezifische Strategien wie Planen, Problemlösen und das Überprüfen des eigenen Fortschritts, sowie konditionales Wissen, also Kenntnisse über adäquate Anwendungsbedingungen der Strategien und des Wissens. Entsprechende Aufgaben erfordern das Interpretieren einer Problemstellung und darauf aufbauend den Vergleich und das begründete Auswählen einer Strategie.

Personen verwenden strategisches Wissen, um zu erkennen, wann bestimmte Prozeduren ausgeführt werden sollten, um die Merkmale einer Aufgabe zu analysieren, um darauf aufbauend zu entscheiden, welche Strategie verwendet werden kann, setzen Ziele und überprüfen den Fortschritt bei der Zielerreichung. Allgemeine Anwendungen sind Planungsstrategien, wie das Zerlegen in Teilaufgaben, Kategorisieren und Priorisieren.

Strategiewissen kommt vor allem bei Aufgaben zum Einsatz, die neuartige, unbekannte Problemstellungen beinhalten.

Klassifizierung kognitiver Prozesse

Reproduktion von Wissen

Bei der Reproduktion wird Wissen so wiedergeben, wie es abgespeichert wurde. Das Wissen muss nicht in einen anderen Sinn- bzw. Bedeutungszusammenhang transferiert, sondern kann so verwendet werden, wie es abgespeichert wurde. Reproduziert werden können alle Wissensarten. Faktenwissen wird zum Beispiel bei der Aufzählung von Vor- und Nachteilen reproduziert; Prozedurales Wissen, wenn Abläufe skizziert werden; Konzeptionelles Wissen ist reproduzierbar, wenn Studierende eine Begriffsdefinition wiedergeben können.

Beispiele für Verben, die auf eine Reproduktion hindeuten: nennen, aufzählen, anführen, benennen, bezeichnen, berichten, erzählen, beschreiben, skizzieren, angeben, darstellen, schreiben, schildern, aussagen.

Naher Transfer von Wissen

Naher Transfer liegt vor, wenn das vorhandene Wissen in einen neuen Bedeutungszusammenhang transferiert werden muss. Ein Erkennungsmerkmal bei Aufgaben, die auf nahen Transfer abzielen, ist, wenn die Aufgabensituation nur geringfügig von bereits bekannten oder geübten Aufgaben bzw. von der Lernsituation abweicht. Für eine Einschätzung der Differenz zwischen Aufgaben- und Lernsituation sind verschiedene Aspekte bedeutend. Beispielsweise kann die Aufgabe so gestellt sein (mit entsprechenden Hinweisen), dass der Person klar ist, welches Wissen (ob Prozeduren, Faktenwissen oder Konzepte) für die Aufgabenlösung zur Anwendung kommen sollte. Die Aufgabe gibt den Studierenden beispielsweise genaue Handlungsvorgaben. Sie sollen ein Prinzip anwenden, unter Rückgriff auf Vorwissen, das in der Aufgabe explizit genannt wird. Faktenwissen kann in den Kontext einer Problemstellung transferiert werden, indem entschieden wird, ob der entsprechende Fakt im Kontext der Problemstellung richtig oder relevant ist.

Ein weiterer Aufgabentyp, der auf nahen Transfer abzielt, sind Aufgaben, die ein kleinschrittiges Ausführen von Verfahren oder Algorithmen erfordern, bei denen klar ist, welche Handlungen zur Lösung entsprechender Aufgaben nötig sind und welches Wissen dafür benötigt wird. Die Angaben sind so eingegrenzt, dass aus der gewählten Formulierung hervorgeht, welches Lösungsprinzip Anwendung findet. Ein weiteres Erkennungsmerkmal ist das kleinschrittige Ausführen von Verfahren und Algorithmen (z.B. das schrittweise Vorgehen bei Dreisatzaufgaben).

Weiter Transfer von Wissen

Weiter Transfer liegt vor, wenn vorhandenes Wissen in unbekanntem Situationen eingesetzt werden muss. Es ist nicht sofort erkennbar, welches Wissen zum Einsatz kommt. Das vorhandene Wissen muss nach Prinzipien, Regeln und/oder Bedingungen durchsucht werden, die sich auf die entsprechende Situation anwenden lassen. Im Vergleich zur nächsten Stufe (kreatives Problemlösen) ist das Wissen allerdings in der Form vorhanden bzw. leicht abrufbar, in der es zur Anwendung kommen sollte. Wenn bei einer unbekanntem Sachaufgabe zum Beispiel das Problem verstanden wurde, können bekannte Konzepte oder Algorithmen zur Lösung herangezogen werden.

Kreatives Problemlösen

Beim kreativen Problemlösen ist unklar, welches Wissen zur Anwendung kommt. Das zur Bearbeitung entsprechender Aufgaben benötigte Wissen muss erst zusammengefügt werden. Die Studierenden müssen auf unterschiedliche Wissensarten zurückgreifen und einen Problemlöseprozess entwerfen (Problem finden bzw. definieren, Lösungen entwerfen, umsetzen und bewerten).

Das "kreative Problemlösen" lässt sich von Transferaufgaben unterscheiden. Während bei Reproduktion, sowie nahem und weitem Transfer, das Wissen bereits in der Form vorhanden ist, wie es in der Aufgabe zur Anwendung kommen soll, sind beim kreativen Problemlösen Aufgaben denkbar, bei denen zunächst einmal neues Wissen geschaffen werden muss, um eine Aufgabe lösen zu können. Die Kognitionspsychologie konnte zeigen, dass bei solchen Problemlöseaufgaben das vorhandene Wissen sogar ein Hindernis darstellen kann (z. B. funktionale Fixierung, vgl. Anderson 1989). Allenfalls können bestimmte Problemlöseheuristiken genutzt werden. Die Subkategorie kreative Problemlöseaufgaben deckt im Wesentlichen die oberste Stufe der Dimension kognitive Prozesse bei Anderson & Krathwohl (2001) ab, nämlich das „Erschaffen“ (create). Es handelt sich um

Aufgaben, die von Schülerinnen und Schülern eine kreative Leistung, das heißt, das Zusammenfügen vorhandener Elemente verlangen, so dass neues Wissen entsteht, das für die Lösung der Aufgabe erforderlich ist.

Beispiel: Nach der Teilnahme an der Modulveranstaltung sind die Studierenden in der Lage, sich unter Bezugnahme auf Evidenz aus der Bildungsforschung für bestimmte Vorgehensweisen in ihren Berufsfeldern zu entscheiden.

Bei diesem Beispiel ist nicht klar, auf welche Situationen die Studierenden reagieren sollen. Darüber hinaus ist offen, welche Art von Wissen angewendet werden muss, um entsprechende Aufgaben zu lösen. Jedoch wurde im Seminar sichergestellt, dass die Studierenden wissen, wie sie Evidenz aus der Forschung nutzen können, um auftretende Probleme zu interpretieren.

Offenheit von Aufgaben

Definierte und konvergente Aufgabe

Aufgaben, die einen eindeutigen Arbeitsauftrag bzw. eine klar identifizierbare Fragestellung umfassen. Die Gliederung der Aufgabe ist typischerweise kleinschrittig. Es ist eine einzige Lösung gesucht bzw. richtig (ist nicht zwingend explizit angegeben). Es gibt nur einen Lösungsweg. Die Aufgabe kann zum Beispiel nur mit einem Ansatz (z.B. Berechnung) erfolgen. Aufgaben im Wahlantwortformat mit mehreren möglichen Antworten gehören zu dieser Art Aufgaben, da die Lösung über Ankreuzen erfolgt und keine individuellen Lösungswege ermöglicht werden.

Definierte und divergente Aufgabe

Aufgaben mit einem eindeutigen Arbeitsauftrag bzw. einer eindeutigen Fragestellung, die jedoch mehrere richtige Lösungen bzw. Lösungswege zulassen (in den meisten Fällen werden die Studierenden darauf hingewiesen). Es sind unterschiedliche Möglichkeiten zur Beantwortung möglich (berechnen, schätzen, funktional darstellen). Es kann außerdem freigestellt sein, ob die Antwort per Rechnung, in Prosa oder als Zeichnung erfolgt. Näherungslösungen sind möglich.

Schlecht definierte und divergente Aufgaben

Aufgaben, bei denen eine Problemsituation die einzige Handlungsaufforderung ist. Studierende erhalten Informationen über ein Problem, eine Situation, etc. Unterschiedliche Fragestellungen sind denkbar und damit auch mehrere Lösungen und Lösungswege. Die Problemstellung ist dabei unscharf definiert. Die Lösungen müssen nicht exakt sein, sondern können auch Schätzungen beinhalten. Zur Lösung bedarf es der Integration verschiedener domänenspezifischer Kenntnisse und Fähigkeiten. Eventuell besteht die Notwendigkeit der Erweiterung der Wissensbasis. Benötigte Daten sind gegebenenfalls nicht angegeben, sondern müssen geschätzt werden. Ein weiteres Merkmal ist die methodische Reflektion zum Vorgehen bei der Problemlösung.

Literatur

Anderson, J. R. (2001). Kognitive Psychologie. Heidelberg: Spektrum.

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Addison Wesley Longman, Inc.

Maier, U., Kleinknecht, M., Metz, K. & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 28(1), 84-96. http://blogs.phsg.ch/lernarrangements/files/2014/06/Maier-et-al_2010.pdf (abgerufen: 27.07.2016)

Shavelson, R. J., Ruiz-Primo, M. A., Li, M., & Ayala, C. C. (2003). *Evaluating new approaches to assessing learning*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Schritt 4: Die Festlegung von Bewertungskriterien

Nachdem in den ersten drei Schritten die Erstellung und Überprüfung von kompetenzorientierten Prüfungsaufgaben thematisiert wurde, geht es im vierten Schritt um die Erstellung von Bewertungskriterien, um objektives, reliables und valides Prüfen sicher zu stellen. Dieser vierte Schritt sollte immer parallel mit der Aufgabenerstellung (Schritt 2) erfolgen, wird hier jedoch in einem gesonderten Schritt ausführlich behandelt. Es wird skizziert, wie sich ein umfassendes Bewertungsraster in wenigen Schritten erstellen lässt. Zwei ergänzende Beispiele veranschaulichen diese Schritte und die Bestandteile des Bewertungsrasters.

Bewertungskriterien reichen von der Festlegung, welche Antworten als richtig, teilrichtig oder falsch gewertet werden, bis hin zu genauen Regeln für die Punktevergabe. Die Objektivität, Zuverlässigkeit und Gültigkeit der Leistungseinschätzung wird durch gute und transparente Bewertungskriterien erhöht, da diese neben der Abbildung und Gewichtung von angestrebten Lernergebnissen in einem konkreten Bewertungsmaßstab auch dafür sorgen, dass unterschiedliche Bewerterinnen und Bewerter auf Grundlage dieses Maßstabs zu gleichen Ergebnissen kommen. Um die Kriterien der Bewertung möglichst transparent und objektiv zu gestalten, eignet sich ein sogenanntes Bewertungsraster.

Prinzipien bei der Aufgabenbewertung

Bereits bei der Aufgabenstellung wurden systematische Überlegungen angestellt, wie bestimmte Lernergebnisse auf konkrete Aufgabenanforderungen abgebildet werden können. Es sollten deshalb ausschließlich die Aspekte in den Antworten bewertet werden, die sich auf das entsprechende Lernergebnis beziehen. Dabei sollte darauf geachtet werden, dass jede Aufgabe konkret einem (und nur einem!) angestrebten Lernergebnis zugeordnet werden kann. Wurden zu einem angestrebten Lernergebnis viele kleinere Aufgaben gestellt, so kann es dann ausreichen, zwischen richtig oder falsch zu unterscheiden, wobei im Sinne der Bewertung genau zu charakterisieren ist, was eine richtige oder falsche Lösung ausmacht. Eine Einschätzung des *Niveaus* der Erreichung eines Lernergebnisses erfolgt dann darüber, wie viele Aufgaben ein Studierender zu einem Lernergebnis richtig gelöst hat, also *zwischen den Aufgaben*.

Bei umfassenderen Aufgaben kann es dagegen nützlich sein, eine feinere Abstufung der Niveaus *innerhalb der Aufgabe* vorzunehmen. In diesem Fall kann eine Unterscheidung in „richtig“, „teilweise richtig“ oder „falsch“ sinnvoll sein. Wichtig ist auch hier, dass die Anzahl der Kategorien inhaltlich mit Lernergebnisbezug begründet wird. Vor allem bei der Zuordnung von Punkten für richtige Teillösungsschritte sollte hierauf geachtet werden, damit die Punkte tatsächlich systematisch den zu überprüfenden Leistungsstand abbilden. Das Kriterium ist dann zum Beispiel: man vergibt einen Punkt für die jeweilige Teillösung – also beispielsweise pro mathematischer Umformung. Teillösungsschritte sollten stattdessen zu sinnvollen Bewertungseinheiten mit Bezug zum angestrebten Lernergebnis zusammengefasst werden.

Prinzipielle Überlegungen für die Bewertung beziehen sich also für jede Aufgabe einerseits darauf, in welcher Weise Lösungsaspekte dem beabsichtigten Lernergebnis zugeordnet werden können und auf der anderen Seite auf das Bewertungsniveau, ausgehend von Überlegungen zum notwendigen Differenzierungsgrad der Beurteilung.

Das Bewertungsraster

In einem Bewertungsraster werden die einzelnen Aufgaben mit Antwortkategorien („Richtig“, „Falsch“ etc.), deren charakterisierenden Beschreibung, Ankerbeispielen und der Zuweisung von Werten (z.B. Punkte) tabellarisch dargestellt. Zu jeder Antwortkategorie wird dabei beschrieben, welche Eigenschaften eine Antwort erfüllen muss, um der Kategorie zugeordnet zu werden. Die Beschreibungen beinhalten also Eigenschaften, Charakteristika, Indikatoren, Bewertungs- und Entscheidungsregeln, die eine zweifelsfreie Zuordnung unterschiedlicher aber qualitativ gleichwertiger Antworten ermöglichen. Denkbar ist dabei (1) eine Aufzählung aller möglichen richtigen Antworten oder (2) eine generische Beschreibung von Eigenschaften, die eine Antwort erfüllen muss,

um einer entsprechenden Antwortkategorie zugeordnet zu werden. Neben der Musterlösung, als Beispiel für eine eindeutig richtige Antwort, sollten wo möglich Grenzfälle angeführt werden, die gerade noch einer Kategorie zugeordnet werden können. Darüber hinaus werden Ankerbeispiele angegeben, die eine Zuordnung erleichtern. Indem die Kriterien in tabellarische Form (= Bewertungsraster) gebracht werden, lassen sich die Bewertungskriterien für jede Aufgabe transparent nachvollziehen (vgl. Tabelle 3).

Tabelle 3: Exemplarisches Bewertungsraster

optional ¹ Code	Antwortkategorien	Charakterisierung der Kategorien	Ankerbeispiele	optional Punktevergabe
C2	Richtig	Als richtig werden Antworten gewertet, die....		1 Punkte
C1	teilweise richtig	Als teilweise richtig werden Antworten gewertet, die....		0,5 Punkte
C0	Falsch	Als falsch werden Antworten gewertet, die....		0 Punkte

Schritt für Schritt zum Bewertungsraster

Im Folgenden ist skizziert, wie ein Bewertungsraster in wenigen Schritten erstellt werden kann.

1. Erstellen einer Musterlösung der gesamten Klausur (empfohlen)

Eine Musterlösung dient als Rahmen und liefert Beispiele für „ideale“ Antworten. Diese sollte bereits beim Konzipieren von Aufgaben miterstellt werden, da dadurch einerseits ein bewusstes Vorgehen bei der Aufgabenstellung gefördert wird. Andererseits werden bei der Erstellung der Musterlösung oftmals unklare oder nicht eindeutige Aufgabenstellungen, ein zu hoher Schwierigkeitsgrad der Aufgabe o.ä. sichtbar, so dass die Aufgabe noch vor der Verwendung verändert werden kann.

2. Festlegen von Bewertungskategorien

Im nächsten Schritt ist zu überlegen, wie viele Bewertungsstufen für die jeweilige Aufgabe sinnvoll sind. Wie bereits erläutert, hängt dies unter anderem davon ab, wie spezifisch die Aufgabe gestellt ist und wie differenziert die Leistung jeweils erfasst werden soll. Zum Beispiel kann die Fähigkeit, ein Integral aufzustellen, in mehrere Schritte unterteilt werden. Es wäre dann zu überlegen, ob die Differenzierung in diese Lösungsschritte mit der dahinterliegenden Fähigkeit begründbar ist, indem zum Beispiel jeder Schritt eine Teilfacette der Fähigkeit abbildet. In diesem Fall könnten drei oder mehr Kategorien festgelegt werden, um die Ausprägung der Fähigkeit differenzierter erfassen zu können. Handelt es sich beim sicheren Aufstellen eines Integrals dagegen um eine Mindestanforderung, kann die Lösung auch dichotom bewertet werden – entweder das Integral wurde richtig aufgestellt oder nicht. Ein Aspekt, der bei der Aufstellung von Regeln für die Punktevergabe beachtet werden muss, ist allerdings, dass hierdurch bei den Aufgaben unterschiedliche Bewertungsskalen entstehen können und somit eine Vergabe der Gesamtnote durch einen Punktescore – also ein einfaches Addieren der Punkte über alle Aufgaben - zu diagnostisch falschen Ergebnissen führen kann (siehe Abschnitt zu Punktevergaberegeln).

¹ Das Beispiel enthält zusätzlich die Möglichkeit, den Bewertungskategorien einen Code zuzuweisen, der eine mögliche statistische Auswertung der Prüfungsergebnisse erleichtert. Diese Codes können bei der Korrektur auch anstelle einer ausführlichen Beschreibung der Bewertung verwendet werden. Die Anzahl der Punkte kann ebenfalls in das Bewertungsraster aufgenommen werden, um die Zuweisung von Punkten bei der Korrektur zu erleichtern. Der Code und die Punktevergabe sind unabhängig voneinander und müssen sich nicht entsprechen.

3. Charakterisierung der Kategorien

Im nächsten Schritt muss festgelegt werden, wann eine Lösung mit den jeweiligen Kategorien zu werten ist. Eine solche Charakterisierung beinhaltet deshalb entsprechende Entscheidungsregeln. Diese gehen von Überlegungen aus, die bereits bei der Entscheidung für die Anzahl an Lösungskategorien angestellt wurden, wie zum Beispiel eine lernergebnisorientierte Begründung einzelner Teillösungsschritte. Dabei sollten unbedingt alle Aspekte aufgezählt werden, die in der Lösung enthalten sein müssen, um die jeweilige Kategorie (mindestens) zu erfüllen. Bei Aufgaben, die ein konkretes Ergebnis vorsehen, ist dieser Schritt verhältnismäßig einfach umzusetzen, da oftmals Ergebnisse oder Zwischenergebnisse ausschlaggebend für die Punktevergabe sind. Bei offenen Aufgaben unterscheiden sich die Antworten der Studierenden dagegen oftmals ganz erheblich. In diesem Fall müssen generische Beschreibungen gefunden werden, die Antworten soweit charakterisieren, dass sinngemäß richtige Antworten auch korrekt eingeordnet werden.

4. Erstellen von Ankerbeispielen

Durch die Verwendung von Ankerbeispielen kann die Zuordnung von Lösungen zu den Bewertungskategorien erleichtert werden. Ankerbeispiele können konstruiert oder aus bereits bearbeiteten Prüfungen entnommen werden. Eine weitere Möglichkeit zur Sammlung von Ankerbeispielen besteht darin, die Antworten von Kolleginnen und Kollegen zu sammeln, die beim Proberechnen der Prüfung entstehen.

5. Regeln für Vergabe von Punkten festlegen

Ziel der Punktevergabe ist es, die Leistung eines Studierenden in einem numerischen Wert abzubilden. Oftmals wird dazu lediglich ein Summenscore gebildet und anschließend mittels eines Notenschlüssels die Gesamtnote gebildet. Dabei kann es jedoch zu Abbildungsfehlern kommen, wenn zum Beispiel einzelne Aufgaben unterschiedlich skaliert sind (z.B. dichotome und intervallskalierte Antwortskalen, siehe Schritt zur Erstellung von Bewertungskategorien). So kann es vorkommen, dass eine Aufgabe mit vier Antwortstufen qualitativ gleichbedeutend ist mit einer Aufgabe, die nur richtig oder falsch beantwortet werden kann, für die erste Aufgabe jedoch maximal 3 Punkte vergeben werden, für die zweite jedoch nur einer. Auch hier müssen also Überlegungen stattfinden, in welchem Ausmaß die einzelnen Aufgaben zur Gesamtnote beitragen sollen. Insofern die einzelnen angestrebten Lernergebnisse nicht gleichwertig in Bezug auf die Gesamtleistung sind, kann es hilfreich sein, sich erst einmal zu überlegen, welchen Anteil eine einzelne Aufgabe zur Bewertung eines jeweiligen Lernergebnisses beitragen soll und dann die angestrebten Lernergebnisse entsprechend für die Gesamtnote zu gewichten.

Häufig wird die Punktevergabe in der Hochschulpraxis am Zeitaufwand für die Bearbeitung bemessen, um ein gemeinsames Bewertungskriterium für alle Aufgaben zu haben. Inhaltlich ist eine solche Punktevergabe allerdings zu hinterfragen, da der Zeitaufwand nicht unbedingt dem Anteil der Aufgabe für die Beurteilung eines angestrebten Lernergebnisses entsprechen muss. Einer einfachen, aber aufwendigen Aufgabe zu einem untergeordneten angestrebten Lernergebnis können auf diese Weise mehr Punkte zugesprochen werden als einer kürzeren, aber voraussetzungsreicheren zu einem wichtigen angestrebten Lernergebnis.

Es sollten also immer verschiedene Kriterien für die Punktevergabe mit einbezogen werden.

Weitere Hinweise zur Anwendung des Bewertungsrasters

Sind mehrere Personen an der Korrektur beteiligt, sollte sichergestellt sein, dass jede/r Korrigierende ausreichend mit dem Bewertungsraster und Korrekturhinweisen vertraut ist. Dies kann zum Beispiel durch eine gemeinsame Schulung oder durch Verschicken von Korrekturhinweisen erreicht werden. Sollten während der Korrektur Antworten auftauchen, die nicht eindeutig einer Bewertungskategorie zugeordnet werden können, sollten diese als neue Grenzfälle dem Bewertungsraster hinzugefügt und eindeutige Regelungen festgelegt und an alle kommuniziert werden.

Durch eine stichprobenweise Zweitkorrektur kann die Zuverlässigkeit der Erstkorrektur überprüft werden, insbesondere bei Aufgaben, bei denen sehr unterschiedlichen Antworten möglich sind. Rechtlich ist in Bayern das Zweiprüferprinzip nur auf **nicht** bestandene schriftliche Prüfungen beschränkt (Art. 61 Abs. 3 Satz 2 Nr. 10 BayHschG).

Aufgabenbeispiel

Das beschriebene Vorgehen zur Bewertung von Prüfungsaufgaben auf der Grundlage eines Bewertungsrasters wird an nachfolgendem Beispiel² verdeutlicht. Das Beispiel wurde den Vergleichsarbeiten Mathematik für die 8. Jahrgangsstufe entnommen und bot sich für dieses Informationsblatt an, da die Aufgabe mit einer Kombination aus Berechnungsaufgabe und offener gestellter Aufgabe zwei Aufgabentypen beinhaltet, die an der Universität weitverbreitet sind. Der Fokus liegt darauf, die obigen Ausführungen zur Korrektur und Bewertung von Prüfungsantworten zu verdeutlichen. Daher wurde das fachliche Niveau des Beispiels bewusst einfach gehalten.

Teilaufgabe 1

Die Teilaufgabe 1 zum Kauf eines DVD Players besteht aus einer Berechnungsaufgabe.

Kauf eines DVD-Players

In einem Online-Shop im Internet ist ein Angebot für einen tragbaren DVD-Player zu finden. Der ursprüngliche Preis dieses DVD-Players von 99,99 € wird um 20 % reduziert.

Wenn man den Rechnungsbetrag vom Bankkonto abbuchen lässt, bekommt man auf diesen reduzierten Preis nochmal einen Rabatt von 5 %.

Teilaufgabe 1:

Gib den Preis für den DVD-Player an, wenn man ihn ohne Abbuchung vom Bankkonto bezahlt. Runde auf ganze Cent.

_____ €

Bei Teilaufgabe 1 gibt es nur ein richtiges Ergebnis. Das Bewertungsraster enthält lediglich die zwei Kategorien „richtig“ oder „falsch“. Die Charakterisierung der Bewertungskategorien erfolgt über die Angabe des richtigen Ergebnisses. Damit ist eindeutig festgelegt, dass ausschließlich Antworten, die das genaue Ergebnis enthalten, als richtig gewertet werden. Alle anderen Antworten werden als falsch gewertet (siehe Tabelle 4). Der Rechenweg wird in diesem Beispiel nicht bewertet. Soll der Rechenweg in die Beurteilung einfließen, könnte dies über eine weitere Kategorie, z.B. „teilweise richtig“, realisiert werden.

Tabelle 4: Bewertungsraster Teilaufgabe 1

optional Code	Bewertungskategorien	Charakterisierung der Kategorien	Ankerbeispiele	optional Punktevergabe
C1	Richtig	Als richtig werden Antworten gewertet, die das genaue Ergebnis von 79,99 € oder den gerundeten Wert von 80,00 € enthalten	79,99 € oder 80,00 €	1 Punkte
C0	Falsch	Als falsch werden Antworten gewertet, die abweichende Ergebnisse enthalten.		0 Punkte

² Aufgabe aus den VERA-Vergleichsarbeiten Mathematik, 8. Jahrgangsstufe: <http://www.iqb.hu-berlin.de/system/taskpool/getTaskFile?fileid=9777>; aufgerufen: 04.02.2014

Tabelle 5: Bewertungsraster Teilaufgabe 2

Code	Bewertungs-kategorien	Charakterisierung der Kategorien ³	Ankerbeispiele	Punkte
1	richtig	Als richtig werden Antworten gewertet, die eine richtige Begründung enthalten, in welcher darauf verwiesen wird, dass sich der Preisnachlass von 20 % und der Rabatt von 5 % auf unterschiedliche Grundwerte beziehen . [Anm.: Auch das Berechnen des jeweils zu zahlenden Betrages wird akzeptiert. Es können auch gerundete Werte verwendet werden. Die Lösung wird auch akzeptiert, wenn bei den Zwischenergebnissen mit aufgerundeten Werten weitergerechnet wird.]	<ul style="list-style-type: none"> • Der Preisnachlass von 20 % bezieht sich auf den Grundwert 99,99 €. Der Rabatt von 5 % bezieht sich dann jedoch auf einen niedrigeren Grundwert. • Erst 20 %, dann 5 %: $0,80 * 99,99 \text{ €} = 79,99 \text{ €}$ dann $0,95 * 79,99 \text{ €} = 75,99 \text{ €}$ Sogleich 25 %: $0,75 * 99,99 \text{ €} = 74,99 \text{ €}$ (Stimmt nicht überein!). • Alternative Berechnungsmethoden, z. B. "Dreisatz" 	1 Punkt
0	falsch	Antworten, die keine Begründung, sondern lediglich eine Behauptung enthalten (siehe erstes Ankerbeispiel). Weiterhin alle unvollständigen, fehlerhaften oder falschen Antworten, die die oben genannten Eigenschaften nicht aufweisen, insbesondere der Verweis auf die unterschiedlichen Grundwerte.	<ul style="list-style-type: none"> • Nein, wenn man 25 % abzieht, kostet es mehr, als wenn man 20 % und dann noch 5 % abzieht. • Ja, man kann die beiden Prozentsätze addieren, da sie sich auf denselben Preis beziehen. → falsch • (Grenzfall) Nein, denn wenn ich gleich 25 % abziehe, ist dies mehr, als wenn ich zunächst nur 20 % und dann 5 % abziehe. [Anm.: Hier wird nicht erläutert, worauf diese Tatsache beruht.] • (Grenzfall) Die beiden Grundwerte sind verschieden! 	0 Punkte

In diesem Beispiel wurden die Bewertungskategorien „richtig“ und „falsch“ verwendet, um die Einheitlichkeit der Bewertungskategorien über den Test hinweg zu erhalten. Aus inhaltlicher Sicht könnte für die Begründungsaufgabe ebenso gut das inhaltliche Niveau der Antworten differenziert werden. Betrachtet man das erste Ankerbeispiel in der Bewertungskategorie „falsch“, beinhaltet diese Antwort sinngemäß richtige Anteile. Das fachliche Niveau der Antwort ist jedoch nicht ausreichend, da nicht explizit auf den Grundwert verwiesen wird. Gleiches gilt für das dritte Ankerbeispiel. Das vierte Ankerbeispiel bezieht sich auf den Grundwert, berücksichtigt jedoch nicht die Zahlenwerte. Diese Antworten sind streng genommen nicht falsch, erfüllen jedoch das erwartete fachliche Niveau nicht ausreichend. Eine Differenzierung hinsichtlich des fachlichen Niveaus würde in diesem Fall die Leistung eventuell sogar treffender klassifizieren. Eine weitere Überlegung könnte sein, diese Grenzfälle über eine Kategorie „teilweise richtig“ von Antworten, die keine richtigen Anteile enthalten, zu unterscheiden. Dadurch würden die Testergebnisse Informationen über Testpersonen enthalten, die den Anforderungen zwar nicht vollständig aber zumindest teilweise gerecht werden. Der diagnostische Wert der Testergebnisse würde dadurch erhöht.

³ Korrekturhinweise aus den VERA-Vergleichsarbeiten Mathematik, 8. Jahrgangsstufe: <http://www.iqb.hu-berlin.de/system/taskpool/getTaskFile?fileid=9779>, aufgerufen: 04.02.2014

Literatur

Anderson, L.W. & Krathwohl, D.R., Hrsg. (2001). A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's

Vergleichsarbeiten (VERA). Bildungsserver Berlin-Brandenburg. <https://www.iqb.hu-berlin.de/vera/aufgaben/ma1> (abgerufen: 14.07.2016)

Bayerische Hochschulgesetz (2006).BayHSchG. <http://www.gesetze-bayern.de/Content/Document/BayHSchG-61?AspxAutoDetectCookieSupport=1> (abgerufen: 28.07.2016)

Schritt 5: Die Auswertung von Prüfungsergebnissen

Nachdem in den ersten vier Schritten dargestellt wurde, wie sich eine kompetenzorientierte Prüfung inklusive Bewertungskriterien gestalten lässt, geht es im fünften Schritt um die umfassende Analyse der Prüfungsergebnisse. Deshalb geht dieser Schritt auf wichtige Voraussetzungen ein, die gegeben sein müssen, um solch eine Analyse durchzuführen. Anschließend werden zentrale statistische Kennwerte erläutert und wichtige Hinweise für die Analyse von Prüfungsergebnissen gegeben. Es wird zudem kurz erläutert, welche Rückschlüsse man auf Basis der Ergebnisse auf die Prüfung, einzelne Aufgaben aber auch auf die zugrundeliegende Lehrveranstaltung ziehen kann. Im letzten Abschnitt wird die Erstellung einer Aufgabendatenbank auf Basis von Aufgabencharakteristika beschrieben.

Mit der Korrektur der Prüfung ist der Prüfungsprozess noch nicht abgeschlossen. Die Analyse von Prüfungsdaten kann ausgesprochen hilfreich zur Bewertung der Qualität einer Prüfung bzw. einzelner Aufgaben sein. Prüfungsdaten erlauben nämlich nicht nur die Bewertung der Studierenden, sondern sind auch die Basis für empirisches Feedback zur Qualität der Prüfung, der Aufgaben und der Lehrveranstaltung. Prüfungsdaten fällt damit eine maßgebliche Bedeutung für eine systematische Verbesserung der Prüfung zu.

Die vorliegende Handreichung gibt einen wesentlichen Überblick über die Möglichkeiten der Analyse von Prüfungsdaten. Zusätzlich beschreibt sie konkrete Möglichkeiten zur Umsetzung in Microsoft-Excel. Als weiteres Hilfsmittel steht zu diesem Dokument eine Vorlage für Excel zur Verfügung, die - nach Eingabe und Anpassung der individuellen Prüfungsdaten –automatisch alle wichtigen Kennwerte berechnet.

Voraussetzung: Differenzierte Datenerfassung

Die zentrale Voraussetzung für die genannten Möglichkeiten der Auswertung ist die differenzierte Erfassung von Prüfungsaufgaben auf Aufgabenebene. Es ist nicht ausreichend, ausschließlich das Gesamtergebnis für jeden Studierenden zu erfassen. Stattdessen müssen die Punktezahlen für jede Aufgabe/Teilaufgabe und für jeden Studierenden festgehalten werden.

Wie differenziert die Aufgaben dabei erfasst werden, ist unter anderem auch davon abhängig, wie differenziert die dahinterstehenden Anforderungen im Vorfeld ausgearbeitet wurden. Oftmals stellt auch das Lösen/Nicht-Lösen von Teilschritten eine wichtige Information in Bezug auf die angenommenen Fähigkeiten dar. Entscheidend sollte also der Informationsgehalt sein, der mit jeder weiteren Differenzierung gewonnen wird. Zum Beispiel bietet sich bei umfangreichen, aber wenigen Aufgaben eine Erfassung auf Teilaufgabenebene an. Sind die Aufgaben dagegen eher kleinschrittig konzipiert, sind Daten auf Ebene der gesamten Aufgabe ausreichend.

In Tabelle 6 ist beispielhaft eine Datenmatrix für eine Prüfung mit 3 Aufgaben und 100 Studierenden skizziert.

Tabelle 6: Datenmatrix für eine Prüfung mit drei Aufgaben und 100 Prüfungsfällen

	Aufgabe 1	Aufgabe 2a	Aufgabe 2b	Aufgabe 3	Summe
Studierender 1	3	2	4	1	10
Studierender 2	3	3	5	3	14
Studierender 3	1	0	2	2	5
.					
.					
.					
Studierender 100	2	3	5	2	12

Wichtige Kennwerte

Neben einfachen Maßzahlen wie Mittelwert, Standardabweichung und Minima/Maxima sind besonders die Aufgabenschwierigkeit sowie die Bestimmung der Trennschärfe und Korrelationsmaße aussagekräftig mit Hinblick auf die Qualität einzelner Aufgaben, die Gesamtprüfung, Personenfähigkeiten sowie die Lehrveranstaltung. In Tabelle 7 sind alle relevanten Kennwerte aufgelistet. Dazu werden für die einzelnen Kennwerte kurz angegeben, worüber diese Informationen liefern können. Der anschließende Abschnitt erläutert die jeweiligen Kennwerte dann genauer.

Tabelle 7: Kennwerte und analytischer Nutzen

Kennwerte	Analytischer Nutzen
Mittelwert, Standardabweichung und Minima/Maxima	<ul style="list-style-type: none"> ➤ Erste Eindrücke über die Prüfung und Aufgaben gewinnen ➤ Informationen über das Leistungsniveau der Studierenden ➤ Informationen zur Leistungsdifferenzierung
Aufgabenschwierigkeiten	<ul style="list-style-type: none"> ➤ Identifikation zu schwieriger oder zu einfacher (Teil)Aufgaben ➤ Abgleich der erwarteten mit der tatsächlichen Schwierigkeit einzelner (Teil)Aufgaben ➤ Sortieren von Aufgaben zu einer bestimmten Anforderungsdimension nach Schwierigkeitsgrad ➤ Rückschlüsse auf Fähigkeits- bzw. Kompetenzstufen
Trennschärfe, Diskriminationsindex und Korrelationsmaße	<ul style="list-style-type: none"> ➤ Rückmeldungen zur Leistungsdifferenzierung ➤ Grad der Repräsentation einer oder mehrerer Dimensionen durch einzelne Aufgaben ➤ Fragen zur Plausibilität der angenommenen Fähigkeits- und Kompetenzstruktur

Mittelwert, Standardabweichung und Minima/Maxima

Zentrale statistische Kennwerte wie Mittelwert, Standardabweichung und Minima/Maxima eignen sich insbesondere dazu, sich einen Überblick über die Prüfungsergebnisse zu schaffen und zu ersten Einschätzungen zum Schwierigkeitsgrad der Aufgaben und dem Leistungsniveau sowie der Leistungsdifferenzierung der Personen zu gelangen.

Minimum und Maximum

Anhand der *Minima und Maxima* lässt sich ablesen, ob Aufgaben von keinem oder von allen Personen gelöst werden konnten. Mit dem Minimum wird dabei die geringste bei einer Aufgabe erreichte Punktzahl bezeichnet. Das Maximum benennt die maximal erreichte Punktzahl bei einer Aufgabe. Wurden bei einer Aufgabe zum Beispiel nur maximal 60% der erzielbaren Punkte erreicht, liegt also das Maximum bei 6 von 10 Punkten, so liegt die Vermutung nahe, dass die Aufgabe zu schwierig ist (für einen genaueren Kennwert siehe Abschnitt zur Aufgabenschwierigkeit) oder aber Merkmale enthält, die die Aufgabe auch für leistungsstarke Personen nicht komplett lösbar machen. Insofern die Aufgabe auch nach einer eingehenden Überprüfung nicht als zu schwierig erscheint, so sollte überlegt werden, wie in der Lehrveranstaltung besser auf diese speziellen Anforderungen vorbereitet werden kann.

Anwendung

In MS-Excel lassen sich das Minimum und Maximum ähnlich der Summenfunktion bestimmen. Das heißt, es gilt ein Zielfeld auszuwählen und dann im Register „Start“ in der Gruppe „Bearbeiten“ auf den Optionenpfeil der Schaltfläche „AutoSumme“ zu klicken und die entsprechenden Zellen zu wählen (alternativ können manuell die Befehle MIN() oder MAX() verwendet werden, wobei zwischen den Klammern jeweils ein Eingabebereich gewählt werden muss).

Es bietet sich dabei an, die Kennwerte für einzelne Aufgaben jeweils unter den Aufgabenspalten berechnen zu lassen. So sind die wichtigsten Kennwerte für die jeweiligen Aufgaben auch optisch den jeweiligen Aufgaben zugeordnet.

Arithmetisches Mittel

Mit dem arithmetischen Mittel (=Mittelwert oder Durchschnittswert) lässt sich die bei einer Aufgabe im Durchschnitt erreichte Punktzahl angeben.

Bei Aufgaben, die lediglich mit richtig oder falsch bewertet wurden, also auf einem dichotomen Skalenniveau beruhen, ist der Mittelwert gleichbedeutend mit der Aufgabenschwierigkeit. Diese entspricht dann der Lösungswahrscheinlichkeit der Aufgabe und kann auch entsprechend interpretiert werden, sprich: je geringer der Wert, desto schwieriger die Aufgabe (siehe entsprechender Abschnitt).

Der Mittelwert gibt einen ersten Hinweis auf das Leistungsniveau der Studierenden bei den jeweiligen Aufgaben, wobei er natürlich immer vor dem Hintergrund der maximal erreichbaren Punktzahlen interpretiert werden sollte. Die Aufgabenschwierigkeit ist also hier die aussagekräftigere Kennzahl. Ebenso sollte beim Mittelwert beachtet werden, dass er von der Streuung der Werte unabhängig ist (siehe nächster Abschnitt).

Anwendung

In MS-Excel lässt sich der Mittelwert einfach als die mittels Summenfunktion (Schaltfläche oder SUM()) als die Summe aller Punkte zu einer Aufgabe), dividiert durch die Anzahl an Prüfungsfällen, bestimmen. Die Anzahl an Prüfungsfällen kann entweder manuell eingegeben oder mittels der ANZAHL()-Funktion berechnet werden.

Standardabweichung

Die *Standardabweichung* gibt die durchschnittliche Streuung der Punktzahlen um ihren Mittelwert an und gibt erste Hinweise darauf, wie differenziert die jeweilige Aufgabe die Leistung der Studierenden misst, wie „gut“ sie also die gewählte Skala ausschöpft. Streut eine Aufgabe, für die 12 Punkte vergeben werden, zum Beispiel sehr gering um einen Mittelwert von 6 Punkten, so bedeutet dies, dass sich die Punktezahlen der Studierenden zum Beispiel vornehmlich im Bereich von 5-7 Punkten bewegen. In diesem Fall könnte es sein, dass die Aufgabe zum Beispiel aus einem sehr (zu) einfachen und einem sehr (zu) schwierigen Teil besteht. Entsprechend könnte die Aufgabe zukünftig um die Teile gekürzt werden, die zu schwierig oder zu einfach sind. Beim Lösen der Aufgabe kann so zusätzlich Zeit gespart werden.

Anwendung

In MS-Excel lässt sich die Standardabweichung mittels der Formel

=STABW(ZELLE:ZELLE)

berechnen. Die Varianz (als quadrierte Standardabweichung) erhält man über die Formel

=VARIANZ(ZELLE:ZELLE).

Die Standardabweichung ist als die mittlere Abweichung einfacher zu interpretieren, die Varianz wird für weitere Kennwerte als Zwischenschritt benötigt.

Aufgabenschwierigkeiten

Die Aufgabenschwierigkeit entspricht der Wahrscheinlichkeit, dass ein Studierender eine Aufgabe richtig löst. Bei dichotomen Aufgaben entspricht sie deshalb dem Anteil der Studierenden, die die Aufgabe richtig gelöst haben. Allgemein ist die Aufgabenschwierigkeit als das Verhältnis der durchschnittlich bei einer Aufgabe erzielten Punktzahl im Verhältnis zur Maximalpunktzahl (Aufgabemittelwert / Maximalpunktzahl) definiert. Sie ist also auf einen Bereich zwischen 0 und 1 normiert, wobei ein höherer Wert für eine leichtere (!) Aufgabe steht.

Dadurch können Aufgaben mit unterschiedlichen Bewertungsskalen (Punktezahlen) verglichen werden. Grundsätzlich kann bei Werten > 0.8 von leichten Aufgaben, bei Werten von 0.2 bis 0.8 von mittelschweren Aufgaben und bei Werten von < 0.2 von Aufgaben mit einer hohen Schwierigkeit ausgegangen werden, wobei je nach Kriterium hierbei auch individuelle Bereiche festgelegt werden können.

Mit den Aufgabenschwierigkeiten lassen sich grundlegend erst einmal leichte Aufgaben von schwierigen unterscheiden und zu leichte und zu schwierige Aufgaben identifizieren. Bei der Konzeption einer Aufgabe wird dabei immer auch ein vermuteter Schwierigkeitsgrad zugrunde gelegt, der auf bestimmten Annahmen über die Lösung der Aufgabe beruht. Die empirische (=statistische) Aufgabenschwierigkeit kann deshalb vor allem dafür herangezogen werden, den vermuteten (=erwarteten) Schwierigkeitsgrad auf seine Plausibilität hin zu überprüfen. So kann es vorkommen, dass eine als mittelschwerig angenommene Aufgabe tatsächlich zu leicht oder zu schwierig ausfällt. Die Aufgabenschwierigkeiten sind deshalb immer auch *vor dem Hintergrund von Schwierigkeitsannahmen* zu betrachten und nicht isoliert. Dafür bietet es sich zum Beispiel an, die berechneten Aufgabenschwierigkeiten in einer Tabelle mit den erwarteten Schwierigkeiten (geschätzte Werte oder Werte aus vergangenen Prüfungen) gegenüberzustellen.

Anwendung

Für die Bestimmung der Aufgabenschwierigkeiten sollte in MS-Excel in einem Schritt unter jeder Aufgabenspalte in einer Zelle eingetragen werden, wie viele Punkte bei der Aufgabe maximal erreicht werden konnten.

Die Aufgabenschwierigkeiten können dann folgendermaßen berechnet werden:

Aufgabenschwierigkeit = Aufgabenmittelwert/Maximalpunktzahl

Eine interessante Möglichkeit der Hierarchisierung von Aufgaben bietet sich dann, wenn bei der Konzeption von Aufgaben darauf geachtet wurde, diese entlang ganz bestimmter Anforderungsdimensionen (=angestrebter Lernergebnisse) zu gestalten (zu einer empirischen Überprüfung dieser „Dimensionierung“ siehe nächster Abschnitt). In diesem Fall lassen sich Aufgaben nach diesen Dimensionen gruppieren (zum Beispiel Aufgabe 1, 3 und 5 überprüfen die Fähigkeit, mathematisch zu integrieren) und anhand ihrer empirischen Aufgabenschwierigkeiten innerhalb der Dimension hierarchisieren. Auf diese Weise können Niveaustufen unterschieden und interpretiert werden, auch in Hinsicht darauf, in welcher Reihenfolge die dafür notwendigen Teilfähigkeiten vermittelt werden sollten, also wie diese aufeinander aufbauen. Die Hierarchisierung von Aufgaben anhand ihrer Schwierigkeitsgrade bietet also eine gute Grundlage, um fachdidaktische Überlegungen über Kompetenz- und Fähigkeitsstufen anzustellen.

Die Aufgabenschwierigkeit ist kein absolutes Charakteristikum einer Aufgabe, sondern immer in Bezug auf die geprüfte Stichprobe zu betrachten. Außerdem ist Vorsicht bei der Interpretation von Aufgabenschwierigkeiten geboten, wenn ein nicht unerheblicher Teil der Studierenden zum Beispiel eine Aufgabe aufgrund von Zeitmangel nicht bearbeitet hat. Die geringe Lösungswahrscheinlichkeit liegt dann nicht in Schwierigkeit generierenden Merkmalen der Aufgabe, sondern der Gesamtpfprüfung begründet.

Trennschärfe, Diskriminationsindex und Korrelationsmaße

Eine Prüfung muss zwischen Personen mit stark und wenig ausgeprägten Fähigkeiten und Kompetenzen unterscheiden können. Entsprechend lässt sich jede Aufgabe dahingehend betrachten, inwieweit sie zur Differenzierung beiträgt. Ideal wäre für die Überprüfung (im Sinne der externen Validität) ein Außenkriterium (z.B. messbarer Erfolg in beruflichen Handlungssituationen), mit dem das Abschneiden bei einer Aufgabe mehr oder weniger gut korreliert. Ein solches Außenkriterium steht aber in Prüfungen nicht zur Verfügung. Entsprechend wird in der Regel die Korrelation zu einem „Hilfskriterium“ bestimmt, nämlich dem Gesamtergebnis der Prüfung.

Trennschärfe

Das Maß für die Korrelation einer Aufgabe mit dem Gesamtergebnis wird als **Trennschärfe** bezeichnet. Eine hohe Trennschärfe liegt dann vor, wenn Studierende mit einer hohen Gesamtpunktzahl auch bei der jeweiligen Aufgabe eine hohe Punktzahl erzielen und umgekehrt. Die Trennschärfe gibt damit an, wie gut eine Aufgabe zwischen guten und schlechten Kandidaten unterscheidet. Von einer guten Trennschärfe wird in der Regel bei Werten über 0.4 gesprochen, eine akzeptable Trennschärfe liegt bei Werten zwischen 0.2 und 0.4 vor. Auch negative Trennschärfewerte sind möglich. In diesem Fall bedeutet dies, dass schlechte Studierende bei einer Aufgabe sogar besser abschneiden als gute Studierende.

Anwendung

Auch die Trennschärfe lässt sich in MS-Excel relativ einfach bestimmen, da die Trennschärfe die Korrelation zwischen der Einzelskala einer Aufgabe mit der Gesamtskala angibt. Entsprechend kann die Trennschärfe für Aufgabe 1 mit dem Befehl

```
KORREL(STUD1_A1:STUD100_A1;STUD1_SUMME:STUD100_SUMME)
```

berechnet werden, wobei STUD1_A1 für die vom ersten Studierenden bei Aufgabe 1 erzielten Punkte stehen und STUD1_SUMME für die von Studierenden 1 insgesamt erzielten Punkte. Korreliert werden also die Leistungen bei einzelnen Aufgaben mit der Gesamtleistung über alle Studierenden hinweg.

Soll die Trennschärfe in Bezug auf eine bestimmte Anforderungsdimension bestimmt werden, so ist anstelle der Spalte über die Gesamtleistung eine Spalte über die erzielten Punkte bei den Aufgaben aus dieser Dimension zu bilden.

Problematisch am Konzept der Trennschärfe ist, dass in der Regel das *Gesamtergebnis* als Kriterium herangezogen wird. Die Trennschärfe ist also nur unter der Bedingung sinnvoll zu interpretieren und als Entscheidungshilfe zu sehen, dass tatsächlich alle Aufgaben auf eine Fähigkeitsdimension zurückzuführen sind. Dies ist in der Regel aber bei Prüfungen nicht der Fall. So kann es durchaus vorkommen, dass sich die Stärken von Studierenden zwischen verschiedenen Anforderungsdimensionen insoweit unterscheiden, dass einzelne Aufgaben nicht mehr unbedingt alle gut mit dem Gesamtscore korrelieren müssen. Wird also zum Beispiel zwischen vier angestrebten Lernergebnissen unterschieden, so kann es Sinn ergeben, die Aufgaben vor allem in Bezug auf die einzelnen Lernergebnisse zu analysieren, also in Bezug auf vier verschiedene Kriterien bzw. Skalen. Entsprechend kann es in diesem Fall sinnvoller sein, die Trennschärfe einer einzelnen Aufgabe in Bezug auf den Gesamtscore aller anderen Aufgaben zu einem bestimmten Lernergebnis zu bestimmen. Praktisch eignet sich hierfür eine Bestimmung der *internen Konsistenz* einer Skala, da bei gängiger Statistiksoftware (zum Beispiel SPSS) die Trennschärfe des Items in Bezug auf die untersuchte Skala mit ausgegeben wird.

Diskriminationsindex

Eine alternative Maßzahl dafür, wie gut eine Aufgabe zwischen Personen differenziert, ist der sogenannte **Diskriminationsindex**. Für seine Bestimmung werden für jede Aufgabe drei Gruppen über die erzielten Punkte gebildet. In der Regel werden dazu die oberen 27 % und die unteren 27 % bestimmt (stattdessen können aber auch zum Beispiel Terzile verwendet werden) und jeweils einer Gruppe zugeordnet. Anschließend wird die Aufgabenschwierigkeit in jeder dieser Gruppen bestimmt und anschließend die Differenz gebildet. Die Differenz gibt dann an, wie gut die Aufgabe differenziert. Im Gegensatz zur Trennschärfe gibt der Diskriminationsindex nicht die Schärfe der Trennung an, sondern die Stärke.

Anwendung

Die Berechnung des Diskriminationsindex mit MS-Excel ist etwas aufwendiger, da hierfür eine entsprechende Gruppenvariable gebildet werden muss. Ein geeignetes Vorgehen ist die Bestimmung des oberen und unteren 27%-Quantils für jede Aufgabe über die Excel-Funktion:

=QUANTIL.EXKL(B4:B23;0,27), wobei B4:B23 die jeweilige Datenspalte der Aufgabe ist.

Anschließend werden die Ausprägungen der Gruppenvariablen definiert, zum Beispiel 1 für die „schwache“ und 2 für die „starke“ Gruppe.

Anschließend sollte in einer leeren Datenmatrix (Datenmatrix kopieren und Aufgabenwerte löschen) berechnet werden, welcher Gruppe jeder Fall für jede Aufgabe angehört. Hierfür eignet sich die WENN-DANN-Funktion:

=WENN(B4<B\$25;1;WENN(B4>B\$26;3)), wobei mit B\$25 und B\$26 die Zellen markiert werden, in denen die Schwellenwerte bestimmt wurden. In diesem Beispiel wurde ein exklusives Kriterium gewählt, das heißt, es werden alle Werte der Gruppe zugeordnet, die nicht gleich dem Schwellenwert sind. Je nach Aufgabenpunktzahl und Fallzahl kann auch ein inklusives Kriterium gewählt werden.

Anschließend sollte für die beiden Gruppen der Mittelwert gebildet werden. Hierfür eignet sich die SUMMEWENN-Funktion:

=(SUMMEWENN(B\$4:B\$23;\$L27;Daten!B\$4:B\$23))/(ZÄHLENWENN(B\$4:B\$23;L27)), wobei L27 hier den Zellen der Gruppenausprägungen entspricht.

Anschließend kann der Schwierigkeitsgrad innerhalb der Gruppen auf bekannte Weise bestimmt werden:

=B27/B\$24

Der Diskriminationsindex berechnet sich abschließend aus der Differenz der beiden Aufgabenschwierigkeiten für die starke und die schwache Gruppe.

Reliabilität und interne Konsistenz

Mit der **internen Konsistenz** (in der Regel in Form von Cronbachs α) wird dagegen die Reliabilität einer Messung eingeschätzt. Auch diese Kennzahl ist eher als Annäherung zu verstehen, da die tatsächliche Bestimmung der Reliabilität – also wie zuverlässig die Prüfung misst – nur mit einer Testwiederholung oder einer Paralleltestung realisierbar wäre. Es kann aber mathematisch gezeigt werden, dass Cronbach's α ein unteres Maß für die Reliabilität darstellt, die tatsächliche Reliabilität also sogar eher noch höher ausfällt. Cronbach's α gibt an, inwieweit eine bestimmte Aufgabe eine Gruppe von Aufgaben (oder eben auch die Gesamtprüfung) repräsentiert. Dies bedeutet nicht, dass die Aufgabe auch tatsächlich das misst, was sie messen soll. Zum Beispiel können auch alle Aufgaben etwas anderes messen und eine hohe interne Konsistenz aufweisen. Cronbachs α ist zudem von der Aufgabenanzahl abhängig: Der Wert steigt entsprechend, je mehr Aufgaben in Bezug auf ein Lernergebnis oder in der Gesamtprüfung enthalten sind. Dies ist deshalb der Fall, da mit jeder zusätzlichen Aufgabe mehr Information zur Verfügung steht und die Zuverlässigkeit der Leistungseinschätzung erhöht wird.

Eine Analyse der internen Konsistenz einer Skala (im Sinne einer Reihe von Aufgaben, die eine bestimmte Anforderungsdimension messen sollen) mit Cronbachs α setzt in allen Fällen voraus, dass eine theoretische Annahme darüber besteht, welche Aufgaben eine ähnliche Fähigkeit messen. Zum Beispiel können alle Aufgaben, die „mathematisches Integrieren“ messen sollen, auf ihre interne Konsistenz hin überprüft werden. Wird für die Analyse eine Statistiksoftware wie SPSS verwendet, werden bei der Berechnung von Cronbach's α mehrere Werte ausgegeben: Der Kennwert selbst, als Gesamtschätzung für die Konsistenz der gewählten Aufgaben und für jede Aufgabe ein alternativer Wert Alpha, für den Fall, dass diese Aufgabe aus der Skala bzw. Prüfung entfernt würde. Würde sich durch das Weglassen einer Aufgabe Cronbach's α erhöhen, so ist zu vermuten, dass diese Aufgabe etwas anderes misst als die übrigen Aufgaben. Durch die Bestimmung der internen Konsistenz von sämtlichen Aufgaben, die ein spezifisches Lernergebnis abbilden sollen, lassen sich also auch Hinweise darauf finden, inwieweit diese Annahme plausibel ist – mit der Einschränkung, dass die Ähnlichkeit der Aufgaben natürlich auch in einem anderen Aspekt liegen kann. In diesem Sinne ist Cronbach's α keine Maßzahl für die Einschätzung der Validität, die weiterhin gut begründet sein muss. Eine Möglichkeit, ähnliche Aufgaben zu identifizieren, wenn nur wenige theoretische Annahmen gemacht wurden (oder um diese zu überprüfen), wird im nächsten Abschnitt beschrieben.

Anwendung

Auch Cronbach's α lässt sich mit MS-Excel bestimmen. Dafür müssen einige Zwischenwerte berechnet werden. Benötigt werden:

1. Summe der Einzelvarianzen jeder Aufgabenskala ($\sum_{j=1}^k \sigma_j^2$): Zeilensumme der Einzelvarianzen
2. Anzahl der Aufgaben (k): Manuell abzählen oder über ANZAHL(AUFGABE1:AUFGABEJ)
3. Gesamtvarianz (σ^2): Varianz der Spalte „Summe“

Die einzelnen Werte können am besten jeweils einzeln in Zielfeldern berechnet werden und stehen dann auch für weitere Analysen zur Verfügung.

Anschließend kann Cronbach's α berechnet werden. In diesem Fall wird dafür die Varianz über alle Aufgaben mit der Summe der Einzelvarianzen ins Verhältnis gesetzt und durch die Anzahl an Variablen gewichtet:

Cronbach's $\alpha = (k/(k-1)) * (1 - \sum_{j=1}^k \sigma_j^2 / \sigma^2)$, wobei für k, $\sum_{j=1}^k \sigma_j^2$ und σ^2 die jeweiligen Zellen mit den entsprechenden Zwischenwerten anzugeben sind.

In diesem Fall wird Cronbach's α über *alle* Variablen, also unter Einbezug aller Aufgaben, bestimmt. Die Auswertung kann auch für einzelne Indizes erfolgen, das heißt, zum Beispiel nur über alle Aufgaben, die für ein bestimmtes angestrebtes Lernergebnis stehen. Entsprechend sind hier die Summe der Einzelvarianzen und die Gesamtvarianzen nur über die selektierten Aufgaben zu bilden und für die Berechnung zu verwenden.

Etwas aufwendiger ist die Bestimmung des veränderten Cronbach's α , das heißt, der Veränderung des Wertes unter der Bedingung, dass eine bestimmte Aufgabe weggelassen wird. Hierfür muss für jede Aufgabe Cronbach's α erneut *ohne Einbezug der jeweiligen Aufgabenvariablen* berechnet werden. Steigt der Wert nach Weglassen der jeweiligen Aufgabenvariablen, so bedeutet das, dass die entsprechende Aufgabe möglicherweise etwas anderes misst als die anderen Aufgaben.

Tipp: Steht eine Statistiksoftware zur Datenauswertung zur Verfügung, so sollte diese auf jeden Fall genutzt werden, da die Berechnung deutlich komfortabler durchgeführt werden und außerdem gleichzeitig die Interkorrelationsmatrix ausgegeben werden kann (siehe nächster Abschnitt).

Inter-Item-Korrelationen

Mittels einer **Interkorrelationsmatrix**, die zum Beispiel im Zuge der Bestimmung von Cronbach's α mit ausgegeben werden kann, können Anhaltspunkte für das Gruppieren von Aufgaben gefunden werden. In einer Interkorrelationsmatrix werden sämtliche Korrelationen zwischen einzelnen Aufgaben dargestellt. Eine hohe Korrelation zwischen zwei Aufgaben ist ein möglicher Indikator dafür, dass diese Aufgaben etwas Ähnliches messen. Insofern kann spalten- oder zeilenweise für jede Aufgabe entschieden werden, zu welchen anderen Aufgaben Ähnlichkeiten bestehen (Beispiele siehe Excel-Sheet zur Datenanalyse).

Im Idealfall lassen sich so klar Gruppen von Aufgaben anhand der Stärke ihrer Korrelationen bilden. In der Praxis finden sich dabei wahrscheinlich Aufgaben, die mit mehreren anderen Aufgaben hoch korrelieren, die aber ihrerseits wiederum andere Verwandtschaften aufweisen. In diesem Fall können meistens keine klar abgrenzbaren Gruppen gebildet werden. Insofern ist die Analyse der Korrelationsmatrix als heuristisches Vorgehen zu betrachten, um erste Anhaltspunkte über das Zusammenspiel von Aufgaben zu finden. Eine Methode, die auf der Interkorrelationsmatrix von Aufgaben aufbaut und die zum Ziel hat, Aufgaben zu Faktoren zu gruppieren, denen gemeinsame Dimensionen zugrunde liegen, ist die Faktorenanalyse. Da diese jedoch relativ aufwendig umzusetzen ist, wird sie deshalb hier nicht näher erläutert.

Ergänzung: Gruppenspezifische Analysen

Die genannten Kennwerte lassen sich natürlich auch **gruppenspezifisch** auswerten, das heißt, es können zum Beispiel über die gesamten Ergebnisse hinweg drei Gruppen gebildet werden, die jeweils das obere, mittlere und untere Leistungsdrittel der Studierenden enthalten. Steigen in diesem Fall die *Mittelwerte* zum Beispiel über die Gruppe an, so differenziert die Aufgabe gut.

Erstellung einer Aufgabendatenbank auf Basis von Aufgabencharakteristika

Neben dem Nutzen der Diagnostik, der Optimierung der Aufgaben, der Gesamtprüfung und der Lehrveranstaltung bietet die Analyse von Prüfungsdaten noch einen mittel- bis langfristigen Vorteil, der im Folgenden kurz beschrieben werden soll: die Möglichkeit zur Anlage einer Aufgabendatenbank.

Zweck einer Aufgabendatenbank

Eine Aufgabendatenbank dient dem Archivieren von Aufgaben nach festgelegten Kriterien. Aufgaben können so für zukünftige Prüfungen nutzbar gemacht werden. Ebenso können Aufgaben über die Veranstaltung hinaus mit Kollegen und Kolleginnen geteilt werden. Aufgaben können so – in veränderter Form – auch in anderen Prüfungen Anwendung finden. In einer geteilten Aufgabendatenbank können sich Lehrende Ideen zur Aufgabengestaltung holen. Nicht zuletzt sorgt das Archivieren anhand spezifischer Kriterien für eine bessere Vergleichbarkeit.

Voraussetzungen für die Nutzung einer Aufgabendatenbank

Die Voraussetzung ist, dass die Aufgaben nach denselben Standards erstellt und Prüfungsergebnisse anschließend anhand derselben Kriterien ausgewertet werden. Es sollte also – vor allem wenn die Datenbank gemeinsam geführt wird – ein verbindlicher Standard vereinbart werden. Ebenso sollte ein gemeinsamer Server zur Verfügung stehen, sodass ein entsprechendes Dokument (z.B. ein Excel-Sheet) von allen bearbeitet werden kann. Vor der wiederholten Verwendung bietet es sich an, Aufgaben erst einmal zu sammeln, da eine zu rasche Wiederholung bestimmter Aufgaben mit nur geringen Variationen zu Lerneffekten führen kann.

Tipps zum Vorgehen beim Aufbau einer Aufgabendatenbank

Als Basis eignet sich auch hier ein Excel-Sheet. Auf dem entsprechenden Server kann – bei einer geteilten Datenbank – für jede Dozierende und jeden Dozierenden ein entsprechendes Dokument angelegt werden. Innerhalb können die einzelnen Blätter – bei einer Veranstaltung – Aufgaben für einzelne angestrebte Lernergebnisse oder – bei mehreren Veranstaltungen – Aufgaben für die jeweilige Klausur beinhalten.

Die Aufgaben selbst können separat in einem Unterordner abgelegt und im Excel-Sheet jeweils direkt verlinkt werden. Dasselbe gilt für Musterlösungen und Bewertungsraster.

In der Tabelle können die Aufgaben zeilenweise aufgelistet werden. In den Spalten können wichtige Informationen zu den jeweiligen Aufgaben hinterlegt werden, wie zum Beispiel:

- Aufgabencode zur eindeutigen Identifizierung: Sollte bei einem geteilten Aufgabenpool gemeinsam festgelegt werden, z.B. Dozierendenkürzel+Veranstaltungsnummer+Aufgabennummer.
- Verlinkung auf externe Dokumente mit der Aufgabe, einer Musterlösung oder Bewertungskriterien.
- Angestrebtes Lernergebnis: Welchem angestrebten Lernergebnis wird die Aufgabe zugeordnet?
- Didaktischer Kommentar: Wie ist die Aufgabe aufgebaut, um das angestrebte Lernergebnis zuverlässig zu erfassen?
- Logische Aufgabenanalyse: Wurde eine logische Aufgabenanalyse durchgeführt und auf welche Weise bzw. mit welchem Ergebnis?
- Veränderungsmöglichkeiten: Wie lässt sich die Aufgabe geschickt variieren für weitere Klausuren? Welche Verbesserungsmöglichkeiten sind angedacht?
- Wann wurde die Prüfungsaufgabe das letzte Mal verwendet?
- Statistische Kennziffern für einzelne Aufgaben:
 - ✓ Maximalpunktzahl für Aufgabe
 - ✓ Geschätzte Bearbeitungsdauer (eventuell durch Probelösen ermittelt)
 - ✓ Aufgabenschwierigkeit
 - ✓ Trennschärfe
 - ✓ Diskriminationsindex
- Angaben für die Gesamtklausur (eventuell auf separatem Blatt):
 - ✓ Genaue Beschreibung der Veranstaltung
 - ✓ Angestrebte Lernergebnisse
 - ✓ Durchschnittliche Anzahl von Prüfungsfällen
 - ✓ Angabe der Interkorrelationsmatrix
 - ✓ Cronbachs α für die Gesamtklausur oder für einzelne Teildimensionen

Literatur

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.

Möltner A; Schellberg D.; Jünger, J. (2006). Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild. Themenheft Prüfungen*, 23(3):Doc53. http://www.ectaveo.ch/Mediathek/2015/08/Moeltner_Quantitative_Analysen.pdf (abgerufen: 08.08.2016)

Herausgegeben von

ProLehre

Susanne Klatten –
Stiftungslehrstuhl für
Empirische
Bildungsforschung

Technische Universität
München
Augustenstraße 44
80333 München

Technische Universität
München
Marsstr. 20-22
80335 München

Autorenschaft

Dr. Nadine Schlomske-
Bodenstein

Dr. Alexandra Strasser

Dr. Christoph Schindler

Florian Schulz