

Prüfungen

Definition

Mit ‚Prüfung‘ meinen wir im folgenden Text jede Form der bewerteten Leistungserhebung oder –messung im Hochschulkontext.

Funktion

Prüfungen erfüllen eine Vielzahl unterschiedlichster Funktionen. Aus gesellschaftlicher Sicht stellen sie wichtige Initiationsriten dar, sie regeln den Zugang zu Bildungsressourcen, zu Berufsfeldern und -privilegien und damit indirekt auch zu Ansehen, Wohlstand und sozialem Status. Insbesondere sind sie auch rechtliche Legitimation solcher Ungleichheiten. Aus ökonomischer Sicht sollen sie die gezielte Entwicklung und den effizienten Einsatz von Humankapital ermöglichen. Aus bildungspolitischer Sicht sollen sie eine aussagekräftige Differenzierung Auszubildender hinsichtlich ihres Potentials oder ihrer Kompetenzen gewährleisten und damit eine selektive Zulassung zu passenden Bildungsangeboten sowie eine Zertifizierung erworbener Bildung ermöglichen. Darüber hinaus sind sie, aus Sicht der Lehre, wichtige Garanten und Lenker studentischen Lernens. Aus Sicht der Lernenden stellen sie zu überwindende Hürden in ihrer Lebensplanung dar, sind Leitplanken ihres Bildungsalltags, Gelegenheiten erworbene Fähigkeiten zu demonstrieren oder gefürchtete Momente potentiellen Scheiterns. Sie sind ein überaus starker Motivator und ein entscheidender Katalysator persönlicher Entwicklung. Andererseits können sie Entwicklung auch behindern, durch übermässigen Stress, das Verstärken negativer Selbstkonzepte, bis hin zur Begünstigung pathologischer Entwicklungen wie Depressionen oder Suchterkrankungen. Angesichts der zahlreichen Schlüsselrollen, welche Prüfungen in unserer Kultur zukommen, ergibt sich eine hohe Verantwortung der Prüfenden, Prüfungen nach bestem Wissen und Gewissen zu gestalten, unter Einbezug prüfungsdidaktischer und messtheoretischer Methoden, gemäss dem aktuellen Stand der Wissenschaft und Praxis.

Gütekriterien

Prüfungen stellen Instrumente zur Messung studentischer Leistungen dar. Um aussagekräftige Ergebnisse gewährleisten zu können, muss eine Prüfung einer Reihe von **Gütekriterien** genügen: Die Messung soll unabhängig von der messenden Person wie auch der Untersuchungssituation sein (**Objektivität**), sie soll verlässlich und präzise sein (**Reliabilität**) und sie soll gültig sein – d.h. sie soll tatsächlich und ausschliesslich das interessierende Merkmal messen – und nicht irgendetwas anderes (**Validität**). Des Weiteren soll die Messung keine Personen oder Personengruppen bevor- oder benachteiligen (**Fairness**). Die Kandidaten sollen ihre Ergebnisse nicht verfälschen können (**Unverfälschbarkeit**). Der Aufwand für die Messung soll im Verhältnis zu deren Nutzen stehen (**Ökonomie**). Die Kandidaten sollen im Bild sein darüber, was auf welche Weise geprüft wird (**Transparenz**). Die Messung soll von den Kandidaten oder Drittpersonen als aussagekräftig und gültig *empfunden* werden (**Augenscheinvalidität**). Und schliesslich soll die Messung für die Kandidaten zumutbar sein

(**Zumutbarkeit**). Das wichtigste Gütekriterium ist *immer* die Validität, denn ohne Gültigkeit verliert eine Messung ihre Berechtigung.

Messung

Gegenstand einer Prüfung im Hochschulbereich sind im Allgemeinen Leistungen in den Bereichen **kognitive** Fähigkeiten (**Intelligenz**), fachrelevantes Wissen oder fachrelevante **Kompetenzen**. Anders als bei physikalischen Grössen, wie z.B. dem Körpergewicht, handelt es sich hierbei um nicht direkt beobachtbare **latente Eigenschaften**. Diese können nur indirekt gemessen werden, z.B. indem sie aus dem Verhalten – in unserem Kontext dem Lösen einer Prüfung – erschlossen werden (**Psychometrie**). Als erstes benötigen wir eine Vorstellung der latenten Eigenschaft(en), welche wir erheben möchten, um diese anschliessend verbindlich zu definieren (**Konstrukt**). Im Prüfungskontext geschieht dies durch die Vereinbarung von **Lernzielen**. Prüfungen messen also das Erreichen von Lernzielen. Diese müssen einer Messung zugänglich gemacht werden, indem passende Fragen und/oder Aufgaben (**Items**) entwickelt werden (**Operationalisierung**). In der ‚Prüfung‘ werden diese Aufgaben dann von den Prüflingen bearbeitet. Die Durchführung der Prüfung stellt dabei die eigentliche Messung dar, die bearbeitete Prüfung liefert die **Rohdaten**. Diese werden anschliessend regelhaft bewertet – z.B. durch den Abgleich mit vorgegebenen Referenzantworten (**Musterlösung**) – und auf diskrete Variablen abgebildet (‚Punktvergabe‘/**Quantifizierung**). Aus der Gesamtheit der bewerteten Aufgaben/Punkte wird üblicherweise ein Prüfungsgesamtwert (‚Gesamtpunktzahl‘) berechnet. Um eine einfache und einheitliche Interpretation der Ergebnisse zu ermöglichen, wird dieser schliesslich auf eine allgemeine, normative (**Noten**-)Skala abgebildet (**Normierung**). Bei **kriteriumsorientierten** Prüfungen erfolgt die Normierung durch den Vergleich der Prüfungsergebnisse mit einem vorgängig definierten, inhaltlich begründeten Leistungsmaßstab. Wird andererseits normiert durch einen Vergleich der Prüfungsergebnisse mit den Resultaten einer Vergleichsstichprobe, dann spricht man von einer **normorientierten** Prüfung. Die Normierung entspricht der Eichung eines Messinstruments. Der gesamte beschriebene Vorgang, von der Definition der zu messenden latenten Eigenschaften, bis zur Normierung der Prüfung, ist dabei so zu gestalten, dass die daraus resultierende Messung die weiter oben besprochenen Gütekriterien so gut wie möglich erfüllt. Da die Gütekriterien nie vollständig erfüllt werden können, unterliegt jede Messung Fehlern bzw. Ungenauigkeiten. Fehler, welche auf mangelnde Reliabilität, also Messgenauigkeit, zurückzuführen sind, nennt man **Messfehler**. Sie sind unsystematisch und folgen einer Zufallsverteilung. Alle weiteren Fehler, welche auf Mängel in den übrigen Gütekriterien zurückzuführen sind, sind systematischer Natur und werden durch **Störvariablen** verursacht.

Didaktik

Aus didaktischer Sicht unterscheidet man verschiedene Formen von Prüfungen gemäss ihrer Funktion. **Summative Prüfungen** dienen der Selektion und Zertifizierung von Kandidaten. Liegt dabei der Schwerpunkt auf der Einschätzung von Entwicklungspotentialen bzw. zukünftig zu erwartenden Leistungen, so spricht man von **prognostischen Prüfungen**. Bei **diagnostischen Prüfungen** andererseits sollen gegenwärtige Leistungen eingeschätzt werden. In Abgrenzung zu den summativen Prüfungen stehen die **formativen Prüfungen**. Sie sollen Informationen über vollzogene Lernfortschritte liefern und helfen die Gestaltung von Lehre und Lernen zu strukturieren und verbessern. Prüfungen erfüllen darüber hinaus eine zentrale Funktion in der

Gewährleistung der Lernmotivation (extrinsische **Motivation**) wie auch der Steuerung des Lernverhaltens. Deshalb ist es wichtig, Prüfungen und Lehre aufeinander abzustimmen (**Alignment**). Formal kann man zwischen **Powertests** und **Speedtests**, auch Schnelligkeitstests und Niveautests, unterscheiden. Speedtests beinhalten einfache Aufgaben, welche unter Zeitdruck gelöst werden müssen. Powertests beinhalten schwierige Aufgaben, welche möglichst ohne Zeitdruck gelöst werden sollen. Bei Speedtests ergibt sich die Prüfungsschwierigkeit aus der geforderten Bearbeitungsgeschwindigkeit, bei Powertests aus der Aufgabenkomplexität. Auf der Itemebene schliesslich unterscheidet man zwischen **geschlossenen Items** und **offenen Items**. Bei ersteren wird den Prüflingen eine endliche Anzahl von Lösungsmöglichkeiten zur Auswahl gegeben (z.B. Multiple Choice Fragen). Bei offenen Items muss die Lösung frei generiert werden (z.B. ein Aufsatz, Vortrag). Items, welche Eigenschaften beider Fragetypen vereinen, nennt man **halboffene Items** (z.B. Lückentext).

Taxonomie

Prüfungssitems lassen sich inhaltlich durch zwei Hauptaspekte klassifizieren (**Taxonomie von Bloom**): Erstens den (Lern-) Inhalt, auf welchen sich ein Item bezieht, und zweitens die zur erfolgreichen Bearbeitung des Items benötigten (kognitiven) Prozesse. In einer Überarbeitung der Taxonomie von Bloom wird der Aspekt ‚Prozesse‘ weiter aufgespalten in eine ‚Wissensdimension‘ und eine ‚kognitive Prozessdimension‘. Die Wissensdimension orientiert sich dabei lose an der Klassifizierung unterschiedlicher Gedächtnisformen in der kognitiven Psychologie. Die kognitive Prozessdimension soll eine Hierarchie hinsichtlich der Komplexität der zur Aufgabenlösung erforderlichen kognitiven Prozesse bilden. Die Taxonomie von Bloom hat sich in der didaktischen Praxis seit gut fünfzig Jahren etabliert und als praxistauglich erwiesen. Dies obwohl viele ihrer grundlegenden Annahmen und Konzepte im Konflikt mit Befunden und Methoden der psychologischen und neurowissenschaftlichen Forschung stehen. In der Tat haben diese Forschungsgebiete einen anderen Blickwinkel auf den Prüfungsprozess. Die Psychometrie liefert insbesondere eine andere Sicht auf die Dimensionalität einer Messung durch eine bestimmte Prüfung ‚Prüfung X‘. Die verschiedenen Prüfungssitems messen hier Informationsanteile des Konstrukts ‚Prüfung X‘. Durch statistische Analyse der erhobenen Daten können dabei die Merkmalsdimensionen (**Faktoren**) des untersuchten Konstrukts ‚Prüfung X‘ eruiert werden. Die Ausprägungsgrade dieser Merkmalsdimensionen können anschliessend für jeden Prüfling einzeln berechnet werden. Analog zur Taxonomie von Bloom ist eine Prüfung also ein mehrdimensionales Konstrukt. Der Unterschied zur Taxonomie von Bloom besteht aber darin, dass diese ihre Merkmalsdimensionen einfach von Vorherein postuliert. Eine empirische Validierung bleibt sie schuldig. Bei einem psychometrischen Verfahren hingegen werden die Dimensionen des Konstrukts ‚Prüfung X‘ *gemessen*, d.h. sie ergeben sich empirisch aus der Datenerhebung und können anschliessend interpretiert werden. Die **Neuropsychologie** liefert insbesondere einen anderen Blickwinkel auf das Thema Komplexität. Das Gehirn ist ein äusserst dynamisches Organ, welches sich zeitlebens verändert: Lernprozesse bewirken eine physische Neustrukturierung der beteiligten Hirnregionen und manifestieren sich bereits nach sehr kurzer Zeit (**Plastizität**). Eine klare Trennung zwischen ‚kognitiven Prozessen‘ und ‚Inhalten‘ gibt es dabei nicht. Stattdessen besteht das Gehirn aus neuronalen Netzen, welche durch ihr Zusammenspiel Funktionen erfüllen. Kognitive Inhalte und Prozesse – ‚Gedächtnis‘ – sind implizit durch solche Netze und deren Aktivierung manifest. Diese Netze entwickeln sich im Gehirn nach Bedarf, das heisst gemäss ihrer Nutzung. Gleichzeitig organisiert sich das Gehirn nach dem Prinzip der Effizienz: Häufige Prozesse

benötigen mit zunehmender Übung immer weniger Ressourcen. Beide Effekte sind in einem Vergleich zwischen Experten und Novizen empirisch nachweisbar: Novizen zeigen eine grössere und unspezifischere Aktivierung von Hirnregionen als Experten und andererseits sind bei Experten aufgabenrelevante Hirnregionen stärker ausgeprägt als bei Novizen. Vereinfacht gesagt ist Expertise dadurch charakterisiert, dass selbst komplexe Aufgaben durch einfache Gedächtnisabrufe gelöst werden können. Umgekehrt kennzeichnen sich Novizen dadurch, dass sie selbst einfache Aufgaben nur mittels starker Aktivierung zahlreicher neuronaler Netze lösen können. Diese Beobachtungen stehen dem impliziten Bloom'schen Postulat, dass sich Expertise durch das Beherrschen kognitiver Prozesse zunehmender Komplexität kennzeichne, diametral entgegen. Bei Expertise oder bei erworbenen Fähigkeiten spricht man in der Psychologie auch von **kristalliner Intelligenz**. Der Bloom'sche Komplexitätsbegriff steht dem Begriff der **fluiden Intelligenz** näher. Fluide Intelligenz beinhaltet logisches Denken und die Fähigkeit neue Probleme zu lösen.

Alle drei Sichtweisen haben ihre Berechtigung. Die Bloom'sche Taxonomie ist eine Heuristik, welche Prüfern hilft, didaktisch sinnvolle Prüfungen zu entwerfen. Die Psychometrie zeigt, was und wie wir mit Prüfungen messen können. Die (Neuro)Psychologie schliesslich hilft uns, Lernprozesse und ihre Korrelate in Form von Wissen oder Expertise zu verstehen.

Glossar

Alignment (,Constructive Alignment')

Der Begriff ,Constructive Alignment' wurde von John B. Biggs eingeführt. Er beruht auf einem konstruktivistischen Modell des Lernens und zielt auf eine einheitliche Gestaltung der Lehre ab. Er beinhaltet zwei zentrale Konzepte: (1) Lernende *konstruieren* Bedeutung beim Lernen, indem sie neue Erfahrungen mit bestehendem Wissen in Beziehung setzen. (2) Der Unterrichtende gestaltet das Lernen durch die Vorgabe klarer Ziele, die Gestaltung zielführender Aktivitäten sowie einer passende Evaluation der Zielerreichung: Lernziele, Unterricht und Evaluation sind ,aligned'.

Augenscheinvalidität

Nebengütekriterium. Meint das Ausmass, mit welchem ein Laie ein Instrument oder eine Messung als valide *empfindet*. Es handelt sich also um die subjektive Gültigkeit einer Messung. Die Augenscheinvalidität ist deshalb keine Form der Validität im eigentlichen Sinne. Beispiel: Das Abfragen von Faktenwissen *erscheint* einem Chemieprofessor als valide Methode zur Einschätzung der Eignung von Prüflingen für das weitere Chemiestudium. Das Abfragen von Faktenwissen hat für ihn hohe Augenscheinvalidität.

Diagnostische Prüfung

Prüfung, welche gegenwärtige Leistungen oder Fähigkeiten einschätzen soll. Beispiel: Abschlussprüfung.

Fairness

Nebengütekriterium. Eine Prüfung soll keine Person oder Personengruppen in systematischer Weise bevorzugen oder benachteiligen (Lernzielrelevante Faktoren ausgenommen). Beispiel: Eine Physikprüfung, welche in Satzaufgaben stets Bezug auf die Schweizer Geschichte nimmt, ist gegenüber nicht-Schweizern nicht fair.

Faktor

Die Faktoranalyse ist ein statistisches Verfahren, welches aus vielen verschiedenen manifesten Variablen, welche auf beobachtbaren Items basieren, auf wenige zugrunde liegende latente Eigenschaften (Faktoren) schliesst. Beispiel: In der Intelligenzdiagnostik lassen sich zwei Hauptfaktoren der Intelligenz nachweisen: Die kristalline und die fluide Intelligenz.

Fluide Intelligenz

Die Fähigkeit Probleme zu lösen, sich Gegebenheiten anzupassen oder zu lernen. Fluide Intelligenz ist weitgehend angeboren, erreicht ihr Maximum nach dem zwanzigsten Lebensjahr und nimmt danach langsam ab.

Formative Prüfung

Formative Prüfungen dienen der Verbesserung von Lehre und Lernen. Die Erhebung von Kompetenzen und Lernfortschritten ermöglicht eine bedürfnisgerechte Gestaltung der weiteren Lehr- und Lernprozesse.

Geschlossene Frage

Frage mit vorgegebenen Antwortmöglichkeiten. Richtige Antworten müssen objektiv und eindeutig als solche erkennbar sein. Die richtige Antwort wird *erkannt*.

Beispiel: Multiple Choice Fragen, Sortieraufgabe.

Gütekriterien

Verbindliche Qualitätskriterien für psychometrische Messungen. Man unterscheidet Haupt- und Nebengütekriterien.

Halboffene Frage

Frage mit eingegrenzten aber nicht vorgegebenen Antwortmöglichkeiten.

Beispiel: Lückentext.

Hauptgütekriterien

Die Hauptgütekriterien sind Objektivität, Reliabilität und Validität. Validität setzt Reliabilität voraus und Reliabilität Objektivität.

Beispiel: Das Wägen stellt eine objektive und reliable Messung dar. Will man aber die Körpergrösse messen, so stellt Wägen keine valide Art der Messung dar.

Intelligenz

Es existieren unterschiedliche Definitionen von Intelligenz. Für unsere Zwecke können wir Intelligenz als *kognitive Leistungsfähigkeit* definieren. Es existieren zahlreiche Unterteilungen von Intelligenz. Die für uns wichtigste ist diejenige in fluide und kristalline Intelligenz. Fluide Intelligenz beinhaltet die Fähigkeit Probleme zu lösen, sich Gegebenheiten anzupassen oder zu lernen. Sie ist weitgehend angeboren, erreicht ihr Maximum nach dem zwanzigsten Lebensjahr und nimmt danach langsam ab. Kristalline Intelligenz beinhaltet alles, was wir erlernt haben. Sie nimmt über das Leben hinweg – potentiell bis ins hohe Alter – stetig zu.

Item

Einzelne Aufgabe einer Prüfung. Jedes Item erhält in der Prüfungsbewertung einen (Punkt)Wert.

Kognition

Prozesse und Inhalte des ‚Denkens‘ im weitesten Sinne: Wissen, Lernen, Wahrnehmen, Aufmerksamkeit, Planen, Schlussfolgern, Erinnern, Problemlösen, ...

Kompetenz

Die Fähigkeit bestimmte Aufgaben zu lösen.

Konstrukt

Ein empirisch nicht direkt messbarer Sachverhalt, welcher indirekt aus messbaren Sachverhalten erschlossen werden muss. Man spricht auch von einem latenten Konstrukt, bzw. latenten Variablen oder latenten Eigenschaften. Der Prozess des Erschliessens ist die Operationalisierung.

Kristalline Intelligenz

Alles, was wir erlernt haben. Kristalline Intelligenz nimmt über das Leben hinweg – bis ins hohe Alter – stetig zu.

Kriteriumsorientierte Prüfung

Prüfung, welche anhand eines definierten Leistungskriteriums normiert wurde. Bei kriteriumsorientierten Prüfungen hängt das Bestehen der Prüfung davon ab, wie gut man diese im Vergleich zu einem definierten Leistungskriterium gelöst hat.

Beispiel: 70% der maximal erzielbaren Punkte werden für ein Bestehen der Prüfung benötigt.

Latente Eigenschaft

Nicht direkt beobachtbare Eigenschaft einer Person (z.B. Intelligenz, Kompetenz, Ängstlichkeit, Leistungsmotivation). Eigenschaft muss indirekt aus Verhaltensbeobachtungen oder Verhaltenskorrelaten erschlossen werden.

Beispiel: Intelligenz, Extraversion, Lesekompetenz, Lernmotivation.

Lernziele

Lernziele definieren die spezifischen angestrebten Ergebnisse, welche durch das studentische Lernen erreicht werden sollen. Lernziele sind ein zentrales Werkzeug zur Gestaltung und Steuerung von Lehre, Lernen und Prüfen.

Messfehler/‘Rauschen‘

Zufällige bzw. unsystematische Messfehler. ‚Ungenauigkeit‘ einer Messung. Unsystematische Messfehler betreffen die Reliabilität einer Messung.

Motivation

Die Prozesse der Initiierung, Steuerung und Aufrechterhaltung von Aktivitäten.

Beispiel: Lernmotivation, Leistungsmotivation.

Musterlösung

Bewertungshilfe für offene und halboffene Items. Beinhaltet prototypische ‚beste‘ Lösungen von Prüfungsisems und sind Bewertungsmaassstab für die Quantifizierung der Prüfungslösungen der Prüflinge.

Nebengütekriterien

Weitere Gütekriterien neben den Hauptgütekriterien.

Neuropsychologie

Wissenschaft, welche sich mit den neuronalen Korrelaten menschlichen Handelns und Erlebens befasst.

Normierung

Abbildung erhobener Prüfungsleistungen auf eine Referenzskala anhand eines definierten Leistungskriteriums (Kriteriumsorientierung) oder einer definierten Vergleichsstichprobe (Normorientierung).

Normorientierte Prüfung

Prüfung, welche anhand einer definierten Vergleichsstichprobe normiert wurde. Bei normorientierten Prüfungen hängt das Bestehen der Prüfung davon ab, wie gut man diese im Vergleich zur Vergleichsstichprobe gelöst hat.

Beispiel: Nur 50 Prüflinge dürfen bestehen, da nur so viele Ausbildungsplätze zur Verfügung stehen.

Objektivität

Hauptgütekriterium. Meint die Unabhängigkeit einer Messung von der Untersuchungssituation und/oder der Untersuchenden Person. Die Objektivität ist ein quantifizierbares Mass.

Beispiel: Assistent A bewertet Prüfungen nach anderen Kriterien als Assistent B.

Offene Frage

Frage ohne vorgegebene Antwortmöglichkeiten. Die objektive und reliable Abgrenzung in bessere und schlechtere Antworten gestaltet sich im Allgemeinen schwierig. Die Antwort muss *produziert* werden.

Beispiel: Aufsatz.

Ökonomie

Nebengütekriterium. Meint die Verhältnismässigkeit des Aufwands einer Prüfung im Vergleich zu ihrem Nutzen.

Operationalisierung

Die ‚Messbarmachung‘ einer latenten Eigenschaft. Die Entwicklung geeigneter Fragen oder Aufgaben, zusammen mit einem Bewertungsschema, um eine indirekte Messung der gesuchten latenten Eigenschaft zu ermöglichen.

Beispiel: Mit welchen Aufgaben kann ich das Erreichen des Lernziels X überprüfen?

Plastizität

Die Fähigkeit des Gehirns, sich entsprechend neuer Anforderungen (oder Einschränkungen) umzustrukturieren und anzupassen. Entgegen landläufigen Vorstellungen entwickelt und verändert sich das Gehirn bis ins hohe Alter.

Beispiel: Bei Taxifahrer findet ein Grössenzuwachs von Gehirnarealen statt, welche für die räumliche Orientierung zentrale Aufgaben übernehmen.

Powertest

Test mit anspruchsvollen Aufgaben, welche ohne Zeitdruck gelöst werden können. Die Tests werden so konstruiert, dass selbst die Besten an der Aufgabenschwierigkeit und nicht am Zeitdruck scheitern. Die Testschwierigkeit ergibt sich aus den Aufgabenschwierigkeiten.

Prognostische Prüfung

Prüfung, welche zukünftige Leistungen einschätzen soll.

Beispiel: Basisprüfung.

Prüfung

Mit ‚Prüfung‘ meinen wir jede Form der bewerteten Leistungserhebung oder –messung im Hochschulkontext.

Psychometrie

Teilgebiet der Psychologie, welches sich mit dem Messen latenter Eigenschaften befasst.

Quantifizierung

Meint die regelhafte Zuordnung einer Bewertung, in der Regel in Form eines Punktwerts, zu einem gelösten Prüfungssitem.

Reliabilität

Hauptgütekriterium. Meint die *Genauigkeit*, im Sinne von Reproduzierbarkeit, einer Messung. Reliabilität setzt Objektivität voraus. Die Reliabilität ist ein quantifizierbares Mass.

Beispiel: Ein Metermass mit Millimeterangabe ermöglicht eine reliablere Messung, als eines mit Zentimeterangabe.

Rohdaten

Messdaten, bevor sie bewertet, interpretiert oder aggregiert worden sind.

Skala

Relationssystem in welches Messergebnisse abgebildet werden.

Beispiele: Geschlecht (m/w), Temperatur (°C), Platzierung bei einem sportlichen Wettkampf (1./2./...), Punktwert für eine Aufgabe (1, 2, 3, ...), Noten (1-6).

Skalenniveau(s)

Nominal-, Ordinal-, Intervall- und Proportionalskala bilden die verschiedenen Skalenniveaus. Die Nominalskala bildet Kategorien ab. Sie erlaubt Aussagen über Gleichheit und Verschiedenheit. Die Ordinalskala bildet Rangreihenfolgen ab. Sie erlaubt Aussagen der Art grösser-kleiner, besser-schlechter, etc.. Bei Intervallskalen sind Differenzen interpretierbar, was Aussagen der Art $A-B=C-D$ ermöglicht. Bei Proportionalskalen sind auch Quotienten/Proportionen interpretierbar, was Aussagen der Art $A/B=C/D$ ermöglicht. Daten auf höheren Skalenniveaus lassen sich immer auf einem tieferen Skalenniveau interpretieren, z.B. darf man proportionalskalierte Daten auch als ordinalskalierte Daten interpretieren. Keinesfalls aber umgekehrt. Prüfungsnoten sind ordinalskaliert!

Beispiel: Nominalskala: männlich und weiblich sind verschieden. Ordinalskala: Der erstplatzierte war besser als der Zweitplatzierte. Intervallskala: 30°C ist um gleichviel wärmer wie 25°C, wie 23°C wärmer wie 18°C ist. Proportionalskala: 30m sind doppelt so lang wie 15m.

Speedtest

Test mit vielen, vergleichsweise einfachen Aufgaben, welche unter Zeitdruck gelöst werden müssen. Die Tests werden so konstruiert, dass selbst die Besten nicht alle Aufgaben in der vorgegebenen Zeit lösen können. Die Testschwierigkeit ergibt sich aus der Zeitanforderung.

Störvariable

Störfaktoren, welche zu systematischen Fehlern bzw. Verzerrungen in der Messung der interessierenden Merkmalsdimension führen. Störvariablen beeinträchtigen die Validität einer Messung.

Beispiel: Findet eine Prüfung in zwei verschiedenen Räumen statt und ist einer der beiden sehr lärmig, so stellt der Prüfungsraum bzw. der Lärm eine Störvariable dar.

Summative Prüfung

Ergebnisorientierte Prüfung, in Abgrenzung zu formativen Prüfungen. Im Mittelpunkt steht die abschliessende Einschätzung der Erreichung von Lernzielen.

Beispiel: Abschlussprüfung, Basisprüfung, Semesterendprüfung.

Taxonomie von Bloom

Klassifizierungsheuristik für Lernziele. Die überarbeitete Fassung von D.R. Krathwohl unterscheidet zwei Eigenschaften von Lernzielen: (a) den Inhalt und (b) Prozesse oder Arbeitsschritte, welche mit diesen Inhalten vorgenommen werden sollen. Diese werden weiter in zwei Dimensionen unterteilt: (1) eine ‚Wissensdimension‘ und (2) eine ‚kognitive Prozessdimension‘.

Transparenz

Nebengütekriterium. Alle relevanten Eigenschaften einer Prüfung wie Prüfungsmethode, zu prüfende Kompetenzen, Leistungserwartung und Kriterium sollen den Prüflingen bekannt gemacht werden.

Beispiel: Lernziele, Prüfungsform, Bewertungsschema sowie die Art der zu erwartenden Prüfungsaufgaben sollten den Prüflingen bekannt sein.

Unverfälschbarkeit

Nebengütekriterium. Die Prüfungsergebnisse sollen von den Prüflingen nicht verfälscht werden können.

Beispiel: Betrugsmöglichkeiten durch ‚Spicken‘ verletzen das Kriterium der Unverfälschbarkeit.

Validität

Hauptgütekriterium. Meint die objektive *Gültigkeit* einer Messung und ist damit das wichtigste aller Gütekriterien: Misst die Prüfung tatsächlich das, was sie zu messen vorgibt? Validität setzt Objektivität und Reliabilität voraus. Validität lässt sich im Allgemeinen nicht quantifizieren.

Beispiel: Das Wägen stellt eine objektive und reliable Messung dar. Will man aber die Körpergröße messen, so stellt Wägen keine valide Art der Messung dar.

Zumutbarkeit

Nebengütekriterium. Eine Prüfung muss für die Prüflinge zumutbar sein. Hierbei handelt es sich in erster Linie um ein ethisches Kriterium.