

# Leitfaden für das Planen, Durchführen und Auswerten von Prüfungen an der ETHZ

Balthasar Eugster, lic. phil.  
Dr. Leonard Lutz

3. Juli 2003

(Überarbeitete Version vom 9. Dezember 2004)

## Inhaltsübersicht

<b>Zusammenfassung</b>	Seite	2
<b>Checklisten für das Planen, Durchführen und Auswerten von Prüfungen</b>		4
1. Schriftliche Prüfungen		4
1.1 Konstruktion/Planung von schriftlichen Prüfungen		4
1.2 Durchführung von schriftlichen Prüfungen		6
1.3 Bewertung und Auswertung von schriftlichen Prüfungen		7
2. Mündliche Prüfungen		8
2.1 Konstruktion/Planung von mündlichen Prüfungen		8
2.2 Durchführung von mündlichen Prüfungen		9
2.3 Bewertung und Auswertung von mündlichen Prüfungen		11
3. Allgemeine Leitfragen für andere Formen von Leistungskontrollen		12
<b>Anhang</b>		13
1. Prüfungen als Teil des Curriculums		13
2. Testtheoretische Grundlagen zur Planung, Durchführung und Auswertung von Prüfungen		16
3. Testtheorie und die verschiedenen Formen von Prüfungen und Leistungskontrollen		21
<b>Literatur</b>		22

## Zusammenfassung

Prüfungen sind Leistungsmessungen bzw. –bewertungen. Sie vergleichen studentische Fähigkeiten und Kompetenzen mit einem äusseren Referenzsystem, indem sie ein Prüfungsergebnis entweder normorientiert in Beziehung zu den Resultaten der übrigen Kandidaten/-innen oder kriteriumsorientiert in Beziehung zu einem von den anderen Individuen unabhängigen Massstab setzen. Normorientierte Prüfungen liefern Rangfolgen und eignen sich allenfalls für Aufnahmeprüfungen mit vorgegebenen Quoten. Demgegenüber sollten eigentliche Kontrollen studentischer Leistungen kriteriumsorientiert erfolgen, d.h. der Bewertungsmaßstab und die Zuordnung von Noten- zu Punktwerten ist vor der Korrektur festzulegen.

Im Rahmen des Curriculums einer Lehrveranstaltung oder eines gesamten Studienganges bilden die Lernziele den Bewertungsmaßstab für die Leistungsbewertung. Lernziele und Prüfungen stehen dabei in einer Wechselwirkung: Prüfungen müssen sich auf Lernziele beziehen (sind also eigentliche Lernzielkontrollen), und umgekehrt müssen Lernziele so formuliert werden, dass aus ihnen prüfbare Kompetenzen abgeleitet werden können.

Selbstverständlich messen Prüfungen Leistungsmerkmale nicht mit absoluter Exaktheit, sondern unterliegen stets einem Messfehler. Um den Messfehler möglichst klein zu halten, müssen Prüfungen verschiedenen Gütekriterien genügen, welche in der klassischen Testtheorie beschrieben werden: Die Ergebnisse von Prüfungen sollten zunächst unabhängig sein von den untersuchenden Personen (Objektivität). Die Prüfenden (Examinatoren/-innen) dürfen die Messung während der Durchführung nicht beeinflussen (Durchführungsobjektivität), so muss auch die Auswertung, d.h. die Vergabe von Punktwerten unabhängig von der korrigierenden Person erfolgen (Auswertungsobjektivität), und schliesslich darf auch die Zuordnung von Noten- zu Punktwerten zwischen verschiedenen beteiligten Personen nicht variieren (Interpretationsobjektivität).

Weiterhin muss eine Prüfung das zu prüfende Merkmal mit hoher Genauigkeit messen (Reliabilität). Genau ist ein Prüfungsergebnis, wenn der Messfehler klein ist bzw. der wahre Wert einen hohen Anteil des Testergebnisses ausmacht. Die Reliabilität einer Prüfung hängt von der sorgfältigen Konstruktion der einzelnen Prüfungsfragen und vor allem von der Anzahl der (Teil-)Fragen ab.

Schliesslich bemisst sich die Güte eines Tests an seiner Gültigkeit (Validität). Nur wenn eine Prüfung das misst, was sie zu messen vorgibt, kann sie gültige Aussagen über den Leistungsstand der Studierenden machen. In Bezug auf den Inhalt sollte ein Test die zu prüfenden Lernziele repräsentativ abdecken (Inhaltsvalidität), er muss gültige Rückschlüsse auf den Ausprägungsgrad der in den Lernzielen beschriebenen Kompetenzen ermöglichen (Konstruktvalidität) und im Sinne der Selektionsfunktion von Prüfungen zu Aussagen über künftige Leistungen beitragen (Kriteriumsvalidität).

Neben diesen Hauptgütekriterien der klassischen Testtheorie kann die Qualität einer Prüfung aufgrund weiterer testtheoretischer (Neben-)Kriterien bestimmt werden: Ein Test bedarf eines Bezugssystems (z.B. einer genau definierten Notenskala), um die individuellen Testergebnisse interpretieren zu können (Normierung). Darüber hinaus sollten für eine Prüfung mehrere Varianten vorliegen, welche Vergleiche zulassen (Vergleichbarkeit). Und nicht zuletzt ist bei der Konstruktion, Durchführung und Auswertung einer Prüfung auf die Verhältnismässigkeit von Ressourcen und Material zu achten (Ökonomie).

Die testtheoretischen Gütekriterien gelten für alle Arten von Bewertungen studentischer Leistungen. Gerade im Zuge der Bologna-Reformen werden vielfältige Formen von Lei-

stungskontrollen (im Sinne klassischer Prüfungen wie auch neuer Formen der Leistungsbewertung) zum Einsatz gelangen, welche nur dann aussagekräftige und vergleichbare Aussagen über studentische Leistungen ermöglichen, wenn sie den testtheoretischen Anforderungen genügen.

Die Testtheorie beschreibt die Rahmenbedingungen für die Konstruktion, Durchführung und Auswertung von Prüfungen. In der konkreten Prüfungsarbeit sind Dozierende mit vielen kleineren und grösseren Fragen und Hindernissen konfrontiert, welche die Qualität von Leistungskontrollen beeinflussen. In den folgenden Checklisten sind einige wichtige Aspekte zusammengefasst: Die verschiedenen Leitfragen geben Hinweise, wie die testtheoretischen Gütekriterien umgesetzt werden können. Dabei wird das Hauptaugenmerk auf schriftliche und mündliche Prüfungen gelegt, doch sind abschliessend auch zentrale Leitfragen zur Qualitätsbeurteilung von anderen Formen der Leistungskontrolle festgehalten. Selbstverständlich beinhalten die Checklisten nicht abschliessende Zusammenstellungen aller relevanten Aspekte, die aus den testtheoretischen Gütekriterien abgeleitet werden können. Umgekehrt wird die einzelne Prüfung nicht alle Einzelaspekte umfassend berücksichtigen können. Prüfungen sind immer auch Kompromisse zwischen dem idealtypischen Lernzielkatalog und der – von vielen Rahmenbedingungen und Sachzwängen geprägten – Realität konkreter Lehr-Lern-Prozesse an universitären Hochschulen.

In Anhang zu den Checklisten finden sich Ausführungen zu den curriculum- und testtheoretischen Zusammenhängen von Leistungskontrollen. Sie bilden das Bezugssystem für die verschiedenen Leitfragen und Aspekte der konkreten Prüfungsarbeit.

Dieser Leitfaden dient den Examinatorinnen und Examinatoren als Ratgeber für das Planen, Durchführen und Auswerten von Prüfungen und hat somit den Status einer Empfehlung.

## Checklisten für das Planen, Durchführen und Auswerten von Prüfungen

Die verschiedenen Aspekte und Leitfragen der Checklisten basieren auf testtheoretischen Grundlagen und Ergebnissen der empirischen Hochschulforschung. Daher finden sich in den Checklisten Hinweise auf die entsprechenden Ausführungen im Anhang dieses Leitfadens. So verweist z. B. die Bezeichnung "→ Lernziele: A1.3" auf die Erläuterungen in Kapitel 1.3 des Anhangs.

### 1 Schriftliche Prüfungen

#### 1.1 Konstruktion/Planung von schriftlichen Prüfungen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Lernziele	1.1 Sind für die Lehrveranstaltung Lernziele vorhanden? → Lernziele: A1.3	<input type="checkbox"/>
	1.2 Sind aus den Lernzielen prüfbare (beobachtbare) Kompetenzen abzuleiten?	<input type="checkbox"/>
	1.3 Sind die Lernziele (in ihrem Inhalt und ihrer Strukturierung) den Studierenden zu Beginn der Lehrveranstaltung bekannt gegeben worden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
2. Allgemeiner Aufbau der Prüfung	2.1 Lässt sich aus den Lernzielen eine sinnvolle inhaltliche Gliederung der Prüfung ableiten?	<input type="checkbox"/>
	2.2 Decken die Prüfungsteile inhaltlich die Gesamtheit der Lernzielbereiche ab? → Konstruktvalidität: A2.3.2	<input type="checkbox"/>
	2.3 Entsprechen die den einzelnen Prüfungsteilen zugeordneten Prüfungsformen (z.B. MC-Fragen, Offene Fragen etc.) den in den entsprechenden Lernzielen implizierten kognitiven Prozessen (Denkprozesse wie z.B. Wissen, Verstehen, Anwenden, Analysieren, Synthese Erstellen/Erarbeiten, Beurteilen)? → Kognitive Prozesse: A2.3.1	<input type="checkbox"/>
3. Zusammenhang von Prüfung und Unterrichtsdurchführung	3.1 Konnten alle geprüften Lernziele im Unterricht behandelt oder im Selbststudium erarbeitet werden?	<input type="checkbox"/>
4. Inhaltliche Aspekte der einzelnen Prüfungsfragen	4.1 Auf welches Lernziel/welche Lernziele bezieht sich die einzelne Prüfungsfrage? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	4.2 Entspricht der kognitive Prozess, welcher zur Lösung der Aufgabe nötig ist, dem kognitiven Prozess, der im entsprechenden Lernziel impliziert ist (z.B. Wissen, Verstehen, Anwenden, Analysieren, Synthese erbringen, Beurteilen)? → Kognitive Prozesse: A2.3.1	<input type="checkbox"/>

## 1.1 Konstruktion/Planung von schriftlichen Prüfungen (Fortsetzung)

5. Formale Aspekte der einzelnen Prüfungsfragen	5.1	Ist die Aufgabe präzise formuliert? Sind Teilfragen klar voneinander abgegrenzt? Werden nicht vorausgesetzte Fachwörter vermieden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	5.2	Ist die Frage so formuliert, dass die Kandidaten/-innen den Zusammenhang zu den Lernzielen erkennen können?	<input type="checkbox"/>
	5.3	Wissen die Kandidaten/-innen, in welcher Form sie zu antworten haben? Passt die verlangte Darstellungsform der Ergebnisse zu den mit der Aufgabe verbundenen Denkprozessen? Ist die angegebene oder vorgesehene Bearbeitungszeit angemessen (Zeitdruck ist nur dann angezeigt, wenn er Teil des Lernzieles ist)?	<input type="checkbox"/>
	5.4	Gibt es für die Kandidaten/-innen Angaben, in welcher Weise die Bewertung erfolgt (z.B. ob sie zehn oder nur fünf Argumente anführen müssen), d.h. welche Leistungen von ihnen erwartet werden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	5.5	Ist angegeben, welches Gewicht die Aufgabe in Relation zur Gesamtprüfung hat (max. Punktzahl der Aufgabe)? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	5.6	Sind mit Blick auf die Korrektur Musterlösungen formuliert? Passen diese Musterlösungen mit der verlangten Darstellungsform der Antworten überein? → Auswertungsobjektivität: A2.1.2	<input type="checkbox"/>
	5.7	Bei MC-Fragen: Ist bei jeder Frage klar, um welchen MC-Fragetyp es sich handelt und wie sie bewertet wird? Ist eine angemessene Zahl von Wahlantworten vorgegeben (3-5)? Sind die einzelnen Wahlantworten voneinander unabhängig, d.h. schliessen sie sich nicht gegenseitig aus? Werden Hinweiswörter wie "immer", "vollständig" oder "in keinem Fall" vermieden, welche auf falsche Lösungen deuten?	<input type="checkbox"/>
6. Vorevaluation der Prüfung	6.1	Bleibt neben der Summe der Bearbeitungszeiten aller Aufgaben eine gewisse Reservezeit? → Reliabilität: A2.2.4	<input type="checkbox"/>
	6.2	Sind alle Aufgaben von einer Fachperson gegen gelesen und auf ihre Lösbarkeit geprüft worden?	<input type="checkbox"/>
	6.3	Wird eine angemessene Anzahl von Lernzielen in der Prüfung geprüft? → Validität: A2.3.1, 2.3.2	<input type="checkbox"/>

## 1.2 Durchführung von schriftlichen Prüfungen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Vorinformationen	1.1 Sind die Studierenden über die Inhalte, den Ablauf und die Bewertung der Prüfung sowie die erlaubten Hilfsmittel schriftlich informiert worden (Merkblatt)? → Durchführungsobjektivität: A2.1.1	<input type="checkbox"/>
	1.2 Sind bei Prüfungen zu Lehrveranstaltungen, an welchen verschiedene Dozierende beteiligt sind, alle Studierenden durch gleiche Unterlagen informiert worden?	<input type="checkbox"/>
	1.3 Sind die Vorgaben der Allgemeinen Verordnung über Leistungskontrollen sowie allfälliger spezieller Verordnungen befolgt worden?	<input type="checkbox"/>
2. Deckblatt	2.1 Sind auf dem Deckblatt der Prüfung alle nötigen Informationen aufgelistet?	<input type="checkbox"/>
3. Infrastruktur	3.1 Haben alle Kandidaten/-innen die gleichen Voraussetzungen in Bezug auf Arbeitsplatz, Ruhe, Lichtverhältnisse etc.?	<input type="checkbox"/>
4. Prüfungsaufsicht	4.1 Ist dafür gesorgt, dass sich alle Aufsichtspersonen gleich verhalten? → Durchführungsobjektivität: A2.1.1	<input type="checkbox"/>
	4.2 Ist vorgesorgt, dass bei unvorhergesehenen Ereignissen (etwa bei Unlösbarkeit einer Aufgabe) alle Kandidaten/-innen die gleichen Informationen erhalten (z.B. wenn die Prüfung in verschiedenen Räumen gleichzeitig stattfindet)?	<input type="checkbox"/>

### 1.3 Bewertung und Auswertung schriftlicher Prüfungen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Beurteilungskriterien	1.1 Sind die Musterlösungen vor Beginn der Korrektur erstellt? → Kriteriumsorientierte Prüfungen: A1.2	<input type="checkbox"/>
	1.2 Sind genaue schriftliche Anleitungen für die Korrektur vorhanden? Wissen alle Korrektoren/-innen, für welche Antworten sie wie viele Punkte vergeben sollen? → Auswertungsobjektivität: A2.1.2	<input type="checkbox"/>
	1.3 Gibt es einheitliche Vorgaben, wie die Punkte und weitere Hinweise auf den Prüfungsbogen vermerkt werden (die Kandidaten/-innen sollten bei der Prüfungseinsicht erkennen können, weshalb sie nicht die volle Punktzahl erhalten haben)?	<input type="checkbox"/>
2. Benotung	2.1 Ist die Zuteilung der Notenwerte zu den erreichten Punktzahlen vor Beginn der Korrektur festgelegt? → Interpretationsobjektivität: A2.1.3	<input type="checkbox"/>
	2.2 Ist vor Beginn der Korrektur festgelegt, wie die Benotung bei aussergewöhnlichen Ereignissen (z.B. notwendige Streichung einer Aufgabe) angepasst wird?	<input type="checkbox"/>
	2.3 Ist gewährleistet, dass die Benotung kriteriumsorientiert und nur im Ausnahmefall normorientiert erfolgt? → Referenzsystem: A1.2	<input type="checkbox"/>
3. Korrektur	3.1 Ist ein einheitliches Vorgehen bei der Korrektur gewährleistet? Wird die Prüfung, wenn möglich, Aufgabe um Aufgabe (nicht Kandidat/-in um Kandidat/-in) korrigiert? → Auswertungsobjektivität: A2.1.2	<input type="checkbox"/>
	3.2 Ist das Vorgehen klar, wie bei einer nötigen Anpassung der Korrektur einer bestimmten Aufgabe vorgegangen wird?	<input type="checkbox"/>
	3.3 Erfolgt (bei kleineren Zwischenprüfungen zumindest mit Stichproben) eine Zweitkorrektur? Erfolgt die Zweitkorrektur wenn möglich durch eine andere Person? → Auswertungsobjektivität: A2.1.2	<input type="checkbox"/>
4. Umgang mit den Prüfungsergebnissen	4.1 Werden die Prüfungsergebnisse so aufbewahrt, dass die Korrektoren/-innen von Prüfungen desselben Prüfungsblockes davon keine Kenntnis erhalten?	<input type="checkbox"/>
5. Analyse der Prüfungsergebnisse	5.1 Werden die Ergebnisse der einzelnen Prüfungsfragen erfasst und ausgewertet?	<input type="checkbox"/>
	5.2 Werden die Ergebnisse vergleichbarer Prüfungen zusammengetragen und ausgewertet? → Kriteriumsvalidität: A2.3.3; Normierung: A2.4	<input type="checkbox"/>
	5.3 Ist gewährleistet, dass aufgrund der Prüfungsanalysen Anpassungen am Fragepool und allenfalls an den Lernzielen vorgenommen werden? → Prüfungen und Curriculum: A1.3	<input type="checkbox"/>
6. Prüfungseinsicht	6.1 Ist die Einsicht in die Ergebnisse und die Korrektur der Prüfung organisiert?	<input type="checkbox"/>

## 2 Mündliche Prüfungen

### 2.1 Konstruktion/Planung von mündlichen Prüfungen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Lernziele	1.1 Sind für die Lehrveranstaltung Lernziele vorhanden? → Lernziele: A1.3	<input type="checkbox"/>
	1.2 Sind aus den Lernzielen prüfbare (beobachtbare) wissenschaftliche Handlungen abzuleiten?	<input type="checkbox"/>
	1.3 Sind die Lernziele (in ihrem Inhalt und ihrer Strukturierung) den Studierenden zu Beginn der Lehrveranstaltung bekannt gegeben worden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
2. Allgemeiner Aufbau der Prüfung	2.1 Wie kann gewährleistet werden, dass die mündliche Prüfung die Gesamtheit der Lernziele in einer angemessenen Breite abdeckt?	<input type="checkbox"/>
	2.2 Wird darauf geachtet, dass nicht Fragen (insbesondere weiterführende Nachfragen) einfließen, welche über die Lernziele hinausführen? → Konstruktvalidität: A2.3.2	<input type="checkbox"/>
	2.3 Sind verschiedenen Kombinationen von vorbereiteten Fragen im Sinne von Prüfungsserien vorhanden? → Vergleichbarkeit: A2.5; Paralleltest-Reliabilität: A2.2.2	<input type="checkbox"/>
3. Allgemeine Vorbereitung	3.1 Sind die zentralen Fragen/Aufgaben, welche das Prüfungsgespräch leiten und strukturieren, schriftlich vorbereitet? → Validität: A2.3; Objektivität: A2.1; Reliabilität: A2.2	<input type="checkbox"/>
	3.2 Ist dafür gesorgt, dass die vorgesehenen und vorbereiteten Fragen auch tatsächlich gestellt werden (dass die Prüfung nicht bei einer Frage hängen bleibt)?	<input type="checkbox"/>
4. Inhaltliche Aspekte der einzelnen Prüfungsfragen	4.1 Welchem Lernziel ist die einzelne Prüfungsfrage inhaltlich zuzuordnen? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	4.2 Entspricht der kognitive Prozess, welcher zur Lösung der Aufgabe/Frage nötig ist, dem kognitiven Prozess, der im entsprechenden Lernziel impliziert ist (z.B. Wissen, Verstehen, Anwenden, Analysieren, Synthese Erstellen/Erarbeiten, Beurteilen)? → Kognitive Prozesse: A2.3.1	<input type="checkbox"/>
5. Formale Aspekte der einzelnen Prüfungsfragen	5.1 Ist die Aufgabe präzise formuliert? Sind Teilfragen klar voneinander abgegrenzt? Werden nicht vorausgesetzte Fachwörter vermieden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	5.2 Ist die Frage so formuliert, dass die Kandidaten/-innen den Zusammenhang zu den Lernzielen erkennen können?	<input type="checkbox"/>
	5.3 Wissen die Kandidaten/-innen, in welcher Form sie zu antworten haben (mit einer kurzen prägnanten Antwort oder längeren Ausführungen)? Passt die verlangte Darstellungsform der Ergebnisse zu den mit der Aufgabe verbundenen Denkprozessen?	<input type="checkbox"/>
	5.4 Gibt es für die Kandidaten/-innen Angaben, in welcher Weise die Bewertung erfolgt (z.B. ob sie zehn oder nur fünf Argumente anführen müssen), d.h. welche Leistungen von ihnen erwartet werden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	5.5 Sind mit Blick auf die Bewertung Musterlösungen formuliert? → Auswertungsobjektivität: A2.1.2	<input type="checkbox"/>

(Fortsetzung nächste Seite)



## 2.1 Konstruktion/Planung von mündlichen Prüfungen (Fortsetzung)

6. Vorevaluation der Prüfung	6.1 Sind alle Aufgaben von einer Fachperson gegen gelesen und auf ihre Lösbarkeit geprüft worden?	<input type="checkbox"/>
	6.2 Wird eine angemessene Anzahl von Lernzielen in der Prüfung geprüft?	<input type="checkbox"/>

→ Validität: A2.3.1, 2.3.2

## 2.2 Durchführung von mündlichen Prüfungen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Vorinformationen	1.1 Sind die Studierenden über die Inhalte, den Ablauf und die Bewertung der Prüfung sowie die erlaubten Hilfsmittel schriftlich informiert worden (Merkblatt)? → Durchführungsobjektivität: A2.1.1	<input type="checkbox"/>
	1.2 Sind bei Prüfungen zu Lehrveranstaltungen, an welchen verschiedenen Dozierende beteiligt sind, alle Studierenden durch gleiche Unterlagen informiert worden?	<input type="checkbox"/>
	1.3 Sind die Vorgaben der Allgemeinen Verordnung über Leistungskontrollen sowie allfälliger spezieller Verordnungen befolgt worden?	<input type="checkbox"/>
2. Prüfungsorganisation	2.1 Ist der Prüfungsplan so gestaltet, dass die Bewertung eines Kandidaten/einer Kandidatin vorgenommen werden kann, bevor der nächste Kandidat/die nächste Kandidatin geprüft wird? → Kriteriumsorientierte Prüfungen: A1.2	<input type="checkbox"/>
	2.2 Sind genügend Pausen für die Prüfenden eingeplant, sodass auch allfällige Verzögerungen wettgemacht werden können?	<input type="checkbox"/>
	2.3 Ist gewährleistet, dass das Prüfungsgespräch nicht durch Telefonklingeln, Klopfen etc. gestört wird?	<input type="checkbox"/>
	2.4 Ist die Sitzordnung so gewählt, dass sich die Kandidaten und die Prüfenden nicht wie bei einer Gerichtsverhandlung, sondern in angenehmer Gesprächsdistanz gegenüber sitzen? Sitzen allfällige Gäste nicht im Rücken, sondern seitlich der Kandidaten/-innen?	<input type="checkbox"/>
	2.5 Ist ein einheitliches Verfahren für die Zuteilung der Fragen (Fragepakete) auf die Kandidaten/-innen vorgesehen (wenn möglich Zufallszuteilung)?	<input type="checkbox"/>
	2.6 Beginnt das einzelne Prüfungsgespräch pünktlich zur vorgesehenen Zeit?	<input type="checkbox"/>
	2.7 Werden alle Kandidaten/-innen während der vorgesehenen Zeit geprüft?	<input type="checkbox"/>

(Fortsetzung nächste Seite)

## 2.2 Durchführung von mündlichen Prüfungen (Fortsetzung)

3. Prüfungsgespräch	3.1	Werden die Kandidaten mit Namen begrüßt und der Beisitzer/die Beisitzerin vorgestellt?	<input type="checkbox"/>
	3.2	Werden die Kandidaten/-innen zu Beginn kurz über den Ablauf der Prüfung orientiert?	<input type="checkbox"/>
	3.3	Wird darauf geachtet, dass zu Beginn des Gesprächs eine Frage gestellt wird, welche einen gewissen Antwortspielraum offen lässt und nicht die Reproduktion von auswendig gelerntem Wissen verlangt?	<input type="checkbox"/>
	3.4	Spricht während des Gesprächs hauptsächlich der Kandidat/die Kandidatin und nicht der/die Prüfende?	<input type="checkbox"/>
	3.5	Wird trotz vorbereiteter Fragen ein eigentliches Prüfungsgespräch erreicht und nicht eine Frageliste im Stile einer Checkliste abgearbeitet?	<input type="checkbox"/>
	3.6	Werden die Fragen inhaltlich klar und in ruhigem, aber deutlichem Sprechen gestellt?	<input type="checkbox"/>
	3.7	Wird darauf geachtet, dass nicht mehrere Teilfragen in einen Satz gepackt werden? → Inhaltsvalidität: A2.3.1	<input type="checkbox"/>
	3.8	Werden Suggestivfragen vermieden?	<input type="checkbox"/>
	3.9	Lässt der/die Prüfende die Kandidaten/-innen aussprechen?	<input type="checkbox"/>
	3.10	Wird es vermieden, zu lange auf der richtigen Beantwortung einer Frage zu beharren? Wird den Kandidaten allenfalls die Möglichkeit geboten, am Schluss der Prüfung auf nicht beantwortete Fragen zurückzukommen?	<input type="checkbox"/>
	3.11	Werden ironische oder gar zynische Bemerkungen vermieden?	<input type="checkbox"/>
	3.12	Werden unklare Rückmeldungen wie etwa die Verwendung des Prädikates "gut" bei falschen Antworten vermieden?	<input type="checkbox"/>
	3.13	Wird der Übergang zu einem neuen Thema sowohl für die Kandidaten/-innen wie auch die Protokoll führenden Beisitzer/-innen deutlich markiert? → Konstruktvalidität: A2.3.2	<input type="checkbox"/>
	3.14	Wird das Gespräch in angemessener Form beendet und der Kandidat/die Kandidatin verabschiedet?	<input type="checkbox"/>

## 2.3 Bewertung und Auswertung von mündlichen Prüfungen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Beurteilungskriterien	1.1 Ist sich der/die Prüfende bewusst, dass die Bewertung durch subjektive Wahrnehmungen (etwa des Aussehens und der Kleidung der Kandidaten/-innen, der Vorkenntnisse über die Kandidaten/-innen, der Reihenfolge der Kandidaten/-innen) beeinflusst wird? → Objektivität: A2.1	<input type="checkbox"/>
	1.2 Sind für die vorbereiteten Fragen Musterlösungen erstellt?	<input type="checkbox"/>
	1.3 Liegen Angaben vor, welche Kriterien für die Zuteilung einer bestimmten Punktzahl oder Bewertung erfüllt sein müssen? → Kriteriumsorientierte Prüfungen: A1.2	<input type="checkbox"/>
	1.4 Wird darauf geachtet, dass die Bewertungsskala für die einzelnen zu beurteilenden Prüfungsteile nicht zu fein differenziert ist (mehr als vier Intervalle sollten nicht vorgesehen sein)? → Auswertungsobjektivität: A2.1.2	<input type="checkbox"/>
2. Benotung	2.1 Ist die Zuteilung der Notenwerte zu den erreichten Punktzahlen bzw. die Verrechnung der einzelnen Bewertungen vor Beginn der Prüfung festgelegt? → Interpretationsobjektivität: A2.1.3	<input type="checkbox"/>
	2.2 Wird die Beurteilung der Leistung eines Kandidaten/einer Kandidatin unmittelbar nach Abschluss des Prüfungsgesprächs vorgenommen? Wird nach Abschluss der gesamten Prüfung aller Kandidaten/-innen die Notwendigkeit einer Anpassung der Noten geprüft? → Referenzsystem: A1.2	<input type="checkbox"/>
	2.3 Ist gewährleistet, dass die Benotung kriteriumsorientiert und nur im Ausnahmefall normorientiert erfolgt?	<input type="checkbox"/>
	2.4 Ist gewährleistet, dass die Benotung bei vergleichbaren Prüfungen (insbesondere bei Prüfungen innerhalb eines Prüfungsblocks) nach demselben Muster erfolgt? → Interpretationsobjektivität: A2.1.3	<input type="checkbox"/>
3. Protokoll	3.1 Erhalten die Beisitzer/-innen vor Beginn der Prüfungen klare und einheitlich Anweisungen, wie die Prüfungsgespräche protokolliert werden sollen. → Durchführungsobjektivität: A2.1.1	<input type="checkbox"/>
	3.2 Ist aufgrund der Protokolle nachvollziehbar, weshalb die Prüfung mit einer bestimmten Note bewertet wurde? Sind die Beurteilungen der einzelnen Prüfungsteile ersichtlich? → Interpretationsobjektivität: A2.1.3	<input type="checkbox"/>
4. Umgang mit den Prüfungsergebnissen	4.1 Werden die Prüfungsergebnisse so aufbewahrt, dass die Korrektoren/-innen von Prüfungen desselben Prüfungsblockes davon keine Kenntnis erhalten?	<input type="checkbox"/>
5. Analyse der Prüfungsergebnisse	5.1 Werden die Ergebnisse der einzelnen Prüfungsfragen erfasst und ausgewertet?	<input type="checkbox"/>
	5.2 Werden die Ergebnisse vergleichbarer Prüfungen zusammengetragen und ausgewertet? → Kriteriumsvalidität: A2.3.3; Normierung: A2.4	<input type="checkbox"/>
	5.3 Ist gewährleistet, dass aufgrund der Prüfungsanalysen Anpassungen am Fragenpool und allenfalls an den Lernzielen vorgenommen werden? → Prüfungen und Curriculum: A1.3	<input type="checkbox"/>
6. Prüfungseinsicht	6.1 Ist die Einsicht in die Ergebnisse und die Bewertung der Prüfung organisiert?	<input type="checkbox"/>

### 3 Allgemeine Leitfragen für andere Formen von Leistungskontrollen

Aspekte	Leitfragen	<input checked="" type="checkbox"/>
1. Objektivität	1.1 Wie wird die Durchführungsobjektivität der Leistungskontrolle gewährleistet? → A2.1.1; A3	<input type="checkbox"/>
	1.2 Welche Massnahmen garantieren die Auswertungsobjektivität der Leistungskontrolle? → A2.1.2; A3	<input type="checkbox"/>
	1.3 Ist die Interpretationsobjektivität der konkreten Leistungskontrolle garantiert? → A2.1.3; A3	<input type="checkbox"/>
2. Reliabilität	2.1 Sind Auswertungsverfahren vorgesehen, welche die Reliabilität, insbesondere die innere Konsistenz der Leistungskontrolle, analysieren? → A2.2; A3	<input type="checkbox"/>
3. Validität	3.1 Werden durch die Leistungskontrolle die der Lehrveranstaltung (oder entsprechender Teile davon) zugrunde liegenden Lernziele angemessen überprüft, sodass die Inhaltsvalidität der Leistungskontrolle nachgewiesen ist? → A2.3.1; A3	<input type="checkbox"/>
	3.2 Erlaubt die Leistungskontrolle Aussagen über das Erlangen der in den Lernzielen enthaltenen Kompetenzen (Konstruktvalidität)? → A2.3.2; A3	<input type="checkbox"/>
	3.3 Können die Ergebnisse der Leistungskontrolle mit anderen ähnlichen Leistungskontrollen im Sinne eines Aussenkriteriums verglichen werden (Kriteriumsvalidität)? → A2.3.3; A3	<input type="checkbox"/>

## Anhang

### 1. Prüfungen als Teil des Curriculums

#### 1.1 Funktionen von Prüfungen

Im Hochschulalltag haben Prüfungen verschiedene Funktionen: Sie sind u. a. Leistungsmessungen, Selektionsinstrumente, Druckmittel für den Unterricht, Bühne für Selbstdarstellungen, Motivation für Lernleistungen, Stabilisatoren hochschultypischer Hierarchien, Rückmeldungen über den Lehrerfolg der Dozierenden, Reproduktionsmechanismen des Wissenskanons, Zugangspforten zu höheren gesellschaftlichen Positionen etc.<sup>1</sup> Prüfungen beeinflussen wesentlich das Lernverhalten der Studierenden (es wird gelernt, was an der Prüfung kommt) und wirken sich auch auf die Lehre spürbar aus (für jede Stoffsequenz muss deren Prüfungsrelevanz aufgezeigt werden).

Wenn Prüfungen Aussagen über den Leistungsstand von Studierenden ermöglichen sollen, dann stellt sich die grundsätzliche Frage: "Was und wie kann mit Prüfungen überhaupt gemessen werden?" Die darin angesprochene Messbarkeit von individuellen Merkmalsausprägungen verweist auf eine zweifache Herausforderung im Umgang mit Prüfungen: Zum einen ist die Bedeutung von Prüfungen innerhalb von Lehr-Lern-Prozessen zu klären, und zum anderen sind die qualitativen Anforderungen an aussagekräftige Prüfungen festzulegen.

Darüber hinaus erlangt der Umgang mit Prüfungen im Kontext der Bologna-Reformen zusätzliche Relevanz. Die Studiengestaltung im System von Bachelor-, Master- und Doktorentsabschlüssen ermöglicht neue Lehr- und Lernformen und zielt auf ein erweitertes Kompetenzgefüge ab, wofür auch neue Arten der Bewertung von Leistungen zu entwickeln sind.

Die "Allgemeine Verordnung über Leistungskontrollen an der Eidgenössischen Technischen Hochschule Zürich" (AVL ETHZ)<sup>2</sup> umschreibt eine "Prüfung" als ein "Verfahren, mit dem die Beherrschung des Lehrstoffs einer ein- oder zweisemestrigen Lehrveranstaltung kontrolliert und mit einer Note bewertet wird" (Art. 2 lit. f AVL ETHZ). Prüfungen sind zusammen mit schriftlichen Arbeiten die hauptsächlichen Formen von "Leistungskontrollen", welche jedes Verfahren meinen, "mit dem die Leistung von Studierenden gemessen und bewertet wird" (Art. 2 lit. e AVL ETHZ).

---

<sup>1</sup> Siehe dazu etwa bei Bauer (2002) und Prahl (1995).

<sup>2</sup> AVL ETHZ vom 10. September 2002.

## 1.2 Referenzsystem von Prüfungen

Diese zweistufige Begriffsbestimmung zeigt an, dass es bei Prüfungen um den qualifizierbaren und quantifizierbaren Nachweis von Leistungen in Bezug auf das Beherrschen von Lehr- bzw. Lernstoff geht. Weil 'studentische Leistungen' wie auch das 'Beherrschen von Stoff' zunächst wenig bestimmte Formeln bezeichnen, bedarf es aus didaktischer Sicht einer Präzisierung der juristischen Definition. Beide Konzepte legen nicht absolute Grössen fest, welche in sich einen Bewertungsmaassstab enthalten. Was eine gute Leistung ist, kann nur mit Bezug auf ein äusseres Referenzsystem bestimmt werden. Was es also braucht, um eine bestimmte Note zu erreichen, muss durch eine Setzung der Prüfenden vorgegeben werden. Je nach Art des Referenzsystems ist eine Prüfung entweder

- **normorientiert**, d.h. sie setzt das individuelle Testergebnis in Bezug zu den Resultaten der anderen Studierenden und bestimmt die Position eines einzelnen Ergebnisses innerhalb einer Gruppe; oder sie ist
- **kriteriumsorientiert**, d.h. sie vergleicht die individuelle Leistung der einzelnen Studierenden mit einem von den anderen Individuen unabhängigen Maassstab (z.B. vordefinierten Lernzielen).<sup>3</sup>

Normorientierte Tests sind angezeigt bei Aufnahmeprüfungen, wo es hauptsächlich darum geht, einen bestimmten Prozentsatz der besten Kandidaten/-innen auszuwählen. Der weit verbreitete Einsatz von normorientierten Tests im normalen (Hochschul-)

Unterricht ist allerdings fragwürdig, interessiert doch bei einer Fachprüfung nicht in erster Linie der Rang eines Kandidaten/einer Kandidatin, sondern vielmehr das Erlangen angepeilter Kompetenzen. Konkret heisst dies: In einer Prüfung sollten die Noten nicht *nach* der Korrektur aufgrund von Rangverteilungen vergeben werden (z.B.: Durchschnitt bei 4.6, die besten erhalten eine 6.0, die schlechtesten eine ungenügende Note, unabhängig wie gut oder schlecht ihre Leistung wirklich ist). Es ist zu empfehlen, die Zuordnung von Noten- zu Punktwerten *vor* der Korrektur aufgrund eines Bewertungsmaassstabes (Kriterium) festzulegen (allenfalls muss dieser beim Eintreten von aussergewöhnlichen Umständen wie z.B. fehlerhaften Aufgaben angepasst werden).<sup>4</sup>

---

<sup>3</sup> Dieses in der Literatur zu Prüfungen meist verwendete Begriffspaar "normorientiert – kriteriumsorientiert" präzisiert Klauer durch die Unterscheidung von "realnormorientierten Tests" (was "normorientierten Tests" entspricht) und "idealn timerorientierten Tests" ("kriteriumsorientierten Tests") gegenüber (Klauer 1983, 693f.), denn die vorherrschende Unterscheidung "norm- vs. kriteriumsorientiert" kann zu Missverständnissen führen und setzt eine strenge Unterscheidung von Kriterien/Maassstäben (Leistung als Bezugsgrösse) einerseits und Normen (Resultate aller geprüften Individuen als Bezugsgrösse) andererseits voraus (siehe etwa bei Gage und Berliner (1996, 585ff.).

<sup>4</sup> Zu den testtheoretischen Grundlagen kriteriumsorientierter, d.h. idealn timerorientierter, Prüfungen siehe bei Klauer (1987).

### 1.3 Prüfungen und Lernziele

Dozierende kommen also nicht umhin, möglichst genau die erwarteten Leistungen zu definieren, denn was eine gute Leistung ist, lässt sich nicht aus dem vermittelten bzw. gelerten Stoff selber ableiten. Prüfungen und deren Bewertung ergeben sich nicht direkt aus der wissenschaftlichen Sache, sondern sind nur im Kontext der konkreten Ausgestaltung von Lehr-Lern-Prozessen zu realisieren. Selbstverständlich orientiert sich dieser Prozess am Stand des jeweiligen wissenschaftlichen Wissens, doch ist er immer auch eine didaktische Festlegung. Prüfungen sind also eng verknüpft mit der grundlegenden didaktischen Entscheidung, was im Rahmen einer Lehrveranstaltung von den Studierenden gelernt werden soll. Dadurch sind sie notwendiger Teil eines umfassenden Curriculums, welches sich in seiner Grundstruktur etwa wie folgt darstellen lässt:

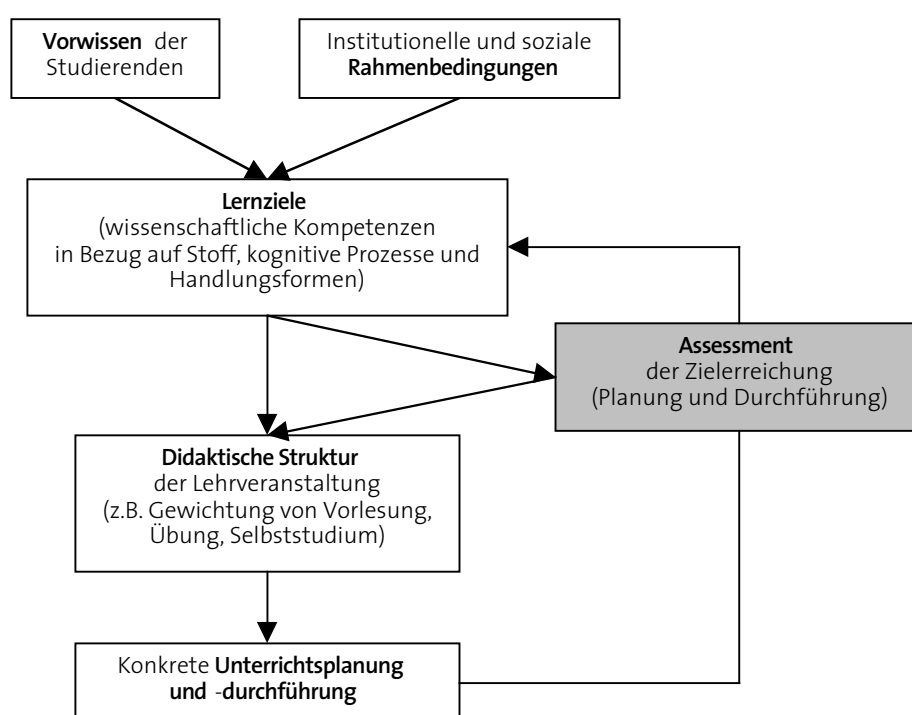


Abb. 1: Grundstruktur eines Curriculums (vereinfachte Darstellung)

Prüfungen sollten also immer auf Lernziele bezogen sein. Mehr noch: Lernziele sind die notwendigen Bedingungen für (kriteriumsorientierte) Prüfungen. Diese Beziehung gilt aber auch umgekehrt: Lernziele sind didaktisch nur dann sinnvoll, wenn sie auch überprüfbar sind. So muss bei der Formulierung von Lernzielen immer auch mit überlegt werden, wie das Erreichen dieser Ziele beobachtet und gemessen werden kann (bzw. "könnte", wenn keine formelle Prüfung vorgesehen ist). Je nachdem, welche Ziele erreicht werden sollen, sind die Art der Prüfung und die einzelnen Prüfungsaufgaben auszuwählen. So kann etwa eine komplexe Beurteilungskompetenz nicht mit Hilfe von Multiple-Choice-Aufgaben nachgewiesen werden.

## 2. Testtheoretische Grundlagen zur Planung, Durchführung und Auswertung von Prüfungen

Prüfungen messen die von ihnen zu erfassenden Leistungsmerkmale nicht mit absoluter Exaktheit. Mit den komplexen Problemstellungen rund um die Messbarkeit von Persönlichkeits- und Leistungsmerkmalen befasst sich die Testtheorie. Gemäss dem ersten Axiom der klassischen Testtheorie<sup>5</sup> besteht ein Testergebnis stets aus der Addition des "wahren Wertes" und des Messfehlers. Für jeden Test ist folglich ein solcher Messfehler zu vermuten, der aber durch sorgfältiges Planen, Durchführen sowie Auswerten der Prüfungen möglichst klein gehalten werden soll. Um die Unzulänglichkeiten der Messung zu minimieren, sind die drei in der klassischen Testtheorie beschriebenen Hauptgütekriterien zu berücksichtigen. Ein Test muss demnach objektiv, reliabel und valide sein. Daneben sollte ein Test – im Sinne von Nebenkriterien – normiert, vergleichbar und ökonomisch sein (Lienert&Ratz 1998, 7ff.).

### 2.1 Objektivität

Mit der Objektivität eines Tests wird die Unabhängigkeit seiner Ergebnisse von der untersuchenden Person geprüft. Gelangen verschiedene Prüfer/-innen zu denselben Beurteilungen, weist ein Test hohe Objektivität aus.

Weil der verfälschende Einfluss der Testanwender/-innen in verschiedenen Prozessphasen wirksam werden kann, unterscheidet man üblicherweise drei Aspekte der Objektivität:

**2.1.1 Durchführungsobjektivität:** Das Verhalten der prüfenden Person während des Tests sollte das Ergebnis nicht beeinflussen, sodass es nicht darauf ankommt, wer die Prüfung durchführt.

Bei schriftlichen Prüfungen müssen daher genaue Instruktionen zuhanden der Aufsichtspersonen vorliegen, welche auch deren Verhalten bei Rückfragen regeln. Die Durchführungsobjektivität ist aber insbesondere bei mündlichen oder praktischen Prüfungen von zentraler Bedeutung. Um dem Sinn mündlicher Examina – dem Führen eines eigentlichen Prüfungsgesprächs – gerecht zu werden und doch die Objektivität der Prüfung zu garantieren, ist ein transparentes und reproduzierbares Vorgehen während der Prüfung detailliert zu planen. Die zentralen Hauptfragen eines Prüfungsgesprächs sollten schriftlich vorbereitet und der Prüfungsablauf möglichst präzise protokolliert werden. Genau zu beachten sind auch die Nachfragen oder Hilfestellungen, welche im Verlaufe eines Prüfungsgesprächs an die Kandidaten/-innen gerichtet werden.

**2.1.2 Auswertungsobjektivität:** Auch die Auswertung, d.h. die Zuteilung von Punktwerten zu den abgegebenen Antworten muss unabhängig von der prüfenden Person erfolgen. Für schriftliche Prüfungen, die von verschiedenen Personen korrigiert werden, müssen genaue Korrekturanleitungen entwickelt werden, damit eine einheitliche Vergabe der Punkte gesichert ist. Dazu gehören auch die bereits bei der Prüfungskonstruktion zu erstellenden Musterlösungen. Es empfiehlt sich daher stets eine Zweitkorrektur (falls möglich) durch eine andere Person. Die Auswertungsobjektivität wird zudem erhöht, wenn die einzelnen Prüfungsbereiche, welchen ein Punktwert zugeordnet wird, nicht zu gross gehalten sind

---

<sup>5</sup> Neben der klassischen Testtheorie, welche für die konkrete Planung und Durchführung von Prüfungen von institutionalisierten Leistungsprüfungen von besonderer Bedeutung ist, muss für spezielle Fragestellungen auch die probabilistische Testtheorie berücksichtigt werden. Zum Verhältnis von klassischer und probabilistischer Testtheorie siehe dazu etwa bei Fischer (1975) oder bei Fischer (1983).



und bei der Korrektur nicht nur die Punktzahl, sondern auch ein kurzer begründender Kommentar festgehalten wird.

Speziell zu beachten ist die Auswertungsobjektivität bei mündlichen Prüfungen<sup>6</sup>: Ideal ist die gleichwertige Leistungsbeurteilung durch zwei Experten/-innen. Zudem erleichtert ein genaues Prüfungsprotokoll die Bewertung, welche unmittelbar nach Ende der individuellen Prüfung und nicht erst nach mehreren Kandidaten/-innen erfolgen sollte. Die Objektivität kann beträchtlich erhöht werden, wenn die einzelnen zu bewertenden Prüfungseinheiten anhand einer nicht zu feinen Skala beurteilt werden: Mehr als vier Bewertungsstufen sollten vermieden werden, weil sonst die Streuung grösser ist als die Skaleneinheit.

**2.1.3 Interpretationsobjektivität:** Eine hohe Objektivität der Testinterpretation bringt zum Ausdruck, dass das Prüfungsergebnis von verschiedenen Person nicht unterschiedlich gedeutet wird.

Bei Prüfungen ist die Interpretationsnorm die von der Prüfungsordnung vorgegebene Notenskala. Zumindest innerhalb eines Studienganges sollte dabei auf eine möglichst einheitliche Zuordnung von Noten- zu Punktwerten geachtet werden. Es verfälscht die Interpretation von Prüfungsergebnissen, wenn bei einem Dozenten 80% der maximalen Punktzahl für die Note 6.0 ausreichen, während bei einer anderen Dozentin für dieselbe Note 95% der Punkte erreicht werden müssen.

## 2.2 Reliabilität

Ein Test ist reliabel (zuverlässig), wenn er das geprüfte Merkmal mit einem hohen Mass an Genauigkeit misst. Hohe Zuverlässigkeit meint im Sinne des ersten Axioms der klassischen Testtheorie einen kleinen Messfehler, d.h. einen hohen Anteil des wahren Werts am Testergebnis.<sup>7</sup> Die Reliabilität eines Tests kann grundsätzlich auf verschiedene Arten erfasst bzw. berechnet werden:

**2.2.1 Retest-Reliabilität (Stabilität):** Dabei wird derselbe Test denselben Kandidaten/-innen nach einem Zeitintervall erneut vorgelegt. Korrelieren die Testergebnisse in hohem Masse, weist der Test eine hohe Retest-Reliabilität auf.

Im pädagogischen Kontext hat diese Form der Reliabilitätsbestimmung keine grosse Bedeutung, da sie höchstens bei einer Testwiederholung nach einer nicht bestandenen Prüfung eingesetzt werden kann – dann aber gerade nicht mit der Stabilität des wahren Testwertes gerechnet wird.<sup>8</sup>

**2.2.2 Paralleltest-Reliabilität:** Diese Form der Reliabilität wird bestimmt, indem den Kandidaten/-innen zwei verschiedene Prüfungen vorgelegt werden, welche beide dasselbe Merkmal (dieselben Lernziele bzw. Kompetenzen) prüfen sollen. Zeigen beide Prüfungen ähnliche Ergebnisse, erweist sich die Paralleltest-Reliabilität als hoch.

**2.2.3 Innere Konsistenz:** Während für die ersten beiden Arten der Reliabilitätsprüfung mehrere Testdurchführungen bzw. verschiedene Test nötig sind, kommt die Prüfung der inneren Konsistenz mit nur einem Test aus, was der Prüfungsrealität institutionalisierter Aus-

---

<sup>6</sup> Siehe dazu die Ausführungen bei Stary (2002).

<sup>7</sup> Da sich nach den Axiomen 2 bis 5 der klassischen Testtheorie die Messfehler bei wiederholter Messung ausgleichen sowie vom wahren Wert, von anderen Merkmalen und untereinander unabhängig sind (nicht korrelieren), erzeugen sie bei mehreren Messungen unsystematische Abweichungen, welche die sogenannte Fehlervarianz konstituieren. Reliabilität lässt sich daher auch als der Anteil der wahren Varianz an der beobachteten Varianz bestimmen (Bortz & Döring 1995, 181f.).

<sup>8</sup> Ein bei einer Prüfungswiederholung verbessertes Ergebnis kann testtheoretisch aufgrund einer Leistungssteigerung (Instabilität des gemessenen Merkmals) oder niedriger Reliabilität der Prüfungen zustande kommen.

bildungsgänge besser entspricht. Entweder wird dazu die Gesamtheit aller Aufgaben einer Prüfung in zwei Hälften gesplittet (Testhalbierung) und die Ergebnisse beider Testhälften miteinander verglichen; oder es wird jede einzelne Aufgabe der Prüfung als unabhängiger Teil aufgefasst (Konsistenzanalyse) und mit Hilfe komplexer Berechnungsmethoden mit den anderen Aufgaben verglichen.<sup>9</sup>

**2.2.4 Reliabilität und die Konstruktion von Prüfungen:** Die theoretischen Überlegungen zur Sicherung der Reliabilität, insbesondere der inneren Konsistenz von Tests deuten Konsequenzen für die Konstruktion von Prüfungen an:

Eine Prüfung muss sehr sorgfältig zusammengesetzt sein. Neben der prinzipiellen Lösbarkeit einer Aufgabe trägt auch der Schwierigkeitsindex<sup>10</sup> einer Aufgabe zu ihrer Reliabilität bei. Es empfiehlt sich, nicht eine zu grosse Anzahl Aufgaben mit extremer Schwierigkeit in eine Prüfung einzubauen, da Aufgaben mit einem mittleren Schwierigkeitsindex in der Regel eine höhere Reliabilität der Prüfung ergeben (Lienert&Raatz 1998, 213). Am wirksamsten kann die Reliabilität jedoch über die Länge der Prüfung verbessert werden. Durch das Hinzufügen von Fragen wird die Reliabilität erhöht, wobei natürlich im Sinne der Ökonomie der Prüfung (siehe unten Kapitel 2.6) eine Prüfung nicht unbegrenzt verlängert werden kann und soll.

Im Weiteren wird die Reliabilität einer Prüfung mit Multiple-Choice-Fragen auch durch den Zufallsfaktor bei Antwortalternativen beeinflusst. Zufallsabhängige Aufgaben (insbesondere Fragen mit nur zwei Antwortalternativen) vermindern die Reliabilität (Lienert&Raatz 1998, 208).

Schliesslich dürfte auch die Testzeit Auswirkungen auf die Testreliabilität haben. Es gibt offenbar für eine bestimmte Prüfung eine optimale Zeitdauer, deren Verkürzung bzw. Verlängerung sich ungünstig auf die Reliabilität auswirkt. Als Faustregel wird in der Literatur empfohlen, die Testzeit so zu wählen, dass 80% der Kandidaten alle Aufgaben bearbeiten können (Lienert&Raatz 1998, 209).

## 2.3 Validität

Wenn ein Test eine hohe Objektivität bzw. Reliabilität aufweist, ist damit noch nicht gesichert, dass er eine Aussage über die zu messenden Merkmale zulässt. Denn erst wenn er wirklich das misst, was er zu messen vorgibt, liefert er gültige, d.h. valide Messergebnisse. Wie bei den beiden andern Hauptgütekriterien lassen sich verschiedene Aspekte der Validität unterscheiden, welche für die Handhabung von Prüfungen von grosser Relevanz sind. Diese Aspekte beschreiben jedoch nicht scharf abgrenzbare Dimensionen der Validität, vielmehr fokussieren sie aus unterschiedlichen Blickwinkeln auf die zentralen Merkmale der Testgültigkeit (Metzger & Nüesch 1997, B3ff.):

**2.3.1 Inhaltsvalidität:** Deckt ein Test die Fülle der zu prüfenden Lernziele repräsentativ ab, kann er als inhaltlich valide bezeichnet werden. Dies bezieht sich sowohl auf die verschiedenen Themen wie auch auf die Komplexität der zur Erreichung der Lernziele nötigen kognitiven Prozesse.<sup>11</sup>

<sup>9</sup> Diese erfolgen meist mit Hilfe des Alpha-Koeffizienten von Cronbach.

<sup>10</sup> Der (statistische) Schwierigkeitsindex bezeichnet den prozentualen Anteil der auf eine Aufgabe entfallenden richtigen Antworten innerhalb einer Stichprobe. Davon zu unterscheiden ist die psychologische Schwierigkeit, also die Einschätzung durch die Kandidaten, ob eine Aufgabe schwierig zu lösen sei.

<sup>11</sup> Die Komplexität von kognitiven Prozessen bzw. den damit verbundenen Handlungsmustern lässt sich etwa mit Hilfe einer Taxonomie von Lernzielen und Prüfungsaufgaben erfassen, wie sie Bloom et al. in der zweiten Hälfte des 20. Jahrhunderts entwickelt haben (Bloom et al. 1956). Nach Bloom werden in der Reihenfolge zunehmender Komplexität folgende Stufen unterschieden: Wissen – Verstehen – Anwenden – Analysieren – Synthesen Erstellen/Erarbeiten – Beurteilen.

So ist etwa eine Prüfung nicht inhaltsvalid, wenn sie die bloße Reproduktion auswendig gelerntes Wissen verlangt, während die Lernziele ausführliche Analysen von vielschichtigen Zusammenhängen beschreiben. Damit Prüfungsaufgaben und –fragen sich auf genau bestimmbare Themenbereiche und entsprechende kognitive Prozesse beziehen, müssen sie formal angemessen ausgestaltet sein. Eine komplette Prüfungsaufgabe sollte – als Minielemente – eine präzise Aufgabenstellung (verschiedene Teilaufgaben deutlich voneinander abgegrenzt), klare Angaben über die Form der Ergebnisdarstellung (z.B. "in Stichworten", "in vier bis fünf Sätzen" oder "in Form einer Tabelle mit folgenden Elementen...") sowie Angaben über den Beurteilungsmassstab (Angabe der erreichbaren Punkte und die erwartete Qualität der Lösung) enthalten. Bereits bei der Erstellung der Prüfung sollten unbedingt die richtigen Lösungen in der verlangten Darstellungsform notiert werden. Dabei zeigen sich oft Konstruktionsmängel der Aufgabe, welche noch rechtzeitig behoben werden können. Zugleich lässt sich dadurch auch die Validitätsprüfung einfacher durchführen, denn die Inhaltsvalidität einer Prüfung lässt sich nicht numerisch berechnen, sie muss aufgrund von (subjektiven) Expertenurteilen nachgewiesen werden. Unterstützt wird dieser Nachweis durch eine einheitliche Lernzielsystematik (in Bezug auf den Stoff und die kognitive Komplexität der Lernziele), welcher die einzelnen Prüfungsaufgaben zugeordnet werden können. Prüfungsaufgaben sind immer nur Hilfsmittel, um die Lernzielerreichung zu beurteilen. Ein Kandidat/Eine Kandidatin sollte anhand der Aufgaben seine Kompetenz in Bezug auf die vorgegebenen Ziele unter Beweis stellen können – und nicht seine Fähigkeit, verwirrend formulierte Fragen zu enträtseln!

**2.3.2 Konstruktvalidität:** Nicht immer lassen sich Lernziele operationalisieren, d.h. die darin ausgedrückten Kompetenzen sind nicht direkt beobachtbar und unmittelbar zu überprüfen. Die Konstruktvalidität eines Tests bezieht sich auf diesen Gesichtspunkt, indem sie Auskunft darüber gibt, ob mit Hilfe der Prüfungsaufgaben gültige Rückschlüsse auf den Ausprägungsgrad der im Lernziel beschriebenen Kompetenzen möglich sind. Mit anderen Worten: Eine Prüfung ist konstruktvalid, wenn sie Hypothesen bestätigt, die sich aus den Lernzielen ableiten lassen. So sollte ein Test in der Gesamtheit der von ihm erforderten Fähigkeiten ein möglichst differenziertes Bild des angestrebten Kompetenzprofils liefern. Auf curricularer Ebene zwingt diese testtheoretische Forderung zu einer konsequenten Lernzielkonstruktion. Kompetenzen dürfen dabei nicht mit leeren Worthülsen umschrieben, sondern müssen als theoretische Konstrukte verstanden werden, die sich in einer Prüfung als Komplex messbarer Merkmale definieren lassen müssen. Bei der Entwicklung konkreter Prüfungen ist sorgfältig auf das Gesamtgefüge der Aufgaben zu achten. Diese sind nicht isoliert aneinander zu reihen, vielmehr müssen sie als Wechselspiel verschiedener Aspekte eines Gesamtkonstrukts zusammengefügt werden.

**2.3.3 Kriteriumsvalidität:** Im Sinne ihrer Selektionsfunktion dienen Prüfungen auch zur Vorhersage künftiger Leistungen. Insbesondere Ergebnisse von Zulassungsprüfungen, aber auch die Noten des Mobilitätsscharniers "Bachelorabschluss" sollen zur Auswahl der "richtigen" (Master-)Studierenden beitragen. Durch den späteren Vergleich mit nachfolgenden Leistungskontrollen lässt sich die prognostische Validität der Noten aus den Zulassungsprüfungen bzw. aus dem Bachelorstudium bestimmen und so für die Konstruktion ähnlicher Prüfungen nutzen.

Ohne verschiedene Messzeitpunkte abwarten zu müssen, kann die Übereinstimmungsvalidität einer einzelnen Prüfung mit einem Aussenkriterium bestimmt werden: Beispielsweise können im Rahmen eines Prüfungsblockes die Prüfungsergebnisse eines ähnlichen Faches zur Validierung beigezogen werden. Selbstverständlich liefern solche Vergleiche nur Indizien für die Validierung; sie unterstreichen aber, dass eine Prüfung nie als isoliertes Ereignis mit absoluten Ergebnissen zu interpretieren ist.

## 2.4 Normierung

Ein Test ist normiert, wenn für die Interpretation der Ergebnisse ein Bezugssystem für die Einordnung der individuellen Resultate definiert ist. Bei Leistungskontrollen ist ein solches Bezugssystem mit der sich von 1.0 bis 6.0 erstreckenden Notenskala gegeben. Dieses Notensystem ist im Kontext des schweizerischen Bildungssystems derart selbstverständlich, dass seine Eigentümlichkeiten oft übersehen werden. Insbesondere ist die Bedeutung der Grenze zwischen "bestanden" und "nicht bestanden" zu beachten: Diese Trennlinie liegt nicht auf dem Skalenmittelwert (3.5), sondern bei 4.0, wodurch die beiden Werte "bestanden" und "nicht bestanden" unterschiedlich grosse Skalenbereiche umfassen. Dies deutet an, dass es sich bei einer Notenskala oft nicht um eine sogenannte Intervallskala (bei welcher identische Notenabstände identischen Leistungsunterschieden entsprechen<sup>12</sup>) oder gar um eine Verhältnisskala (bei welcher zwei Notenverhältnisse dem Verhältnis der entsprechenden Leistungen entsprechen<sup>13</sup>) handelt. Eine Notenskala ist demgegenüber meist vom Typ einer Ordinalskala, welche Rangfolgen zum Ausdruck bringt. Gleiche Noten bedeuten nicht unbedingt gleiche Leistung und eine Kandidatin mit einer 6 ist nicht doppelt so gut wie ein Kandidat mit der Note 3.

## 2.5 Vergleichbarkeit

Wenn von einem Test mehrere Paralleltestformen vorliegen und validitätsähnliche Tests vorliegen, weist er eine hohe Vergleichbarkeit auf.

Es empfiehlt sich für alle Prüfungen verschiedenen Varianten auszuarbeiten, welche bei auftretenden Güteproblemen eine einfachere und besser überprüfbare Verbesserung der Tests ermöglicht.

## 2.6 Ökonomie

Ein Test sollte eine verhältnismässige Durchführungsdauer beanspruchen, nicht zu viel Material verbrauchen, für die Kandidaten/-innen einfach zu handhaben und für die Prüfenden einfach auszuwerten sein.

---

<sup>12</sup> Ein typisches Beispiel für eine Intervallskala stellt die Celsiustemperaturskala dar.

<sup>13</sup> Z.B. erfolgt die Messung von Weitsprungleistungen mit Hilfe des Metermasses auf einer Verhältnisskala.

### 3. Testtheorie und die verschiedenen Formen von Prüfungen und Leistungskontrollen

Die beschriebenen testtheoretischen Anforderungen gelten grundsätzlich für alle Formen von Prüfungen und Leistungskontrollen, unabhängig ob es sich dabei um eine wissenschaftliche Arbeit etwa im Sinne einer Masterarbeit, eine schriftliche bzw. mündliche Prüfung oder um eine andere Art und Weise der Zielüberprüfung handelt. Auf den ersten Blick könnte die stark formalisierte Testtheorie aber vermuten lassen, Leistungskontrollen würden nur dann den Gütekriterien genügen, wenn sie ihrerseits in eine möglichst strenge Form – wie etwa bei Multiple-Choice-Prüfungen – gegossen sind. Dabei wird aber der dargestellte curriculare Aspekt von Leistungskontrollen verkannt: Zunächst geht es bei der Leistungsmessung um den Nachweis der Lernzielerreichung. Leistungskontrollen müssen so konzipiert – und auch rechtlich institutionalisiert – sein, dass sie für die Bewertung der Lernzielerreichung im Sinne der Gütekriterien geeignet sind. Wenn es sich erweist, dass z.B. das Realisieren einer Forschungsrecherche oder das Verfassen eines Sitzungsprotokolls mit Blick auf die Lernziele relevante Handlungsformen sind, dann empfehlen sich diese auch als Prüfungsformen. Gerade weil sich Kompetenzen an wenigen, isolierten Zeitpunkten nur sehr eingeschränkt nachweisen lassen, gehen die mit den Bologna-Studienreformen verbundenen hochschuldidaktischen Konzepte von einer grösseren Zahl verschiedener Formen von Leistungskontrollen aus. Neben Prüfungen im klassischen Sinne werden vermehrt auch Studienleistungen während der Lernprozesse in die Leistungsbeurteilung mit einzubeziehen sein, sodass die Studierenden im Verlaufe des Studiums eine Art "Portfolio" über ihren individuellen Lernprozess und dessen Lernergebnisse zusammenstellen.<sup>14</sup> Dieses Verständnis differenzierter Leistungsmessung schlägt sich auch in der Konzeption des Kreditpunktesystems (ECTS) aus. Damit steigen die Anforderungen an die Qualität der Leistungsbeurteilung, denn für die unterschiedlichen Formen der Leistungskontrolle sind Standards zu definieren, welche die Aussagekraft von einzelnen Portfoliobestandteilen garantieren. Qualität bestimmend sind auch dann die testtheoretischen Gütekriterien. Mit anderen Worten: Für jede Art Prüfung und Leistungskontrolle sind der Nachweis der Objektivität, Reliabilität und Validität im Sinne einer Qualitätssicherung zu dokumentieren. Angesichts einer stärkeren Modularisierung von Lehr-Lern-Prozessen ist insbesondere der Kriteriumsvalidität von einzelnen Leistungskontrollen besondere Aufmerksamkeit zu schenken. Innerhalb eines Studienganges sollten daher die verschiedenen Leistungskontrollen miteinander verglichen und systematisch ausgewertet werden.

---

<sup>14</sup> Siehe dazu etwa bei Huber (2001).

## Literatur

- Allgemeine Verordnung über Leistungskontrollen an der Eidgenössischen Technischen Hochschule Zürich (AVL ETHZ) vom 10. September 2002.
- Bauer, U (2002). Das Prüfen ausländischer Studierender. Probleme und Lösungen, In B. Berendt, H.-P. Voss, J. Wildt (Hrsg.), Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten (Loseblatt-Ausg: Teil H3.1). Berlin: Raabe.
- Bloom, B. S., Engelhardt, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. (1956). Taxonomy of educational objectives. The classification of educational goals: Handbook 1, Cognitive domain. New York: McKay.
- Bortz, J. & Döring, N. (1995). Forschungsmethoden und Evaluation (2., vollständig überarbeitete und aktualisierte Auflage). Berlin, Heidelberg, New York: Springer.
- Fischer, G.H.. (1974). Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen. Stuttgart, Wien: Hans Huber.
- Fischer, G.H. (1983). Neuere Testtheorie, In H. Feger & J. Bredenkamp (Hrsg.), Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 3: Messen und Testen (S. 604-692). Göttingen, Toronto, Zürich: Hogrefe.
- Gage, N.L. & Berliner, D.C. (1996). Pädagogische Psychologie (5., vollständig überarbeitete Auflage). Weinheim: Beltz PVU.
- Huber, L. (2001). Lehren, Lernen, Prüfen: Probleme und Chancen von Credit-Systemen. In U. Welbers (Hrsg.), Studienreform mit Bachelor und Master. Gestufte Studiengänge im Blick des Lehrens und Lernens an Hochschulen. Modelle für die Geistes und Sozialwissenschaften (S. 43-59). Neuwied/Kriftel: Luchterhand.
- Klauer, K.J. (1983). Kriteriumsorientierte Tests, In H. Feger & J. Bredenkamp (Hrsg.), Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 3: Messen und Testen (S. 693-726). Göttingen: Hogrefe.
- Klauer, K.J. (1987). Kriteriumsorientierte Tests. Göttingen: Hogrefe.
- Lienert, G.A. & Raatz, U. (1998). Testaufbau und Testanalyse (6. Auflage). Weinheim: Beltz PVU.
- Metzger, Chr. & Nüesch, Ch. (1997): Qualitätsleitfaden zur Evaluation von Prüfungen an der Universität St. Gallen. St. Gallen: Universität St. Gallen, Institut für Wirtschaftspädagogik.
- Prahl, H.-W. (1995): Prüfungen, In L. Huber (Hrsg.), Ausbildung und Sozialisation in der Hochschule, Enzyklopädie Erziehungswissenschaft, hrsg. von D. Lenzen, Bd. 10 (S. 438-450). Stuttgart, Dresden: Klett-Verlag für Wissen und Bildung.
- Stary, J. (2002): "Doch nicht durch Worte allein..." Die mündliche Prüfung, In B. Berendt, H.-P. Voss, J. Wildt (Hrsg.), Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten (Loseblatt-Ausg: Teil H2.1). Berlin: Raabe.