

3 Prüfen, Testen, Evaluieren: Grundlagen

Es ist kein Fremdsprachenunterricht denkbar, in dem die Lehrenden nicht einen Großteil der Zeit bewusst oder auch unbewusst, formell oder auch informell evaluieren und diagnostizieren. Die Informationen, die sie dadurch bekommen, werden an die Lernenden oder ihre Eltern rückgemeldet und sind für die weitere Planung und Gestaltung des Unterrichts wichtig. Diagnostische Kompetenzen sind ein zentrales Merkmal einer professionellen Lehrerpersönlichkeit. Um diagnostische Kompetenzen zu entwickeln, braucht man allerdings ein beträchtliches Grundlagenwissen. Die wichtigsten Aspekte erarbeiten wir in diesem Kapitel.

Kapitel 3 zeigt Ihnen

- verschiedene Formen von Evaluation, z.B. Prüfungen im Unterricht, Prüfungen externer Testanbieter, Evaluationen durch Lehrende und Lernende,
- wie Sie die Gütekriterien für Prüfungen bei der Erstellung eigener Prüfungen und bei der Interpretation von Prüfungsergebnissen berücksichtigen können,
- den Aufbau und die Inhalte einiger standardisierter Prüfungen und wie Sie diese Informationen für Ihre eigene Prüfungspraxis nutzen können,
- welche Aufgabenformate in Prüfungen möglich sind,
- wie Sie beurteilen können, ob sich bestimmte Aufgabenformate und Prüfungen für Ihren Kontext und Ihre Zielsetzungen eignen.

3.1 Grundlegende Typen der Evaluation

Wir stellen Ihnen zunächst einmal die wichtigsten Typen von Evaluation und Prüfungen vor. Eine Kenntnis dieser Typen erlaubt Ihnen eine bessere Einschätzung Ihrer eigenen Praxis. Wir gehen dabei davon aus, dass es Ihnen leichter fällt, über Ihre Prüfungspraxis nachzudenken, wenn Sie verschiedene Formen der Evaluation unterscheiden können und ihre Funktionen kennen.

Wenn Sie dieses Teilkapitel bearbeitet haben, dann

- kennen Sie den Unterschied zwischen formeller und informeller Evaluation,
- kennen Sie typische Merkmale standardisierter Prüfungen,
- kennen Sie Merkmale bezugsgruppenorientierter und kriteriumsorientierter Evaluation,
- kennen Sie Unterschiede zwischen formativer und summativer Evaluation,
- kennen Sie Unterschiede zwischen Fremd-, Selbst- und Peer-Evaluation,
- können Sie die Bedeutung der vorgestellten Typen von Evaluation und Prüfungen für Ihren Unterricht einschätzen.

formelle und informelle Evaluation

Bei formellen Formen der Evaluation handelt es sich zumeist um Prüfungen und Tests. **Formelle Prüfungen und Tests** sind in der Regel das Ergebnis langwieriger und aufwändiger Bemühungen von Spezialisten. Beispiele sind die Zertifikatsprüfungen des *Goethe-Instituts*, die Prüfungen des *Österreichischen Sprachdiploms (ÖSD)*, der *Test Deutsch als Fremdsprache (TestDaF)* oder auch Englischtests wie die *Cambridge-Prüfungen*, *TOEFL* oder *IELTS*. Formelle Prüfungen müssen bestimmten, weithin akzeptierten **Gütekriterien** genügen; auf diese gehen wir in Teilkapitel 3.4 noch genauer ein.

Bei **informellen Prüfungen und Tests** handelt es sich dagegen um weniger aufwändige und weniger anspruchsvolle Verfahren. Trotzdem müssen auch informelle Prüfungen bestimmten Qualitätsanforderungen genügen. Beispiele für informelle Prüfungen sind Lernfortschrittstests, die eine Lehrkraft selbst für eine Klasse / einen Kurs erstellt, oder

selbst erstellte Prüfungen, mit denen sich die Lehrkraft einen Überblick über den Lernstand einer neu übernommenen Lernergruppe verschaffen will.

Standardisierung

Ein typisches Merkmal formeller Prüfungen ist die **Standardisierung**. Das Ziel der Standardisierung ist, die Bedingungen, unter denen die Prüfung gemacht wird, möglichst weitgehend vergleichbar zu machen, um so gültige und faire Ergebnisse zu erhalten. Bei einer standardisierten Prüfung ist z.B. genau geregelt, welche Texte und Aufgabenformate (Aufgabentypen) eingesetzt werden sollen, welche Reaktionen aufseiten der Lernenden als Belege für welche Kompetenz zu werten sind, wie die Reaktionen genau gewertet werden sollen und in vielen Fällen auch, welchem Kompetenzniveau (z.B. B1 des GER) bestimmte Leistungen entsprechen.

Wenn Sie schon einmal an einer standardisierten Prüfung teilgenommen haben, z.B. an einer Prüfung des *Goethe-Instituts* oder dem *TestDaF*, dann ist Ihnen sicherlich aufgefallen, dass die Durchführungsbestimmungen sehr streng sind. Im Anschluss an die Prüfung haben Sie dann eine Bescheinigung oder ein Zertifikat erhalten, auf dem die erreichte Punktzahl und/oder eine Einstufung auf einem inhaltlich beschriebenen Kompetenzniveau (z.B. B1 oder A2) angegeben sind.

bezugsgruppenorientierte Evaluation

Viele Evaluationen und insbesondere Prüfungen zielen lediglich darauf ab, die Teilnehmenden in eine Rangfolge zu bringen und Aussagen zu formulieren wie z.B.: „Prüfungsteilnehmerin A ist im Vergleich zu Prüfungsteilnehmer B besser.“ oder „Teilnehmerin C gehört zu den fünf leistungsstärksten Prüfungsteilnehmenden der Gruppe.“ Ein solches Vorgehen bezeichnet man als bezugsgruppenorientierte Evaluation oder häufig als normorientierte Evaluation. Bei einer bezugsgruppenorientierten Evaluation wird die Leistung jeweils relativ zu den Leistungen der übrigen Prüflinge in der jeweiligen Gruppe oder relativ zu einer externen Bezugsgruppe (z.B. alle Schülerinnen und Schüler einer bestimmten Schulstufe) charakterisiert. Man spricht hier deshalb von einer sozialen Bezugsnorm. Die Orientierung an einer sozialen Bezugsnorm kennen Sie vielleicht aus Ihren Ländern, wenn es darum geht, einen begehrten Studienplatz oder einen Arbeitsplatz zu bekommen. Eine bezugsgruppenorientierte Evaluation findet häufig auch in der Unterrichtspraxis statt, z.B. wenn Sie die besten Kandidaten für ein Stipendium nach Deutschland auswählen wollen. Mit dieser Art der Beurteilung ist allerdings ein entscheidendes Problem verbunden: Die Bewertung der Leistung als „gut“ oder „schlecht“, als „besser“ oder „schlechter“ hängt vom Leistungsstand in der jeweiligen Gruppe sowie von der Schwierigkeit der jeweiligen Prüfungsaufgaben ab. Eine Aussage, an welcher Position eine Person in einer Gruppe steht, bietet weder ihr selbst noch anderen Informationen darüber, in welchen Situationen die geprüfte Person sprachlich handeln kann und welche spezifischen sprachlichen Kompetenzen sie hat.

Gehört also jemand in einer Prüfung zu den besten zehn Prozent, dann muss dies keineswegs heißen, dass sie/er eine hohe fremdsprachliche Kompetenz aufweist. Denn zum einen ist es möglich, dass die anderen Prüfungsteilnehmenden ein besonders niedriges Niveau haben. Zum anderen hat sie/er vielleicht an einer Prüfung teilgenommen, die weit unter ihrem/seinem Niveau lag und für sie/ihn vergleichsweise leicht war.

kriteriumsorientierte Evaluation

Häufig will man nicht nur relative Aussagen bezogen auf den Leistungsstand in der jeweiligen Lerner- oder Prüfungsgruppe, sondern inhaltliche Aussagen zu spezifischen Kompetenzen machen. In diesem Fall muss man eine (zusätzliche) Beurteilung auf der Basis von vorher festgelegten Kriterien durchführen, d.h. eine sogenannte **kriteriumsorientierte/kriteriale Evaluation**. Kriterien könnten z.B. die Breite des eingesetzten Wortschatzes oder die situative Angemessenheit der Äußerungen sein. Eine kriteriumsorientierte Evaluation ist absolut, d.h. nicht von der Leistung einer Gruppe abhängig, und damit von unmittelbarer inhaltlicher Aussagekraft. Arbeitet man mit Kann-Beschreibungen, dann handelt es sich stets um eine kriteriumsorientierte Evaluation. Ist das Kriterium als Lernziel oder Lehrziel formuliert, dann spricht man auch von lernziel- bzw. lehrzielorientierter Evaluation. Eine kriteriale, an transparenten Lernzielen orientierte Evaluation kann

für viele Lernende eine motivierende Funktion haben, da sie genau wissen, was von ihnen erwartet wird. Gerade bei schwächeren Lernenden kann aber eine kriteriumsorientierte Evaluation auch demotivierend wirken, wenn alle Lernenden an den gleichen Kriterien gemessen werden. Dem können die Lehrenden z.B. dadurch entgegenwirken, dass sie den individuellen Fortschritt ihrer Lernenden besonders lobend herausstellen. So könnte die Lehrkraft einen Bereich hervorheben, den eine Lernerin / ein Lerner schon besser bearbeitet hat als in der letzten Prüfung, und dafür einen Pluspunkt bei der Bewertung geben. Außerdem ist es im Rahmen einer eher informellen Evaluation natürlich möglich, binnendifferenzierend vorzugehen und für Lernende auf unterschiedlichen Kompetenzstufen jeweils unterschiedliche Aufgaben und Bewertungsmaßstäbe einzusetzen.

Ein Prüfungsergebnis kann zumeist sowohl bezugsgruppenorientiert als auch kriteriumsorientiert interpretiert werden: So erfolgt die Notengebung im Unterricht häufig gleichzeitig bezugsgruppenorientiert und kriterial. Auch die Skalenwerte vieler standardisierter Prüfungen lassen sich sowohl in Bezug auf die Leistungsverteilung in der jeweiligen Bezugsgruppe als auch inhaltlich interpretieren, wenn es z.B. um die Bewältigung bestimmter Anforderungen geht.

Aufgabe 14

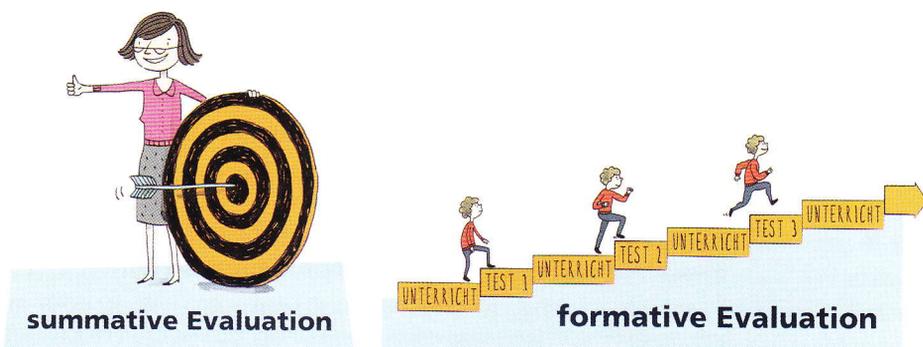
Beschreiben Sie Situationen, in denen Sie bezugsgruppenorientiert bzw. kriteriumsorientiert evaluieren.

Bezugsgruppenorientiert gehe ich vor, wenn

Kriteriumsorientiert gehe ich vor, wenn

summative und formative Evaluation

Die Bewertung von Leistungen kann punktuell und produkt-/ergebnisorientiert am Ende eines Lernabschnitts stattfinden (summativ) oder kontinuierlich und prozessorientiert in den Unterricht integriert werden (formativ). Allerdings kann man einer Evaluation in Form einer Prüfung zumeist nicht unmittelbar ansehen, ob sie summativ oder formativ ist. Das hängt von dem Ziel ab, mit dem die Prüfung eingesetzt wird.



Die **summative Evaluation** kann als interne Evaluation innerhalb einer Schule oder eines anderen Bildungsträgers stattfinden oder auch als externe Evaluation, etwa im Rahmen einer öffentlichen Rechenschaftspflicht. Typische Beispiele summativer Evaluation sind die am Ende eines Halbjahres oder Jahres in der Schule vergebenen Zeugnisnoten oder auch landesweite Tests zur Überprüfung, inwieweit bestimmte Standards am Ende eines bestimmten Bildungsabschnitts erreicht wurden (siehe auch Teilkapitel 1.3 und 2.3). Summative Evaluation zielt damit auf (punktuelle) Qualitätskontrolle und ist häufig verbunden mit Funktionen wie Notengebung, Zuweisung (z.B. zu einem bestimmten Kurs) oder Auslese (z.B. im Hinblick auf ein Studium). Entsprechend wird in der englischsprachigen Literatur die summative Evaluation zuweilen auch als *assessment of learning* charakterisiert, d.h. als Evaluation des Lernerfolgs.

Bei der **formativen Evaluation** werden die Ergebnisse dagegen direkt in die weitere Planung des Unterrichts einbezogen. Sie dienen dazu, den Unterricht zu optimieren und die Kompetenzen der Lernenden weiterzuentwickeln. Deshalb wird sie treffend auch als *assessment for learning*, d.h. als Evaluation im Dienste des Lernens charakterisiert. Ein zentrales Instrument der formativen unterrichtlichen Evaluation ist die Beobachtung. Formative Evaluation erfolgt üblicherweise kriteriumsorientiert und vermeidet einen bezugsgruppenorientierten Vergleich der Lernenden untereinander. Zudem werden die Lernziele und Bewertungskriterien transparent gemacht und soweit möglich von den Lehrenden und Lernenden gemeinsam verantwortet. Während es sich bei summativen Evaluationen häufig um Tests oder Prüfungen handelt, die für den Geprüften mit weitreichenden Konsequenzen verbunden sind (High-Stakes-Tests), ist dies bei der formativen Evaluation in der Regel nicht der Fall.

Im Grunde ist jede Art der lernprozessbegleitenden Diagnose eine Art von formativer Evaluation. Wenn Sie also z.B. bei der Bearbeitung einer Aufgabe beobachten, dass Lernende nicht adäquat über Vergangenes erzählen können und diese Information zur Rückmeldung an die Lernenden oder auch für die Planung Ihres weiteren Unterrichts nutzen, dann ist das bereits ein Beispiel für eine formative Evaluation. Im Gegensatz zum standardisierten Prüfen ist eine formative Evaluation häufig nicht vorher geplant und findet oft auch als dialogische und kooperative Interaktion zwischen der Lehrkraft und den Lernenden statt. Wir werden auf diesen Aspekt noch im Zusammenhang mit der dynamischen Evaluation (siehe Teilkapitel 6.1.2) zu sprechen kommen.

Schließlich sollten Rückmeldungen beim formativen Evaluieren deskriptiv formuliert sein, d.h. es sollten Stärken und Schwächen des Lernenden beschrieben werden, damit die Lernenden ihre eigenen Fortschritte, aber auch mögliche Probleme wahrnehmen und reflektieren können. Die Rückmeldungen sollten möglichst nicht mit einer Benotung verbunden sein, sondern vielmehr auch auf eine Steigerung der Motivation zielen.

Wenn man mehrfach formativ testet, kann man die Dynamik der Entwicklung von lerner-sprachlichen Kompetenzen besser berücksichtigen. Man geht heutzutage davon aus, dass Lernerkompetenzen sich nicht linear, sondern dynamisch entwickeln und dass als Folge auch die Lernaltersprache ein dynamisches System darstellt. Lernerkompetenzen sind gekennzeichnet durch Phänomene wie plötzliche Rückschritte, plötzliche Fortschritte im Zuge grundlegender Umstrukturierungen des Systems oder langfristige Verfestigungen von fehlerhaften Strukturen (Fossilierungen). Eine einmalige (summative) Testung erlaubt nur eine Momentaufnahme: Werden Lernende z.B. gerade in einer Phase getestet, in der ihre Kompetenzen und ihre Lernaltersprache instabil sind, dann ergibt sich ein in Teilen verfälschtes Bild der tatsächlichen Kompetenzen. Ein mehrfaches (formatives) Testen kann dem entgegenwirken.

Aufgabe 15

Welche der folgenden Szenarien können eher als **summative Evaluation** und welche eher als **formative Evaluation** gelten? Kreuzen Sie an.

Szenario	summative Evaluation	formative Evaluation
1. Li Yang nimmt in Deutschland an einem Vorbereitungskurs für das <i>Zertifikat B1</i> teil. Am Ende des Kurses legt sie die Prüfung für das <i>Zertifikat B1</i> erfolgreich ab.	<input type="checkbox"/>	<input type="checkbox"/>
2. Die Lehrerin weist den Schüler Omar nach einer teilweise inkorrekten Äußerung auf seinen Fehler hin, indem sie die Äußerung noch einmal fragend und mit Betonung auf dem Fehler wiederholt: „Du hast gestern nach München gefahren?“. Omar korrigiert sich selbst („Ah ja, ich bin nach München gefahren.“) und erhält dafür eine Bestätigung und ein Lob der Lehrerin: „Ja genau, gut. Du bist nach München gefahren.“	<input type="checkbox"/>	<input type="checkbox"/>

Szenario	summative Evaluation	formative Evaluation
3. Ein Lehrer setzt während eines Kurses selbst entwickelte, informelle Vokabeltests ein, um festzustellen, ob die Lernenden den bearbeiteten Wortschatz gelernt haben. Die Ergebnisse bespricht er mit den Lernenden; er wiederholt noch einmal schwierige Wörter und gibt Empfehlungen für das Weiterlernen.	<input type="checkbox"/>	<input type="checkbox"/>

Zusammenfassend können wir festhalten, dass sich formative und summative Evaluation anhand von fünf Merkmalen unterscheiden lassen:

formative Evaluation	summative Evaluation
Die Evaluation ist in den Unterricht integriert und findet regelmäßig statt.	Die Evaluation steht am Ende eines Lernabschnitts.
Die Evaluation ist häufig interaktiv, dialogisch und kooperativ.	Die Evaluation ist eine Momentaufnahme der Leistungen des Lernenden durch die Lehrkraft.
Die Ergebnisse der Evaluation werden von der Lehrkraft für die Optimierung der Lehrprozesse genutzt.	Die Ergebnisse der Evaluation dienen der Lehrkraft zur Information, ob und inwieweit die Lernziele des jeweiligen Lernabschnittes von den Lernenden erreicht wurden.
Soweit es in der Unterrichtssituation sinnvoll ist, werden die Ergebnisse so an die Lernenden zurückgemeldet, dass diese selbst ihr Lernen optimieren können.	Die Ergebnisse werden an die Lernenden als ein abschließender Lernertrag zurückgemeldet. Die Ergebnisse bilden meist die Grundlage für Auslese und Einstufung der Lernenden in Leistungsniveaus.
Rückmeldungen sind in der Tendenz deskriptiv und zu meist nicht mit einer Benotung verbunden.	Rückmeldungen erfolgen durch Bewertung und Benotung.

Wir kommen nun zu den wichtigen Unterschieden zwischen **Fremd-, Selbst- und Peer-Evaluation**. Es hängt natürlich stark von Ihrer spezifischen Situation und von der Gesamtplanung des Unterrichts ab, wann Sie die einzelnen Formen einsetzen: wann Sie also z.B. das Ablegen externer Prüfungen unterstützen, wann Sie als Lehrkraft bewerten, wann Sie Lernenden eine gegenseitige Bewertung vorschlagen und wann Sie Formen der Selbstevaluation anregen.

Fremdevaluation

Alle standardisierten Tests und Prüfungen (z.B. *DSD*, Zertifikatsprüfungen des *Goethe-Instituts*, *ÖSD*, *TestDaF*) und zum großen Teil auch die Prüfungen, die Sie in Ihrem Unterricht einsetzen, werden mit dem Ziel der Fremdevaluation genutzt. Dabei gilt: Eine externe Institution oder auch Sie als Lehrkraft beurteilen die sprachlichen Kompetenzen der Prüflinge. Häufig erhalten die Prüflinge nach einer Fremdevaluation ein offizielles Zertifikat, eine Bescheinigung oder im Fall einer schulischen Fremdevaluation eine Note. Wird eine unterrichtliche Fremdevaluation von einer externen Institution – wie etwa im Fall der *PISA-Studien* – durchgeführt, spricht man auch von externer Evaluation. Externe Evaluationen sind zumeist formell. Führen Lehrende Tests und Prüfungen in eigener Verantwortung im Unterricht durch, spricht man von interner Evaluation. Interne Evaluationen können sowohl formell als auch informell sein.

Selbstevaluation

Eine typische Selbstevaluation hat folgende Merkmale: Sie kann sowohl von der Lehrkraft initiiert werden als auch von den Lernenden selbst. Dazu können u.a. auch vorgefertigte Selbsteinstufungstests eingesetzt werden. Die Ergebnisse geben Auskunft über den eigenen Lernstand und helfen dabei, den weiteren Lernprozess zu steuern.

Lernende sollen in einer Selbstevaluation auch ihr eigenes Lernen reflektieren und Nutzen für weiteres Lernen daraus ziehen. Einige in diesem Zusammenhang wichtige Fragen, die zur Selbstreflexion anregen sollen, haben wir im Folgenden aufgelistet:



Selbstevaluation	
Was kann und weiß ich?	_____
Welches sind meine besonderen Stärken und Schwächen?	_____
Wo muss/will ich mich verbessern?	_____
Wie kann ich dabei vorgehen?	_____
Worauf muss/will ich dabei achten (z.B. Zeit, Schwierigkeit)?	_____
Wie überprüfe ich, ob mir mein Vorgehen dabei geholfen hat, mich zu verbessern?	_____
Welche weiteren Entscheidungen treffe ich und wie kontrolliere ich den Erfolg?	_____

Wenn Lernende ihre Kompetenzen selbst evaluieren, dann können die entsprechenden Informationen auch für die Lehrkraft sehr wichtig sein. Denn manchmal schätzt man z.B. bei Lernenden, die im Unterricht eher zurückhaltend sind, als Lehrkraft die mündliche Kompetenz eher niedrig ein. Eine abweichende Selbsteinschätzung von Lernenden kann dafür sensibilisieren, sie in Zukunft genauer zu beobachten.

Peer-Evaluation

In Teilkapitel 2.3 haben Sie selbst schon Erfahrungen mit einer Peer-Prüfung gemacht. Sie sollten sich gegenseitig Fragen stellen und anschließend gemeinsam überlegen, was Sie schon gut können, wo Sie möglicherweise noch Defizite haben und was Sie sich unbedingt merken wollen. Sie sollten also mit Ihnen gleich gestellten Partnern (Peers) eine Sie betreffende Evaluation durchführen. Die Funktion von Peer-Evaluationen besteht vor allem darin, dass Lernende sich gegenseitig dabei helfen, über ihre sprachlichen Kompetenzen und Möglichkeiten zum Weiterlernen nachzudenken. Peer-Beurteilungen sind damit ebenso wie die Selbstevaluation ein wichtiges Verfahren für die Entwicklung der Fähigkeit zum selbstreflexiven Lernen.

Sowohl Selbst- als auch Peer-Evaluationen werden häufig im Rahmen einer formativen Evaluation in Situationen eingesetzt, die für die Lernenden keine weitreichenden Konsequenzen haben.

Aufgabe 16



Welche Formen der Selbst- und Peer-Evaluation verwenden Sie in Ihrem Unterricht und welche Erfahrungen haben Sie damit gemacht?

a) Notieren Sie.

Formen der Selbst- bzw. Peer-Evaluation im Unterricht	Erfahrungen

b) Tauschen Sie sich auch mit Ihren Kolleginnen und Kollegen aus.

Sie haben nun verschiedene Evaluationsverfahren kennengelernt. Überprüfen Sie abschließend, was Sie gelernt haben.

Aufgabe 17

Ergänzen Sie den Lückentext mit den Begriffen aus dem Schüttelkasten.

Selbstevaluation • bezugsgruppenorientierte Evaluation • Peer-Evaluation • kriteriumsorientierte oder kriteriale Evaluation • formative Evaluation

- A. Wenn ich als Lehrkraft die sprachlichen Fähigkeiten der Lernenden kontinuierlich beurteile und ihnen Rückmeldung (Feedback) gebe, dann nennt man das _____.
- B. Wenn Lehrende einen Test einsetzen, weil sie die drei besten Deutschlernenden an einer Schule identifizieren wollen und sie dann für den Wettbewerb „Deutscholympiade“ anmelden wollen, dann handelt es sich um eine _____.
- C. Wenn Lehrende einen Test einsetzen und die Leistungen mit Blick auf Merkmale wie Inhalt, Verständlichkeit und Korrektheit beurteilen, dann handelt es sich um eine _____.
- D. Wenn Lernende sich gegenseitig Rückmeldung zu ihren Leistungen geben, dann handelt es sich um eine _____.
- E. Wenn Lernende Aussagen über ihr eigenes Können anhand von (vorgegebenen) Kann-Beschreibungen vornehmen, dann nennt man dies _____.

Literatur
zum Weiterlesen

Weitere Hinweise zum Unterschied zwischen Fremdevaluation, Selbstevaluation und Peer-Evaluation finden Sie bei Harsch (2009), Kleppin (2008) und Roche (2010). Informationen zur formativen Evaluation fremdsprachlicher Kompetenzen, die auch für den Unterricht relevant sind, enthalten Dlaska/Krekeler (2009) und Kieweg (2010). Die formative und die summative Beurteilung wird in den Kapiteln 9.3.4 und 9.3.5 des GER (Europarat 2001) beschrieben.

Wichtige Charakteristiken eines *assessment for learning* (Evaluation im Dienste des Lernens) hat u.a. die *Assessment Reform Group* (2002) in ihrem Bericht *Assessment for Learning: 10 principles. Research-based principles to guide classroom practice* formuliert. Sie finden diesen grundlegenden Bericht im Internet, wenn Sie den Titel in Ihre Suchmaschine eingeben.

Zusammenfassung

Sie haben in diesem Teilkapitel gelernt, die wichtigsten Evaluationsformen zu unterscheiden, die Ihnen in Ihrer Praxis begegnen können:

- Fremdevaluationen in Form von externen Prüfungen, auf die Sie Ihre Lernenden vorbereiten,
- summative oder formative Evaluationen in Ihrem Unterricht, die Sie selbst erstellen und bewerten,
- bezugsgruppenorientierte oder/und kriterienorientierte Evaluationen,
- Evaluationen, in denen die Lernenden sich selbst oder sich gegenseitig Rückmeldung zu ihren Leistungen geben.

Diese unterschiedlichen Evaluationsformen haben also bestimmte Funktionen und jeweils Vor- und Nachteile, die Sie kennen sollten, um Ihre eigene Prüfungs- und Evaluationspraxis zu reflektieren und angemessen auszugestalten. Machen Sie sich also immer bewusst, welchen Zweck Sie mit einer Evaluation verfolgen und welche Form der Evaluation dafür am geeignetsten ist.

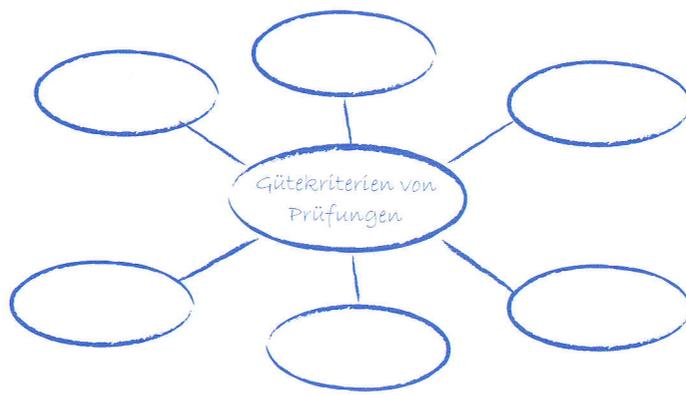
3.4 Gütekriterien von Prüfungen und Tests

Mit diesem Teilkapitel möchten wir erreichen, dass Sie

- wichtige Qualitätsmerkmale von Prüfungen und Tests, sogenannte Gütekriterien, kennen, wie z.B. Validität, Reliabilität, Objektivität, Fairness,
- Prüfungen mithilfe der Gütekriterien im Hinblick auf ihre Qualität einschätzen können,
- beurteilen können, inwieweit Sie die Gütekriterien auch in Ihren selbst erstellten Prüfungen berücksichtigen wollen und können.

Aufgabe 21

Welche Gütekriterien von Prüfungen kennen und beachten Sie schon in der eigenen Prüfungspraxis?



Wir stellen Ihnen nun wichtige Gütekriterien für Prüfungen und Prüfungsaufgaben vor. Sicherlich haben Sie in Ihrer Mindmap schon einige der wesentlichen Merkmale erfasst.

Validität (Gültigkeit)

Beim Einsatz von Prüfungen müssen wir uns stets fragen, ob und in welchem Umfang die Aussagen und Entscheidungen, die wir anhand der beobachteten Leistungen der Prüflinge treffen, gerechtfertigt sind. Anstelle von gerechtfertigt kann man auch von **valide** (gültig) sprechen.

Dabei geht es zunächst um unsere Interpretationen der beobachteten Prüfungsleistungen im Hinblick auf das zu messende Konstrukt. Im Fall des Konstrukts „Leseverstehen“ kann dies z.B. heißen: Können wir auf der Basis unserer Interpretation der beobachteten Leistung bei der Beantwortung von Verständnisfragen zu Lesetexten hinreichend valide Aussagen zu den Leseverstehenskompetenzen der Prüfungsteilnehmenden in bestimmten Realsituationen formulieren?

Weiterhin müssen wir uns fragen: Inwieweit können wir auch die Entscheidungen, die wir anhand unserer Interpretationen der Prüfungsergebnisse treffen, rechtfertigen? Können wir davon ausgehen, wenn es um Versetzungen, Zulassungen, Einstufungen usw. geht, dass unsere Prüfungen eine valide Entscheidungsgrundlage darstellen und dass die getroffenen Maßnahmen so weit wie möglich fehlerfrei sind? Auch dies ist ein Aspekt der **Validität** von Prüfungen und Prüfungsaufgaben.

Die Validität gilt als **das zentrale Gütekriterium** von Prüfungen. Das Gütekriterium Validität kann allerdings nur dann an ein Prüfungs- oder Testformat angelegt werden, wenn man die jeweilige Zielsetzung und den Einsatzbereich kennt. Eine Aussage wie: „Diese Prüfung ist valide“ ist ohne Angabe des Verwendungszwecks und des Testkonstrukts deshalb nicht gerechtfertigt bzw. eine unzulässige Vereinfachung. Außerdem kann eine Prüfung mehr oder minder valide sein. So ist der *Deutsch-Test für Zuwanderer (dtz)*, der sich

auf Niveaustufe B1 bezieht, im akademischen Kontext deutlich weniger valide als der *TestDaF*, der speziell für diesen Kontext erstellt wurde.

Ein weiterer Aspekt der Validität bezieht sich auf die inhaltliche Gültigkeit der Prüfungs- und Testaufgaben, z.B. vor dem Hintergrund eines bestimmten Curriculums oder Lehrplans. Valide bedeutet in diesem Zusammenhang, dass die Testaufgaben das Curriculum oder den Lehrplan möglichst gut repräsentieren sollten. Eine Curriculum-basierte Prüfung sollte z.B. keine grammatischen Phänomene beinhalten, die im Unterricht noch nicht vorgekommen oder im Curriculum gar nicht vorgesehen sind. Dieser gerade im unterrichtlichen Kontext wichtige Aspekt wird als Inhaltsvalidität oder speziell bezogen auf Curricula auch als curriculare Validität bezeichnet.

Tests validieren

Die Validierung einer Prüfung, d.h. der Nachweis, dass eine Prüfung eine zufriedenstellende Gültigkeit aufweist, ist in der Regel ein anspruchsvolles Unternehmen und bedarf einer komplexen Argumentation. Entwickelt eine Lehrkraft eine Prüfung oder auch einzelne Aufgaben, dann ist ein Nachweis der Validität sicherlich nur ansatzweise möglich. Trotzdem sollten auch Lehrende stets kritisch reflektieren, ob die eingesetzten Prüfungsaufgaben für die jeweilige Zielsetzung hinreichend valide sind.

Zuweilen wird argumentiert, dass Prüfungsformate, die die Verwendung von Sprache außerhalb der Prüfungssituation (d.h. in authentischen Realsituationen) weitgehend simulieren, automatisch valide seien. Aber auch in Bezug auf solche authentischen Aufgabenformate gilt, dass man möglichst präzise spezifizieren muss, was genau gemessen werden soll, und dass auch genau begründet werden muss, warum eine bestimmte Prüfungsaufgabe verwendet wird und warum die Aufgabe auf eine bestimmte Weise bewertet wurde. Der Austausch mit Kolleginnen und Kollegen ist in diesem Zusammenhang sehr hilfreich.

Reliabilität (Zuverlässigkeit)

Reliabilität bezieht sich auf unterschiedliche Aspekte der Zuverlässigkeit bzw. Verlässlichkeit von Messinstrumenten und Prüfungsergebnissen. Da es sich bei Reliabilität um ein nicht einfaches Konzept handelt, geben wir Ihnen zunächst ein Beispiel in Form eines Gedankenexperiments: Stellen Sie sich vor, Sie wollen die genaue Größe von mehreren Personen messen. Sie messen mit einem flexiblen Zentimetermaßband aus Gummi. Um ganz sicher zu sein, wiederholen Sie die Prozedur kurz darauf noch einmal. Sie vergleichen die Ergebnisse und stellen erhebliche Unterschiede fest. Natürlich ist Ihnen sofort klar: Die tatsächliche Größe der einzelnen Personen kann sich in der Kürze der Zeit nicht verändert haben. Also muss es wohl an dem flexiblen Maßband aus Gummi liegen, das für diese Messung kein geeignetes Instrument ist. Anders ausgedrückt: Die Ergebnisse des Messvorgangs mit diesem Instrument sind unzuverlässig, also nicht hinreichend reliabel.



Doch nehmen wir ein Beispiel aus dem Unterrichtskontext. Sie haben wahrscheinlich schon einmal die Erfahrung gemacht, dass Sie, wenn Sie sich eine Schreibaufgabe ein zweites Mal angesehen haben, zu einer etwas anderen Einschätzung der Leistung gekommen sind als beim ersten Mal.

Aufgabe 22

Welche Gründe können dazu führen, dass man bei einer zweiten Durchsicht einer Schreibaufgabe zu einer abweichenden Einschätzung kommen kann?

Mögliche Gründe können sein:

Wenn Sie selbst bei einer zweiten Durchsicht einer Schreibaufgabe zu einer anderen Einschätzung der Leistung kommen oder wenn ein Zweitkorrektor ein anderes Urteil als Sie abgibt, dann geht es in beiden Fällen um die Beurteilerreliabilität. Wenn dieselben Beurteilenden stets weitgehend gleich streng bzw. milde und nach den gleichen Kriterien urteilen, d.h. ein konsistentes Urteil abgeben, dann spricht man auch von Intrarater-Reliabilität. Wenn verschiedene Beurteilende zu einer ähnlichen oder gleichen Einschätzung kommen, dann spricht man von Interrater-Reliabilität.

Objektivität

Im Hinblick auf das Gütekriterium der **Objektivität** kann man drei Aspekte unterscheiden:

- Je weniger das Prüfungsergebnis von den Durchführungsbedingungen und speziell von den Prüfenden abhängt, desto größer ist die Durchführungsobjektivität der Prüfung / der Aufgabe.
- Je weniger das Ergebnis von den Auswertungs- und Bewertungsbedingungen abhängt, desto größer ist die Auswertungsobjektivität. Geschlossene Aufgabenformate wie Mehrfachwahlaufgaben sind völlig auswertungsobjektiv (sie können vom Computer ausgewertet werden) und werden deshalb auch als objektive Aufgabenformate bezeichnet. Entsprechend werden Aufgabentypen, bei denen die Bewertenden einen subjektiven Ermessensspielraum haben, wie z.B. das Schreiben eines Aufsatzes, auch als subjektive Aufgabenformate bezeichnet.
- Der dritte Aspekt der Objektivität ist die sogenannte Interpretationsobjektivität. Hier geht es darum, inwieweit unterschiedliche Personen bei der Interpretation der Ergebnisse der Auswertung und Bewertung zu übereinstimmenden Interpretationen kommen. So hat z.B. eine Prüfung eine perfekte Interpretationsobjektivität, wenn mit einem bestimmten Prüfungsergebnis (Punktwert) automatisch die Zuweisung zu einer bestimmten Lernergruppe oder auch die Vergabe einer bestimmten Note verbunden ist. Kommen dagegen z.B. zwei Lehrende bei einem bestimmten Punktwert in ein und demselben Aufsatz zu deutlich unterschiedlichen Noten, liegt eine unzureichende Interpretationsobjektivität vor.

Die Standardisierung einer Prüfung oder eines Tests zielt in der Regel auf eine Erhöhung aller drei genannten Typen von Objektivität. Insbesondere wenn es um die Auswertung von Aufgaben mit einem subjektiven Ermessensspielraum und/oder um die Vergabe von Noten oder Berechtigungen geht, sprechen manche Autoren auch von Bewertungsobjektivität.

Objektivität und Reliabilität sagen für sich allein genommen allerdings nur relativ wenig über den Wert einer Prüfung aus. Sie sind vor allem deswegen wichtige Gütekriterien, weil es sich um notwendige Voraussetzungen für die Validität handelt. Konkret bedeutet dies, dass eine wenig objektive Prüfung auch nicht hinreichend reliabel sein kann und eine wenig reliable Prüfung auch nicht zu hinreichend validen und fairen Messergebnissen führen kann.

Fairness

Bei dem Gütekriterium der **Fairness** geht es um die Gerechtigkeit bei Prüfungen. Dabei denken Sie wahrscheinlich als Erstes an eine gerechte Bewertung und Notengebung oder auch an eine Durchführung, bei der alle die gleichen Chancen haben. Oben haben wir

bereits erwähnt, dass es z.B. nicht vorkommen darf, dass einige Teilnehmende allein aufgrund des Abstandes zur Hörquelle (Lautsprecher) oder aufgrund einer größeren Vertrautheit mit den Aufgabenformaten bei ansonsten gleichen Fähigkeiten Aufgaben besser lösen als andere Teilnehmende. Auch bei der Erstellung einer Prüfung kann es schon zu Fairnessproblemen kommen, z.B. im Hinblick auf die Vertrautheit mit einem Thema. Ebenso kann es bei der Bewertung von Prüfungsleistungen zu Gerechtigkeitsproblemen kommen.

Aufgabe 23

Nennen Sie einige Beispiele möglicher Fairnessprobleme und begründen Sie.

Beispiel	Ein mögliches Fairnessproblem kann entstehen, weil ...

Bestimmte Textinhalte können bestimmte Lernergruppen wie z.B. weibliche oder männliche Prüfungsteilnehmende systematisch benachteiligen. Einige Themen können auch bestimmte Gruppen beleidigen oder kränken und dadurch einen Einfluss auf die Ergebnisse haben (konstruktirrelevante Varianz). Solche Themen werden häufig auch als Tabuthemen bezeichnet, also Themen, die nicht in Prüfungen vorkommen sollten. Die Fairness von Prüfungen ist eng verbunden mit der Validität und kann als ein spezifischer Aspekt der Validität angesehen werden.

Aufgabe 24

Welche Themenbereiche könnten für Ihre Lernergruppen in Prüfungen als Tabuthemen gelten? Warum?

Washback-Effekt
für den Unterricht

Natürlich sollten Prüfungsaufgaben immer so gestaltet sein, dass sie eine möglichst positive Auswirkung auf den Unterricht haben. Wenn z.B. nur Grammatikaufgaben in Prüfungen vorkommen, dann werden sich viele Lehrende in ihrem Unterricht vor allem auf die Vermittlung von Grammatik beziehen. Auch die Lernenden bereiten sich dann vor allem auf die entsprechenden Prüfungsaufgaben zur Grammatik vor und vernachlässigen die Kompetenzen, die in der Prüfung nicht erfasst werden, wie z.B. situationsangemessenes, kohärentes Schreiben. Der negative und der positive Effekt von Prüfungen auf den Unterricht wird – das haben wir in Teilkapitel 1.3 schon erwähnt – häufig als Washback- oder Backwash-Effekt bezeichnet. Darüber hinaus werden diese Begriffe in einem weiteren Sinn verwendet und beziehen sich dann auch auf die Rückwirkungen von Prüfungen, etwa auf die Curriculumentwicklung oder die Zulassungspolitik von Hochschulen (wie im Fall des *TestDaF*). Der Washback kann ebenso wie die Fairness als ein Aspekt der Validität gesehen werden.

Authentizität

Im Hinblick auf die **Authentizität** als Gütekriterium sind drei ihrer zentralen Merkmale für Sie als Lehrkraft besonders interessant:

- Die sprachliche Authentizität, insbesondere der verwendeten Texte. Texte sollten aus authentischen Quellen stammen wie z.B. aus einer deutschen Jugendzeitschrift. Falls Sie den Text vereinfachen oder kürzen müssen, sollten Sie darauf achten, dass er seine sprachliche Authentizität so weit wie möglich behält.
- Die situationelle Authentizität der Prüfungsaufgaben, d.h. die Aufgabe sollte einer realen zielsprachlichen Verwendungssituation entsprechen.
- Die in der Prüfungssituation geforderten sprachlichen Handlungen sollen den im zielsprachlichen Verwendungskontext vorkommenden sprachlichen Handlungen entsprechen.

Praktikabilität

Hierunter können Sie sich sicherlich schon vieles vorstellen. Wenn Sie z.B. gern einen Online-Test im Unterricht einsetzen wollen, aber nur eingeschränkt Zugang zum Internet haben, dann ist der Einsatz eines Online-Tests für Sie vermutlich nicht hinreichend prakti-

kabel. Insgesamt gilt: Immer wenn die für eine Prüfung notwendigen Ressourcen die vorhandenen Ressourcen überschreiten, ist die **Praktikabilität** einer Prüfung unzureichend. Die Praktikabilität kann sich auf die Herstellung, die Durchführung und die Auswertung von Prüfungen beziehen.

Nützlichkeit

Ein weiteres Gütekriterium für Prüfungen und andere Evaluationsinstrumente ist ihre **Nützlichkeit** (auch Zweckmäßigkeit; engl. *usefulness*). Nützlichkeit gilt häufig als übergeordnetes Gütekriterium und bestimmt den Stellenwert der anderen Kriterien. Je nach Zielsetzung einer Prüfung wird man z.B. die Validität, Reliabilität, Authentizität und den Washback-Effekt unterschiedlich gewichten. So wird man für einen informellen Test im Unterrichtskontext die Reliabilität eher geringer gewichten. Dagegen ist es bei einem wichtigen (High-Stakes-)Zulassungstest nicht akzeptabel, wenn z.B. eine Erhöhung der Praktikabilität oder auch Authentizität zugleich zu einer deutlichen Verringerung der Reliabilität führen würde.

Transparenz

Spätestens seit Erscheinen des GER wird **Transparenz** immer häufiger als ein Qualitätsmerkmal von Prüfungen genannt. Damit ist gemeint, dass sich alle Adressaten von Prüfungen (die Teilnehmenden selbst, im Fall von jungen Lernenden auch die Eltern, mögliche Nutzerinnen und Nutzer der Prüfungsergebnisse wie Arbeitgeber oder Hochschulen) über Zielsetzung, Testkonstrukt, Prüfungsaufbau, Aufgabentypen, Qualitätsmerkmale, Art der Ergebnisermittlung, Interpretation der Ergebnisse usw. ausreichend informieren können. Für Ihre Prüfungen im Unterrichtskontext ist vor allem von Bedeutung, dass die Lernenden wissen, nach welchen Kriterien z.B. ihre Schreib- und Sprechleistungen beurteilt werden.

Trennschärfe

Bisher haben wir die Gütekriterien schwerpunktmäßig im Hinblick auf einen gesamten Test vorgestellt. **Trennschärfe** bezieht sich dagegen auf einzelne Teilaufgaben, auf sogenannte Items oder Gruppen von Items. Mit dem Begriff **Item** werden in der Regel (dekontextualisierte) Einzelaufgaben bezeichnet. Items sollten so gestaltet sein, dass sehr kompetente Prüflinge diese häufiger richtig lösen als weniger kompetente. Ist dies der Fall, dann ist das Item hinreichend trennscharf. Wenn nämlich weniger kompetente Prüflinge ein Item häufiger richtig lösen als sehr kompetente Prüflinge, dann könnte es sein, dass das Item etwas anderes testet, als man eigentlich testen wollte.

Schwierigkeit/ Leichtigkeit

Soll zwischen Lernenden so differenziert werden, dass diese in eine Rangfolge gebracht werden können, muss bei der Konstruktion einer Prüfung darauf geachtet werden, dass Items nicht so schwer sind, dass sie von keinem Prüfling gelöst werden können, und nicht so einfach, dass alle sie lösen können. Denn wenn keiner oder andererseits alle ein Item lösen können, dann liefert dieses keine Informationen zur Leistungsdifferenzierung innerhalb der Lernergruppe. Im Rahmen einer bezugsgruppenorientierten Evaluation sind Items wertlos, die von allen Teilnehmenden oder von keinem gelöst werden. Im Hinblick auf eine kriteriumsbezogene Evaluation können solche Items jedoch durchaus informativ sein, da sie anzeigen, dass das jeweilige Kriterium (z.B. ein Lernziel) von allen Prüfungsteilnehmenden erreicht wurde.

Viele Testanbieter wie das *Goethe-Institut* oder auch das *TestDaF-Institut* machen mittlerweile auf ihrer Homepage transparent, wie ihre Prüfungen aussehen und welchen Qualitätsmerkmalen sie genügen. Hinweise zum Aufbau der Goethe-Prüfungen finden Sie z.B. unter <http://www.goethe.de/lrn/prj/pba/bes/deindex.htm> und in den dort herunterladbaren ausführlichen Handbüchern, wie z.B. dem informativen Handbuch zum neuen *B1-Zertifikat* (Glaboniat/Perlmann-Balme/Studer 2013).



Literatur zum Weiterlesen

Vertiefende Informationen zu den Gütekriterien von Prüfungen finden Sie in Grotjahn (2008, 2010). Bachman/Palmer (2010) diskutieren ausführlich anhand einer Vielzahl von Beispielen das Gütekriterium der Nützlichkeit (Zweckmäßigkeit) sowie das Vorgehen bei der Validierung von Sprachprüfungen. Plassmann (2011) beschreibt am Beispiel des

Ein wesentlicher Bestandteil der Prüfungsspezifikationen ist die Beschreibung der zu verwendenden Aufgaben in Form von mehr oder minder standardisierten und detaillierten **Aufgabenspezifikationen**. Auch beim Prüfen im Unterricht sind Prüfungs- und Aufgabenspezifikationen, in denen das Testkonstrukt, die Prüfungsziele sowie die Aufgabenformate angegeben werden, wichtige Hilfsmittel, um adäquate Prüfungen zu erstellen. Wenn Sie z.B. zusammen mit anderen Kolleginnen und Kollegen Aufgaben für parallele Prüfungen erstellen, die über Ihre Klasse / Ihren Kurs hinaus für einen Leistungsvergleich eingesetzt werden sollen, dann ist sicherzustellen, dass sich die Prüfungen möglichst wenig unterscheiden. Hier können schriftliche Prüfungsspezifikationen und genaue Spezifikationen der zu entwickelnden Aufgaben einen wichtigen Beitrag zur Qualitätssicherung leisten.

3.7 Aufgabenformate in Prüfungen

Wenn Sie dieses Teilkapitel bearbeitet haben, dann

- kennen Sie wichtige Aufgabenformate, die in Prüfungen verwendet werden,
- können Sie entscheiden, welche Formate Sie auch in Ihrem Unterricht einsetzen, um möglichst adäquat bestimmte Kompetenzen zu überprüfen.

Sie haben bereits eine Reihe von **Aufgabenformaten** kennengelernt. Mit einigen sind Sie sicherlich auch als Prüfungsteilnehmende in Berührung gekommen. Andere setzen Sie vielleicht selbst als Lehrende und Prüfende im Unterricht ein. Bevor wir auf die zentrale Frage eingehen, was man mit bestimmten Aufgabenformaten überprüfen kann, möchten wir Sie bitten, sich anhand einer alphabetisch geordneten Liste von Aufgabenformaten bewusst zu machen, wie viele Formate existieren und welche davon Sie schon kennen.

Aufgabe 27

Markieren Sie in der Tabelle die Aufgabenformate, die Sie entweder als Prüfling oder als Prüferin/Prüfer kennen.

Aufgabenformate	kenne ich als Prüfling	kenne ich als Prüferin/Prüfer
1. Auffinden von Fehlern in einem Text (Fehlersuche)	<input type="checkbox"/>	<input type="checkbox"/>
2. Beschreibung einer Bildgeschichte	<input type="checkbox"/>	<input type="checkbox"/>
3. Beurteilung der Angemessenheit einer sprachlichen Äußerung z.B. in Bezug auf ihre Höflichkeit	<input type="checkbox"/>	<input type="checkbox"/>
4. Cloze-Test (Lückenaufgabe, bei der in einem längeren Text z.B. jedes 5. oder jedes 7. Wort gelöscht wurde)	<input type="checkbox"/>	<input type="checkbox"/>
5. C-Test (Lückenaufgabe, bei der bei jedem 2. Wort die 2. Hälfte gelöscht wurde)	<input type="checkbox"/>	<input type="checkbox"/>
6. dichotome oder Zweifachauswahl (z.B. Ja/Nein, richtig/falsch)	<input type="checkbox"/>	<input type="checkbox"/>
7. Diktat	<input type="checkbox"/>	<input type="checkbox"/>
8. Fragen zum Textinhalt in Stichworten beantworten (Hörtext, Hör-Seh-Text oder Lesetext)	<input type="checkbox"/>	<input type="checkbox"/>
9. Informationstransfer (eine verbale Anweisung wie z.B. eine Wegbeschreibung grafisch wiedergeben)	<input type="checkbox"/>	<input type="checkbox"/>
10. Interview im Rahmen einer mündlichen Prüfung	<input type="checkbox"/>	<input type="checkbox"/>
11. Kommentar zu einem vorgegebenen Stimulus schreiben (z.B. Zitat, Grafik, Statistik)	<input type="checkbox"/>	<input type="checkbox"/>
12. Lückendiktat (Prüflinge füllen beim Hören eines Textes die vorgegebenen Lücken aus)	<input type="checkbox"/>	<input type="checkbox"/>

Aufgabenformate	kenne ich als Prüfling	kenne ich als Prüferin/Prüfer
13. Lückentext mit einer vorgegebenen Liste von Auswahlmöglichkeiten ergänzen	<input type="checkbox"/>	<input type="checkbox"/>
14. Multiple Choice (Auswahl aus zumeist drei bis vier Möglichkeiten, von denen meist nur eine richtig ist)	<input type="checkbox"/>	<input type="checkbox"/>
15. mündliche Reaktionen auf durch Tonträger gesteuerte vorgegebene Stimuli (z.B. auf eine Aufforderung reagieren wie beim Mündlichen Ausdruck im <i>TestDaF</i>)	<input type="checkbox"/>	<input type="checkbox"/>
16. Rollenspiel, Simulation einer Handlungssituation	<input type="checkbox"/>	<input type="checkbox"/>
17. Sprachmittlung (sinngemäßes Übertragen eines mündlichen oder schriftlichen Textes von der Ziel- in die Erstsprache oder in umgekehrter Richtung)	<input type="checkbox"/>	<input type="checkbox"/>
18. Textfragmente in eine sinnvolle Reihenfolge bringen	<input type="checkbox"/>	<input type="checkbox"/>
19. trichotome oder Dreifachauswahl (z.B. trifft zu / trifft nicht zu / kommt im Text nicht vor)	<input type="checkbox"/>	<input type="checkbox"/>
20. überflüssige Wörter oder Ausdrücke in einem Text finden	<input type="checkbox"/>	<input type="checkbox"/>
21. Übersetzung in das Deutsche oder aus dem Deutschen (Hin- bzw. Herübersetzung)	<input type="checkbox"/>	<input type="checkbox"/>
22. Wortschatztest wie Übersetzungsäquivalente, Synonyme, Antonyme, Zuordnungen zwischen Paraphrasen und Wörtern finden, Kollokationen vervollständigen	<input type="checkbox"/>	<input type="checkbox"/>
23. Zuordnungsaufgabe (z.B. Überschriften oder Titel kurzen Texten zuordnen)	<input type="checkbox"/>	<input type="checkbox"/>
24. Zusammenfassen eines Textes	<input type="checkbox"/>	<input type="checkbox"/>

3.7.1 Aufgabenformate und Testkonstrukt

Auch wenn Sie bestimmte Aufgabenformate schon selbst im Unterricht eingesetzt haben, haben Sie sich möglicherweise noch nicht intensiver damit auseinandergesetzt, welche Kompetenzen man mit dem jeweiligen Format genau erfassen kann.

Aufgabe 28



Welche Kompetenzen überprüfen Sie mit den Aufgabenformaten, die Sie regelmäßig einsetzen?

- Denken Sie an 2 bis 3 Aufgabenformate, die Sie oft einsetzen, und notieren Sie diese bitte in der mittleren Spalte der Tabelle. Sie können dafür auch noch einmal die Liste in Aufgabe 27 durchgehen.
- Überlegen und notieren Sie: Welche Kompetenzen möchten Sie überprüfen? Welche Kompetenzen überprüfen Sie tatsächlich?

Kompetenz, die ich prüfen möchte (intendiertes Testkonstrukt)	eingesetztes Aufgabenformat	Kompetenz/en, die ich tatsächlich mit dem Aufgabenformat prüfe
<i>orthografische Kompetenz</i>	<i>Diktat</i>	<i>orthografische Kompetenz, Hörverstehenskompetenz, lexikalische Kompetenz</i>

- Tauschen Sie sich mit Ihren Kolleginnen und Kollegen über Ihre Antworten aus.

	Spektrum	Korrektheit	Flüssigkeit	Interaktion	Kohärenz
B1	Verfügt über genügend sprachliche Mittel, um zurechtzukommen; der Wortschatz reicht aus, um sich, wenn auch manchmal zögernd und mit Hilfe von Umschreibungen, über Themen wie Familie, Hobbys und Interessen, Arbeit, Reisen und aktuelle Ereignisse äußern zu können.	Verwendet verhältnismäßig korrekt ein Repertoire gebräuchlicher Strukturen und Redeformeln, die mit eher vorhersehbaren Situationen zusammenhängen.	Kann sich ohne viel Stocken verständlich ausdrücken, obwohl er/sie deutliche Pausen macht, um die Äußerungen grammatisch und in der Wortwahl zu planen oder zu korrigieren, vor allem, wenn er/sie länger frei spricht.	Kann ein einfaches, direktes Gespräch über vertraute oder persönlich interessierende Themen beginnen, in Gang halten und beenden. Kann Teile von dem, was jemand gesagt hat, wiederholen, um das gegenseitige Verstehen zu sichern.	Kann eine Reihe kurzer, einfacher Einzelelemente zu einer zusammenhängenden, linearen Äußerung verknüpfen.
A2	Verwendet elementare Satzstrukturen mit memorierten Wendungen, kurzen Wortgruppen und Redeformeln, um damit in einfachen Alltagssituationen begrenzte Informationen auszutauschen.	Verwendet einige einfache Strukturen korrekt, macht aber noch systematisch elementare Fehler.	Kann sich in sehr kurzen Redebeiträgen verständlich machen, obwohl er/sie offensichtlich häufig stockt und neu ansetzen oder umformulieren muss.	Kann Fragen stellen und Fragen beantworten sowie auf einfache Feststellungen reagieren. Kann anzeigen, wann er/sie versteht, aber versteht kaum genug, um selbst das Gespräch in Gang zu halten.	Kann Wortgruppen durch einfache Konnektoren wie <i>und</i> , <i>aber</i> und <i>weil</i> verknüpfen.

(Europarat 2001, S. 38)

	Spektrum	Korrektheit	Flüssigkeit	Interaktion	Kohärenz
Leistung von Denise					
Leistung von Potiguar					

Wenn Sie das vollständige Bewertungsraster interessiert, können Sie es in Kapitel 3.3 des GER (Europarat 2001, S. 37f.) finden. Der entsprechende Link lautet:

<http://www.goethe.de/Z/50/commeuro/303.htm#sprechen>



ein weiteres Beispiel

Das vorherige Raster ist relativ grob und eignet sich vor allem zur Einstufung auf den Niveaustufen des GER. Es reicht allerdings nicht, wenn Sie in Ihrem Unterricht mündliche Leistungen bewerten und benoten wollen. Zu diesem Zweck geben wir Ihnen ein Bewertungsraster, das für den schulischen Kontext entwickelt wurde und sich auf die mündliche Sprachproduktion für die Stufe A2/A2+ bezieht. Das Raster stammt von der Website des Bildungsservers des ISB (Staatsinstitut für Schulqualität und Bildungsforschung, München) und ist auch abgedruckt und im Hinblick auf eine transparente Notengebung kommentiert in Grotjahn/Kleppin (2010, S.11). Analoge Raster für die Stufen A1/A1+, B1/B1+ und B2/B2+ finden Sie auf dem gleichen Server, wenn Sie in Ihre Suchmaschine folgende Wortkette eingeben: alp/multiplikatoren isb. Sie können Ideen daraus für Ihre eigene Bewertungspraxis nutzen – auch auf anderen GER-Stufen.

Natürlich müssen Sie nicht immer ein so komplexes Raster zugrunde legen. Sie können z.B. auch nur zwei Punkte der Skala bei jedem Kriterium beschreiben. Wenn Sie also z.B. pro Kriterium maximal 5 Punkte vergeben, dann beschreiben Sie, was ein Prüfling für 2 Punkte und was er für 4 Punkte leisten muss. Die Vergabe von 0, 1, 3 und 5 Punkten erfolgt dann relativ zu den gewählten Ankerpunktwerten.

Außerdem können Sie auch jedes Kriterium anders gewichten. Darüber hinaus können bei den **Deskriptoren** (Beschreibungen der Leistungen) auch feinere Abstufungen in Form von halben Punkten vergeben werden. Wie Sie genau vorgehen, hängt von Ihren Zielsetzungen und Ihrem spezifischen Unterrichts- und Prüfungskontext ab.

Zusammenfassung Die Vermittlung und das Training von Test-Taking-Strategien kann Ihren Lernenden dabei helfen, strategisch klug an Prüfungen heranzugehen. Außerdem können Sie damit die Gefahr verringern, dass Test-Taking-Strategien das Prüfungsergebnis in konstruktirrelevanter Weise beeinflussen.

5.9 Optimierung der eigenen Prüfungspraxis

Wenn Sie dieses Teilkapitel bearbeitet haben, dann wissen Sie,

- wie Sie Ihre Prüfungspraxis reflektieren können,
- in welchen Bereichen Sie etwas verändern könnten.

Abschließend wollen wir Ihnen Leitfragen zu Ihrer Prüfungspraxis auf den Weg geben, und zwar in Form einer Checkliste.

Fragen zu meiner Prüfungspraxis

		trifft zu
1.	Orientiere ich mich bei meinen Prüfungen an den relevanten Curricula und Vorgaben und nicht nur z. B. an den eingesetzten Materialien, die möglicherweise nicht immer oder nicht mehr den aktuellen Vorgaben entsprechen?	<input type="checkbox"/>
2.	Überprüfe ich möglichst breit die Kompetenzen, die ich auch in meinem Unterricht fördern möchte? Wenn also ein wesentliches Ziel meines Unterrichts die Entwicklung kommunikativer Kompetenz ist, setze ich z. B. zusätzlich zu schriftlichen Prüfungen auch vergleichsweise aufwändige mündliche Prüfungen ein?	<input type="checkbox"/>
3.	Nutze ich eine Vielzahl von möglichen Aufgabenformaten, um auch den individuellen Unterschieden meiner Lernenden gerecht zu werden?	<input type="checkbox"/>
4.	Nutze ich – so weit möglich – auch Aufgaben, die von den Prüflingen als echte lebensweltliche Aufgaben wahrgenommen werden (z. B. Aushandeln von für die Lernenden bedeutsamen Entscheidungen)?	<input type="checkbox"/>
5.	Sind meine Anforderungen in Prüfungen möglichst realistisch, damit meine Lernenden z. B. nicht demotiviert werden?	<input type="checkbox"/>
6.	Setze ich auch formative Verfahren ein, damit meine Lernenden ihren Lernprozess verfolgen können und ihre Lernfortschritte wahrnehmen können?	<input type="checkbox"/>
7.	Sind meine Rückmeldungen zu Prüfungsergebnissen so transparent und informativ, dass meine Lernenden auch im Hinblick auf ihr weiteres Lernen profitieren?	<input type="checkbox"/>
8.	Gebe ich in hinreichendem Maße positive, motivierende Rückmeldungen zu den Prüfungsergebnissen? Sage ich z. B., was gut oder auch besser als bei einer vorangehenden Prüfung ist?	<input type="checkbox"/>
9.	Berücksichtigen meine Bewertungskriterien auch (curriculare) Vorgaben wie Handlungsorientierung oder kommunikativen Erfolg?	<input type="checkbox"/>
10.	Kennen meine Lernenden die Bewertungskriterien, die ich bei Prüfungen zu produktiven Kompetenzen verwende (Transparenz)?	<input type="checkbox"/>
11.	Rege ich meine Lernenden auch zur Selbstevaluation an, damit sie in Zukunft besser ihr eigenes Lernen steuern können?	<input type="checkbox"/>
12.	Bespreche ich mit meinen Lernenden auch Prüfungen, Vorgehen in Prüfungen oder auch Ängste, damit Lernende in Zukunft mit Prüfungen besser umgehen können?	<input type="checkbox"/>