

# XIV. Leistungsmessung und Leistungskontrolle

## 142. Kompetenzmodelle und Bildungsstandards für Deutsch als Fremd- und Deutsch als Zweitsprache

1. Einleitung
2. Zu den Begriffen *Kompetenz* und *kommunikative Kompetenz*
3. Kompetenzmodelle
4. Bildungsstandards
5. Bilanz
6. Literatur in Auswahl

### 1. Einleitung

Bildungsstandards konkretisieren Bildungsziele, denen schulisches Lernen folgen soll, in Form von Kompetenzanforderungen. Von der sog. Kompetenzorientierung erwartet man eine Qualitätsverbesserung des schulischen Fremdsprachenunterrichts.

In diesem Beitrag wird zuerst kurz rekapituliert, was unter den Schlüsselbegriffen *Kompetenz* und *kommunikative Kompetenz* verstanden wird (Abschnitt 2). Gegenstand von Abschnitt 3 sind Modellierungen kommunikativer Kompetenz. Genauer diskutiert wird das Kompetenzmodell des *Europäischen Referenzrahmens*, weil davon der grösste Einfluss auf Bildungsstandards in den Fremdsprachen ausgeht. Abschnitt 4 stellt grundlegende Konzepte von Bildungsstandards vor, informiert über die einschlägigen Projekte in den deutschsprachigen Ländern und führt summarisch einige Risiken und Chancen von Standards an.

### 2. Zu den Begriffen *Kompetenz* und *kommunikative Kompetenz*

In der Diskussion um Bildungsstandards in den deutschsprachigen Ländern wurde der sonst sehr verschieden gebrauchte Begriff *Kompetenz* besonders durch die Expertise von Klieme et al. (2003) für das deutsche Bundesministerium für Bildung und Forschung (BMBF) und die deutsche Kultusministerkonferenz (KMK) geschärft und geprägt. Im Klieme-Gutachten wird Kompetenz (in Übereinstimmung mit Weinert 2001) definiert als

die bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. (Klieme et al. 2003: 72)

Kompetenz wird hier als *kognitive Problemlösungsfähigkeit* aufgefasst, die v.a. durch drei Merkmale näher bestimmt ist: Erstens werden Kompetenzen als *etwas Erlerntes und Erlernbares* konzipiert: „Kompetenz hat wohl eine materielle Basis in den Genen, aber sie entwickelt sich aufgrund von Lernprozessen (...) je verschieden“ (Oelkers und Reusser 2008: 26). Zweitens werden Kompetenzen auf wissensbasierte Fähigkeiten *in bestimmten Handlungsbereichen (Domänen) und Kontexten* bezogen (im Gegensatz zur klassischen Vorstellung von Intelligenz als kontextfreier kognitiver Disposition) (vgl. Oelkers und Reusser 2008: 24). Und drittens bedeutet Kompetenz eine Verbindung von *Wissen* und *Können*, die mit Einstellungen, Werten und Motiven interagiert.

Dieser Kompetenzbegriff steht dem soziolinguistisch begründeten von Dell Hymes (1972) viel näher als dem universalgrammatischen von Chomsky (1965). Chomskys Kompetenz ist auf die Beherrschung eines abstrakten grammatischen Regelsystems beschränkt, und dieses System wird klar vom effektiven Sprachgebrauch (Performanz) als etwas Zweitrangigem abgegrenzt. Hymes stellte diesem Kompetenzbegriff eine umfassend konzipierte *kommunikative Kompetenz* entgegen, die sowohl Wissen (*tacit knowledge*) als auch die Fähigkeit zum Sprachgebrauch (*ability to use*) umfasst, wobei auch diese Sprachgebrauchsfähigkeit auf der Ebene der Kompetenz verortet wird und also nicht mit effektivem Gebrauch gleichzusetzen ist. Hymes' *ability to use* umfasst zunächst, wie bei Chomsky, grammatisches Urteilsvermögen, geht dann aber weit darüber hinaus: Eingeschlossen wird die Fähigkeit, Sprache auch kontextspezifisch und sozial angemessen zu produzieren und zu rezipieren.

Einen noch umfassenderen Kompetenzbegriff als Dell Hymes setzt der *Gemeinsame Europäische Referenzrahmen für Sprachen* (Europarat 2001, in der Folge abgekürzt GER) an. Dieser Kompetenzbegriff wird hier zitiert und im nächsten Abschnitt genauer erläutert.

Kompetenzen sind die Summe des (deklarativen) Wissens, der (prozeduralen) Fertigkeiten und der persönlichkeitsbezogenen Kompetenzen und allgemeinen kognitiven Fähigkeiten, die es einem Menschen erlauben, Handlungen auszuführen.

(GER 2001: 21)

### 3. Kompetenzmodelle

Drei Modellierungen kommunikativer Kompetenz sind in der Fremdsprachendidaktik und im Bereich des *language testing* besonders einflussreich geworden: das Modell von Canale und Swain (1980; Canale 1983), dasjenige von Bachman und Palmer (1996; Bachman 1990) und das Modell des GER (2001). Das Modell von Canale und Swain und der GER sind deskriptive Modelle. Bachmanns Modell ist hingegen als Funktionsmodell intendiert, indem strategische Kompetenzen eine zentrale Stellung einnehmen. Aus Platzgründen wird hier nur der Ansatz des GER genauer besprochen. Zur vergleichenden Diskussion der drei Modelle siehe u. a. Lenz (2006), Harsch (2006), Schneider und North (2000) und die dort zitierte Literatur.

Mit dem GER (2001) liegt ein sehr umfassendes und differenziertes *Kategoriensystem zur Beschreibung von kommunikativen Kompetenzen* vor, das verschiedene Tendenzen und Entwicklungen aufnimmt, systematisiert und z.T. weiter ausdifferenziert. Sehr klar herausgearbeitet wird im GER die *Handlungsorientierung* von Kompetenz. Das ist eine

deutliche Parallele zu Klieme (vgl. oben, Abschnitt 2), wobei die Handlungsorientierung im GER mit dem Konzept kommunikativer Aufgaben (*tasks*) und deren Bewältigung verknüpft ist. Im GER werden Sprachenlernende in erster Linie als sozial Handelnde betrachtet, die „unter bestimmten Umständen und in spezifischen Umgebungen und Handlungsfeldern kommunikative Aufgaben bewältigen müssen“ (GER 2001: 21). Zur Bewältigung dieser Aufgaben greifen Lernende auf eine Vielzahl von Kompetenzen (!) zurück, darunter *allgemeine Kompetenzen*, die bei sprachlichen und nicht sprachlichen Handlungen eingesetzt werden (*savoir, savoir faire, savoir être* und *savoir apprendre*; vgl. GER 2001: 22–23 und Abschnitt 5.1.1–5.1.4) und *kommunikative Sprachkompetenzen*, welche zum Handeln mit Hilfe von spezifischen sprachlichen Mitteln befähigen (linguistische, soziolinguistische und pragmatische Kompetenzen; GER 2001: 24–25 und Abschnitt 5.2). Weiterhin gehören *sprachliche Aktivitäten* resp. kommunikative (Sprach-) Aktivitäten (die Terminologie ist hier schwankend) zum Ansatz des GER, worunter „die Ausübung der kommunikativen Sprachkompetenz eines Menschen in einem bestimmten Lebensbereich“ verstanden wird und zu denen rezeptive, produktive, interaktive und sprachmittelnde Aktivitäten sowie damit einhergehende Strategien gezählt werden (GER 2001: 21, 25 und Abschnitt 4.4).

Die Beschreibung der *kommunikativen Sprachkompetenzen* im GER ist Bachmans *Sprachwissen (language knowledge)* recht ähnlich, doch es zeigen sich auch Unterschiede, u. a. im Pragmatikverständnis der beiden Ansätze. Strategien des Sprachgebrauchs spielen im GER eine ebenso wichtige Rolle wie bei Bachman, doch werden sie im GER auch konkret und in der gleichen Weise ausformuliert und sogar skaliert, wie die kommunikativen Aktivitäten selbst. Beispielsweise werden im Anschluss an die Skalen zur mündlichen und schriftlichen Interaktion drei Skalen zu Interaktionsstrategien präsentiert: „Sprecherwechsel“, „Kooperieren“ und „um Klärung bitten“ (GER 2001: 88–89).

Die grosse Wirkung des GER ging und geht allerdings nicht von diesem umfassenden Beschreibungssystem für Kompetenzen aus, auch nicht von der durchgängigen Fragestruktur, mit der die Kompetenzen für die verschiedenen Benutzer des GER (z. B. Curriculumsverantwortliche) präsentiert werden, sondern von den illustrativen Beispiel-Skalen, d. h. von denjenigen Kompetenzbeschreibungen oder „Deskriptoren“ („Kann ...“), die in einem Schweizer Forschungsprojekt empirisch validiert und skaliert, d. h. einem von sechs Niveaus (A1 bis C2) zugeordnet werden konnten (vgl. Schneider und North 2000) und die aus diesem Projekt in den GER übernommen wurden. Skalierte Deskriptoren bietet der GER sowohl für „kommunikative Aktivitäten und Strategien“ als auch für „kommunikative Sprachkompetenzen“ an. Erstere setzen konsequent die Handlungsorientierung um (*was können Lernende auf einem Niveau typischerweise tun?*) und sind zu Einzelskalen wie z. B. „Gespräche zwischen Muttersprachlern verstehen“ zusammengestellt, die ihrerseits den klassischen vier Fertigkeiten (hier dem Hörverstehen) zugeordnet sind. Demgegenüber fokussieren die Deskriptoren zu den kommunikativen Sprachkompetenzen Merkmale der sprachlichen Qualität der Handlungen (*wie gut kann jemand auf einem Niveau sprachlich handeln?*) und sind zu Einzelskalen wie z. B. „Grammatische Korrektheit“ oder „soziolinguistische Angemessenheit“ zusammengestellt. – Zur Kritik am GER, die indessen kaum das Kompetenzmodell, sondern u. a. Aspekte der Skalen betrifft, vgl. Bausch et al. (2003) und dagegen Schneider (2003).

Festzuhalten ist, dass ein erheblicher Unterschied zwischen dem GER als umfassendem Beschreibungssystem für Kompetenzen und dem GER als Sammlung illustrativer Beispielskalen besteht und dass die starke Wirkung des GER zu einem grossen Teil auf

letzteren basiert – und/oder auf den (nie empirisch validierten) Konkretisierungen und Ergänzungen dieser Skalen in *Profile Deutsch* (Glaboniat et al. 2005), mit denen die GER-Skalen oft verwechselt oder in eins gesetzt werden.

Wie GER-Deskriptoren für spezielle Kontexte und Zielgruppen adaptiert und weiter entwickelt werden können, zeigen beispielsweise die folgenden Veröffentlichungen: für junge Lernende in der öffentlichen Schule – Lenz und Studer (2004); für das berufsorientierte Fremdsprachenlernen – Vogt (2007); für den Bereich Deutsch als Zweitsprache – das *Rahmencurriculum für Integrationskurse Deutsch als Zweitsprache* (Goethe-Institut 2008).

Das Rahmencurriculum illustriert sehr gut, wie das Konzept von Sprachenlernenden als sozial Handelnden für die heterogene Gruppe „MigrantInnen in Deutschland“ umgesetzt werden kann. Im Zuge einer umfassenderen Bedarfserhebung wurden Handlungsfelder bestimmt, die entweder für die gesamte Zielgruppe oder aber nur für *eine* Gruppe von MigrantInnen relevant sind (z. B. Umgang mit der Migrationssituation vs. Betreuung und Ausbildung von Kindern). Als Lernziele werden Sprachhandlungen präsentiert, die innerhalb dieser Handlungsfelder bedeutsam sind. Die Lernziele verstehen sich als Maximalangebot, aus denen für drei verschiedene Migrationsgruppen ausgewählt werden kann. Die Sprachhandlungen selbst, etwa „Auskunft geben“, werden a) durch Kann-Beschreibungen konkretisiert (z. B. „Kann einfach und kurz von seinen/ihren Erfahrungen berichten, z. B. über Unterstützung durch Familienangehörige in Deutschland.“), b) der jeweils vorrangigen Aktivität zugewiesen (hier dem Sprechen) und schliesslich c) immer auf demjenigen Niveau situiert, auf dem die Handlungen zuerst sinnvoll umgesetzt werden können (hier A2; insgesamt werden die GER-Niveaus A1, A2 und B1 berücksichtigt). – Zu bedenken ist dabei, dass die Deskriptoren des Rahmencurriculums eine *Achievement-Perspektive* spiegeln. Dies schränkt ihre direkte Verwendbarkeit für Tests, die eine Aussenperspektive an das Gelernte herantragen (*proficiency tests*), ein.

Vor diesem Hintergrund soll nun ein Blick auf Bildungsstandards für Fremdsprachen in den deutschsprachigen Ländern geworfen werden. Die Grundfrage von Bildungsstandards lautet: Über welche Kompetenzen müssen die SchülerInnen zu einem bestimmten Zeitpunkt in der Schullaufbahn verfügen, wenn wichtige Bildungsziele der Schule als erreicht gelten sollen?

#### 4. Bildungsstandards

In den angloamerikanischen und in wenigen europäischen Ländern (z. B. Grossbritannien, Niederlande) sind Standards in unterschiedlicher Form seit längerem feste Steuerungsgrössen im Bildungssystem. In den deutschsprachigen Ländern dagegen sind Bildungsstandards ein Kernelement aktueller Reformentwicklungen, für die sich v. a. drei Gründe bezeichnen lassen (vgl. u. a. Timm 2006): erstens die grossen internationalen Leistungsvergleichsstudien wie TIMSS und PISA, zweitens die Diskussion um Qualitätsmanagement, die zunehmend auch auf pädagogischem Gebiet geführt wird, und drittens die bereits 1991 vom Europarat beschlossene Entwicklung eines Europäischen Referenzrahmens (zum GER siehe oben, Abschnitt 3).

Vor dem Hintergrund v. a. der angloamerikanischen Erfahrungen und bezogen auf das Klieme-Gutachten (vgl. oben, Abschnitt 2) geht man in den deutschsprachigen Ländern davon aus, dass das Setzen von Zielen, welche die Schulen erreichen sollen, für das

Erreichen dieser Ziele effektiver ist als die Steuerung (allein) auf der Basis von Lehrplänen und Lehrmitteln. Formuliert werden daher Erwartungen an die *Lernergebnisse* der SchülerInnen (*performance standards*) und man nimmt an, dass Regulierungen auf der Seite der Ergebnisse (der SchülerInnen) bzw. des Outputs (der Schulen) einen wichtigen Beitrag zur Qualitätsentwicklung im Bildungswesen leisten können. Komplementär dazu soll den Schulen in verschiedenen Bereichen eine grössere Autonomie zugestanden werden, zum Beispiel auf unterrichtsmethodischer Ebene. Allerdings ist die gebräuchliche Formel „Regulierung des Outputs statt Inputsteuerung“ in doppelter Weise ungenau, denn einerseits wird, parallel zur Entwicklung von Standards, an neuen Lehrplänen und Lehrmitteln gearbeitet, also auf der Input-Seite (zum Verhältnis von Standards, Curricula und Lehrmitteln vgl. Bausch et al. 2005), und andererseits entspricht die Einführung von Bildungsstandards erst in Kombination mit externen *Monitoringverfahren*, im Zuge derer systematisch erfasste Lernergebnisse an die Standards rückgebunden bzw. mit diesen verglichen werden, einem outputorientierten Steuerungskonzept (siehe u. a. Labudde 2007). Und schliesslich ist auch der Begriff *Output* selbst zu präzisieren: Gemeint ist eine Ergebnisorientierung im Sinne einer Orientierung an *vergleichbaren Zielvorgaben*, die *als wesentlich erachtete Fachinhalte* explizieren. Das ist ein grundsätzlicher Unterschied zu Schulnoten: Zwar ist auch das traditionelle Notenschema ergebnisorientiert, jedoch sind Schulnoten auf Gruppennormen bezogen und ihnen fehlt der Bezug auf ein fachinhaltliches Lernzielkriterium, das über den Rahmen eines zufällig zusammengesetzten Klassenzimmers hinaus gültig und akzeptiert ist. Insofern geben Noten kaum Aufschluss über die in einem Fach oder einem Lerngebiet erreichten tatsächlichen Kompetenzen. Genau das sollen Standards in Verbindung mit Monitoringverfahren leisten: Sie sollen aufzeigen, was die SchülerInnen am Ende einer Lernperiode *tatsächlich wissen und können*, wobei dieses Wissen und Können auf Kernbereiche ausgewählter Fächer bezogen wird.

Insgesamt ergeben sich somit sehr hohe Ansprüche an die Entwicklung von Bildungsstandards: Bildungsstandards sollen fachliche Kerninhalte als vergleichbare Lehrziele vorgeben. Sie sollen auf Kompetenzmodellen basieren, die, Klieme et al. (2003: 74) folgend, sowohl die verschiedenen Anforderungen beschreiben, deren Bewältigung von den SchülerInnen erwartet wird (Kompetenzmodell als *Komponentenmodell*; primär eine didaktische Aufgabe), als auch Abstufungen dieser Anforderungen aufzeigen, auf denen die Schülerleistungen situiert werden können (Kompetenzmodell als *Stufenmodell*; eine komplexe empirisch-psychometrische Aufgabe, die über die Entwicklung von Testaufgaben zu den Kompetenzkomponenten läuft). Und schliesslich sollen Bildungsstandards niveaubezogene Vorgaben machen, was gleichzeitig kontroverse pädagogische und bildungspolitische Fragen aufwirft.

In Bezug auf diese hohen, wohl generell nur partiell erfüllbaren Ansprüche und bei der konkreten Ausgestaltung der Bildungsstandards sind die deutschsprachigen Länder bisher recht unterschiedliche Wege gegangen. Für Einzeldarstellungen vgl. Oelkers und Reusser (2008); aktuelle Tendenzen besprechen u. a. Harsch (2007) (Deutschland) sowie Lenz und Studer (2008) (Schweiz). Ein grosser Teil der Informationen über die Standards, einschliesslich der gesetzlichen Grundlagen und Beispielen für Testaufgaben, finden sich jetzt auch auf den Websites der für die Entwicklung der Standards zuständigen Institutionen, für Deutschland: <http://www.iqb.hu-berlin.de/bista>; für Österreich u. a.: <http://www.bifie.at/bildungsstandards>; für die Schweiz: <http://www.edk.ch/dyn/20692.php>.

Im Bereich der Fremdsprachen wurden bis jetzt Bildungsstandards für Englisch und Französisch sowie – in der Schweiz – auch für Deutsch entwickelt. Die Standards fo-

kussieren die Abschlussphase der obligatorischen Schule und betreffen die erste (in der Schweiz auch die zweite) schulische Fremdsprache. (Parallel dazu wurden auch Standards für die Schulsprache Deutsch entwickelt, die – und das wird noch zu wenig diskutiert – auch SchülerInnen betreffen, die Deutsch als zweite oder dritte Sprache lernen.)

Die derzeit vorliegenden Standards in Deutschland und in Österreich sind als *Regelstandards* konzipiert. Mit diesen Standards, deren empirische Abstützung erst begonnen hat, wird ein durchschnittliches Anforderungsniveau fokussiert (in Deutschland z. B. A2/B1 für die klassischen vier Fertigkeiten; ohne Niveaueingabe wurden von der KMK auch Standards für sprachliche Mittel sowie für interkulturelle und methodische Kompetenzen gesetzt). Im Gegensatz dazu hat man sich in der Schweiz für *Basisstandards* entschieden. Basisstandards zielen auf ein Mindestniveau, das (fast) alle SchülerInnen erreichen können sollten. Anders als in Deutschland und in Österreich wurden im Schweizer Projekt bereits in der Erarbeitungsphase der Standards repräsentative und kleinere empirische Untersuchungen durchgeführt, die zwei Ziele hatten: Sie dienten einerseits sowohl der Validierung von Aspekten des gewählten Kompetenzmodells (das sich an den GER anlehnt; Lenz und Studer 2008) als auch der Testaufgaben, durch die das Modell operationalisiert wurde, und andererseits der Feststellung der in den Schülerpopulationen tatsächlich vorhandenen Kompetenzen. Aufbauend darauf wurde in Zusammenarbeit mit Fachdidaktikern ein Expertenvorschlag für Standards ausgearbeitet, der in den Kategorien des Kompetenzmodells formuliert ist und der sich bezüglich der Anforderungsniveaus an den in den Untersuchungen festgestellten Leistungen orientiert. Beispielsweise bewegt sich der Vorschlag für einen Basisstandard in DaF am Ende der obligatorischen Schulzeit (7 Lernjahre in der Westschweiz) im Bereich des Niveaus A2, wobei beim Sprechen und bei den rezeptiven Fertigkeiten höhere Erwartungen angesetzt werden als beim Schreiben. Im Schweizer Standard-Projekt erfolgte auch eine ausführliche Auseinandersetzung mit interkulturellen und methodischen Kompetenzen, jedoch wurde bewusst darauf verzichtet, für diese (überfachlichen!) Kompetenzbereiche Bildungsstandards vorzuschlagen, die im Sinne von separaten *outcomes* und in der Art der Fertigkeiten überprüfbar sind.

Kritisch diskutiert werden Bildungsstandards für den Fremdsprachenunterricht im Sammelband von Bausch et al. 2005 (vgl. auch Timm 2006). Zu den hauptsächlich ange-mahnten Risiken der Standards gehören: die Reduktion von Lehrzielen und weiter des Lehrens und Lernens auf das, was sich in Sprachprüfungen erfassen lässt (verbunden oft mit dem *Teaching-to-the-test*-Argument), die Nivellierung von Leistungen (mit der Variante „Nivellierung nach unten“ bei Basisstandards), die Einzelfach-Orientierung, die Vermischung von Testfunktionen.

Zu einigen dieser Punkte gibt es durchaus gute Gegenargumente (vgl. u. a. Schneider 2007). So können Tests zu Bildungsstandards, die handlungsorientierte Aufgaben u. a. zum Hörverstehen und zum Sprechen umfassen, auch einen wünschbaren Reformeffekt auf grammatik- und wortschatzlastige Schul(übertritts)prüfungen haben und einen aufgabenorientierten Unterricht stärken. Andere Punkte dagegen sind nicht einfach von der Hand zu weisen, z. B. ist die ‚Arbeitsteilung‘ bisheriger Standard-Projekte mit Projekten für die Schulsprache hier und die Fremdsprachen dort nicht geeignet, die Entwicklung einer ‚eigentlichen‘ (d. h. integrativen) Mehrsprachigkeit zu fördern.

Auf der anderen Seite sind Bildungsstandards zweifellos auch mit Chancen verbunden. Z. B. können Standards zu mehr curricularer Kohärenz in den Schulen führen, weil der Bezug der Standards auf gestufte Kompetenzmodelle die schulstufenübergreifende

Curriculumsarbeit erleichtern sollte. Kaum kontrovers ist auch, dass Bildungsstandards durch die Fokussierung tatsächlicher Kompetenzen die Durchlässigkeit von sonst häufig impliziten Erwartungen verbessern können, wohingegen die Hoffnung, dass kompetenzorientierte Standards zu mehr Bildungsgerechtigkeit führen, verschieden eingeschätzt wird.

## 5. Bilanz

Für eine genauere Einschätzung der Projekte zu den Bildungsstandards in den deutschsprachigen Ländern ist es zweifellos noch zu früh. Feststellen lässt sich einstweilen die Tendenz, Bildungsstandards zu überschätzen. Ob die Standards zur intendierten Qualitätsverbesserung des Fremdsprachenunterrichts beitragen, hängt nicht nur von den Standards ab, sondern insbesondere auch davon, was auf Seiten der Aus- und Weiterbildung von Lehrpersonen, der Lerngelegenheiten und -bedingungen sowie der Lehrpläne und Lehrmittel getan wird.

## 6. Literatur in Auswahl

- Bachmann, Lyle F.  
1990 *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, Lyle F. und Adrian S. Palmer  
1996 *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bausch, Karl-Richard, Eva Burwitz-Melzer, Frank G. Königs und Hans-Jürgen Krumm (Hg.)  
2005 *Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand*. Tübingen: Narr.
- Bausch, Karl-Richard, Herbert Christ, Frank G. Königs und Hans-Jürgen Krumm (Hg.)  
2003 *Der Gemeinsame Europäische Referenzrahmen in der Diskussion*. Tübingen: Narr.
- Canale, Michael  
1983 On some dimensions of language proficiency. In: John W. Oller (Hg.), *Issues in Language Testing Research*, 333–342. Rowley, Mass.: Newbury House.
- Canale, Michael und Merrill Swain  
1980 Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1: 1–47.
- Chomsky, Noam  
1965 *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Europarat  
2001 *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Glaboniat, Manuela, Martin Müller, Paul Rusch, Helen Schmitz und Lukas Wertenschlag  
2005 *Profile Deutsch*. Berlin: Langenscheidt.
- Goethe-Institut und Bundesministerium des Innern  
2008 Rahmencurriculum für Integrationskurse Deutsch als Zweitsprache. München: Fell. Online: <http://www.integration-in-deutschland.de/> [Suche „Rahmencurriculum für Integrationskurse“] (18. 06. 2009).
- Harsch, Claudia  
2006 Der Gemeinsame europäische Referenzrahmen: Leistung und Grenzen. Die Bedeutung des Referenzrahmens im Kontext der Beurteilung von Sprachvermögen am Beispiel des

semikreativen Schreibens im DESI-Projekt. Online-Publikation: <http://opus.bibliothek.uni-augsburg.de/volltexte/2006/368/> (18. 06. 2009)

Harsch, Claudia

2007 Was können die fremdsprachlichen Bildungsstandards der KMK leisten? *Praxis Fremdspracheunterricht* 6: 2–11.

Hymes, Dell H.

1972 On Communicative Competence. In: Janet B. Pride und Janet Holmes (Hg.), *Sociolinguistics. Selected Readings*, 269–293. Harmondsworth: Penguin.

Klieme, Eckhard, Hermann Avenarius, Werner Blum, Peter Döbrich, Hans Gruber, Manfred Prenzel, Kristina Reiss, Kurt Riquarts, Jürgen Rost, Heinz-Elmar Tenorth und Helmut J. Vollmer

2003 *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn: BMBF.

Labudde, Peter (Hg.)

2007 *Bildungsstandards am Gymnasium – Korsett oder Katalysator?* Bern: h.e.p. verlag ag.

Lenz, Peter

2006 Überlegungen zur Sprachkompetenzbeschreibung und Testvalidierung im Projekt HarmoS/Fremdsprachen. *Bulletin Suisse de Linguistique appliquée* 84: 191–227.

Lenz, Peter und Thomas Studer

2004 Sprachkompetenzen von Jugendlichen einschätzbar machen. *Babylonia* 12: 21–25.

Lenz, Peter und Thomas Studer

2008 Zur Entwicklung der Basisstandards in den Fremdsprachenfächern. *Beiträge zur Lehrerbildung* 26: 361–371.

Oelkers, Jürgen und Kurt Reusser

2008 *Expertise: Qualität entwickeln – Standards sichern – mit Differenz umgehen*. Bonn: BMBF.

Schneider, Günther

2003 Der Europäische Referenzrahmen und die Mehrsprachigkeit. In: Gerhard Neuner und Ute Koithan (Hg.), *Tagungsbeiträge und Arbeitsberichte zum Thema „Mehrsprachigkeit im Bereich Deutsch als Fremdsprache“*, 87–108. Kassel: Kassel University Press.

Schneider, Günther

2007 Der Referenzrahmen und Bildungsstandards für Fremdsprachen. Feindbild, Vorbild, Wunschbild? *Babylonia* 15: 9–13.

Schneider, Günther und Brian North

2000 *Fremdsprachen können – was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Chur: Rüegger.

Timm, Johannes-Peter (Hg.)

2006 *Fremdsprachenlernen und Fremdsprachenforschung: Kompetenzen, Standards, Lernformen, Evaluation*. Tübingen: Narr.

Vogt, Karin

2007 Anpassung von Skalen und Deskriptoren des Gemeinsamen Europäischen Referenzrahmens. *Zeitschrift für Fremdsprachenforschung* 18: 43–66.

Weinert, Franz E.

2001 Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: Franz E. Weinert (Hg.), *Leistungsmessungen in Schulen*, 17–31. Weinheim: Beltz.

*Thomas Studer, Freiburg (Schweiz)*

## 143. Testen und Prüfen von Sprachkenntnissen

1. Einleitung
2. Testen und Prüfen im internationalen Vergleich
3. Bildungsstandards und internationale Leistungsstudien
4. Der Gemeinsame europäische Referenzrahmen für Sprachen und die Qualität von Sprachprüfungen
5. Funktionen des Prüfens und Testens
6. Computerbasierte Testverfahren
7. Literatur in Auswahl

### 1. Einleitung

Eine Prüfung, gesehen von einem unbekanntem Maler im 18. Jahrhundert: Ein junger Mann steht in der Mitte eines Raumes – das Haupt demutsvoll gesenkt, die Hände ungeschickt vom Körper gestreckt. Ihm gegenüber eine vielköpfige Jury aus perückengeschmückten Figuren. Mit einer Mischung aus Schadenfreude und Verachtung blicken die Herren Prüfer auf den armen Prüfling im Kreuzverhör. Klar wird hier, dass Prüfen kein wertfreier Vorgang ist, sondern mit Macht zu tun hat. Bevor die Geprüften einen würdigen Platz in der Gesellschaft zugewiesen bekommen, haben sie sich diesem schmerzhaften *rite de passage* zu unterziehen.

Nicht nur in der Vergangenheit hat das Thema Leistungsmessung negative Gefühle aller Schattierungen in uns hervorgerufen. Da Tests und Prüfungen in vielen Schulen zum Taktgeber des Unterrichts geworden sind, belasten sie auch heute das Klima des Lehrens und Lernens. Das hat damit zu tun, dass hier zwei Funktionen des Prüfens und Beurteilens im Widerstreit miteinander liegen: die Entwicklungsfunktion und die Steuerungs- bzw. Auswahlfunktion.

Während die Entwicklungsfunktion die Evaluation der Lernentwicklung zum Ziel hat, um daraus Informationen für weiteres Vorgehen zu gewinnen, somit also eine zutiefst pädagogische Aufgabe beinhaltet, geht es bei der Steuerungsfunktion darum, gesellschaftlich relevante Entscheidungen wie Versetzungen, Übergänge im Schulsystem oder den Eintritt in Berufswege und Studiengänge zu begründen. Wo begrenzte Kapazitäten zur Verfügung stehen – Stichwort Numerus-clausus-Fächer an den Hochschulen – mündet diese Steuerung in eine Auslese. Der negative Beigeschmack, den die Begriffe Leistungsmessung und -kontrolle heutzutage bei vielen fortschrittlichen Pädagogen haben, rührt daher, dass im Schulalltag die Steuerungsfunktion die Entwicklungsfunktion häufig überlagert. Sie erkennen allzu deutlich, dass Tests als unerwünschten Nebeneffekt eine Einengung der Lernziele mit sich bringen. Gelernt wird häufig nur noch der Prüfungsstoff. Im Erwachsenenunterricht dagegen ändert sich das Bild allmählich. Hier tritt der Aspekt der Freiwilligkeit stärker in den Vordergrund. Leistungskontrollen erfolgen im Rahmen des Schulunterrichts meist auf unfreiwilliger Basis. Tests und Prüfungen in der Erwachsenenbildung erfüllen dagegen zunehmend die Funktion einer Serviceleistung, für die vom Kunden ein expliziter Bedarf angemeldet wurde. Hinzu kommt noch ein weiterer wesentlicher Unterschied: Während Prüfungsinhalte und -verfahren im Schulsystem noch weitestgehend von nationalen Traditionen geprägt sind, orientiert man sich im Erwachse-

nenbildungsbereich zunehmend an transnationalen Standards. Wenn in jüngster Zeit in deutschen Gymnasien auch vereinzelt bereits die Prüfungen internationaler Anbieter von Zertifikaten angeboten werden, dann zeigt sich, dass diese Entwicklung von der Erwachsenenbildung in das Schulsystem hinüberleitet.

## 2. Testen und Prüfen im internationalen Vergleich

Die ersten Fremdsprachentests im modernen Sinne entstanden in den Vereinigten Staaten vor dem Ersten Weltkrieg (Spolsky 1995: 50–51). Ebenfalls in den USA begann die Tradition von Sprachtests zum Zweck der Reglementierung der Einwanderströme, die noch heute etwa in Australien lebendig ist.

Seit dem Beginn dieses Jahrtausends wurden in den Vereinigten Staaten immer wieder neue Tests entwickelt, bei denen höchster Wert auf empirisch belegbare Ergebnisse, Effizienz und kommerzielle Verwertbarkeit gelegt wurde. Das zur Universität von Princeton gehörige Institut *English Testing Service (ETS)* ist der Herausgeber des in den 1960er Jahren entwickelten und heutzutage teilnehmerstärksten Fremdsprachentests der Welt, des *TOEFL (Test of English as a Foreign Language)*. Ein weiterer Meilenstein in der Entwicklung von Fremdsprachenprüfungen wurde in Großbritannien gelegt. Das *University of Cambridge Examination Syndicate (UCLES)* widmet sich seit dem Jahr 1913 der Aufgabe, Prüfungen für Englisch als Fremdsprache zu entwickeln, um in den Schulen der Commonwealthländer ein einheitliches Bildungsniveau zu gewährleisten. Bereits im Jahre 1924 legte diese mit der Universität assoziierte Institution das zu jener Zeit mit einem anderen Namen versehene *Certificate of Proficiency in English (CPE)* vor, eine Prüfung, der ein umfassendes Sprachverständnis jenseits von Vokabellisten und Übersetzungen zugrunde liegt. Während in Europa der humanistisch-skeptizistische Ansatz mit der Bevorzugung von offenen Aufgaben zum Teil bis heute anhält, schlug sich der Einfluss der rationalistisch-empirischen Schule, wie sie in den USA zu Hause war, im Fach Deutsch in der Entwicklung des *Zertifikats Deutsch als Fremdsprache (ZDaF)* Ende der 1960er Jahre nieder. Nicht nur die ungleich größere Bedeutung des Englischen als Fremdsprache, sondern darüber hinaus die in England und den Vereinigten Staaten verankerte Tradition von zentral gestellten Prüfungen hat dafür gesorgt, dass angelsächsischen Institutionen auch heute noch eine Vorreiterrolle in der Entwicklung von Fremdsprachentests zukommt.

Das in den Vereinigten Staaten vorherrschende Misstrauen gegenüber intuitiv korrigierten Aufsätzen sowie die Skepsis hinsichtlich der Objektivität von Lehrkräften bei der Beurteilung der Leistungsfähigkeit ihrer eigenen Schüler teilen vor allem Lehrende und auch zahlreiche Fachleute im deutschsprachigen Raum nicht. In Deutschland herrschte bis vor kurzer Zeit vielfach sogar ein Vorbehalt gegenüber zentral verordneten Prüfungen. Deutsche Autoren werfen den englischsprachig orientierten Testkonstruktoren gern eine Überbewertung der vom Behaviorismus geprägten Formen der Leistungsmessung vor, und wenden dagegen ein, dass sie die ganzheitliche Lernerpersönlichkeit außer Acht lasse (Macht 1995: 283). Doch diese Ablehnung weicht allmählich der Erkenntnis, dass gemeinsame, objektive Maßstäbe notwendig sind, will man Leistungen vergleichen.

### 3. Bildungsstandards und internationale Leistungsstudien

Fragt man in den Ländern der Europäischen Union nach, welches sprachliche Niveau die Schulabgänger erreichen, fällt es Lehrkräften und Schulleitern häufig nicht leicht, dies in wenige Worte zu fassen. Deutsche Schüler, die nach ihrem Leistungsniveau in einer Fremdsprache gefragt werden, antworten zum Beispiel: „Ich habe eine Zwei im Abitur, Leistungskurs“. Für die Welt jenseits des eigenen nationalen Bildungssystems sind solche Angaben kaum aussagekräftig und damit inadäquat für die Mobilität in Europa und der Welt. Mit dem Voranschreiten der Globalisierung und Mobilität von Arbeitskräften nimmt der Bedarf an Transparenz zu. Nicht nur ein internationaler Vergleich, sondern schon der Vergleich der Bildungsabschlüsse verschiedener Bundesländer ist nicht eins zu eins möglich. Aus diesem Grund erarbeiten die Bundesländer so genannte Bildungsstandards für die verschiedenen Schulfächer. Bezogen auf die modernen Fremdsprachen beschreiben die Bildungsstandards den Grad des Sprachkönnens, den die Lernenden in verschiedenen Schularten und Klassenstufen erreichen sollen. Die Urheber dieser Idee internationaler Leistungsvergleiche bzw. eines einheitlichen Bezugsrahmens arbeiten für internationale Institutionen – die OECD und den Europarat. Die OECD setzt auf die normative Kraft eines standardisierten Tests, dem sich die Schülerinnen und Schüler aller teilnehmenden Länder unterziehen. Der Europarat setzt auf die normative Kraft eines gemeinsamen Bezugsrahmens.

Die OECD überträgt die Denkweise der Ökonomie auf den Bildungssektor. Schulleistungen haben den Stellenwert einer Ressource. Sie sind Voraussetzung für die wirtschaftliche Leistungsfähigkeit eines Landes. Mittels standardisierter Tests werden Leistungen von 15-jährigen Schülerinnen und Schülern gemessen. Die Studie vergleicht so unterschiedliche Bildungssysteme wie die Japans, Finnlands, Mexikos und Deutschlands vermittels ein und derselben Messlatte und stellt danach eine Rangordnung her.

Die beiden bisher durchgeführten PISA-Untersuchungen beschäftigten sich mit den Fächern Naturwissenschaften, Mathematik und Lesefähigkeit. Um Fremdsprachenkenntnisse ging es dagegen bereits bei *DESI (Deutsch Englisch Schülerleistungen International)*. Allerdings beschränkte sich diese Studie entgegen dem Namen mit den Schülerleistungen der 9. Klasse in allen Schularten auf Deutschland. Die Erhebung *Surveylang* im Auftrag der EU-Kommission soll dagegen im Jahr 2011 europaweit die Fremdsprachenkenntnisse von 15-Jährigen in Englisch, Deutsch, Französisch, Spanisch und Italienisch untersuchen.

### 4. Der Gemeinsame europäische Referenzrahmen für Sprachen und die Qualität von Sprachprüfungen

Die Strategie des Europarats ist offener angelegt als die internationalen Vergleichsstudien. Es geht um das Etablieren eines gemeinsamen Bezugssystems für alle europäischen Sprachen. Dazu dient die Publikation *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen (GER)*, der zunächst auf Englisch (2000) und 2001 in deutscher Sprache erschien. Der *Referenzrahmen* stellt eine Art Werkzeugkasten dar. Mit Hilfe von Skalen definiert er,

- was man unter Beherrschung einer Fremdsprache versteht,
- wie gut diese Beherrschung ist, d. h. welche Stufen der Sprachbeherrschung es gibt.

In dem Grundlagenwerk *Profile Deutsch* haben Experten aus den deutschsprachigen Ländern die für alle Sprachen zutreffenden Kann-Beschreibungen des *Referenzrahmens* für die deutsche Sprache interpretiert (Glaboniat et al. 2005). Außerdem enthält *Profile Deutsch* Auflistungen aller sprachlichen Mittel, die zur Erfüllung dieser Aktivitäten notwendig sind.

Dem Ziel der Transparenz hinsichtlich von Fremdsprachenkenntnissen will man durch ein kohärentes Beschreibungssystem näher kommen. Sprachliches Können definiert sich innerhalb eines sechsstufigen Systems – A1, A2, B1, B2, C1, C2. Die vier deutschsprachigen Testinstitutionen – Goethe-Institut/Zentrale München, Österreichisches Sprachdiplom/Wien, telc GmbH/Frankfurt und TestDaF-Institut/Hagen – haben ihre international angebotenen Zertifikate sofort nach Einführung des *Referenzrahmens* auf die Stufen A1 bis C2 hin orientiert. Gleichzeitig haben sie die Qualitätsansprüche, die an Prüfungen angelegt werden, erhöht. Im sogenannten *Manual* (2003) gibt ein vom Europarat eingesetztes Autorenteam dazu detaillierte Vorgaben für die methodischen Schritte der Zuordnung einer Prüfung zu einer Niveaustufe. Dabei geht es darum, eine Behauptung, z. B. „das ist eine Prüfung auf dem Sprachniveau B1“ durch qualitative und empirische Daten zu untermauern.

Der Europarat veröffentlicht zu den produktiven Fertigkeiten Schreiben und Sprechen inzwischen auch illustrative Beispiele von Kandidatenbeispielen zu den verschiedenen Niveaustufen in den verschiedenen Sprachen. Mit diesen prüfungsunabhängigen und damit neutralen Beispielen werden die Interpretationsspielräume eingegrenzt. Die erste Publikation von mündlichen Beispielen auf Deutsch zeigt erwachsene Lernende (Bolton u. a. 2008). Solche Beispiele ermöglichen ein Training von Lehrkräften. Außerdem lassen sich im Verfahren des sog. *Benchmarking* lokale Prüfungsleistungen auf den Referenzrahmen eichen.

Um Qualität in Bezug auf Validität und Reliabilität zu gewährleisten, wird zunehmend mehr in die Garantie der statistisch ermittelten Zuverlässigkeit der Ergebnisse investiert. Das heißt: Die in Prüfungen verwendeten Aufgaben werden vor dem Einsatz umfangreichen Erprobungen unterzogen, um ihre Qualität zu garantieren. Erst wenn sich in den Erprobungen gezeigt hat, dass die Aufgaben vom Schwierigkeitsgrad angemessen, unmissverständlich und trennscharf sind, finden sie Aufnahme in die endgültige Fassung der Prüfung.

Die Vorgaben des *Referenzrahmens* haben bei denjenigen Institutionen, die zum Teil ja schon seit Jahrzehnten Sprachzertifikate ausgeben, zu einer Neufassung der Urkunden geführt. Deren Aussagekraft hat sich zum einen durch Angaben zu den erzielten Ergebnissen in den verschiedenen Fertigkeiten – Stichwort: Profil sprachlichen Könnens – vereinheitlicht und verbessert. Als zentrale Aussage weist das Zeugnis aus, welche Kompetenzstufe in den vom Referenzrahmen definierten Fertigkeiten der/die Geprüfte erreicht hat. Damit gelingt es, EU-weit zukünftig Prüfungsergebnisse zu vergleichen.

Die breite Rezeption des Referenzrahmens hat dazu geführt, dass bereits Anfänger sich ihr erreichtes Niveau gern durch ein Zertifikat bestätigen lassen. Während traditionell eine „offizielle“ bzw. „umfassende“ Sprachprüfung am Ende des Sprachlernprozesses vorgesehen war, setzt sich damit das Konzept vom Erreichen der einzelnen Stufen durch. Ähnlich wie bei den asiatischen Kampfsportarten arbeitet sich der Lernende im Verlauf

des unter Umständen lebenslangen Lernens von Gürtel zu Gürtel hoch. Dabei ist jeder Gürtel ein klar definiertes Ziel. Auf einmal zum schwarzen Gürtel zu gelangen, wäre ein demotivierend langer Weg.

## 5. Funktionen des Prüfens und Testens

Leistungsmessung in Form von Tests und Prüfungen ist zunächst neutral zu verstehen, als Sammlung von Informationen über den Kenntnisstand der geprüften Personen. Diese Informationssammlung wird mit einem bestimmten Ziel vor Augen vorgenommen. Ein solches Ziel kann darin bestehen, Entscheidungen über das Lehr- bzw. Kursprogramm zu begründen. Ein anderes Ziel wäre, eine Voraussage über die sprachliche Leistungsfähigkeit einer Person in der realen Welt zu treffen. Je nachdem, wer die Leistungsmessung auswertet und verwendet, unterscheidet man Selbstevaluation und Evaluation durch andere. Je nachdem, zu welchem Zweck im letzteren Fall die gesammelte Information verwendet werden soll, unterscheidet man als wichtigste Funktionen Ein- und Weiterstufung, Eignung und Zulassung, Lernerfolgskontrolle, Abschlussprüfung und Sprachstandsmessung.

Diese unterschiedlichen Ziele der Leistungsmessung bedingen die jeweils gewählten Formen, die in Prüfungen bzw. Tests eine Rolle spielen. Die abwechselnde Verwendung der Begriffe *Test* und *Prüfung* bedeutet keine strukturelle Differenzierung. Vielmehr werden beide Begriffe im Deutschen mittlerweile häufig synonym gebraucht. Format, Auswahl der Fertigkeiten, Prüfungsdauer sowie Aufgabentypen lassen sich erst entscheiden, wenn klar ist, zu welchem Zweck ein Test verwendet wird. Das bedeutet im Umkehrschluss, dass nicht jeder Aufgabentyp bzw. jeder Testaufbau für jede Funktion gleich gut geeignet ist.

Im Folgenden sollen die oben genannten Prüfungsarten kurz charakterisiert und diskutiert werden. Fünf Fragen helfen, die Prüfungsarten voneinander abzugrenzen:

1. Worauf bezieht sich der Inhalt der Prüfung (z. B. auf die Inhalte des vorausgegangenen Unterrichts)?
2. Für wen sind die Ergebnisse bestimmt (z. B. für die geprüfte Person, die Lehrkräfte bzw. für außen stehende Entscheidungsträger)?
3. Welche Entscheidungen werden auf der Basis der Ergebnisse getroffen (z. B. Zulassung zum Studium an einer Hochschule)?
4. Wie aufwendig ist das Verfahren im Hinblick auf die Faktoren Zeit, Arbeitsmittel etc.?
5. Wie aufwendig ist das Verfahren im Hinblick auf die vorgesehenen Korrekturverfahren?

Die Fragen 1, 2 und 3 lassen sich unter das Kriterium der Validität von Tests bzw. Prüfungen subsumieren. Prüfungen und Tests nennen wir valide, wenn sie angemessene, sinnvolle und nützliche Schlussfolgerungen zu den vorher definierten Zielen und Intentionen zulassen. Bei einer Prüfung, deren Ziel das Feststellen von sprachlichen Fähigkeiten auf einem bestimmten Sprachniveau ist, sollte dieses Niveau durch ein externes Bezugssystem wie den GER klar definiert sein. Darüber hinaus sollten andere Kenntnisse und Fähigkeiten, wie z. B. Intelligenz, Weltwissen oder Konzentrationsfähigkeit, nicht die ausschlaggebenden Faktoren sein. Ein Kursabschlussstest ist nur dann valide, wenn

er mit den Zielen, die im Lehrplan formuliert sind und dem Unterricht zugrunde liegen, übereinstimmen.

Mit der Frage 1 ist ein weiteres Gütekriterium verbunden, das die Qualität von Tests und Prüfungen bestimmt, nämlich die Forderung, dass die gewählten Testverfahren die Formen und Inhalte des vorausgegangenen Unterrichts angemessen widerspiegeln und u. U. sogar positive Rückwirkungen auf den Unterricht haben. In der englischsprachigen Literatur ist dieser Aspekt mit dem Terminus *back-wash-* bzw. *wash-back*-Effekt eingeführt (Wall und Alderson 1993). So ist es gelungen, den wenig kommunikativen Fremdsprachenunterricht durch die Einführung des Referenzrahmens und (mündlicher) Prüfungen zu modernisieren. Der Weg einer Lehrplanreform ist vergleichsweise langwieriger. Freilich gibt es auch eine Kehrseite der Medaille. Die formulierten Ziele, eine bestimmte Niveaustufe zu erreichen, sollen – ja müssen – in immer knapperer Zeit, mit weniger Lehrenden und weniger Geld erreicht werden. Denn bei der fortwährenden Reform von schulischen Lehrplänen treten neue Fächer, wie z. B. Computereinsatz in Konkurrenz zu den traditionellen Fächern, wozu die zweiten Fremdsprachen zählen. Dort, wo Deutsch als zweite Fremdsprache – wenn auch oft mit reduziertem Stundendeputat – erhalten geblieben ist, stehen die Lehrkräfte unter dem Druck, ihren Unterricht zu optimieren. Prüfungen erhöhen den Effizienzdruck auf Lehrkräfte und Schüler, Lern- und Prüfungsstrategien stehen zunehmend im Mittelpunkt des Unterrichts. Prüfungsvorbereitung bedeutet damit Einüben von Strategien. Die Fähigkeit zum Aktivieren von Vorwissen hilft beim Entschlüsseln von Lese- oder Hörtexten oft mehr als die Kenntnis der einen oder anderen Vokabel, der Einsatz von Ratetechniken bei geschlossenen Auswahlantworten als letzte Rettung ist hinlänglich als *test wiseness* bekannt.

Die Fragen 4 und 5 berühren die Reliabilität und die Praktikabilität. Das Kriterium der Reliabilität beinhaltet die Forderung nach der Zuverlässigkeit der Leistungsmessung. Ein reliabler Test misst sprachliche Fähigkeit zuverlässig und genau. Der – in jedem Test – enthaltene Messfehler hält sich in engen, akzeptablen Grenzen. Die Zuverlässigkeit von Testergebnissen hängt in hohem Maße von der eingesetzten Bewertungsmethode ab. Wird das Testergebnis durch das subjektive Urteil eines oder mehrerer Prüfer ermittelt, ist Zuverlässigkeit ein besonders wichtiges Thema. Wünschenswert ist eine möglichst hohe Übereinstimmung zwischen verschiedenen Bewertern sowie eine möglichst hohe Stabilität der Ergebnisse eines Bewerbers zu verschiedenen Zeitpunkten.

Idealerweise sollte ein Test wie ein Metermaß funktionieren: So wie dieses auch bei wiederholter Messung eines Gegenstandes immer die gleichen Maße anzeigt, so sollte ein Test für eine bestimmte Leistung immer die gleiche Punktzahl oder Note ergeben. In der Praxis der Testkonstruktion berührt das Kriterium der Reliabilität Fragen wie die Anzahl von Aufgaben, d. h. wie viele einzelne Messungen muss man durchführen, um zuverlässige Ergebnisse zu erhalten, oder die Frage der Korrektur- und Bewertungsverfahren, d. h. bei welchen Verfahren können sich Messfehler einschleichen, die das Ergebnis verfälschen, und dergleichen.

Das Kriterium der Praktikabilität bezieht sich auf den Bedarf an Zeit, Raum und personellen Ressourcen, wobei sowohl an die Korrekturzeiten als auch an die Qualifizierung der Korrigierenden bzw. Prüfenden zu denken ist. Leistungskontrolle sollte möglichst zeitökonomisch sein. Erfahrungswerte aus der Praxis zeigen, dass 90 Minuten ohne Pause für die geprüften Personen das Maximum an Belastung darstellen. Geht die gesamte Testzeit darüber hinaus, wie etwa bei breit angelegten Feststellungsprüfungen, sollte die Dauer für einzelne Testteile dieses Maß nicht überschreiten. Nach jedem Testteil gibt es dann Pausen.

## 5.1. Ein- und Weiterstufung

Einstufungstests helfen bei der Einteilung von Lernenden in Gruppen unterschiedlichen Niveaus und unterschiedlicher Vorkenntnisse. Mit Hilfe dieser sehr weit verbreiteten Prüfungsart werden die sprachlichen Vorkenntnisse der Einzustufenden ermittelt. Bei Weiterstufungstests, wie sie in Institutionen mit einem ausdifferenzierten Kursangebot üblich sind, soll nach dem Besuch eines Kurses festgestellt werden, welchen weiteren Lernweg im Kursangebot die Teilnehmenden nehmen sollen.

Die besondere Bedingung der Ein- und Weiterstufungstests liegt in der Forderung nach strenger zeitlicher Begrenzung: Ein mehrstündiger Einstufungstest wäre u. U. eine Überforderung für die Teilnehmenden und würde dadurch möglicherweise die Ergebnisse sogar verfälschen, wäre also nicht zuverlässig. Denn nicht selten müssen sich die Einzustufenden aus logistischen Gründen unmittelbar bei der Anreise an den Kursort und damit in manchen Fällen unmittelbar nach langen Flugreisen dem Test unterziehen.

Mit Rücksicht auf die zeitliche Beschränkung sind Ein- und Weiterstufungstests in der Praxis häufig als Stichprobenkontrollen angelegt. Testkonstrukteure orientieren sich beim Erstellen eines Ein- bzw. Weiterstufungstests an Kurs- bzw. Lehrplänen für die entsprechenden Stufen. Aufgaben zum Lesen, zu Wortschatz und Grammatik und kurze Schreibaufgaben kommen am häufigsten zum Einsatz. Seltener enthalten Einstufungstests Teile zum Hörverstehen, was vor allem technische Gründe hat. Die mündliche Kommunikationsfähigkeit ermittelt eine Lehr- bzw. Prüfungskraft in der Regel in einem authentischen Gespräch zur Feststellung der Personalien, persönlicher Interessen und Lernziele.

### 5.1.1. Testverfahren zur Einstufung

Da die Testergebnisse innerhalb kurzer Zeit verfügbar sein müssen, wird bei der Wahl der Aufgabentypen weitgehend auf korrekturfremdliche Verfahren zurückgegriffen. Deshalb bestehen Einstufungstest meist aus geschlossenen Aufgaben. Im Gegensatz zu offenen Aufgabentypen ist die sprachliche Reaktion bei geschlossenen Aufgaben nicht frei ausgeführt. Vielmehr beschränkt sich die Aktivität der Geprüften bei geschlossenen Aufgaben auf das Auffinden, Ankreuzen, Ordnen, Zuordnen oder Hervorheben der richtigen Lösungen. Die Bewertung solcher Aufgaben ist zeitsparend und praktisch unabhängig vom subjektiven Urteil des Korrektors. Geschlossene Aufgaben können mit Hilfe von Schablonen oder Computerprogrammen schnell ausgewertet werden. Sie eignen sich zur Überprüfung der rezeptiven Fertigkeiten, denn sie überprüfen lediglich das Erkennen der richtigen Lösung.

Eine klassische geschlossene Aufgabe ist die Multiple-Choice-Aufgabe. Sie besteht in der Regel aus einem einleitenden Satz bzw. einer einleitenden Frage und mehreren (häufig vier, gelegentlich aber auch nur drei, seltener mehr als vier) Auswahlmöglichkeiten. Eine der Auswahlantworten ist die richtige Lösung, alle anderen dienen als sogenannte Distraktoren und sind falsch. Um zu unterstreichen, dass die Anordnung von richtigen und falschen Auswahlantworten dem Zufallsprinzip unterliegt, wird meist eine alphabetische Reihenfolge gewählt.

Von ihren Befürwortern sind Multiple-Choice-Aufgaben wegen der Möglichkeit geschätzt, zu realistischen Voraussagen über den Schwierigkeitsgrad und die Trennschärfe

der einzelnen Aufgaben zu gelangen. Diesen Aufgabentyp setzen sowohl ETS und UCLES als auch das niederländische Centraal Instituut voor Toetsontwikkeling (CITO) in Fremdsprachentests regelmäßig ein. Auf der anderen Seite steht die Kritik an der Validität der Multiple-Choice-Aufgabe. Zum einen wird beanstandet, dass die Aufgaben in der Praxis häufig schlecht konstruiert, d. h. unnötig kompliziert sind und die Distraktoren bewusst auf die falsche Fährte lenken. Dadurch werden den Geprüften Fehler geradezu untergeschoben. Verfechter von offenen Aufgabentypen kritisieren überdies, dass mit Multiple-Choice nur das Erkennen der richtigen Lösung überprüft und somit eine relativ geringe Leistung verlangt wird. Geübte nutzen zudem den Faktor Ratewahrscheinlichkeit aus, d. h. die Möglichkeit, auch durch zufälliges Ankreuzen noch einen gewissen Prozentsatz an richtigen Lösungen zu erzielen. Bei einem Test, der genügend Aufgaben umfasst und eine Bestehensgrenze von über 50 Prozent setzt, ist die Erfolgswahrscheinlichkeit eines Teilnehmenden durch Raten statistisch gesehen jedoch sehr gering.

Bei Einstufungstests taucht neben der Multiple-Choice-Aufgabe eine Vielzahl von weiteren geschlossenen Aufgabentypen auf. Bei dem folgenden Beispiel handelt es sich um textunabhängige Einzelaufgaben, in denen es um das Erkennen der richtigen Struktur bzw. der geeigneten Ausdrucksweise geht: Herstellen der richtigen Wortfolge

*der Fehler / er / gehabt / hätte / mehr / nicht / passiert / wäre / Zeit*

Lösung: Hätte er mehr Zeit gehabt, wäre der Fehler nicht passiert.

Unbefriedigend an dieser Art von kontextunabhängigen Aufgaben ist, dass jeglicher kommunikative Rahmen fehlt. Sie wirken genauso trocken und wirklichkeitsfern wie die Lückentexte der strukturalistischen Ära. In der strukturalistischen Testtheorie ging man davon aus, dass die Beherrschung der Fremdsprache sich nicht direkt, d. h. in realen Verwendungssituationen überprüfen lässt, sondern nur auf dem Weg über die Messung isolierter Elemente der Sprache. Die Summe dieser isolierten Elemente erlaube Rückschlüsse auf die Fähigkeit des Lernenden, Sprache auch in realen Situationen außerhalb des Unterrichts verwenden zu können. Im Zuge der sogenannten „kommunikativen Wende“ in der Fremdsprachdidaktik wurde deutlich, dass Wortschatz- und Strukturen-tests nur einen Aspekt der Kommunikationsfähigkeit in der Fremdsprache erfassen (Canale und Swain 1980; Bachman und Palmer 1996). Solche Aufgaben überprüfen nur die Wissenskomponente der Sprache, d. h. die Fähigkeit, Regeln der Grammatik und Vokabeln richtig einzusetzen. Nicht überprüft werden dagegen die diskursive und die soziokulturelle Kompetenz. Um zum Beispiel verschieden strukturierte Lesetexte zu bearbeiten oder um die Informationen eines gehörten Gesprächs zu notieren, benötigen die Geprüften auch Wissen darüber, wie Texte aufgebaut sind.

Ein Verfahren, das die lexikalische und grammatische Kompetenz mit Hilfe von kontextualisierten Aufgaben erfassen will und dazu kurze Lesetexte auf zeitökonomische Weise einbezieht, ist der Cloze-Test. Das Verfahren wurde in den 1950er Jahren zunächst als Instrument zur Bestimmung der Schwierigkeit von Texten entwickelt (Taylor 1953). Daraus entstand in der Folge ein Verfahren zum Überprüfen des Leseverstehens. Es geht von der Annahme aus, dass die lexikalische und grammatische Kompetenz ein Indikator für die Sprachbeherrschung ist und über das Lesen zu erfassen ist. Ein klassischer Cloze-Test besteht aus einem längeren Text, bei dem in festgelegten regelmäßigen Abständen Wörter gelöscht sind (z. B. jedes fünfte Wort). Die Lücke ist zu ergänzen. Im Gegensatz zum Lückentext herkömmlicher Art, in dem die Lücken nach didaktischen Gesichts-

punkten gelöscht werden, um ganz gezielt bestimmte Themen der Grammatik bzw. bestimmte Wortschatzbereiche zu überprüfen, handelt es sich beim Cloze-Test um eine streng mechanische Tilgung, wobei die Tilgungsfrequenz im ganzen Text beibehalten wird. Im folgenden Beispiel wurde jedes achte Wort gelöscht (Abb. 143.1). Der erste Satz, der den Kontext vorgibt, bleibt bei einem Cloze-Test immer intakt.

*Immer wieder hatte der Richter Gerhard Sch. am Starnberger Amtsgericht mit Fahrraddieben zu tun gehabt - bis er selbst zum Opfer wurde: Direkt vor dem Gerichtsgebäude machten sich freche \_\_\_\_\_ mit seinem roten Fahrrad aus dem Staub. \_\_\_\_\_ legte sich der Richter persönlich auf die \_\_\_\_\_. Weil er nicht damit rechnete, sein Fahrrad \_\_\_\_\_ wieder zu sehen, hatte Gerhard Sch. von \_\_\_\_\_ Anzeige abgesehen.*

Lösungen: (Fahrrad-)Diebe, Daraufhin (Danach), Lauer, jemals (nochmals)

Abb. 143.1: Cloze-Test

In diesem Testverfahren wird die Tatsache genutzt, dass Informationen in einem Text durch mehrere Signale realisiert werden, die sich wechselseitig ergänzen. Eine Kenntnis von den Aufbaukriterien eines Textes ist zum Lösen dieser Aufgabe also unabdingbar. Das mechanische Tilgungsprinzip sorgt beim klassischen Cloze-Verfahren dafür, dass ein Querschnitt von sprachlichen Phänomenen getestet wird. Alle Wortarten sind in den Lücken gelöscht und nicht nur solche, mit denen Lernende des Deutschen besondere Schwierigkeiten haben.

Gleichzeitig treten die Grenzen seiner Verwendbarkeit im Rahmen eines Einstufungstests deutlich hervor. Definiert man Klassen- bzw. Kursstufen nach der Beherrschung bestimmter grammatischer Strukturen, muss ein lehrzielvalider Test genau diese Strukturen testen. Das Zufallsprinzip des Cloze-Tests ist hier eher kontraproduktiv. Forschungsergebnisse zeigen überdies, dass bei sogenannten „natural cloze“-Tests, bei denen weder auf die Auswahl der Texte noch der Lücken Rücksicht genommen wird, Vorbehalte bezüglich der Validität angebracht sind (Brown 1993). Deshalb wurde als Alternative zum klassischen Cloze-Verfahren der modifizierte Cloze-Test entwickelt, bei denen die Lücken mit Bedacht gewählt werden. Damit können lexikalische oder grammatische Aspekte gezielt abgefragt werden. Flexibel ist das Cloze-Verfahren im Hinblick auf den Schwierigkeitsgrad. Je nachdem, ob eine leichtere oder schwierigere Version gewünscht wird, können Auswahlantworten als Schüttelkasten oder Multiple-Choice-Auswahl vorgegeben bzw. keine Vorgaben für die Antworten gemacht werden.

Ein weiteres Testverfahren, das mit ganzen Texten und kontextualisierten Aufgaben arbeitet, ist der sogenannte C-Test. Das Verfahren wurde in den 1980er Jahren entwickelt. Es zielt auf das Überprüfen der allen Fertigkeiten zugrundeliegenden sprachsystematischen Kompetenz ab. Lesekompetenz steht als Mittlerfertigkeit im Mittelpunkt, eine hohe Kompetenz im Verständnis von Textkohärenz spielt in einem C-Test eine wichtige Rolle. Genauso wie das Cloze-Verfahren beruht es auf dem Konzept der reduzierten Redundanz. Beim C-Test ist jeweils die Hälfte von jedem zweiten Wort – vom Wortende ausgehend – gelöscht. Hier ein Beispiel von der Homepage des TestDaF-Instituts (2008). „Fit für TestDaF“ soll eine erste grobe Einschätzung darüber liefern, ob Interessenten schon über eine ausreichende Sprachkompetenz verfügen, um bei der Prüfung TestDaF erfolgreich abzuschneiden (Abb. 143.2).

Die Europäische Union gewinnt immer mehr an Bedeutung. Schon ein  Blick in  die  
 Zeit  beweist es : Die Einigung  Europas ist  seit den  90er  
 Jahren  von ein  Randthema zu  einem wichtig  und zugleich  heftig  
 umstritten  Gegenstand politischer  Auseinandersetzungen geworden . Nunmehr  
 hat  sich die  EU von  einem wirtschaftlich  orientierten Zusammenschluss

zu einer handlungsfähigen Vereinigung mit einem gewählten Parlament entwickelt.

Lösungen: in / n / tung / s / gung / st / en / ren / nem / u / tigen / eich / ttenen / schen / rden /  
 at / ie / on / aftlich /chluss

Abb. 143.2: C-Test

Die Vorteile des C-Testverfahrens liegen in der Testökonomie – mit geringem Aufwand erhält man eine relativ große Menge an Daten. Zudem sind C-Tests vergleichsweise leicht zu entwickeln und zu bewerten. Die wissenschaftliche Diskussion merkt allerdings kritisch an, dass es dem C-Test an Augenscheinvalidität mangelt. Bei einer Versuchsreihe mit Englischlernenden in Mexiko hat sich zum Beispiel gezeigt, dass mit dem Verfahren nicht vertraute Prüfungsteilnehmende erhebliche Schwierigkeiten damit haben können und nicht einsehen, dass ihre sprachlichen Fähigkeiten darin adäquat zum Ausdruck kommen (Jafarpur 1995: 194; Grotjahn 1992).

## 5.2. Eignung und Zulassung

Von Einstufungstests zu trennen sind Eignungstests. Diese wurden entwickelt, um die Erfolgsaussichten einer Person auf einem bestimmten Lerngebiet zu prognostizieren. In den Vereinigten Staaten gibt es seit den 1930er Jahren Versuche, Tests zu entwickeln, um damit Entscheidungen über Zulassung oder Ausschluss von Lernenden beim Fremdsprachenunterricht an High Schools zu rechtfertigen (Michel 1936: 275–287). Im deutschsprachigen Raum ist der TestAS als neueres Beispiel für Eignungstests zu nennen. Dieser zentrale, standardisierte Test prognostiziert die Studierfähigkeit ausländischer Studienplatzbewerber für die Studienfächer, also z. B. geisteswissenschaftliche Fächer, Ingenieurwissenschaften, Mathematik und naturwissenschaftliche Fächer. Der Test kann sowohl in deutscher als auch in englischer Sprache abgelegt werden. Die deutschen bzw. englischen Sprachkenntnisse werden in Form eines Screening mit einem C-Test geprüft, um festzustellen, inwieweit fehlende Sprachkenntnisse die Leistung im Studierfähigkeitstest beeinträchtigt haben. Die Testergebnisse sollen es Hochschulen ermöglichen, Bewerberinnen und Bewerber mit verschiedenartigen Schulabschlüssen miteinander zu vergleichen, sie in eine Rangreihenfolge zu bringen und die Studierenden zu ihren Studienangeboten auszuwählen.

Nur um Zulassung geht es bei den Aufnahmeprüfungen, die ausländische Studienplatzbewerber an deutschen Hochschulen absolvieren müssen. Denn nach dem Hochschulrahmengesetz haben Studienbewerber mit nichtdeutscher Muttersprache den Nachweis hinreichender Sprachkenntnisse zu erbringen. Dabei sollen sie zeigen, dass sie genügend Sprachkenntnisse mitbringen, um an dem Vorlesungs- und Seminarprogramm teilnehmen zu können. Diese Aufnahmeprüfungen sind für die Geprüften mit einem ho-

hen Risiko verbunden, da aufgrund des Prüfungsergebnisses die folgenreiche Entscheidung über ihre Zulassung getroffen wird.

Sprachliche Aufnahmeprüfungen für ausländische Studienplatzbewerber bieten das TestDaF-Institut sowie eine Reihe von Studienkollegs an Hochschulen an. Der TestDaF sowie die *Deutsche Sprachprüfung für den Hochschulzugang ausländischer Studienbewerber (DSH)* sind eigens für diesen Zweck entwickelte Prüfungen, wobei erstere zentral erstellt und ausgewertet wird. Die DSH ist eine dezentral erstellte und ausgewertete Prüfung. Unter Vorgabe einer gemeinsamen Rahmenordnung, in der die Prüfungsinhalte umrissen werden, kann jede Hochschule eine eigene Aufnahmeprüfung erstellen. Um eine gegenseitige Anerkennung zwischen den lokalen, in ihren Prüfungsaufgaben also abweichenden DSH-Prüfungen unter den Hochschulen zu erreichen, registriert inzwischen die HRK (Hochschulrektorenkonferenz) in Zusammenarbeit mit dem FaDaF (Fachverband Deutsch als Fremdsprache) alle DSH-Prüfungen.

### 5.3. Selbstevaluation

Im Kontext des *selbstgesteuerten Lernens* haben in jüngster Zeit auch die Begriffe Leistungsmessung und -bewertung eine neue Konnotation erhalten. Die Lernenden sollen in die Lage versetzt werden, ihre Leistungen selber einzuschätzen. Damit verlagert sich der Schwerpunkt von der Lehrkraft hin zu den Lernenden, die selber Informationen über ihren Kenntnisstand und Lernfortschritt sammeln, um weitere Lernschritte gezielt zu steuern. Die Beteiligung des Lernenden an den Verfahren der Leistungsmessung fördert bei ihm die Einsicht in das Lernziel und die nötigen Fähigkeiten. Wichtigster erster Schritt beim selbstgesteuerten Lernen ist die Selbstdiagnose, d. h. die Evaluation und Analyse eigener Stärken und Schwächen. Daran schließt sich idealerweise ein gezielteres d. h. strategisch richtiges Lernen an, bei dem systematisch Lücken gefüllt werden. Das gezielte Training von Lerntechniken, z. B. zur Verbesserung der Schreibfertigkeit durch bewusste Korrekturgänge (Rampillon 1996: 38–39) kann dabei zur Verbesserung der Resultate eingesetzt werden.

Selbstevaluation als Analyseinstrument ist weniger eine Frage der Aufgabentypen als vielmehr eine Frage der Organisation. Die Lernenden erhalten nicht nur die Aufgaben, sondern die dazugehörigen Antworten bzw. Lösungen sowie Erläuterungen in Form einer Lernnavigation.

Dem Gedanken der Selbsteinschätzung sowie dem europäischen Gedanken verpflichtet ist DIALANG, ein frei im Internet verfügbares computerbasiertes Evaluationsinstrument, das es Fremdsprachenlernenden in 15 Sprachen ermöglicht, sich selbst einzuschätzen. DIALANG wurde im Auftrag der Europäischen Kommission entwickelt. Grundlage dieses Systems ist eine Datenbank von kalibrierten Aufgaben. Die Testergebnisse sind nach jedem Testmodul (Lesen, Hören, Schreiben, Strukturen, Wortschatz) in Form der Stufen A1 bis C2 ausgedrückt. Zum Einsatz kommt DIALANG beispielsweise an Universitäten zur Einstufung von Studienplatzbewerbern.

### 5.4. Lernfortschritt

Versteht man die Instrumente der Leistungsmessung als ein System, bei dem es von einfachen zu immer komplexeren Prüfungstypen geht, dann lässt sich der Lernfort-

schrittstest am Anfang der Skala einordnen. Es handelt sich dabei um ein Kontrollinstrument, das an geeigneter Stelle während eines Kurses eingesetzt wird, um der Lehrkraft Informationen darüber zu liefern, wie effektiv ihr Unterricht war. Den Geprüften bietet er Informationen darüber, wie effektiv der individuelle Lernprozess war. Der Inhalt der Tests knüpft in der Regel unmittelbar an den in der vorangegangenen Unterrichtsphase bearbeiteten Stoff an, ist somit abhängig vom Kurs- bzw. Lehrplan. Wenn das Lehrwerk solche Tests nicht bereits liefert, erstellt die Lehrkraft diese passend zu den aktuellen Unterrichtsinhalten. Hier stellt sich die Frage nach dem Unterschied zwischen Lehrbuchaufgabe und Testaufgabe. Im Gegensatz zu Übungsaufgaben im Unterricht unterliegen Testaufgaben höheren Ansprüchen bei den Gütekriterien Unabhängigkeit und Eindeutigkeit. Bei einem Test sollten die Geprüften bei jedem Item eine neue Chance erhalten – zur Differenzierung zwischen einer Aufgabe und einem Einzelelement eines Tests hat sich im Deutschen inzwischen die englische Bezeichnung *Item* durchgesetzt. Eine falsch gelöste Aufgabe darf also nicht automatisch einen weiteren Fehler nach sich ziehen. Lehrkräfte greifen häufig zu dem Instrument der offenen Aufgabe vom Typ *Fragen zum (Lese- bzw. Hör-)Text*, da diese zu jedwedem Inhalt einfach zu erstellen sind:

*Welche Rolle spielt es, dass alle Angebote im Jugendzentrum kostenlos sind?*

Die Lösungen zu einer solchen Frage variieren sowohl inhaltlich als auch formal. Bei der Beurteilung bereiten Lösungen, die zu knapp sind oder zu viele grammatische bzw. orthographische Fehler enthalten, Probleme. Das Prüfungsziel Rezeption eines Textes (egal ob Lese- oder Hörtext) gerät in Konflikt mit dem impliziten Prüfungsziel Produktion von frei formulierten schriftlichen Äußerungen. Zwar sollten orthographische und grammatische Fehler bei Aufgaben zur Rezeption keine Rolle spielen, doch fällt es Korrektoren meistens schwer, Lösungen mit der vollen Punktzahl zu bewerten, die zwar inhaltlich adäquat aber formal fehlerhaft sind.

Halboffene Aufgabentypen haben gegenüber diesen völlig offenen Aufgaben den Vorteil, dass die verlangte produktive Leistung innerhalb eng gesteckter Grenzen bleibt. Ein halboffenes Verfahren zum Überprüfen des Leseverstehens ist zum Beispiel die mit Lü-

#### **Der sprachgewandte Vetter**

*Zwei Jahre Arbeit und kaum ein Ergebnis: Matata begreift es einfach nicht. Die Zwergschimpansin sollte "Yerkish" lernen, eine Kunstsprache mit 256 abstrakten Symbolen für Substantive und Verben. Nicht einmal sieben Zeichen konnte sie sich merken. Sue Savage war enttäuscht. Die Wissenschaftlerin an einem Sprachforschungszentrum in Atlanta/USA hatte geglaubt, dass Bonobo-Zwergschimpansen sprachlich mehr auf dem Kasten hätten als gewöhnliche Schimpansen. (...)*

Der Artikel berichtet von einem Experiment, das (0) in Atlanta durchgeführt wird. Eine Wissenschaftlerin versucht dort, Affen (1) beizubringen. Sie verwendet eine Sprache namens (2), die aus Symbolen besteht.

(0) am Sprachforschungsinstitut

(1) \_\_\_\_\_

(2) \_\_\_\_\_

Lösungen: (1) Das Sprechen, (2) Yerkish, Goethe-Institut, Zentrale Mittelstufenprüfung, Modellsatz (1996)

Abb. 143.3: „summary cloze“-Aufgabe

cken versehene Textzusammenfassung, die sogenannte „summary cloze“-Aufgabe. Dabei muss die Textzusammenfassung nicht komplett erstellt werden, was eine Transferleistung vom Lesen zum Schreiben wäre, sondern die Lehrkraft bzw. der Testautor verfasst die mit Lücken versehene Textzusammenfassung, in der die Kerninformationen des Textes abgefragt werden. Die vorgegebenen Lücken reißen den Kontext auf und schränken die Bandbreite möglicher Lösungen inhaltlich ein.

Bei Aufgaben dieses Typs geht es um das Überprüfen der Rezeption. Die Lücken sind daher so gesetzt, dass sie nicht bereits von grammatischen Strukturen determiniert sind (z. B. ein Nomen im Dativ-Plural erforderlich machen). Bei der Korrektur müssen formale Fehler unberücksichtigt bleiben.

## 5.5. Kursabschluss

Am Ende eines Kurses besteht bei einer großen Zahl von erwachsenen Kursteilnehmenden das Bedürfnis nach einer Dokumentation des Fortschritts, den sie im Verlauf des Kurses gemacht haben. Bei fremdfinanzierten Kursen hat der Geldgeber häufig ein Interesse daran, den Erfolg der geförderten Maßnahme in Form eines Abschlusstests zu überprüfen. Ein Beispiel dafür ist die Einführung der weitestgehend durch Steuergelder finanzierten *Integrationskurse* für Zuwanderinnen und Zuwanderer in Deutschland im Jahr 2005. Diese Sprachkurse sind differenziert nach Teilzielgruppen, z. B. schnelle Lernende, Elternkurse, Jugendkurse, Kurse zur Alphabetisierung. Sie umfassen in der Regel 600 Unterrichtseinheiten mit der Möglichkeit einer Verlängerung. Alle Kurse wurden bis 2009 mit dem *Zertifikat Deutsch* abgeschlossen, seit 2009 kommt ein speziell für diesen Zweck entwickelter *Deutschtest für Zuwanderer (dtz)* zum Einsatz.

Intendiert ist eine enge Verbindung mit dem Lernstoff des vorausgegangenen Kurses. Das Bindeglied für die Anbindung der Sprachprüfung an den Kurs ist das Curriculum. Das sog. *Rahmencurriculum* (Goethe-Institut 2007) beschreibt repräsentative Lernziele für die Sprachkurse. Diese Lernziele sind zugleich Prüfungsziele. Geprüft wird ein Kernbereich, der in allen Kursarten Unterrichtsgegenstand und für alle Teilzielgruppen relevant ist. Das sind zum Beispiel Themen wie die Ausbildung und Betreuung von Kindern. Eine Prüfungszielbeschreibung legt eine Definition des sprachlichen Niveaus aufgefächert nach sprachlichen Mitteln im Bereich Sprachhandlungen, Intentionen, Notionen, grammatischen Strukturen und Wortschatz vor und entspricht damit den Anforderungen, die auch an eine Feststellungsprüfung gestellt werden. Rahmencurriculum und Prüfungszielbeschreibung bilden die Grundlage für Unterrichtsmaterialien und Lehrwerke, die in den Kursen eingesetzt werden. Die Kursabschlussprüfung *dtz* unterscheidet sich von Feststellungsprüfungen wie dem weltweit seit dem Jahr 2000 für alle Zielgruppen eingesetzten *Zertifikat Deutsch* durch den engen inhaltlichen Bezug zu den vorausgegangenen Kursen und zu den Bedürfnissen der Zielgruppe.

## 5.6. Sprachstandsmessung

Die Funktion einer Feststellungsprüfung ist es, die sprachlichen Fähigkeiten der Geprüften zu einem bestimmten Zeitpunkt zu bestimmen. In der Regel melden sich Interessenten zu einer solchen Prüfung aus freien Stücken an und bezahlen für diese Dienstleistung

eine Gebühr. Als Gründe, warum sie sich einer Prüfung unterziehen, nennen Lernende in der Reihenfolge ihrer Wichtigkeit:

- a) aus persönlichem Interesse
- b) als Nachweis beruflicher Qualifikation
- c) als Nachweis im Rahmen der Studienplatzbewerbung
- d) als Nachweis im Rahmen des Einreise- und Visarechtes

Was die Prüfungsinhalte angeht, so beschränkt sich die Feststellungsprüfung nicht auf den Stoff eines bestimmten Kurses oder Lehrplans, sondern legt eine detaillierte Prüfungszielbeschreibung zugrunde. Bei einer allgemeinsprachlichen Prüfung orientiert sich diese an der zukünftigen Sprachverwendung im privaten, beruflichen und öffentlichen Leben. Im Falle einer fachsprachlich ausgerichteten Prüfung sind die Prüfungsinhalte auf das jeweilige Fachgebiet, zum Beispiel berufsbezogene Verwendungssituationen, eingegrenzt. Die Prüfungszielbeschreibung gibt Auskunft über Prüfungsziele und -inhalte, sprachliche Funktionen, Intentionen bzw. Handlungsfelder, den zugrunde gelegten Wortschatz und die grammatischen Strukturen, das sprachliche Niveau etc. Seit Veröffentlichung des *Gemeinsamen europäischen Referenzrahmens* und der *Profile Deutsch* beziehen sich solche Prüfungsbeschreibungen für Deutsch als Fremd- und Zweitsprache in der Regel auf die Niveaustufen A1 bis C2 (Grotjahn 2007).

Feststellungsprüfungen umfassen in der Regel Testaufgaben zu allen vier Fertigkeiten, d. h. Leseverstehen, Hörverstehen, Schriftlicher Ausdruck und Mündlicher Ausdruck. Prüfungsaufgaben zu der Fertigkeit der Mediation sowie spezielle Grammatik-Teile sind dagegen selten. Da die Prüfungsergebnisse sich an Endabnehmer, also zum Beispiel Arbeitgeber oder Bildungseinrichtungen, außerhalb der prüfenden Institution richten, entstehen besondere Ansprüche an die *Augenschein*-Validität. Beim Durchsehen der Prüfungsaufgaben sollte der gebildete Laie den Eindruck erhalten, dass Testaufgaben für ihre Zwecke relevante Inhalte und Fertigkeiten überprüfen. Im Zuge der sogenannten kommunikativen Wende spielt dabei die Verwendung (semi-)authentischer Texte und handlungsorientierter Aufgabenformen eine entscheidende Rolle.

Bei der Formulierung von Testaufgaben zum Schreiben etwa sollte die kontextuelle Einbettung realistisch sein. Vor allem muss der Adressat klar sein. Realitätsnahe Schreibaufgaben richten sich nicht an den Prüfenden/Korrigierenden, sondern zum Beispiel an einen Brieffreund. Überdies muss in der Aufgabe definiert sein, welche Form der zu schreibende Text annehmen soll, d. h. welche Textsorte produziert werden und wie lang der Text sein soll. Aus Gründen der Fairness sollte den Geprüften bei einer produktiven Aufgabe, die nach bestimmten Kriterien beurteilt wird, außerdem bekannt sein, worauf bei der Korrektur Wert gelegt wird.

Bei der Feststellungsprüfung ist außer der Augenscheinvalidität die sogenannte „Kontentvalidität“ von Bedeutung. Sie betrifft die Frage der Testziele. Während die Auffassung, dass Feststellungsprüfungen alle vier Fertigkeiten überprüfen sollen, inzwischen historisch gewachsen ist, gibt es keinen Konsens darüber, wie jede einzelne Fertigkeit getestet wird. Dies hängt von den jeweiligen Feinzielen ab. So legt zum Beispiel das Hörverstehen im TestDaF Wert auf das Verstehen von vorlesungsähnlich strukturierten Texten. Eine Prüfung zur Fachsprache Wirtschaftsdeutsch wie das vom Deutschen Volkshochschul-Verband und vom Goethe-Institut gemeinsam entwickelte *Zertifikat Deutsch für den Beruf* (ZDfB) (1995) räumt berufsspezifischen Sprechanlässen und Aspekten inter- bzw. metakultureller Kommunikation eine große Bedeutung ein.

In Feststellungsprüfungen spielen mündliche Prüfungsteile meist eine wesentliche Rolle. Entscheidend für das Testformat ist die Zahl der gleichzeitig Geprüften. In jüngerer Zeit greifen Testentwickler dabei häufiger auf das Format der Paar- oder sogar Gruppenprüfung zurück. Aufgaben zur Problemlösung und Diskussion lassen sich unter gleichberechtigten Teilnehmenden besser bearbeiten als zwischen Teilnehmenden und Prüfenden. Teilnehmende sollen miteinander eigene Gedanken entwickeln und artikulieren. Doch das gleichzeitige Prüfen von mehreren Personen bringt besondere Probleme mit sich. Intellektuell überlegene und reife Persönlichkeiten haben es bei solchen Testformen leichter als Personen, die aus kulturspezifischen oder persönlichen Gründen nicht daran gewöhnt sind, eigene Ideen im Gespräch zu entwickeln. Dieser Unterschied muss sich in den dazugehörigen Bewertungskriterien niederschlagen. Aufgabe und Bewertung gehören also zusammen. Je offener ein Test zur mündlichen Kommunikation angelegt ist, umso größer ist der Handlungs- und Entscheidungsspielraum der Prüfenden. In der Praxis nutzen Prüfende diesen Spielraum allerdings nicht selten in einer Weise, dass von den ursprünglichen Prüfungszielen wenig übrigbleibt. Dadurch werden die Prüfungsergebnisse unzuverlässig. Prüferverhalten sowie die Praktikabilität von Bewertungskriterien sind die beiden zentralen Facetten des komplexen Prüfungsgeschehens bei mündlichen Prüfungen. Sie berühren sowohl die Frage der Validität wie der Reliabilität von Tests. Je freier ein Prüfender die Aufgabenstellung interpretiert oder je unübersichtlicher die Bewertungskriterien, die er heranziehen soll, umso größer die Fehlerquelle. Der Relevanz von Prüferverhalten und der Bewertung von mündlichen Leistungen ist eine Vielzahl von Einzeluntersuchungen gewidmet (Lumley und McNamara 1995).

Auch hinsichtlich der Reliabilität werden an Feststellungsprüfungen die höchsten Anforderungen gestellt. Zuverlässige Tests dieses Typs müssen zunächst eine Mindestzahl an Items anbieten, wobei jedes Item als Einzelmessung zu betrachten ist, das unabhängig von anderen wertvolle Daten liefert. Um gesicherte Aussagen über den Sprachstand machen zu können, sollte beim Lesen und Hören eine Mindestzahl von Aufgaben vorhanden sein. Diese wird gegenwärtig zwischen 20 und 30 gesehen. Eine Konsequenz daraus ist zum Beispiel, dass die von europäischen Prüfungsinstitutionen angebotenen Feststellungsprüfungen in der Regel mehrere Stunden dauern. Die Resultate dieser Sprachstandsmessung sollen differenziert und fundiert genug sein, um den Endabnehmern eine begründete Entscheidung zu ermöglichen, ob die nachgewiesenen Sprachkenntnisse für die anvisierte Tätigkeit ausreichen.

## 6. Computerbasierte Testverfahren

Die Vorgabe, in möglichst kurzer Zeit ein möglichst genaues Bild des sprachlichen Kenntnisstandes der Teilnehmenden zu erhalten, hat in den 1990er Jahren zur Entwicklung von computergestützten Testverfahren geführt. Ein Beispiel dafür ist das sogenannte *BUSINESS LANGUAGE TESTING SYSTEM (BULATS)*. Das Instrument wird in vier Sprachen angeboten: Englisch, Deutsch, Französisch und Spanisch. Ein Prüfender ist bei diesem Computertest nicht beteiligt. Die Ergebnisse werden unmittelbar nach dem etwa 15-minütigen Verfahren in Form eines Protokolls bekanntgegeben. Im Hintergrund dieses Testsystems steht eine Bank von kalibrierten, d. h. erprobten und damit im Schwierigkeitsgrad definierten Aufgaben zu den rezeptiven Fertigkeiten, d. h. Lesen und Hören sowie Strukturen und Wortschatz.

Im Gegensatz zum traditionellen Verfahren in Papierform wird beim computeradaptiven Testsystem nicht jedem Teilnehmenden die gleiche Testbatterie vorgelegt. Der Computer errechnet vielmehr nach jedem gelösten Item, zu welchem Item die geprüfte Person weitergeschickt wird. Hat er das vorhergehende richtig gelöst, wird der Schwierigkeitsgrad erhöht, hat er es falsch gelöst, bekommt er ein Angebot mit geringerem Schwierigkeitsgrad. Vorteile eines computergesteuerten Systems sind neben dem geringen Zeitaufwand die Genauigkeit der Ergebnisse, der Abwechslungsreichtum der kontextualisierten Aufgaben, die auch das Hörverstehen integrieren, sowie die weite Streuung der Aufgaben vom Anfängerniveau bis zu weit fortgeschrittenem Kenntnisstand.

## 7. Literatur in Auswahl

- Bachman, Lyle und Adrian Palmer  
 1996 *Language Testing in Practice*. Oxford: Oxford University Press.
- Bolton, Sibylle, Manuela Glaboniat, Helga Lorenz, Michaela Perlmann-Balme und Stefanie Steiner  
 2008 *Mündlich. Mündliche Produktion und Interaktion Deutsch. Illustration der Niveaustufen des Gemeinsamen europäischen Referenzrahmens*. München: Langenscheidt.
- Brown, James Dean  
 1993 What are the characteristics of natural cloze tests. *Language Testing* 10(2): 93–116.
- Canale, Michael und Merrill Swain  
 1980 Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1: 1–47.
- Council of Europe  
 2003, Sept. *Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg. [www.coe.int/lang](http://www.coe.int/lang).
- Deutscher Volkshochschul-Verband, Goethe-Institut  
 1995 *Das Zertifikat Deutsch für den Beruf. Lernziele, Wortliste, Testmodell, Bewertungskriterien*. Frankfurt a. M.
- Europarat; Council for Cultural Co-operation, Education Committee, Modern Languages Division; Goethe-Institut Inter Nationes u. a. (Hg.)  
 2001 *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin/München: Langenscheidt.
- Glaboniat, Manuela, Martin Müller, Paul Rusch, Helen Schmitz und Lukas Wertenschlag  
 2005 *Profile Deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen, Kannbeschreibungen, Kommunikative Mittel, Niveau A1–A2, B1–B2, C1–C2*. Berlin/München: Langenscheidt.
- Goethe-Institut  
 2007 *Rahmencurriculum für Integrationskurse Deutsch als Zweitsprache*. München. [www.goethe.de/integration](http://www.goethe.de/integration).
- Grotjahn, Rüdiger (Hg.)  
 1992 *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Grotjahn, Rüdiger  
 2007 Testen und Prüfen: Aktuelle Tendenzen. *Neue Beiträge zur Germanistik* 6(2): 19–38.
- Jafarpur, Abdoljavad  
 1995 Is C-Testing superior to cloze? *Language Testing* 12(2): 194–216.
- Lumley, Tom und Timothy McNamara  
 1995 Rater characteristics and rater bias: implications for training. *Language Testing* 12(1): 54–71.

- Macht, Konrad  
 1995 Leistungsmessung und Curriculum. In: Karl-Richard Bausch, Herbert Christ und Hans-Jürgen Krumm (Hg.), *Handbuch Fremdsprachenunterricht*, 282–285. 3. Aufl. Tübingen: Narr.
- Michel, Sister Virgil  
 1936 Prognosis in German. *Modern Language Journal* 20: 275–87.
- Rampillon, Ute  
 1996 Schüler beurteilen sich selbst. Ein Zugang zum selbstgesteuerten Lernen. *Friedrich Jahresheft* 14: 38–39.
- Spolsky, Bernard  
 1995 *Measured Words. The Development of Objective Language Testing*. Oxford: Oxford University Press.
- Taylor, Wilson L.  
 1953 Cloze procedure: a new tool for measuring readability. *Journalism Quaterley* 30: 414–38.
- Wall, Diane und J. Charles Alderson  
 1993 Examining wash-back: the Sri Lanka Impact Study. *Language Testing* 10(1): 41–69.

*Michaela Perlmann-Balme, München (Deutschland)*

## 144. Sprachprüfungen für Deutsch als Fremdsprache

1. Einleitung
2. Testarten
3. Merkmale standardisierter Zertifikatsprüfungen
4. Internationale DaF-Prüfungen
5. Zusammenfassung
6. Literatur in Auswahl

### 1. Einleitung

In einer Zeit der Globalisierung und eines zusammenwachsenden Europas spielen nicht nur Fremdsprachenkenntnisse eine immer größere Rolle, sondern zunehmend auch deren Zertifizierung. Die internationale Mobilität in Berufs- und Bildungskontexten erfordert objektive, aussagekräftige, transparente und vergleichbare Nachweise von Sprachkompetenzen. Entsprechend der erhöhten Nachfrage ist in den letzten Jahren auch der internationale Prüfungsmarkt in Bewegung geraten. Ein deutlicher Meilenstein in Richtung erhöhter Vergleichbarkeit und Transparenz wurde dabei durch das Erscheinen des Gemeinsamen Europäischen Referenzrahmens (= GER, Europarat 2001) gesetzt. Beinahe alle größeren internationalen Zertifikatsanbieter und Testorganisationen haben ihre Sprachprüfungen gemäß den darin festgelegten Referenzniveaus analysiert, adaptiert bzw. in Anlehnung daran neue entwickelt. Dies gilt ganz besonders auch für den deutschsprachigen Prüfungsbereich und betrifft nicht nur allgemeinsprachliche Tests und Prüfungen, sondern auch Prüfungen für besondere Zielgruppen, wie z. B. Prüfungen für Kinder und

Jugendliche, studien- oder berufsbezogene Prüfungen oder spezifische Prüfungen für MigrantInnen (vgl. Art. 145).

Im folgenden Beitrag soll nach einer kurzen Definition verschiedener Test- und Prüfungsverfahren geklärt werden, welche Art von Prüfungen und Zertifikaten im Zentrum dieses Beitrags stehen und durch welche Merkmale sie sich charakterisieren lassen. Anschließend stellt eine Übersicht die aktuelle DaF-Testlandschaft dar, gefolgt von einer kurzen Beschreibung der diversen Prüfungen. In diesem Zusammenhang wird auch auf die Rolle des GER für den Bereich der Sprachzertifizierung eingegangen.

## 2. Testarten

Das Überprüfen von Sprachkompetenzen kann je nach Testziel und Funktion in unterschiedlichster Art und Weise vorgenommen werden. In der Testtheorie unterscheidet man prinzipiell zwischen *diagnostischen* Verfahren, wie z. B. den *Einstufungstests* (*placement tests*), den auf einen Unterricht bzw. Kurs bezogenen *Leistungs- bzw. Lernfortschrittsmessungen* (*achievement tests*) und den an vorgegebenen Kompetenz- bzw. Niveaubeschreibungen orientierten *Feststellungsprüfungen*, die auch als *Niveauprüfungen* bzw. *Qualifikationstests* (*proficiency tests*) bezeichnet werden. Letztgenannte Prüfungen erfassen den Ausprägungsgrad von vorab definierten Anforderungen sprachlicher Handlungsfähigkeit in Bezug auf abschätzbare bzw. mögliche zukünftige Verwendungssituationen. Sie sind als solche, wie später noch ausgeführt wird, meist performanz- bzw. sprachhandlungsorientiert und überprüfen unabhängig von Kursen oder Lehrwerken den momentanen Sprachstand eines Kandidaten, indem sie sich an vorher festgelegten Spezifikationen, Niveau- und Kompetenzbeschreibungen orientieren. Standardisierte, internationale Zertifikate sind den Definitionen zufolge in diese Kategorie der Niveauprüfungen einzuordnen, weshalb im Folgenden ausschließlich auf diese Prüfungen eingegangen wird.

## 3. Merkmale standardisierter Zertifikatsprüfungen

### 3.1. Kommunikatives Testen

Die meisten in diesem Beitrag erwähnten Prüfungen lassen sich als *kommunikativ orientierte* Sprachprüfungen bezeichnen. Kommunikative Prüfungen setzen (vereinfacht ausgedrückt) nicht sprachliches *Wissen*, sondern *Können*, also sprachliche *Handlungsfähigkeit* in den Mittelpunkt und sind daher zum größten Teil auch als *Performanztests* anzusehen. Im Gegensatz zu sog. Kompetenztests (die meist mit theoretischen Modellen „tieferliegender Grundfähigkeiten“ arbeiten) spiegeln Performanztests realitätsnahes ziel-sprachliches Verhalten in realitätsnahen, wahrscheinlichen und relevanten Situationen wider (*real life approach*, vgl. dazu Bachman 1990; Weir 1993; Glaboniat 1998). Vom Verhalten und den Leistungen der Testteilnehmenden in diesen Situationen wird abgeleitet, wie sich eine Person außerhalb des Testzusammenhangs verhalten *könnte*. Dazu werden die TestkandidatInnen in möglichst realitätsnahe und kommunikative Akte involviert und müssen innerhalb einer vorgegebenen Situation und sozialen Rolle unterschiedliche

sprachliche Handlungsaufgaben erfüllen. Die kommunikative Orientierung findet sich dabei aber nicht nur in der Kontextualisierung der Aufgabe, sondern auch in der Authentizität der (Input-)Texte, in der Berücksichtigung der kommunikativen Lernerbedürfnisse, in der repräsentativen Auswahl der Aufgaben, in der Alters-, Sozialisations- und Kulturadäquatheit der Themen sowie in der Berücksichtigung von Kommunikationsstrategien.

Kommunikativ orientierte Sprachprüfungen – das gilt auch für alle hier angeführten – gehen von einem sehr breit gefassten, pragmatisch orientierten Konzept der Sprachkompetenz aus (vgl. v. a. Modell von Bachmann 1990; Glaboniat 1998: 53–62; GER: 21–29, 51–102) und differenzieren in der Regel nach den vier sprachlichen Hauptaktivitäten bzw. Fertigkeiten Leseverstehen, Hörverstehen (= mündliche und schriftliche Rezeption), Schreiben und Sprechen (= mündliche und schriftliche Produktion bzw. Interaktion).

Linguistische Kompetenzen, wie Grammatik und Lexik, werden teilweise implizit mit-erfasst (v. a. innerhalb des Schreibens, Sprechens), teilweise in entsprechenden Subtests eigens getestet (vgl. z. B. „Sprachbausteine“ in der Prüfung B1 Zertifikat Deutsch).

### 3.2. Direktes Testen

Um eine möglichst hohe Validität zu erreichen, versuchen kommunikative Tests möglichst *direkt* zu testen. Dies ist bei produktiven Aktivitäten wie Schreiben und Sprechen relativ leicht zu realisieren und führt zu sehr offenen Aufgaben wie z. B. dem Verfassen von Briefen oder E-Mails im schriftlichen Teil oder dem Führen eines alltäglichen Gesprächs im mündlichen Teil einer Prüfung. Die testtheoretische Herausforderung bei diesen Testformaten liegt v. a. in der Gewährleistung der Auswertungsreliabilität und -objektivität. Bei den rezeptiven Hör- und Leseverstehensaufgaben ist dies genau umgekehrt. Zwar wird grundsätzlich versucht, möglichst direkt zu testen, d. h. sämtliche für Hör- bzw. Lesekompetenz irrelevante Faktoren auszublenden. Dennoch wird letztlich immer ein indirektes Antwortformat gebraucht, auf der die jeweilige – innerlich ablaufende und daher an sich nicht direkt beobachtbare – Verstehensleistung nach außen nachgewiesen wird. Diese Antwortformate können im Grad ihrer Direktheit sehr unterschiedlich sein. Ein Beispiel für nicht-direktes, nicht-valides und nicht-zuverlässiges Testen wäre, dass die TestkandidatInnen Gehörtes in Form einer schriftlichen Zusammenfassung in der Zielsprache wiedergeben müssen. Hinzukommende, verfälschende Faktoren wären hier, 1) die geforderte Merkleistung, 2) die Wiedergabe in der Zielsprache und 3) das alles in schriftlicher Form (sprachliche Korrektheit, Textsortenkompetenz „Wie schreibt man eine Zusammenfassung?“). Im Bemühen um möglichst hohe Direktheit und Reliabilität werden in standardisierten Tests daher meist geschlossene Formate wie Richtig-Falsch oder Multiple Choice verwendet – auch wenn solche Auswahlmöglichkeiten in realen Situationen natürlich nicht vorkommen. Selbstverständlich liegt hier als verfälschender Faktor auch immer noch die Abhängigkeit vom (Lese-)Verstehen der Aufgaben (Items) vor, aber im Vergleich zu den schriftlichen Anforderungen im o. a. Beispiel ist dieser Kompromiss innerhalb des Spannungsfelds zwischen Validität und Reliabilität von Testaufgaben vertretbar.

### 3.3. Qualitätsstandards und die Zuordnung zu den Europaratsstufen

Internationale Sprachprüfungen und Zertifikate werden auch als Qualifikations- oder *high stakes tests* bezeichnet, da sie als Selektionsinstrumente oder im Zusammenhang mit Entscheidungen eingesetzt werden, die weitergehende Konsequenzen für die TestkandidatInnen haben (z. B. Anstellung, Beförderung, Zulassung zur Universität, Staatsbürgerschaft, Visumsverlängerung). Es versteht sich daher von selbst, dass solche Prüfungen höchsten Qualitätsstandards gerecht werden müssen. Das betrifft nicht nur die klassischen Testgütekriterien Validität, Objektivität und Reliabilität, sondern auch andere Qualitätsmerkmale wie Praktikabilität, Fairness, Nützlichkeit und Transparenz. Die meisten der hier angeführten Prüfungsorganisationen stellen diesbezüglich an sich selbst sehr hohe Ansprüche bzw. fühlen sich internationalen Standards verpflichtet, wie z. B. dem *Code of Practice* der *Association of Language Testers in Europe* (ALTE).

Eine der wichtigsten Voraussetzungen für standardisiertes Prüfen ist das *kriterien-* oder *sachorientierte* Messen. Im Gegensatz zu *norm-* oder *individualorientiertem* Vorgehen in nicht standardisierten Verfahren (z. B. in schulischer Leistungsmessung, vgl. dazu Glaboniat 2002), erfordert standardisiertes Testen, dass der Testgegenstand (das Testkonstrukt bzw. das Testkriterium) vorher festgelegt ist und als Bezugsgröße für jede Messung gilt. Demnach wird die Leistung einer Person weder in Bezug auf Leistungen anderer Personen (Norm), noch in Bezug auf eigene, vorangegangene Leistungen (Individuum) bewertet, sondern lediglich in Bezug auf das vorgegebene Lernziel (Kriterium). Nur so ist eine transparente und aussagekräftige Beurteilung der Sprachbeherrschung eines Lernenden gewährleistet. Standardisierte Prüfungen und Tests müssen somit über klar definierte Beschreibungen der gewünschten Kompetenzen und Fähigkeitsgrade und, davon abgeleitet, über transparente Prüfungsziele und Beurteilungskriterien verfügen. Der GER und die anderen in seinem Umfeld stehenden Europaratsinstrumentarien, wie z. B. „Profile deutsch“ (Glaboniat et al. 2005) für die deutsche Sprache, haben in diesem Zusammenhang nicht nur wesentliche Hilfsmittel für kriterienorientiertes Testen geschaffen, sondern – zumindest auf den ersten Blick – auch den Weg zu mehr Transparenz und schneller Vergleichbarkeit geebnet.

Mittlerweile gibt es keine internationale Sprachprüfung mehr, die sich nicht niveaumäßig innerhalb des Bezugssystems des Europarats einordnet und auch die Bezeichnungen A1 bis C2 wählt.

Um die Qualität der Prüfungen und die Gültigkeit in Hinblick auf die Niveauzuordnung zu gewährleisten, unterziehen bzw. unterzogen sich seriöse Prüfungen einer intensiven Auseinandersetzung mit dem GER, die in vielen Fällen zu Revisionen, Adaptionen oder Neuentwicklungen führt/geführt hat. In diesem Zusammenhang erwähnenswert sind v. a. zwei Publikationen des Europarats, die Testinstitutionen bei der Erstellung und Zuordnung von Tests behilflich sein sollten. Während die Publikation „Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Language Examining and Test Development“ (vgl. Council of Europe 2002) eine allgemeine Einführung in den Bereich des Sprachentestens darstellt, unterstützt das „Manual for Relating Language Tests to the CEFR“ (vgl. Council of Europe/Relating Language Tests to the CEFR, Manual 2009) v. a. den Zuordnungsprozess: Durch das Vergleichen, Analysieren und Abgleichen von Prüfungszielen und -inhalten mit den detaillierten Niveaubeschreibungen des GER (und deren sprachlicher Konkretisierung, wie sie z. B. für die deutsche Sprache durch „Profile deutsch“ vorliegt), werden die einzelnen Prüfungen mit

dem GER verlinkt, validiert und evaluiert. Bei aller Euphorie muss aber auch betont werden, dass der GER in der Fachdiskussion durchaus nicht unumstritten ist, dass er viele Schwachstellen aufweist und es eine Menge offene und ungeklärte Fragen gibt (vgl. Bausch et al. 2003). Hinzu kommt, dass die Niveauzuordnung hie und da – sei es bei Lehrwerken oder nicht-standardisierten Tests im Internet – nicht nachvollziehbar ist und Etikettenschwindel angenommen werden muss. Internationale Sprachprüfungen nehmen die Zuordnung in der Regel jedoch sehr ernst, womit für den Bereich des Sprachentestens durch den GER tatsächlich ein wichtiger Meilenstein gesetzt wurde. Es existieren nun zumindest Anhaltspunkte, die die Prinzipien wie Vergleichbarkeit, Transparenz und Kohärenz erstmals wirklich ermöglichen.

## 4. Internationale DaF-Prüfungen

### 4.1. Beispiel *Zertifikat Deutsch als Fremdsprache*

Bevor im Folgenden die gängigsten DaF-Prüfungen vorgestellt werden, soll die weltweit am meisten frequentierte Deutschprüfung, das auf dem Niveau B1 angesiedelte *Zertifikat Deutsch* (kurz: ZD), beispielhaft etwas detaillierter skizziert werden.

Gemäß den Beschreibungen des GER sollen Lernende auf diesem Niveau zur „selbstständigen Sprachverwendung“ in Alltagssituationen fähig sein, d. h. sie können u. a. sprachlich „die meisten Situationen bewältigen, denen man auf Reisen im Sprachgebiet begegnet“ und „sich einfach und zusammenhängend über vertraute Themen und persönliche Interessengebiete äußern“ (GER B1).

Konzipiert ist das ZD für Lernende ab 16 Jahren, für jüngere DeutschlerInnen zwischen 12 und 16 Jahren wird zusätzlich das *Zertifikat Deutsch für Jugendliche* (ZDj) angeboten. Das ZDj unterscheidet sich vom ZD lediglich in der Auswahl der Themen und Texte; Umfang und Format der Aufgaben sowie die Bewertung und letztlich auch die ausgestellte Zertifikatsurkunde sind identisch.

Erstellt und herausgegeben werden beide Prüfungen (ZD und ZDj) in trinationaler Zusammenarbeit vom Goethe-Institut (GI) und der telc GmbH/Deutschland, dem Österreichischen Sprachdiplom Deutsch (ÖSD) und dem Lern- und Forschungszentrum der Universität Freiburg/Schweiz. Der trinationalen Kooperation zu Grunde liegt die pluri-zentrische Sprachauffassung des Deutschen, wonach die Standardvarietäten aus Deutschland, Österreich und der deutschsprachigen Schweiz gleichberechtigt nebeneinander gestellt sind. In den Prüfungen finden sich daher v. a. in den rezeptiven Teilen Hör- und Lesetexte aus den drei Ländern.

Die Aufgabenstellungen, d. h. die in der Prüfung zu bewältigenden Situationen mit unterschiedlichen rezeptiven (Hör- und Lesetexte) und produktiven Anforderungen (Sprech- und Schreibenanlässen) orientieren sich an eigens für das ZD erstellten Spezifikationen, die „Szenarien“, Strategien, Notionen und Themen sowie detaillierte Sprachhandlungs-, Wortschatz- und Grammatikinventare (*Zertifikat Deutsch Lernzielkatalog 1999*: 26 und 61–371) umfassen. Dieser ZD-Lernzielkatalog geht in großen Teilen wiederum auf die 1980 im Umfeld des Europarats erschienene „Kontaktschwelle“ von Baldeger et al. 1980 (bzw. dem 1975 erschienenen „Threshold Level“ von Van Ek) zurück. Somit orientiert sich das ZD nicht wie andere, neuere Prüfungen direkt an den Bestimmungen des GER bzw. Profile deutsch, sondern geht noch auf deren Vorarbeiten zurück.

Auf Grund dieser Zusammenhänge lässt sich zwar generell annehmen, dass sich die meisten der ZD-Spezifikationen mit den Niveaubeschreibungen und Skalen des GER bzw. den Konkretisierungen in *Profile deutsch* decken, eine detaillierte Prüfung bzw. Analyse sowie daraus resultierende Überarbeitung wird in Fachkreisen als längst fällig angesehen. Erste Schritte in diese Richtung wurden 2008 gesetzt, eine Revision dürfte in den nächsten Jahren zu erwarten sein.

Tabelle 144.1 zeigt u. a. den Aufbau des ZD (auch ZDj) mit den einzelnen Prüfungsteilen, deren Aufgabenformate und Gewichtung.

Tab. 144.1: Übersicht über das Zertifikat Deutsch

Übersicht über das Zertifikat Deutsch					
Fertigkeiten/ Kompetenzen	Überprüfungsdomäne/ Subtests	Aufgabentypen/ Antwortformate	Zeit	Punkte	rel. Gewichtung
<b>Schriftliche Prüfung</b>			<b>150'</b>	<b>225</b>	
Leseverstehen			<b>90'</b>	<b>75</b>	<b>25 %</b>
	LV1: Globalverstehen	Zuordnung (5)*		25	
	LV2: Detailverstehen	Mehrfachauswahl (5)		25	
	LV3: Selektives Verstehen	Zuordnung (10)		25	
Sprachbausteine				<b>30</b>	<b>10 %</b>
	SB1: Lexik; v. a. Morpho- syntax	Mehrfachauswahl (10)		15	
	SB2: Lexik; v. a. Semantik	Mehrfachauswahl (10)		15	
Hörverstehen			<b>ca. 30'</b>	<b>75</b>	<b>25 %</b>
	HV1: Globalverstehen	Richtig-Falsch (5)		25	
	HV2: Detailverstehen	Richtig-Falsch (10)		25	
	HV3: Selektives Verstehen	Richtig-Falsch (5)		25	
Schreiben			<b>30'</b>	<b>45</b>	<b>15 %</b>
	Schriftliche Interaktion; In-, halbformeller Antwortbrief	4 Leitpunkte bearbeiten		45	
<b>Mündliche Prüfung</b>			<b>ca. 15'</b>	<b>75</b>	<b>25 %</b>
Sprechen (Mündlicher Ausdruck)	v. a. mündliche Interak- tion, tlw. mündliche Produktion	Paar- oder Einzelprüfung			
	MA 1: Kontaktaufnahme			15	
	MA 2: Gespräch über ein Thema			30	
	MA 3: Gem. eine Aufgabe lösen			30	
<b>ZERTIFIKAT DEUTSCH gesamt</b>				<b>300</b>	

\* = Anzahl der Items pro Subtest

## 4.2. Überblick über internationale DaF-Zertifikate und kurze Beschreibung

In der Tabelle 144.2 finden sich die gegenwärtig am internationalen Prüfungsmarkt angebotenen, standardisierten DaF-Zertifikate. DaZ-Prüfungen, die nur im deutschsprachigen Inland angeboten werden, sind hier bewusst ausgespart (vgl. Art. 145).

Studienspezifische Prüfungen wie *TestDaF*, *DSH* oder *DSD* wurden in die Tabelle 2 nicht aufgenommen, weil sie entweder über mehrere Niveaus gehen (vgl. TestDaF 3–5) oder nicht in dem Ausmaß standardisiert sind wie die anderen.

Tab. 144.2: Standardisierte DaF-Zertifikate

Niv <sup>1</sup> .	Prüfung	Zielgruppe	Träger <sup>2</sup>
A1	Start Deutsch 1	Erwachsene	GI; TELC
	A1 Grundstufe Deutsch 1 (A1 GD 1)	Erwachsene	ÖSD
	Fit in Deutsch 1	Kinder/Jugendliche	GI
	A1 Kompetenz in Deutsch 1 (A1 KID 1)	Kinder/Jugendliche	ÖSD
A2	Start Deutsch 2	Erwachsene	GI; TELC
	A2 Grundstufe Deutsch 2 (A2 GD 2)	Erwachsene	ÖSD
	Fit in Deutsch 2	Kinder/Jugendliche	GI
	A2 Kompetenz in Deutsch 2 (A2 KID 2)	Kinder/Jugendliche	ÖSD
B1	B1 Zertifikat Deutsch (B1 ZD)	Erwachsene	GI; TELC; ÖSD; EDK
	B1 Zertifikat Deutsch für Jugendliche (ZDJ)	Jugendliche (12–16 J.)	GI; TELC; ÖSD; EDK
B2	Goethe Zertifikat B2	Erwachsene	GI
	B2 Mittelstufe Deutsch (B2 MD)	Erwachsene	ÖSD
	telc Deutsch B2 (= früher: Zertifikat Deutsch plus)	Erwachsene	TELC
	Zertifikat Deutsch für den Beruf (ZDfB)	Erwachsene	GI; TELC
C1	Goethe Zertifikat C1	Erwachsene	GI
	C1 Oberstufe Deutsch (C1 OD)	Erwachsene	ÖSD
	telc Deutsch C1	Erwachsene	TELC
	Prüfung Wirtschaftsdeutsch International (PWD)	Erwachsene	GI und Partner <sup>3</sup> ,
C2	Zentrale Oberstufenprüfung (ZOP)	Erwachsene	GI
	Kleines Deutsches Sprachdiplom (KDS)	Erwachsene	GI
	C2 Wirtschaftssprache Deutsch (C2 WD)	Erwachsene	ÖSD
	Großes Deutsches Sprachdiplom (GDS)	Erwachsene	GI

<sup>1</sup> Diese Niveauzuordnung erfolgt meist durch Selbsteinstufungen der Prüfungsanbieter, die auf diversen qualitativen und quantitativen Verfahren (Expertengutachten, Benchmarking-Verfahren, psychometrische Linkingprozesse) basieren.

<sup>2</sup> Trägerorganisationen: GI = Goethe-Institut; TELC = früher: Weiterbildungs-Testsysteme GmbH; ÖSD = Österreichisches Sprachdiplom Deutsch; EDK = Schweizerische Konferenz der kantonalen Erziehungsdirektoren.

<sup>3</sup> GI in Kooperation mit dem Deutschen Industrie- und Handelstag und den Carl Duisberg Centren

#### 4.3. Kurzcharakteristik: internationale DaF-Zertifikate (Prüfungen des GI, ÖSD und TELC)

Auf den Stufen A1 und A2 werden weltweit einerseits die Prüfungen *Start Deutsch 1* und 2 des GI und TELC sowie andererseits die Prüfungen *Grundstufe Deutsch 1* und 2 des ÖSD angeboten. Die einzelnen Prüfungen sind kommunikativ aufgebaut und doku-

mentieren auf dem Niveau A1, dass die Teilnehmenden „vertraute, alltägliche Ausdrücke und ganz einfache Sätze verstehen und verwenden können“ bzw. auf A2-Niveau, dass sie sich u. a. „in einfachen, routinemäßigen Situationen verständigen können“. (Es gibt von *Start 1* und *2* auch die Inlandsvarianten für Zuwanderer *Start Deutsch 1(z)* und *Start Deutsch 2(z)*, die in den vom Bundesamt für Migration und Flüchtlinge (BAMF) geförderten Integrationskursen durchgeführt werden. In Österreich ersetzt die ÖSD-Prüfung *A1 Grundstufe 1* den früheren *Sprachkenntnisnachweis*, eine österreichische DaZ-Prüfung, die bis 2006 als Nachweis zur Erfüllung der Integrationsvereinbarung/Niederlassungsbewilligung ausreichte. Seit 2006 ist dafür das Niveau A2 erforderlich.

Für beide A-Stufen liegen auch spezifische Prüfungen für Kinder und Jugendliche vor. So können junge LernerInnen entweder die Prüfungen des Goethe-Instituts *Fit für Deutsch 1 und 2* (v. a. in Italien und Frankreich angeboten) oder die des *ÖSD KID 1* und *2* absolvieren.

Während sich auf der Stufe B1 wie oben erwähnt nicht nur die größte, sondern auch die am längsten existierende Prüfung, das *Zertifikat Deutsch* (vor 1999: *Zertifikat Deutsch als Fremdsprache*) befindet, sind die Prüfungen der drei großen Testanbieter GI, ÖSD und telc auf den Stufen B2 und C1 relativ neu.

Im Zuge der Zuordnungs- und Anpassungsprozesse der DaF-Prüfungen an den GER wurde die *Zentrale Mittelstufenprüfung (ZMP)* des GI 2008 durch zwei neue Prüfungen abgelöst: das *Goethe Zertifikat B2* einerseits und das *Goethe Zertifikat C1* andererseits. Eine ähnliche Entwicklung erfuhr die frühere ÖSD-Prüfung *Mittelstufe Deutsch*. Sie wurde Ende 2007 ebenfalls durch zwei neue Prüfungen ersetzt, die *B2 Mittelstufe Deutsch* und die *C1 Oberstufe Deutsch* des ÖSD. Auch Telc bietet seit einiger Zeit Prüfungen auf diesen Niveaus an, nämlich *telc Deutsch B2* (früher: *Zertifikat Deutsch plus*) und *telc Deutsch C1*.

Kenntnisse der Niveaustufe C2 können Lernende beispielsweise durch die *Zentrale Oberstufenprüfung (ZOP)*, das *Kleine Deutsche Sprachdiplom (KDS)* und das *Große Deutsche Sprachdiplom (GDS)* unter Beweis stellen. Durch die ZOP wird sprachliches Handeln innerhalb eines breiten Spektrums von Situationen und Themen überprüft. Von den KandidatInnen werden differenzierte sprachliche Mittel, eine starke Nuancierung beim Ausdruck und ein breites Repertoire an idiomatischer Ausdrucksweise erwartet. Der Test für das etwa auf dem gleichen Niveau angesiedelte KDS wird vom GI in Zusammenarbeit mit der Ludwig-Maximilian-Universität München (LMU) entwickelt; so auch das GDS, welches als höchstqualifizierender Abschluss im Bereich DaF außerhalb einer universitären oder einer Dolmetscher-/Übersetzer-Ausbildung gilt.

Im berufsbezogenen Kontext finden sich auf den höheren Stufen folgende Prüfungen:

Auf B2 ist das vom GI angebotene *Zertifikat Deutsch für den Beruf* angesiedelt. Dies ist eine am Arbeitsalltag orientierte Sprachprüfung, die die Kommunikationsfähigkeit in Situationen des täglichen Berufslebens überprüft.

Das GI bietet in Zusammenarbeit mit dem Deutschen Industrie- und Handelstag sowie den Carl Duisberg Centren die *Prüfung Wirtschaftsdeutsch International (PWD)* an. Geprüft wird die Kommunikationsfähigkeit in Geschäftssituationen. Das Zeugnis bescheinigt angemessene Kommunikationskompetenz auf der Niveaustufe C1 in typischen Situationen des Geschäftslebens, also bei Präsentationen, in schriftlichen und mündlichen Verhandlungen, u. Ä.

Eine Prüfung mit ähnlicher Ausrichtung, aber auf noch höherer Stufe ist die *Prüfung C2 Wirtschaftssprache Deutsch (WD)* des ÖSD. Das Diplom dient als Nachweis von kom-

petenter Sprachverwendung auf der höchsten Stufe des GER und erfordert überdies fachsprachliche Kompetenz in Wirtschaftsberufen und wirtschaftsnahen Bereichen. Die Prüfung wurde in Kooperation mit der Wirtschaftskammer Österreich entwickelt.

Die genannten berufsbezogenen Prüfungen werden sowohl von Fach- und Führungskräften als auch von Wirtschaftsstudenten gerne bei Bewerbungen eingesetzt.

#### 4.4. Spezifische Prüfungen für den Studienzugang

Speziell für studienspezifische Zwecke entwickelt wurde der *TestDaF*, eine vom TestDaF-Institut entwickelte und herausgegebene Prüfung zur Eignungs- und Leistungsfeststellung im Hochschulbereich. Der TestDaF wird weltweit angeboten und von allen Hochschulen in Deutschland als Sprachtest für die Zulassung ausländischer Studierender anerkannt. Die TestDaF-Niveaustufe 3 entspricht der Stufe B2, die Stufe TestDaF 4 der Stufe C1 des GER.

Zusätzlich werden an Deutschlands Auslandschulen die DSD-Prüfungen der deutschen Kultusministerkonferenz (KMK) angeboten. Von den beiden Prüfungsstufen *DSD 1* und *DSD 2* garantiert das DSD 2 den Hochschulzugang in Deutschland. Es setzt ungefähr 1600 Stunden Schulunterricht (in der Regel 7 Jahre) voraus und lässt sich – in der seit 2008 vorliegenden, ebenfalls nach dem GER revidierten Version – auf dem Niveau C1 einordnen.

Ebenfalls für die Zulassung zum Hochschulzugang in Deutschland wurde die *Deutsche Sprachprüfung für den Hochschulzugang (DSH)* entwickelt. Die DSH-Prüfungen werden ausschließlich in Deutschland, jeweils vor Beginn eines Studienseesters, angeboten. Der Unterschied zu den oben angeführten Prüfungen, die alle unter hohen Qualitätsanforderungen zentral erstellt und teilweise auch zentral ausgewertet werden, liegt bei der DSH im geringeren Grad ihrer Standardisierung: Zwar gibt es eine für ganz Deutschland geltende Rahmenordnung für die DSH, allerdings gestalten die einzelnen Universitäten die Prüfungen, d. h. Inhalte und Aufgabenstellungen, selbstständig. Ein vergleichbares Modell stellt das UNICert<sup>®</sup>-Zertifikatssystem dar, das auf einer Rahmenvereinbarung deutscher Universitäten und Hochschulen beruht. Träger von UNICert<sup>®</sup> ist der Arbeitskreis für Sprachenzentren (AKS).

Ähnliches gilt auch für universitätsinterne/-eigene Prüfungen oder Tests in Österreich und der deutschsprachigen Schweiz (z. B. *Ergänzungsprüfung Deutsch* in Österreich, der Deutschtest der Universität Bern, die Deutschprüfung der Universität Zürich usw.) oder für die (Abschluss-)Prüfungen der an den Universitäten angebotenen, studienvorbereitenden Deutschkurse (z. B. Vorstudienlehrgänge, Hochschulkurse, Studienkollegs u. Ä.): Sie unterscheiden sich – trotz wachsender Bezugnahme auf den GER – in Schwierigkeitsgrad, Inhalten, Aufbau und v. a. im Ausmaß ihrer Standardisierung und Qualitätssicherung immer noch so sehr voneinander, dass sie nur sehr eingeschränkt vergleichbar sind.

Für den Hochschulzugang stehen somit verschiedene Prüfungen zur Verfügung. Neben den letztgenannten, studienspezifischen Prüfungen werden aber je nach Universität und Hochschule bzw. Studienrichtung auch die anderen, oben erwähnten Prüfungen akzeptiert. Entsprechende Informationen findet man für Deutschland im Internet unter [www.sprachnachweis.de](http://www.sprachnachweis.de). In Österreich werden die ÖSD-Prüfungen an allen Universitäten anerkannt, es werden allerdings je nach Universität unterschiedliche Sprachniveaus

gefordert und diesen Niveaus entsprechend auch die anderen oben angeführten standardisierten Prüfungen akzeptiert. Informationen zur Anerkennung des ÖSD an Österreichs bzw. auch Schweizer und deutschen Universitäten sind unter [www.osd.at](http://www.osd.at) abrufbar. Für die Schweiz findet man die entsprechenden sprachlichen Aufnahmevoraussetzungen im Internet unter „Studieren in der Schweiz“ ([www.crus.ch](http://www.crus.ch)) und in allen drei Ländern selbstverständlich auch auf den entsprechenden Informationsseiten jeder einzelnen Universität.

## 5. Zusammenfassung

Der Überblick über die verschiedenen DaF-Prüfungen spiegelt nicht nur das immer größer werdende Angebot am (internationalen) Prüfungsmarkt wider, sondern zeigt die Prüfungen auch in ihrer jeweiligen Niveaueinordnung. Dies wäre noch bis vor einigen Jahren kaum möglich gewesen, bestand doch seit jeher beim Ein- und Zuordnen von Sprachzeugnissen das fundamentale Problem der fehlenden Bezugsgröße. Da verschiedene Prüfungen naturgemäß verschiedene Inhalte überprüfen, unterschiedliche Schwerpunkte setzen und für unterschiedliche Zielgruppen konzipiert sind, war eine Vergleichbarkeit der Beurteilungen immer sehr schwierig. Erst durch die Entwicklung eines gemeinsamen Referenzrahmens erfolgte eine grundsätzliche Neuorientierung. Es wurde ein gemeinsames Bezugssystem geschaffen, an welchem sich kommunikativ orientierte Prüfungen beim Analysieren, Festlegen oder Interpretieren der eigenen Lern- und Prüfungsziele orientieren können. Wie die bisherige Erfahrung bereits zeigt, wurde durch den GER für den Bereich des Prüfens und Zertifizierens ein durchaus hilfreiches und taugliches Instrument geschaffen, das sowohl innerhalb der Sprachen als auch sprachübergreifend deutlich mehr Vergleichbarkeit, Transparenz und Kohärenz ermöglicht.

## 6. Literatur in Auswahl

Bachman, Lyle F.

1990 *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Baldegger, Markus, Martin Müller und Günther Schneider

1980 *Kontaktschwelle Deutsch als Fremdsprache*. Strassburg: Langenscheidt.

Bausch, Karl-Richard; Herbert Christ, Frank G. Königs und Hans-Jürgen Krumm (Hg.)

2003 *Der Gemeinsame europäische Referenzrahmen für Sprachen in der Diskussion. Arbeitspapiere der 22. Frühjahreskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr.

Council of Europe

2002 *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Language Examining and Test Development. Prepared under the direction of Michael Milanovic (A.L.T.E.)*, Strasbourg: Council of Europe, Language Policy Division. Verfügbar unter <http://www.coe.int/T/DG4/Portfolio/documents/Guide%20October%202002%20revised%20version1.doc> (30. 11. 2009).

Council of Europe

2009 *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF). Manual*. Strasbourg: Council of

Europe, Language Policy Division. Verfügbar unter <http://www.coe.int/T/DG4/Portfolio/documents/Manual%20Revision%20-%20proofread%20-%20FINAL.pdf> (30. 11. 2009).

Europarat

2001 *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren und beurteilen.* Hrsg. vom Goethe-Institut, der KMK, der EDK und dem BMBWK. Berlin etc.: Langenscheidt.

Glaboniat, Manuela

1998 *Kommunikatives Testen im Bereich Deutsch als Fremdsprache.* Innsbruck etc.: Studienverlag.

Glaboniat, Manuela

2002 *Schulnoten versus standardisierte Prüfungen – Gedanken zum Neben- und Gegeneinander schulischer und standardisierter Leistungsmessung im DaF-Bereich.* In: Hans Barkowski und Renate Faistauer (Hg.): ... in *Sachen Deutsch als Fremdsprache*, 217–230. Baltmannsweiler: Schneider.

Glaboniat, Manuela; Martin Müller, Paul Rusch, Helen Schmitz und Lukas Wertenschlag

2005 *Profile deutsch A1–C2 (Version 2.0).* Berlin etc.: Langenscheidt.

Van Ek, Jan

1975 *The Threshold Level*, with an appendix by L.G. Alexander. Strasbourg: Council of Europe.

Weir, Cyril

1993 *Understanding and Developing Language Tests.* New York: Prentice Hall.

Zertifikat Deutsch. Lernziele und Testformat (= Lernzielkatalog)

1999 Hrsg. von Weiterbildungs- und Testsysteme GmbH, Goethe-Institut, Österreichisches Sprachdiplom Deutsch, Schweizerische Konferenz der kantonalen Erziehungsdirektoren.

Quellen für Prüfungen und Prüfungsanbieter (Zugriff am 30. 11. 2009)

Goethe-Institut:

[www.goethe.de](http://www.goethe.de)

ÖSD:

[www.osd.at](http://www.osd.at)

telc:

[www.telc.net](http://www.telc.net)

TestDaf:

[www.testdaf.de](http://www.testdaf.de)

DSD:

[www.dsd-kmk.de](http://www.dsd-kmk.de)

DSH:

Informationen unter [www.daad.de](http://www.daad.de) oder [www.fadaf.de](http://www.fadaf.de)

*Manuela Glaboniat, Klagenfurt (Österreich)*

## 145. Sprachprüfungen für Deutsch als Zweitsprache

1. Einleitung
2. Bundesrepublik Deutschland: *Deutschtest für Zuwanderer (dtz)*
3. Österreich: *Test des Österreichischen Integrationsfonds (ÖIF-Test)*
4. Fazit
5. Literatur in Auswahl

### 1. Einleitung

Sprachprüfungen Deutsch als Fremdsprache haben eine lange Tradition. Es gibt ein Spektrum an standardisierten DaF-Prüfungen für verschiedene Zielgruppen, zu unterschiedlichen Zwecken und auf allen Niveaus des Gemeinsamen europäischen Referenzrahmens (GER) (vgl. Art. 144). Im Gegensatz dazu ist das Angebot an standardisierten Deutsch als Zweitsprache-Prüfungen sehr begrenzt.

Allgemeines Merkmal von Zweitsprachenprüfungen ist, dass sie Sprachkompetenzen von Sprachminderheiten in einem Umfeld überprüfen, in dem die entsprechende Sprache die der Bevölkerungsmehrheit ist, und dass die Prüfungsinhalte Bezug auf dieses Umfeld nehmen. Für Zweitsprachenprüfungen sind ebenso viele Varianten denkbar wie für Fremdsprachenprüfungen, jedoch ist weltweit der zunehmend häufigere Kontext für Zweitsprachenprüfungen der Bereich der Regulierung von Zuwanderung und Einbürgerung. Viele Staaten verlangen von Zuwanderern den Nachweis von Kenntnissen in der Nationalsprache. Begründet wird dies damit, dass hinreichende Sprachkenntnisse die Chancen auf dem Arbeitsmarkt erhöhen und dass über die Sprache eine erfolgreiche Integration in die Aufnahmegesellschaft möglich ist. Oft werden Zweitsprachenprüfungen als Instrument der Zuwanderungsbeschränkung eingesetzt.

Auch in den drei deutschsprachigen Ländern gelten seit einigen Jahren Zuwanderungsgesetze. In Deutschland und in Österreich wurden in diesem Kontext in öffentlichem Auftrag Prüfungen entwickelt, in der Schweiz ist eine offizielle Prüfung zum Nachweis von (deutschen) Sprachkenntnissen nicht vorgesehen.

### 2. Bundesrepublik Deutschland: *Deutschtest für Zuwanderer (dtz)*

In der Bundesrepublik Deutschland trat am 01.01.2005 das neue Zuwanderungsgesetz in Kraft. Es schreibt vor, dass Zuwanderer zur Erteilung der Niederlassungserlaubnis u. a. Sprachkenntnisse nachweisen müssen (vgl. § 9 Zuwanderungsgesetz): „(2) Einem Ausländer ist Niederlassungserlaubnis zu erteilen, (...) wenn er über ausreichende Kenntnisse der deutschen Sprache verfügt.“ Als ausreichend werden Kenntnisse auf dem Niveau B1 des GER angesehen. Erworben werden können diese in so genannten Integrationskursen, die die Eingliederung von MigrantInnen in das gesellschaftliche, wirtschaftliche und kulturelle Leben der Bundesrepublik Deutschland fördern sollen (vgl. § 43 Zuwanderungsgesetz). Integrationskurse bestehen aus einem 45-stündigen Orientierungskurs und einem 600-stündigen Sprachkurs. Sie basieren auf einem eigens für die Zielgruppe der Migran-

tInnen erarbeiteten Rahmencurriculum (Buhlmann et al. 2007). Der Orientierungskurs wird mit einem Test abgeschlossen, der Sprachkurs mit dem „Deutstest für Zuwanderer“ (dtz).

Der dtz wurde im Auftrag des Bundesinnenministeriums vom Goethe-Institut (München) und der Telc GmbH (Frankfurt), einer Tochtergesellschaft des Deutschen Volkshochschulverbandes, gemeinsam entwickelt. Der dtz ist eine standardisierte skalierte Feststellungs- bzw. Kursabschlussprüfung, die zentral erstellt und dezentral an durch das Bundesamt für Migration und Flüchtlinge (BAMF) autorisierten Institutionen durchgeführt wird. Die schriftlichen Prüfungsteile werden zentral (durch die Telc GmbH) ausgewertet, die mündliche Prüfung von lizenzierten PrüferInnen vor Ort durchgeführt und bewertet. Der dtz überprüft sprachliche Kompetenzen auf mehr als einem Niveau, nämlich auf den Stufen A2 und B1 des GER. Er wird in einer Variante für Erwachsene und einer für Jugendliche angeboten.

## 2.1. Zielgruppe der Prüfung

Zielgruppe des dtz sind Zuwanderer ab 16 Jahren, die die lateinische Schrift beherrschen. Die KandidatInnen haben freiwillig oder obligatorisch einen Integrationskurs besucht und schließen den Kurs mit dem dtz ab. Zugelassen sind auch MigrantInnen, die nicht an einem Kurs teilgenommen haben.

## 2.2. Prüfungsinhalte

Die Prüfungsinhalte basieren auf dem „Rahmencurriculum für Integrationskurse Deutsch als Zweitsprache“ (vgl. o.), welches an der (bundesdeutschen) Lebenswirklichkeit der Zuwanderer orientiert ist. Für das Rahmencurriculum wurden Bedarfsanalysen zur Ermittlung sprachlicher Handlungsfelder, die für MigrantInnen relevant sind, durchgeführt. Ermittelt wurden u. a. die Handlungsfelder Ämter und Behörden, Arbeit, Einkauf, Wohnen. Außerdem wurden fünf Handlungsfelder für übergreifende Kommunikationsbereiche definiert, z. B. Gestaltung sozialer Kontakte, Realisierung von Gefühlen, Haltungen und Meinungen. Auf dieser Basis wurden Lernziele formuliert, von denen ein Kernbereich für die Prüfungsinhalte relevant ist (vgl. Goethe-Institut 2009: 30 ff.), beispielsweise „Kann einfache mündliche Anleitungen verstehen, z. B. zum Gebrauch eines Gerätes“ oder „Kann schriftlichen Aufforderungen der Behörden relevante Informationen entnehmen, z. B. Fristen“.

## 2.3. Prüfungsteile

Der dtz besteht aus vier Subtests, in denen die Fertigkeiten, Hören, Lesen, Schreiben und Sprechen getrennt überprüft werden. Der Subtest Hören enthält vier Aufgaben in geschlossenen Formaten mit insgesamt 20 Items. Ziel ist es zu überprüfen, inwieweit die TestteilnehmerInnen in der Lage sind, unterschiedlichen kürzeren Hörtexten selektiv Informationen zu entnehmen. In der ersten Aufgabe werden Telefonansagen oder öffent-

liche Lautsprecherdurchsagen präsentiert, Aufgabe zwei enthält (Kurz-)Informationen aus dem Radio, z. B. einen Wetterbericht. Grundlage der dritten Aufgabe sind vier Gespräche, zu denen jeweils zwei Aufgaben zu lösen sind. In Aufgabe vier bilden verschiedene Meinungsäußerungen zu einem Thema die Textgrundlage, z. B. obligatorischer Kindergartenbesuch. Der Subtest Hören dauert 25 Minuten.

Der Subtest Lesen umfasst fünf Aufgaben in geschlossenem Format mit insgesamt 25 Items. Ziel ist es zu überprüfen, inwieweit die TestteilnehmerInnen in der Lage sind, (semi-)authentische Texte unterschiedlicher Länge global bzw. selektiv zu verstehen. Textgrundlage der ersten Aufgabe sind Verzeichnisse, z. B. Kaufhaustafeln. Die KandidatInnen sollen zeigen, dass sie sich im Text orientieren und ihm Informationen entnehmen können. In der zweiten Aufgabe werden mehrere Kurztexte präsentiert, z. B. Kleinanzeigen, denen spezifische Informationen zu entnehmen sind. In Aufgabe drei sind die KandidatInnen gefordert, Mitteilungen oder Briefe global oder detailliert zu verstehen. Die vierte Aufgabe enthält einen längeren Text, etwa einen Beipackzettel zu einem Medikament, dem selektiv Informationen oder Anweisungen entnommen werden sollen. Im fünften Teil des Subtests Lesen müssen einzelne Wörter in einem Brief ergänzt werden. In dieser Aufgabe sind Elemente von Schreibkompetenz enthalten, da die Wörter und Wortformen adressatengerecht ausgewählt werden müssen. Für den Subtest Lesen sind 45 Minuten vorgesehen.

Im Subtest Schreiben werden den TeilnehmerInnen zwei Aufgaben zur Auswahl vorgelegt. Anhand von vier Leitpunkten ist eine kurze Mitteilung zu verfassen. Die KandidatInnen sollen zeigen, dass sie in der Lage sind, relativ einfache Korrespondenz auf Deutsch zu führen, etwa eine Entschuldigung für Fehlen im Sprachkurs. Der Prüfungsteil Schreiben dauert 30 Minuten.

Der Subtest Sprechen findet in der Regel als Paarprüfung mit zwei PrüferInnen statt. Die KandidatInnen sollen nachweisen, dass sie in Alltagssituationen sprachlich handeln können. Dabei werden monologisches Sprechen und Interaktion gefordert. Die KandidatInnen bearbeiten drei Aufgaben, in denen unterschiedliche Sprachhandlungen zu realisieren sind. In der ersten Aufgabe sollen sie sich vorstellen und auf Fragen des Prüfers reagieren können. Aufgabe zwei fordert von den KandidatInnen, dass sie – auf der Basis von Text- oder Bildimpulsen – über Gegebenheiten im Heimatland berichten und diese mit der Situation in Deutschland vergleichen, z. B. Mülltrennung. In der dritten Aufgabe sollen die KandidatInnen gemeinsam etwas planen oder aushandeln, etwa ein Nachbarschaftsfest. Der Subtest Sprechen dauert ca. 10 Minuten pro KandidatIn.

## 2.4. Bestehensbedingungen

Der dtz ist durch komplexe Bestehensbedingungen gekennzeichnet. Die rezeptiven Subtests Hören und Lesen werden in der Bewertung zusammengefasst. Bei einer maximalen Punktzahl von 45 Punkten wird das Ergebnis des rezeptiven Prüfungsteils als A2-Niveau gewertet, wenn 20 Punkte erreicht worden sind. Ab einer Punktzahl von 33 wird ein Prüfungsergebnis auf der Niveaustufe B1 eingeordnet.

Die produktiven Prüfungsteile werden – kriteriumsorientiert – von zwei Prüfern unabhängig voneinander bewertet. Zur Bewertung des Subtests Schreiben werden die

Kriterien Vollständigkeit, kommunikative Gestaltung, Korrektheit und Wortschatz herangezogen, beim Sprechen die Kriterien Aufgabenbewältigung, Intonation/Aussprache, Flüssigkeit, Korrektheit und Wortschatz. Die Prüfer entscheiden bei jedem der qualitativen Kriterien, ob eine Prüfungsleistung dem Deskriptor für A1, A2 oder B1 zuzuordnen ist. Liegt die Mehrheit der gewählten Deskriptoren bei B1, so wird das Ergebnis des Subtests Sprechen bzw. Schreiben dem Niveau B1 zugewiesen, sonst darunter.

Zum Prüfungsergebnis tragen die Subtests in unterschiedlichem Maße bei: Besonderes Gewicht kommt dem Sprechen zu. Wenn im Subtest Sprechen und in mindestens einem der schriftlichen Prüfungsteile die Stufe B1 erreicht wurde, wird die gesamte Prüfungsleistung mit B1 bewertet.

### 3. Österreich: *Test des Österreichischen Integrationsfonds (ÖIF-Test)*

In Österreich regelt das Bundesgesetz über die Niederlassung und den Aufenthalt in Österreich (NAG), das am 01. 01. 2006 in Kraft trat, die Zuwanderung von Menschen, die nicht aus dem EWR-Gebiet stammen (Drittstaatsangehörige). Wollen sich diese MigrantInnen dauerhaft oder längerfristig in Österreich niederlassen, müssen sie deutsche Sprachkenntnisse auf dem Niveau A2 des GER nachweisen, um „am gesellschaftlichen, wirtschaftlichen und kulturellen Leben in Österreich“ (NAG § 14) teilnehmen zu können. Dieser Nachweis kann über den ÖIF-Test, der den Abschluss eines Integrationskurses (Modul 2) bildet, erbracht werden. Die Prüfung soll dabei „die Besonderheiten des Sprachenlernens von Migranten sowie deren spezifische Lernvoraussetzungen“ (BGBl.II, Absatz I.5) berücksichtigen.

Beim ÖIF-Test, entwickelt vom Österreichischen Integrationsfonds, Fonds zur Integration von Flüchtlingen und Migranten, handelt es sich um eine nach einheitlichen Standards konzipierte Feststellungsprüfung. Die Testsätze werden zentral erstellt, die Prüfungen werden dezentral, an zertifizierten Instituten von zertifizierten PrüferInnen durchgeführt und ausgewertet.

#### 3.1. Zielgruppe der Prüfung

Zielgruppe des ÖIF-Tests sind erwachsene MigrantInnen, die einen dauerhaften Aufenthalt in Österreich anstreben.

#### 3.2. Prüfungsinhalte

Die Inhalte der Prüfung basieren auf dem Rahmencurriculum für die Deutsch-Integrationskurse und sind am Alltag der MigrantInnen („Lebensraum Österreich“) orientiert. Es werden authentische Situationen aus dem sozialen und beruflichen Alltag abgebildet (u. a. Wohnen, Arbeit, Beruf/Ausbildung, Einkauf). Die TeilnehmerInnen sind gefordert, dem A2-Niveau entsprechend situationsadäquat zu agieren und reagieren, u. a. die eigene Meinung und Bedürfnisse zu vertrauten Themen zu äußern, Auskünfte einzuholen oder

zu geben. Sie sollen nachweisen, dass sie authentische Texte im Rahmen von Sprachkenntnissen auf A2-Niveau lesend verstehen und Hörtexten in deutlich artikulierter (österreichischer) Standardsprache folgen können.

### 3.3. Prüfungsteile

Der ÖIF-Test setzt sich aus vier Subtests (Modulen) zusammen, in denen die Fertigkeiten Lesen, Hören, Schreiben und Sprechen getrennt überprüft werden. Das Modul Lesen umfasst drei geschlossene Aufgaben mit insgesamt 10 Items zu unterschiedlichen Kurztexen. Ziel ist zu überprüfen, inwieweit die TestteilnehmerInnen in der Lage sind, einen Text global, selektiv bzw. detailliert zu verstehen. In den beiden ersten Aufgaben werden jeweils mehrere semi-authentische Kleinanzeigen, z. B. Mietangebote oder Stellenanzeigen, präsentiert, denen Aussagen bzw. Personenprofile zugeordnet werden sollen. Die dritte Aufgabe besteht aus einem kurzen (adaptierten) Zeitungstext zu einem Alltagsthema, z. B. Gesundheit, zu dem Mehrfachwahlaufgaben gestellt werden. Das Modul Lesen dauert 30 Minuten.

Das Modul Hören enthält drei Aufgaben in unterschiedlichen Formaten (halb-offen und geschlossen). Ziel ist zu überprüfen, inwieweit die KandidatInnen in der Lage sind, einen kurzen monologischen Hörtext selektiv zu verstehen. In der ersten Höraufgabe sollen Informationen aus einer Mobilbox-Nachricht entnommen und auf einem Notizblatt notiert werden, z. B. Tag und Uhrzeit eines Arzttermins. Basis der zweiten Aufgabe ist eine mündliche Wegbeschreibung, anhand derer die Testteilnehmer den Weg in einem Stadtplan einzeichnen müssen. In der dritten Aufgabe hören die KandidatInnen vier unterschiedliche Kurzmeldungen aus dem Radio, zu denen jeweils eine Mehrfachwahlaufgabe zu lösen ist. Das Modul Hören dauert ca. 20 Minuten.

Der Subtest Schreiben besteht aus zwei Aufgaben. Ziel ist es zu überprüfen, inwieweit die KandidatInnen in der Lage sind, ein einfaches Formular mit personenbezogenen Daten auszufüllen. In der ersten Aufgabe sollen fünf Lücken in einem Formular mit den persönlichen Daten des Testteilnehmers ergänzt werden. Die zweite Aufgabe fordert von den KandidatInnen, fünf Ergänzungen für eine fiktive Person in einem Formular vorzunehmen. Die zweite Schreibaufgabe ist anspruchsvoller als die erste, da sie auf zwei Texten basiert, z. B. einer Rechnung und dem dazu gehörenden Überweisungsvordruck, und ein gewisses Maß an Lesekompetenz voraussetzt. Der Subtest Schreiben dauert 20 Minuten.

Das Modul Sprechen ist eine Einzelprüfung mit zwei PrüferInnen und enthält drei Aufgaben zum monologischen und dialogischen Sprechen. Ziel ist zu überprüfen, inwieweit die KandidatInnen in der Lage sind, über ihre eigene Person Auskunft zu geben und einen einfachen Alltagsdialog zu führen. Für die erste Aufgabe werden dem Prüfling Impulswörter vorgelegt (z. B. Namen, Ausbildung), anhand derer er über sich Auskunft gibt. In der zweiten Aufgabe soll der Kandidat aus drei Bildkarten, die eine Alltagssituation darstellen (z. B. im Bus), eine auswählen, die dargestellte Situation benennen und einen fiktiven Dialog führen. Die Rolle des Gesprächspartners wird dabei von einem Prüfer übernommen. Im Anschluss daran soll der Kandidat, ausgehend vom Bildimpuls, über die eigene Situation berichten. Das Modul Sprechen dauert ca. 10 Minuten.

### 3.4. Bestehensbedingungen

Der ÖIF-Test gilt als bestanden, wenn in jedem Modul die vorgegebene Mindestpunktzahl erreicht wurde. Dabei tragen die Module in unterschiedlichem Maße zum Bestehen bei: Lesen und Schreiben wiegen weniger als Hören und Sprechen. Wird in einem Modul die Mindestpunktzahl nicht erreicht, so muss der gesamte Test wiederholt werden. Der Zielgruppe und dem A2-Niveau entsprechend wird der formalen Richtigkeit ein weit geringerer Stellenwert zugewiesen als dem erfolgreichen kommunikativen Handeln.

## 4. Fazit

Beide Prüfungen sind vor dem Hintergrund ihrer sozialen und politischen Implikationen zu sehen. Sie wurden unter bestimmten gesetzlichen Vorgaben erstellt, finden in einem institutionellen Rahmen statt und tragen zur Regulierung von Zuwanderung bei. Damit gehören der dtz und der ÖIF-Test in die Kategorie der *high stakes tests*: Die Testergebnisse bilden die Basis für weitreichende Entscheidungen seitens der Ausländerbehörden. Werden in Deutschland keine Sprachkenntnisse auf B1-Niveau nachgewiesen, wird die Frist zur Erlangung der Niederlassungserlaubnis u. U. nicht verkürzt oder eine beantragte Staatsbürgerschaft wird (vorerst) nicht erteilt. Prüfungen dieser Tragweite haben hohe Anforderungen an die Testgütekriterien zu erfüllen: Sie sollen objektiv, reliabel, valide, aber auch ökonomisch, fair und transparent sein.

Die an der Entwicklung des dtz beteiligten Institutionen, beide Mitglied der Association of Language Testers in Europe (ALTE), bekennen sich explizit zum *ALTE Code of Practice*, in dem Qualitätsstandards von Sprachtests und deren Entwicklung niedergelegt sind, zu deren Einhaltung sich die ALTE-Mitglieder verpflichten.

Der Österreichische Integrationsfonds ist kein ALTE-Mitglied, jedoch ist auch hier ein Bemühen zu erkennen, Qualitätsstandards zu erfüllen. Der ÖIF-Test ist aufgrund der geschlossenen Aufgabenformate in den rezeptiven Teilen als ökonomisch und auswertungsobjektiv zu bezeichnen. Die stark gesteuerten Aufgaben in den produktiven Teilen, die ausführlichen Durchführungs- und Auswertungsanleitungen sowie die genauen Bewertungskriterien fördern die Auswertungsobjektivität in den produktiven Testteilen.

Für beide Prüfungen sind die Anforderungen klar definiert und dokumentiert. Modellsätze und die Rahmencurricula sind öffentlich zugänglich. Damit sind auch die Anforderungen hinsichtlich der Transparenz erfüllt (zur kritischen Diskussion vgl. Art. 10, Abschnitt 5).

## 5. Literatur in Auswahl

Buhlmann, Rosemarie, Karin Ende, Susan Kaufmann, Angela Kilimann und Helen Schmitz  
2007 *Rahmencurriculum für Integrationskurse Deutsch als Zweitsprache*. München: Goethe-Institut.

Bundesamt für Migration und Flüchtlinge, Goethe-Institut, Telc  
2009 *Deutschtest für Zuwanderer. Modellsatz Jugendintegrationskurs*. München: Goethe-Institut.

Bundesgesetz über die Niederlassung und den Aufenthalt in Österreich (NAG). Bundesgesetzblatt für die Republik Österreich, BGBl.I Nr. 100/2005 geändert durch BGBl.I Nr. 157/2005, BGBl.I Nr. 31/2006 und BGBl.I Nr. 99/2006.

Bundesgesetzblatt für die Republik Österreich. Jahrgang 2005. Teil II. Ausgegeben am 27. Dezember 2005. Nr. 449. Integrationsvereinbarungs-Verordnung – IV–V.

Bundesrepublik Deutschland

Gesetz zur Steuerung und Begrenzung der Zuwanderung und zur Regelung des Aufenthalts und der Integration von Unionsbürgern und Ausländern (Zuwanderergesetz) vom 30. Juli 2004. Bundesgesetzblatt Jahrgang 2004. Teil I Nr. 41, herausgegeben zu Bonn am 5. August 2004.

Europarat

2001 *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lehren, lernen, beurteilen*. Berlin/München: Langenscheidt.

Goethe-Institut/Telc GmbH

2009 *Deutschtest für Zuwanderer A2–B1. Prüfungsziele, Testbeschreibung*. München: Goethe-Institut.

Österreichischer Integrationsfonds

2006 *ÖIF Modelltest. Kommentierter Modelltest 001*. Wien.

Österreichischer Integrationsfonds

2006 *Testcurriculum. ÖIF-Test*. Wien.

Schweizerische Eidgenossenschaft

2009 SR 142.20 Bundesgesetz über die Ausländerinnen und Ausländer vom 16. Dezember 2005 (Stand 1. Januar 2009).

Link

Association of Language Testers in Europe (ALTE):

<http://www.alte.org/cop/index.php> (30. 11. 2009)

*Gabriele Kniffka, Freiburg (Deutschland)*

## 146. Sprachstandsdiagnosen

1. Sprachdiagnostik: Stellenwert in Wissenschaft und Praxis
2. Sprachdiagnostik im bildungspolitischen Kontext
3. Anforderungen an die Qualität von Sprachdiagnostik
4. Szenarien der Sprachdiagnostik und Anforderungen an die dabei eingesetzten Verfahren
5. Fazit
6. Literatur in Auswahl

### 1. Sprachdiagnostik: Stellenwert in Wissenschaft und Praxis

Sprachdiagnostische Tätigkeiten – also das Beobachten, Interpretieren und Bewerten von sprachlichen Fähigkeiten, Leistungen oder Entwicklungen – machen einen beträchtlichen Teil der Lehrarbeit aus. Sie werden begleitend zur alltäglichen Unterrichtsroutine

ausgeführt, und sie finden im Rahmen von expliziten, oft ritualisierten Momenten des Lehr-Lern-Prozesses statt: bei der Einteilung von Lernenden in Gruppen oder zum Abschluss eines Lehrgangs. Dabei ist man auf das Ermitteln und Interpretieren von Indizien (Indikatoren) angewiesen, um ein Urteil abzugeben: Sprachfähigkeit oder Leistung ist nicht messbar wie Größe, Gewicht oder Geschwindigkeit, sondern es muss von beobachteten Phänomenen auf eine ihnen zugrundeliegende allgemeinere Kompetenz geschlossen werden.

Die hohe Praxisbedeutung sprachdiagnostischer Tätigkeiten korrespondiert weder mit einer entsprechend regen Forschung noch mit einer gründlichen Qualifizierung des pädagogischen Personals für die diagnostischen Tätigkeiten. Entwicklungen in den 1970er und 1980er Jahren, hauptsächlich im Kontext der Einschätzung von Deutschkenntnissen zugewanderter Kinder und Jugendlicher, kamen weitgehend zum Erliegen, bis nach der Wende zum 21. Jahrhundert die Aufmerksamkeit für das Thema wieder wuchs – erneut im Kontext von Migration. In der Ausbildung zum Lehramt wird der Bereich der Sprachdiagnostik im Allgemeinen nach wie vor nur gestreift.

Wie kommt es zu dieser Diskrepanz zwischen praktischer Relevanz und wissenschaftlicher Aufmerksamkeit sowie angemessener Qualifikation für die praktischen Tätigkeiten? Zu den Gründen dafür gehört, dass der Bereich der Sprachdiagnostik, einschließlich der Entwicklung der dafür geeigneten Instrumente, traditionell in der Sprachdidaktik und Sprachlehrforschung eher randständig ist. Für den Bereich des fremdsprachlichen Lernens wird die Notwendigkeit, wissenschaftlichen Standards standhaltende Instrumente zur routinemäßigen Überprüfung von Lernprozessen und -Ergebnissen einzusetzen, vielfach nicht gesehen: Die sogenannte Lernzielkontrolle wird gleichsam als dem, was gelehrt wurde, immanent aufgefasst. So werden etwa die Kontroll-Tests zu Lehrbuchlektionen vielfach in den Handreichungen für Lehrende mitgeliefert. Für die Überprüfung fremdsprachlicher Leistungen, die nicht zur Routine des Unterrichtsalltags gehört, hat sich eine Tradition der (semi-)kommerziellen Entwicklung und Verbreitung von Tests etabliert. Ein Beispiel dafür ist die Produktion von Sprachtests und Zertifikaten im Anschluss an den „Gemeinsamen Europäischen Referenzrahmen für Sprachen“, wie sie u. a. von den Goethe-Instituten und ihren internationalen Pendanten betrieben werden. Die Expertinnen und Experten für die Entwicklung von Tests für Zertifizierungszwecke haben sich zu einer spezialisierten Vereinigung *ALTE – The Association of Language Testers in Europe* zusammengeschlossen, zu deren Zielen die Verbesserung und Sicherung von Standards für die Sprachtestentwicklung gehört, die für die Zertifizierung von Sprachkenntnissen verwendet werden (<http://www.alte.org/>).

Hierneben hat sich ein zweiter Bereich der spezialisierten Beschäftigung mit Sprachdiagnostik in der Spracherwerbsforschung etabliert. Hier werden in der Regel wissenschaftliche Ziele der Beobachtung von Sprachentwicklungsverläufen verfolgt. Die darauf bezogenen Instrumente werden üblicherweise nicht mit der Intention einer Praxisrelevanz entwickelt. Eine Ausnahme hiervon bildet die Ermittlung von Störungen in der kindlichen Sprachentwicklung. Hierfür liegen wissenschaftlich geprüfte, standardisierte Tests vor, die von dafür speziell qualifiziertem Personal in Praxiskontexten eingesetzt und ausgewertet werden können. Zu den bekanntesten Instrumenten für das Deutsche gehört der „Heidelberger Sprachentwicklungstest – HSET“ (vgl. Grimm und Schöler <sup>2</sup>1991: 129).

Mit Instrumenten wie diesen ist es möglich, eine auf spezifische, eng umgrenzte Aspekte konzentrierte Diagnostik durchzuführen – wie beispielsweise die Überprüfung des Vorliegens einer Entwicklungsstörung. Sie eignen sich nicht für den Einsatz im üblichen

Lehr-Lern-Kontext, in dem es nicht um das Überprüfen spezieller, klar eingegrenzter sprachlicher Phänomene geht, sondern um eine Diagnostik, die die komplexen Lehr-Lern-Prozesse von Sprache(n) unterstützt.

## 2. Sprachdiagnostik im bildungspolitischen Kontext

Die Entwicklung sprachdiagnostischer Verfahren, die für die Unterstützung im Lehr-Lern-Prozess geeignet sind, ist stärker im Bereich Deutsch als Zweitsprache verankert als im Bereich Deutsch als Fremdsprache. Dies ist aufgrund der Kontextbedingungen, in die die Bereiche eingebettet sind, auch leicht nachvollziehbar. Während für das fremdsprachliche Lernen vom Regelfall des kontrollierten sprachlichen Inputs durch Unterricht ausgegangen wird – auch wenn dies der Realität nur bedingt entspricht –, war bei der Konzeptionierung des zweitsprachlichen Lehrens und Lernens von vornherein die Vorstellung leitend, dass unterrichtlich geleiteter und außerunterrichtlicher, ungezielter sprachlicher Input zusammenkommen. Sprachstand und Sprachentwicklung am Lernenden selbst zu beobachten war naheliegend oder geboten, da es beim Unterrichten – mindestens den Konzeptionen nach – darum gehen musste, den ungesteuerten und den durch Unterricht gesteuerten Spracherwerb miteinander in Einklang zu bringen.

Trotz dieser Notwendigkeit in der Sache war es eine Zeit lang relativ still um die Entwicklung sprachdiagnostischer Verfahren im Kontext des Deutschen als Zweitsprache. Seit Anfang der 2000er Jahre aber ist ein wahrer Entwicklungseifer zu beobachten. Ausschlaggebend dafür war, dass die Forderung nach verstärkten Maßnahmen der Förderung des Deutschen als Zweitsprache im Anschluss an die Ergebnisse der PISA-Studien auf die Tagesordnung kam. Seit der Studie *PISA 2000* wurden zwei sprachbildungsrelevante Ergebnisse breit diskutiert: zum einen der hohe Stellenwert von Sprachfähigkeiten für das schulische Lernen generell; zum anderen der große Nachteil, den Schülerinnen und Schüler mit Migrationsbiographie, die das Deutsche als Zweitsprache lernen, in allen deutschsprachigen Staaten für ihre Bildungschancen besitzen. Zum Ausgleich dieses Nachteils wurden in Deutschland, Österreich und der deutschsprachigen Schweiz Maßnahmen zur Förderung der Deutschkenntnisse etabliert, und zwar im Schwerpunkt im vorschulischen Bildungsraum. Hier hinter steht das – allerdings unzutreffende – Verständnis, dass die Frühförderung des Deutschen als Zweitsprache die sprachlichen Bildungsvoraussetzungen so verbessere, dass eine weitere Förderung an der Bildungskarriere entlang überflüssig werde (zur Kritik dieses Verständnisses, vgl. Gogolin 2007).

Die Entwicklung sprachdiagnostischer Verfahren war eine Folge dieser Ereignisse: Es wurde bildungspolitisch die Notwendigkeit gesehen, Entscheidungshilfen dafür zu erhalten, ob ein Kind einen Anspruch auf Aufnahme in eine (Früh-)Fördermaßnahme erhalten sollte oder nicht. In einigen deutschen Bundesländern wurde der Anspruch in eine Verpflichtung uminterpretiert: Hier sollte mit Hilfe von Sprachstandsmessungen ermittelt werden, ob Kinder aufgrund unzureichender Deutschkenntnisse vor Schuleintritt – spätestens im fünften Lebensjahr – eine Maßnahme der Deutschförderung besuchen *müssen*. Vor diesem Hintergrund kam es vor allem in Deutschland zu sprachdiagnostischen Entwicklungen in bildungspolitischem Auftrag; ein sehr bekannt gewordenes Beispiel hierfür ist das im Land Nordrhein-Westfalen eingesetzte Verfahren „Delfin 4“, mit dem seit 2007 jährlich eine flächendeckende Untersuchung der Deutschkenntnisse der vierjäh-

rigen Landeskindern durchgeführt wird, um „Förderbedürftige“ zu ermitteln und den Fördermaßnahmen zuzuführen (vgl. Fried 2009).

Parallel zu diesen bildungspolitischen Auftragsarbeiten entfaltet sich verstärkt wissenschaftliche Aktivitäten, die darum bemüht sind, gesicherte Grundlagen für die Entwicklung von sprachdiagnostischen Verfahren zu gewinnen und Standards für ihre Qualität zu formulieren. Der Stand der Dinge mit Bezug auf das Deutsche wurde in einem umfassenden Gutachten dokumentiert, in dem neben Übersichten über vorhandene Verfahren und deren Einschätzung ein Referenzmodell für Phasen des Sprachausbaus vorgestellt wird, das der künftigen Entwicklung sprachdiagnostischer Verfahren zugrunde gelegt werden kann; allerdings steht eine empirische Absicherung dieses Modells noch aus (Ehlich 2005). Kennzeichnend für diese grundlegende Arbeit ebenso wie für andere jüngere Entwicklungen ist, dass der Blick nicht nur auf die Lernenden des Deutschen als Zweitsprache gerichtet wird, sondern auch auf jene, die einsprachig im Deutschen leben. Dahinter steht die Vermutung, dass eine zunehmende Zahl einsprachig lebender Kinder kein sprachanregungsreiches Milieu erlebt, so dass auch sie Schwierigkeiten haben, schulischen Sprachanforderungen gerecht zu werden. Allerdings zeigt der Beitrag von Reich (2005) im erwähnten Band klar auf, dass Verfahren unterschiedliche Anforderungen erfüllen müssen – je nachdem, ob sie bei einsprachigen oder bei zwei- bzw. mehrsprachigen Kindern eingesetzt werden.

### 3. Anforderungen an die Qualität von Sprachdiagnostik

Das Problem der Qualität von Verfahren der Sprachdiagnostik, das direkt mit ihrer Aussagekraft verbunden ist, wird in der Bildungspraxis oft unterschätzt. Hier ist es üblich, zwischen sogenannten formellen und informellen Verfahren der Diagnostik zu unterscheiden. Mit *informell* werden Vorgehensweisen bezeichnet, die meist mehr oder weniger spontan entwickelt und eingesetzt werden. Bei diesen ist weder daran gedacht noch praktisch möglich, eine Güteprüfung vorzunehmen. Ohne solche Prüfung aber ist die Aussagekraft der erzielten Ergebnisse ungesichert. Man erhält zwar ein Resultat, aber es gibt keine fundierten Anhaltspunkte für Antworten auf die Frage, was genau dieses Resultat eigentlich bedeutet – anders gesagt: Man erzeugt die Illusion einer relevanten Information über den Sprachstand einer lernenden Person.

Als *formelle* Verfahren werden solche aufgefasst, deren Güte in angemessener Weise kontrolliert wurde und die sich zum wiederholten Einsatz eignen. Sie werden deshalb auch als *standardisierte* Verfahren bezeichnet; damit ist angezeigt, dass sie in einer Form vorliegen, die ihren wiederholten (standardmäßigen) Einsatz erlaubt. Üblich ist es im pädagogischen Alltagsjargon, beide Typen von Verfahren auch als *Tests* zu etikettieren. Dies ist nicht angemessen. Mit dem Begriff Test verbinden sich Ansprüche an eine Güteprüfung, die nicht hintergangen werden dürfen.

Die wissenschaftliche Güteprüfung eines sprachdiagnostischen Verfahrens nach den Regeln der Kunst verbindet sich traditionell mit der Überprüfung von drei hierarchischen Anforderungsbereichen: Objektivität; Reliabilität und Validität (vgl. Artikel 143). Sprachdiagnostische Verfahren gehören zu der größeren Klasse der Verfahren zur Kompetenzfeststellung (vgl. Prenzel, Gogolin und Krüger 2008). Bei ihrer Konstruktion ist es erforderlich, vorweg genau zu bestimmen, an welcher Bezugsnorm das Ergebnis gemessen werden soll. Hier gibt es drei Möglichkeiten:

- eine soziale Norm; sie erlaubt den Vergleich des Resultats mit den Resultaten der Angehörigen einer definierten Gruppe (also z. B. einer Schulklasse);
- eine kriteriale Norm; hier wird der Vergleich mit einem allgemeingültigen Kriterium ermöglicht (also z. B. einer empirisch geprüften Auskunft darüber, über welche Sprachfähigkeiten Kinder einer bestimmten Altersgruppe normalerweise verfügen);
- eine ipsative Norm; hier geht es um den Vergleich von Resultaten einer Person, die entweder zu verschiedenen Zeitpunkten oder zu verschiedenen Kompetenzbereichen überprüft wird. Es kann also entweder um die Frage gehen, welche Entwicklungen bei dieser Person im gemessenen Leistungsbereich über eine Zeitspanne hinweg verzeichnet werden können, oder darum, ob eine Person in den Leistungsbereichen X und Y ähnlich oder unterschiedlich abschneidet. Sprachdiagnostisch relevante Fragen sind z. B., ob eine Person im rezeptiven Bereich, also beim Verstehen, andere Fähigkeiten zeigt als im Produktiven, also bei Sprechen oder Schreiben.

Nicht nur aus formalen, sondern auch aus inhaltlichen Gründen sind sog. informelle Verfahren in der Regel ungeeignet für fundierte Aussagen über Sprachstand oder Sprachentwicklung eines Menschen – mit Ausnahme der Situation, in der die Überprüfung genau identisch ist mit dem sprachlichen Input, der vorab gegeben wurde (also z. B. beim Abfragen von Vokabeln; dies würde aber auch bei großzügiger Auslegung üblicherweise nicht zur Sprachdiagnostik gerechnet).

Zu den Gründen dafür gehört die Komplexität des Geschehens bei der Diagnose sprachlicher Fähigkeiten. Wenn die Verfahren nicht nach den Regeln der Kunst entwickelt sind, besteht zum einen die Gefahr, dass etwas überprüft wird, das mit Blick auf die intendierte Zielsetzung der Prüfung irrelevant ist. Ein Beispiel hierfür ist das in der Praxis oft beobachtbare Vorgehen, dass das Auftauchen spezieller sprachlicher Formen – etwa Artikel oder Pronomen – kontrolliert wird, weil diese der Beobachtung leicht zugänglich sind. Dabei wird die Relevanz der Verwendung dieser Formen für Aussagen über Sprachkompetenz oft unterstellt, aber weder sichergestellt, dass sie für die Sprachentwicklung in einem bestimmten Erwerbsalter auch tatsächlich gegeben ist, noch hinterfragt, ob sie in bestimmten situativen Kontexten als Äußerungen überhaupt erwartet werden können. Zum anderen besteht bei sog. informellen Verfahren die Gefahr, dass die Aufgabe, sprachliche Phänomene zu identifizieren und richtig einzuschätzen, zu vielschichtig und daher nicht angemessen bewältigbar ist. Beispiele hierfür bieten die zahlreichen Beobachtungsverfahren, die in der Literatur beschrieben oder die zum Kauf angeboten werden. Hier ist das Risiko hoch, dass das komplexe Sprachgeschehen, das beobachtet werden soll, verzerrt wahrgenommen wird, weil die Beobachtungsaufgabe selbst die Kompetenz der Beobachtenden übersteigt. Schulz, Kersten und Kleissendorf (2009) diskutieren solche Verzerrungen am Beispiel des Beobachtungsverfahrens SISMIK (Ulich und Mayr 2003) – eines Verfahrens, das von Erzieherinnen in Kindertagesstätten eingesetzt werden soll. Zur Illustration dient eine Beispielaufgabe aus dem Fragebogen zu SISMIK:

„Im Hauptsatz steht das Verb an der richtigen Stelle, z. B. ‚der macht immer so‘, ‚... ich habe Durst‘, ‚... ich muss (auf die) Toilette‘  
 a) nie, b) selten, c) manchmal, d) häufig, e) das Kind bildet keine Sätze.“

Schulz, Kersten und Kleissendorf weisen darauf hin, dass es nicht nur eines gründlichen Wissens über den linguistischen Hintergrund solcher Äußerungen bedarf – etwa: eines

Bewusstseins für die Differenz der Verbstellung in Haupt- und Nebensätzen –, sondern auch einer gezielten Quantifizierung von Situationen, in denen ein beobachtetes Kind entsprechende Äußerungen potentiell tun kann; ohne diese ist die Aufforderung zur Einschätzung der Häufigkeit des Vorkommens unsinnig (vgl. Schulz, Kersten und Kleissendorf 2009: 130). Hinzu kommt, dass die dieser Beobachtungsaufgabe implizite kriteriale Norm fragwürdig ist. Die Beobachtung soll dem mündlichen Sprachverhalten von Kindern gelten. Im Mündlichen ist es aber (nicht nur bei Kindern) durchaus üblich, pragmatisch angemessene satzförmige Äußerungen ohne Verb zu formulieren oder das Verb in einer quasi infinitivischen Form zu verwenden. Schulz, Kersten und Kleissendorf (2009: 129) nennen als Beispiel die Verwendung des Imperativstils: Die Aufforderung „Reinkommen, Jacke ausziehen“ als Äußerung einer Lehrerin, die Kinder bewegen will, vom Pausenhof in die Klasse zu kommen, ist dem Kontext angemessen und in diesem Sinne korrekt. In sprachdiagnostischen Beobachtungen und unzulänglichen Verfahren aber wird nicht selten eine pragmatisch unangemessene Form als die gesuchte angenommen. Eine übliche Testfrage könnte etwa lauten: „Was macht das Kind?“ Als „korrekt“ vorgegeben wird eine Antwort des Typs: „(Das Kind / Es) zieht die Jacke aus“. Pragmatisch angemessen und im Mündlichen geläufig wäre aber auch hier die Antwort: „Jacke ausziehen“.

Es sind mithin zahlreiche Stolpersteine aufgestellt, was die Qualität von sprachdiagnostischen Verfahren und die Aussagekraft ihrer Ergebnisse anbelangt. In der Praxis benötigen Fachkräfte, die sie einsetzen, nicht nur fundiertes linguistisches Wissen, sondern darüber hinaus zumindest Grundkenntnisse der Psychometrie – also des Wissenschaftsbereichs, dessen Methodenrepertoire für die Güteprüfung relevant ist –, damit es nicht zu unbrauchbaren oder gar unsinnigen Resultaten der Sprachdiagnostik kommt.

#### 4. Szenarien der Sprachdiagnostik und Anforderungen an die dabei eingesetzten Verfahren

Der angemessene Einsatz von Verfahren der Sprachdiagnostik in der Praxis ist also von der Güte der Verfahren ebenso wie von der Qualifikation der sie einsetzenden Fachkräfte abhängig. Da das Thema in der Ausbildung bislang randständig ist, wird es vor allem auf breit angelegte Weiterbildung ankommen, um den Bedarf der Praxis zu decken. Entsprechende Aktivitäten sind vorerst vor allem im Bereich der Qualifizierung für den Elementarbereich zu finden; hier folgen die Angebote den bildungspolitischen Prioritätensetzungen (vgl. als Beispiele Jampert et al. 2009; Reich 2008). Eine Hilfestellung für die Praxis sind Szenarien des Einsatzes von Verfahren, in denen die jeweilige Eignung der sprachdiagnostischen Verfahren angezeigt wird.

##### Szenario 1: Screening

Unter Screening (Reihenuntersuchung) versteht man die Testung einer größeren Zahl von Probanden mit dem Ziel, bestimmte Risiken abzuschätzen. Am bekanntesten sind Screenings aus der Medizin: Hier werden sie eingesetzt, um Krankheiten frühzeitig zu diagnostizieren oder Krankheitsrisiken zu ermitteln.

Im Bereich der Sprachdiagnostik ist der Begriff Screening im Zusammenhang mit den bildungspolitischen Aktivitäten populär geworden, die dazu führen sollen, dass Kinder – insbesondere mit Migrationsbiographie – möglichst frühzeitig in Sprachfördermaßnahmen aufgenommen werden. In den meisten deutschen Bundesländern werden inzwischen bei Vier- oder spätestens Fünfjährigen solche Verfahren durchgeführt, teilweise flächendeckend, teilweise nur bei ausgewählten sogenannten Risikogruppen. Ein öffentlich breit und kontrovers diskutiertes Screening-Verfahren ist das erwähnte nordrhein-westfälische „Delfin 4“.

Kennzeichnend für Screenings ist, dass sie zu einer ja- oder nein-Antwort führen müssen: etwa zu der Aussage, ob Förderbedarf besteht oder nicht. Sie müssen ferner sehr ökonomisch, d. h. ohne großen Zeitaufwand durchführbar und auswertbar sein, da es in der Regel um den Einsatz bei größeren Gruppen geht. Voraussetzung für ihren Einsatz ist, dass sie standardisiert und normiert sind – sie müssen also eine den Regeln der Kunst gemäße Güteprüfung durchlaufen haben, und sie müssen die Norm, an der gemessen ein Risiko oder eine Abweichung identifiziert wird, eindeutig erklären.

Der Vorzug von Verfahren für Screenings – vorausgesetzt, sie erfüllen die Qualitätsanforderungen – besteht darin, rasch zu einem begründeten Urteil zu kommen: Die Werte eines Probanden liegen über oder unter einer festgelegten Grenze oder Norm. Dieser Vorzug ist zugleich ihr Nachteil mit Blick auf den Einsatz im Kontext der Sprachförderung, sei es im Unterricht von Gruppen oder im Einzelfall. Die mit screening-geeigneten Verfahren verbundenen Ansprüche erlauben es nicht, differenziert über die Sprachfähigkeiten der Getesteten Auskunft zu geben; dem stehen die Wünsche nach zeitsparendem Einsatz und Eindeutigkeit der Ergebnisformulierung entgegen. Eindeutigkeit wird dadurch erreicht, dass ein zusammenfassendes Maß formuliert wird (z. B. 30 von 100 Punkten = Förderbedarf). Darin aber gehen die differenzierten Auskünfte über die vom Getesteten aktualisierten sprachlichen Mittel verloren, an die bei der Förderung oder im Unterricht angeknüpft werden kann.

## Szenario 2: Diagnostik als Grundlage der Förder- oder Unterrichtsplanung für Deutsch als Zweitsprache

Bei der Sprachdiagnostik, die der Gestaltung von Förderung oder Unterricht zugrunde gelegt werden kann, ist es das Ziel, möglichst genaue Kenntnisse über die sprachlichen Voraussetzungen zu erlangen, die Lernende für das, was gelehrt werden soll, mitbringen. Es soll sich ein Bild von den spezifischen Eigenschaften ergeben, die das sprachliche Wissen und Können des Probanden in einem gegebenen Augenblick und im Hinblick auf ein Ziel mitbringen. Sprachdiagnostik mit diesem Zweck wird auch als pädagogische Diagnostik oder Förderdiagnostik bezeichnet. Im Kontext des Deutschen als Zweitsprache ist Bestandsaufnahme der sprachlichen Fähigkeiten am Anfang einer Maßnahme ein verbreiteter Fall, also z. B. bei Aufnahme in einen Kindergarten oder eine Schulklasse.

Je allgemeiner das Förder- oder Unterrichtsziel gehalten ist, desto breiter muss die Auskunft sein, die sich durch ein Verfahren gewinnen lässt. Sprachliches Wissen und Können umfasst verschiedene sprachliche Teilbereiche; sprachwissenschaftlich unterschieden werden die Bereiche der Lautung (Phonologie), der Bedeutung (Semantik), der Wortbildung (Morphologie) und der Bildung von Satzstrukturen (Syntax). Um eine Aus-

kunft über die Sprachfähigkeiten eines Getesteten zu erlangen, müsste aus jedem dieser Bereiche ein repräsentativer Ausschnitt abgeprüft werden.

Zudem ist nach dem situativen Charakter einer Sprachverwendung zu differenzieren. So ist es unangemessen, eine Norm des Schriftdeutschen anzulegen, wenn die Fähigkeit zur mündlichen Sprachproduktion überprüft werden soll. Außerdem ist zu berücksichtigen, dass die verschiedenen sprachlichen Teilbereiche nicht unbedingt füreinander repräsentativ sind. So kann etwa von der Sprachverstehensleistung eines Getesteten nicht auf seine Fähigkeit zur Sprachproduktion geschlossen werden und umgekehrt. Für schulische Erfolgchancen besonders relevant ist es, dass von der Beherrschung eines sprachlichen Registers nicht auf andere Register geschlossen werden kann. Das Verfügen über alltägliche Verständigungsfähigkeit ist z. B. nicht aussagekräftig im Hinblick auf die Frage, ob ein Lernender die spezifischen Redemittel der schulischen Lernbereiche oder Fächer, also „bildungssprachliche“ Fähigkeiten, besitzt (Gogolin 2006).

Die Komplexität des Sprachgeschehens und die linguistische Expertise, die für seine Einschätzung erforderlich sind, lassen es ratsam erscheinen, in ihrer Qualität geprüfte Verfahren, deren Reichweite genau benannt ist, bei der Sprachdiagnostik im Kontext von Förder- oder Unterrichtsplanung einzusetzen. Der Vorzug dieser Verfahren ist es, dass eine wohlfundierte Interpretation der Ergebnisse und Einschätzung ihrer Aussagekraft möglich ist. Als Nachteil kann angesehen werden, dass ihre Durchführung mit einigem Zeitaufwand verbunden ist. – Bedauerlicherweise ist die Auswahl an solchen Verfahren nicht groß: Zwar existiert eine große Menge an Angeboten, aber nur ein kleiner Teil der Instrumente hat angemessene Güteprüfungen durchlaufen (vgl. Übersicht über den Stand der Entwicklung bei Lengyel, Reich, Roth und Döll 2009). Bei der Auswahl eines Verfahrens zum Deutschen als Zweitsprache ist es ein Qualitätskriterium, ob bei der zugrunde gelegten Norm angemessen berücksichtigt ist, dass sich die Spracherwerbskonstellationen einsprachiger und zweisprachiger Getesteter grundlegend unterscheiden. Schulz, Kersten und Kleissendorf (2009: 133–134) machen auf den wichtigen Aspekt der Dauer des Sprachkontakts aufmerksam, der bei der Interpretation von Diagnoseergebnissen zu beachten ist. Für Lerner des Deutschen als Zweitsprache ist es nicht angemessen, eine an einsprachigen Lernenden entwickelte Altersnorm als Maßstab zu verwenden; vielmehr muss die Dauer der Kontaktzeit mit der Zweitsprache berücksichtigt werden.

### Szenario 3: Diagnostik zum Zwecke grundlegender Aussagen über die Sprachkompetenz Zweisprachiger

Ein weiterer Gesichtspunkt bei der Auswahl eines Verfahrens ist die Zweisprachigkeit selbst: Wenn es das Ziel einer Diagnose ist, grundlegende Aussagen über die *generelle* Sprachfähigkeit der Probanden zu treffen, ist es erforderlich, dass beide Sprachen in die Prüfung einbezogen werden (vgl. hierzu Reich 2005). Spracherwerbstheoretische Erkenntnisse besagen, dass die Sprachfähigkeiten Zwei- oder Mehrsprachiger nicht als voneinander getrennte Systeme aufzufassen sind. Vielmehr stehen sie miteinander in Kontakt, und es besteht die Wahrscheinlichkeit, dass sie einander wechselseitig beeinflussen. So muss ein zweisprachiger Mensch nur einmal das grundlegende Konzept der Zeit erwerben – also beispielsweise das Wissen darüber, dass man etwas als in der Vergangenheit, der Gegenwart oder der Zukunft geschehend ausdrücken kann. Bei weiteren Spra-

chen muss dann nicht mehr das Konzept als solches erworben werden, sondern es geht lediglich noch um die einzelsprachlich unterschiedlichen Weisen, das Konzept zum Ausdruck zu bringen (Tracy 2007). Um also Grundlegendes über die Sprachkompetenz Zweisprachiger zu erfahren, ist es notwendig, ihre Fähigkeiten in beiden Sprachen vergleichend zu beurteilen.

Auch hierbei sollten nach den Regeln der Kunst geprüfte Instrumente zum Einsatz kommen, bei deren Entwicklung berücksichtigt wurde, dass in beiden Sprachen einander funktionale Äquivalente überprüft werden. Bei der Güteprüfung der Verfahren ist zudem das Problem der angemessenen Normierung zu berücksichtigen. Für beide Sprachen gilt: Ein Verfahren, das an einer einsprachig in einsprachigem Kontext lebenden Stichprobe normiert worden ist, ist für den Einsatz bei Kindern, die dieselbe Sprache als Zweisprachige in einer mehrsprachigen Umgebung erlernen, ungeeignet.

Diese Qualitätsansprüche erfüllen im deutschsprachigen Raum bis dato nur einzelne Verfahren (vgl. Reich, Roth und Neumann 2007). Hier besteht eine besonders empfindliche Lücke in Forschung und Entwicklung.

## 5. Fazit

Sprachdiagnostische Tätigkeiten erfordern ein vertieftes Wissen über linguistische Zusammenhänge – etwa über Gesetzmäßigkeiten des Spracherwerbs, über die Gestalt der Sprache selbst, über die Vielfalt, die Funktionsweisen und die Bedeutung der unterschiedlichen Register für Sprachaneignung und Sprachgebrauch. Hierfür müssen Pädagoginnen und Pädagogen gründlich qualifiziert werden. Dies gilt auch für den Fall, dass in ihrer Qualität geprüfte Instrumente zur Verfügung stehen, die bei der sprachdiagnostischen Tätigkeit Hilfestellung leisten. Die Einschätzung der Angemessenheit für einen gegebenen Einsatzzweck, die adäquate Durchführung, Auswertung und Interpretation der Ergebnisse sind nur möglich, wenn hierfür profunde Kenntnisse mitgebracht werden. Die vielfach an sprachdiagnostische Verfahren geknüpfte Erwartung, dass sie ohne weiteres von nur oberflächlich qualifizierten Personen sinnvoll eingesetzt werden können, ist nicht erfüllbar – ebenso wenig wie die Erwartung, mittels ungeprüfter (informeller) Verfahren, deren Einsatz wenig Zeit und Aufwand kostet, verlässliche sprachdiagnostische Ergebnisse zu erzielen. Es sind also erhebliche Investitionen zu leisten, bis davon gesprochen werden kann, dass im sprachdiagnostischen Feld ein zufriedenstellender Stand erreicht ist.

## 6. Literatur in Auswahl

Ehlich, Konrad

- 2005 *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*. Bonn: Bundesministerium für Bildung und Forschung.

Fried, Lilian

- 2009 *Sprachkompetenzmodell Delfin 4. Testmanual*. Dortmund: Technische Universität; auch unter <http://www.delfin4.fb12.uni-dortmund.de/>; Juli 2009.

- Gogolin, Ingrid  
 2006 Bilingualität und die Bildungssprache der Schule. In: Paul Mecheril und Thomas Quehl (Hg.), *Die Macht der Sprachen. Englische Perspektiven auf die mehrsprachige Schule*, 79–85. Münster etc.: Waxmann.
- Gogolin, Ingrid  
 2007 *Institutionelle Übergänge als Schlüsselsituationen für mehrsprachige Kinder. Expertise für das Deutsche Jugendinstitut*. Hamburg: Universität Hamburg.
- Grimm, Hannelore und Hermann Schöler  
 2009 *Heidelberger Sprachentwicklungstest HSET*. Göttingen: Hogrefe; auch unter <http://www.testzentrale.de/?id=1&mod=detail>; Juli 2009.
- Jampert, Karin, Anne Zehnbauer, Petra Best, Andrea Sens, Kerstin Leuckefeld und Mechthild Laier (Hg.)  
 2009 *Kinder-Sprache stärken! Sprachliche Förderung in der Kita: das Praxismaterial*. Weimar/Berlin: Verlag das Netz.
- Lengyel, Drorit, Hans H. Reich, Hans-Joachim Roth und Marion Döll (Hg.)  
 2009 *Von der Sprachdiagnose zur Sprachförderung*. Münster etc.: Waxmann.
- Prenzel, Manfred, Ingrid Gogolin und Heinz-Hermann Krüger (Hg.)  
 2008 *Kompetenzdiagnostik*. (Zeitschrift für Erziehungswissenschaft. Sonderheft. 08/2007). Wiesbaden: VS Verlag.
- Reich, Hans H.  
 2005 Forschungsstand und Desideratenaufweis zu Migrationslinguistik und Migrationspädagogik für die Zwecke des „Anforderungsrahmens“. In: BMBF (Hg.), *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*, 121–169. (Reihe Bildungsreform Band 11). Berlin: Bundesministerium für Bildung und Forschung.
- Reich, Hans H.  
 2008 *Sprachförderung im Kindergarten. Grundlagen, Konzepte und Materialien*. Weimar/Berlin: Verlag das Netz.
- Reich, Hans H., Hans-Joachim Roth und Ursula Neumann (Hg.)  
 2007 *Sprachdiagnostik im Lernprozess. Verfahren zur Analyse von Sprachständen im Kontext von Zweisprachigkeit*. Münster etc.: Waxmann.
- Schulz, Petra, Anja Kersten und Barbara Kleissendorf  
 2009 Zwischen Spracherwerbsforschung und Bildungspolitik: Sprachdiagnostik in der frühen Kindheit. *Zeitschrift für Soziologie der Erziehung und Sozialisation (ZSE)* 29(2): 122–140.
- Tracy, Rosemarie  
 2007 *Wie Kinder Sprachen lernen. Und wie wir sie dabei unterstützen können*. Tübingen etc.: Francke.
- Ulich, Michaela und Toni Mayr  
 2003 *SISMIK. Sprachverhalten und Interesse an Sprache bei Migrantenkindern in Kindertageseinrichtungen*. Freiburg: Herder.

*Ingrid Gogolin, Hamburg (Deutschland)*

## 147. Portfolios und informelle Leistungsdiagnosen

1. Einleitung: Lernen und Beurteilung mit Portfolios
2. Das Europäische Sprachenportfolio als Konzept
3. Das Europäische Sprachenportfolio und informelle Leistungsdiagnosen
4. Das Europäische Sprachenportfolio und eine neue Beurteilungskultur
5. Schluss: Zukunftsaussichten
6. Literatur in Auswahl

### 1. Einleitung: Lernen und Beurteilung mit Portfolios

„Ein Portfolio bezeichnet eine sinnvolle Sammlung von Arbeiten, mit denen Engagement, Leistungen, Erkenntnisse und Entwicklungen in einem oder mehreren Lernbereichen transparent gemacht werden“ (Müller 2005: 9–10). Wie diese Definition impliziert, kann ein Portfolio in dieser allgemeinen pädagogischen Bedeutung dreierlei Funktionen erfüllen. Erstens kann es dazu dienen, den Lernprozess zu strukturieren; so können Lernende z. B. ein Portfolio verwenden, um die Materialien, welche sie für ein bestimmtes Projekt zusammentragen, zu sammeln und zu gestalten. Zweitens kann es die Grundlage für eine Beurteilung liefern, die das Lernen als Produkt ebenso wie als Prozess in den Mittelpunkt stellt; diese Funktion setzt voraus, dass Schüler im Voraus über die Beurteilungskriterien informiert sind. Drittens können Portfolios dazu verwendet werden, Lern-erfolge zu präsentieren.

Die Konzepte des Portfolio-Lernens sowie der Portfolio-Beurteilung erlangten zum ersten Mal in den Vereinigten Staaten Bedeutung, wo sie in Opposition zur Beurteilung durch standardisierte Tests entwickelt wurden. Die Befürworter der Portfolios argumentierten, dass standardisierte Tests wenig dazu beitragen, den Lernvorgang zu unterstützen, sondern im Gegenteil die Annahme förderten, dass Lehren/Lernen und Testen grundverschiedene Aktivitäten seien. Portfolios dagegen ermöglichen es, Lernen und Beurteilung in eine positive Interaktion zueinander zu setzen: Die Beurteilung *des* Lernens kann auch eine Beurteilung *für das* Lernen sein. Sie verkörpern daher eine Philosophie, die mit der Diskussion „formativer Bewertung“, die im Großbritannien der 1990er Jahre (Black und Wiliam 1998) lanciert wurde, sowie mit dem in vielen europäischen Ländern aktuellen Interesse am dialogischen Lernen (z. B. Winter 2004; Alexander 2006; Ruf, Keller und Winter 2008) eng verwandt ist.

In der Sprachpädagogik wurde der Gebrauch von Portfolios, und insbesondere von Lerntagebüchern, in enge Verbindung mit der Förderung von reflexivem Lernen und Lernerautonomie gebracht (z. B. Dam 1995). Wie in anderen pädagogischen Bereichen blieb diese Herangehensweise bisher das Interesse und die Leistung einer Minderheit. Im Lauf der ersten zehn Jahre des 21. Jahrhunderts jedoch hat das Europäische Sprachenportfolio (ESP) des Europarates dem Portfolio-Lernen und der Portfolio-Beurteilung zu einer weiteren Verbreitung verholfen, obwohl seine Wirkung auf den Unterrichtsraum nach wie vor zu wünschen übrig lässt. Dieser Artikel beschäftigt sich aus drei Gründen mit dem ESP: Es ist ausführlicher durchdacht als andere Portfolio-Konzepte und ist auf einzigartige Weise auf den Spracherwerb zugeschnitten; als pädagogisches Werkzeug unterstützt es Formen des reflexiven Lernens auf innovative Art; und es liefert als alter-

natives Beurteilungsinstrument triftige Gründe für die Entwicklung einer neuen Beurteilungskultur, zumindest im Sprachunterricht. Der Artikel ist dementsprechend in drei Hauptabschnitte unterteilt: das ESP als Konzept, das ESP als pädagogisches Werkzeug, und das ESP als Schlüsselkomponente für eine neue Sichtweise auf die Beurteilung von Zweit-/Fremdsprachenkompetenz.

## 2. Das Europäische Sprachenportfolio als Konzept

Das Europäische Sprachenportfolio hat drei obligatorische Komponenten: einen Sprachenpass, eine Sprachenbiographie und ein Dossier. Der Sprachenpass fasst die sprachliche Identität seines Inhabers oder seiner Inhaberin zu bestimmten Zeitpunkten zusammen, indem er erlernte Zweit-/Fremdsprachen, formale Sprachqualifikationen, für den Zweit-/Fremdsprachengebrauch wichtige Erfahrungen und die Beurteilung des Inhabers/der Inhaberin seines/ihrer aktuellen Leistungsstandes in Zweit-/Fremdsprachen kurz dokumentiert. Die Sprachenbiographie dient dazu, Sprachlernziele zu setzen, Fortschritte zu überwachen, Ergebnisse zu beurteilen und über verschiedene Dimensionen des Spracherwerbs und Sprachgebrauchs zu reflektieren, darunter der Aspekt der Interkulturalität und des Lernen-Lernens. Das Dossier kann entsprechend der Bedürfnisse und Prioritäten spezieller Kontexte gestaltet werden; es enthält immer praktisches Beweismaterial, um die Selbstbeurteilung, welche in der Sprachenbiographie und im Sprachenpass dokumentiert ist, zu stützen.

Das ESP ist dahingehend konzipiert, eine protokollierende und eine pädagogische Funktion zu haben. In seiner protokollierenden Funktion ergänzt es die Zertifikate und Diplome, die auf der Basis formaler Prüfungen verliehen werden, indem es zusätzliche Informationen über die Sprachlernerfahrung des Inhabers/der Inhaberin sowie konkrete Nachweise von Zweit-/Fremdsprachenkompetenz und -leistungen liefert. Darüber hinaus erlaubt es dem Inhaber/der Inhaberin, Sprachlernerfahrungen sowohl außerhalb als auch innerhalb des formalen Bildungssystems zu dokumentieren. In seiner pädagogischen Funktion soll das ESP Mehrsprachigkeit fördern, kulturelles Bewusstsein anheben, den Sprachlernprozess für den Inhaber/die Inhaberin transparenter machen und das Entstehen von Lernerautonomie anregen. Die protokollierende und pädagogische Funktion stehen offensichtlich in engem Zusammenhang und gehen im kontinuierlichen Prozess der Selbstbeurteilung, der für einen effektiven ESP-Gebrauch grundlegend ist, ineinander über.

Der Anstoß zur Entwicklung des ESP ging vom Europaratssymposium in Rüşchlikon („Transparency and coherence in language learning in Europe“, Council for Cultural Cooperation 1992) im Jahr 1991 aus. Dort wurde die Empfehlung ausgesprochen, dass der Rat für kulturelle Kooperation des Europarats die Entwicklung eines Gemeinsamen europäischen Referenzrahmens für Sprachen (GER) fördern und einen Arbeitskreis ins Leben rufen sollte, um mögliche Formen und Funktionen eines Europäischen Sprachenportfolios zu erwägen. Selbstbeurteilung im Rahmen des ESP basiert auf dem GER, welcher Zweit-/Fremdsprachenkompetenz in Bezug auf fünf kommunikative Aktivitäten (HÖREN, LESEN, AN GESPRÄCHEN TEILNEHMEN, ZUSAMMENHÄNGEND SPRECHEN, SCHREIBEN) auf sechs Niveaus beschreibt (A1, A2, B1, B2, C1, C2). Checklisten bestehend aus „Ich kann“-Deskriptoren, die nach Aktivität und Kompetenzniveau angeordnet sind, lenken

Zielsetzung, Überwachung und formative Selbstbeurteilung in der Sprachenbiographie, während das sogenannte Raster zur Selbstbeurteilung des GER (Europarat 2001: 36) das Maßsystem darstellt, gegen das der Lernende in periodischen Abständen eine summative Selbstbeurteilung ausführt.

Das Projekt „Language learning for European citizenship“ brachte zwei Entwürfe für den GER und eine Reihe von Vorschlägen bezüglich des ESP (Council for Cultural Cooperation 1997a) hervor. Der Abschlussbericht des Projekts empfahl, dass der GER in die Probephase gehen und das ESP weitergehend entwickelt werden sollte (Council for Cultural Cooperation 1997b: 73). Zwischen 1998 und 2000 wurden ESP-Pilotprojekte in 15 Europarats-Mitgliedstaaten und von drei internationalen Nichtregierungsorganisationen durchgeführt (Schärer 2000); die Prinzipien und Richtlinien, die das ESP und seinen Zweck bestimmen, wurden weiter ausgearbeitet (Council of Europe 2006). Im Jahr 2000 begründete der Europarat ein Validations-Komitee, dessen Rolle darin besteht, ESPs in Empfang zu nehmen und zu entscheiden, ob diese den Prinzipien und Richtlinien folgen; bis zum Sommer 2010 waren 104 ESPs für gültig erklärt worden.

### 3. Das Europäische Sprachenportfolio und informelle Leistungsdiagnosen

Obwohl grundlegend für den effizienten Gebrauch des ESP, war Selbstbeurteilung während der Pilotphase dennoch Anlass zur Sorge und Kontroverse (Schärer 2000: 10). Einige LehrerInnen bezweifelten, dass ihre Schüler in der Lage wären, ihre eigene Kompetenz akkurat einzuschätzen, andere vermuteten, dass Lernende versucht sein könnten, ihre Errungenschaften überzubewerten. Es gibt drei Möglichkeiten, auf diese Befürchtungen einzugehen. Erstens, was die Fähigkeit der Lernenden zur Selbsteinschätzung betrifft, so besteht die Grundlage für die Selbstbeurteilung im ESP in „Ich kann“-Aussagen, die kommunikatives Verhalten beschreiben – z. B. *Ich kann einfache Wendungen und Sätze gebrauchen, um Leute, die ich kenne, zu beschreiben und um zu beschreiben, wo ich wohne* (A1 ZUSAMMENHÄNGEND SPRECHEN; Europarat 2001: 36). Im Großen und Ganzen ist es wahrscheinlich, dass Lernende wissen, ob sie in der Lage sind, die Aufgaben zu bewältigen, die von solchen Deskriptoren spezifiziert werden. Zweitens sollte es, im Hinblick auf die Ehrlichkeit der Lernenden, ein Leichtes sein, erhebliche Diskrepanzen zwischen der Selbstbeurteilung einerseits und ihren (im Sprachenpass dokumentierten) Prüfungsnoten oder dem in ihrem Dossier erbrachten Kompetenznachweis andererseits festzustellen. Drittens neigten die Befürchtungen der LehrerInnen hinsichtlich Selbstbeurteilung dazu, sich auf deren summative Funktion zu konzentrieren, doch ist dies nur ein Teil des Gesamtbildes. Es ist wahr, dass jedes Mal, wenn Lernende ihren Sprachenpass aktualisieren, ihre Selbstbeurteilung nahezu die gleiche Funktion erfüllt wie wenn ein Lehrer zu Ende einer Lernphase eine Prüfung abhält. Doch die Selbstbeurteilung, welche ins Spiel kommt, wenn ein ESP-Inhaber die Sprachenbiographie verwendet, um Lernziele zu identifizieren oder Fortschritte zu überwachen, ist formativ, nicht summativ: Sie unterstützt und leitet das Lernen, indem es stattfindet – sie bestätigt Leistungen, identifiziert aber auch Schwächen sowie Gebiete, die verstärkter Arbeit oder größeren Einsatzes bedürfen. Natürlich rekurriert Selbstbeurteilung, unabhängig davon, ob ihre Funktion summativ oder formativ ist, immer auf denselben Komplex von Wissen, Selbst-

erkenntnis und Fähigkeiten. Das bedeutet, dass die ESP-Inhaber umso geübter in summativer Selbstbeurteilung sein sollten, je mehr sie sich auf formative Selbstbeurteilung einlassen. Dieses Argument impliziert, dass ein effizienter Gebrauch des ESP auf reflexivem Lehren und Lernen beruhen muss, welches von Verfechtern von Portfolio-Lernen und -Beurteilung (Müller 2005), formativer Bewertung (Black und Wiliam 1998) and dialogischem Lernen (Winter 2004) befürwortet wird.

Die Bildungsprojekte des Europarats im Allgemeinen und seine Fremdsprachenprojekte im Besonderen haben von jeher die Bedeutung der Lernerautonomie hervorgehoben. Lernende können dann als autonom bezeichnet werden, wenn sie explizit Verantwortung für ihr eigenes Lernen auf sich nehmen und diese Verantwortung in dem kontinuierlichen Bestreben praktizieren, zu verstehen, was, warum, wie und mit welchem Erfolg sie lernen (siehe z. B. Holec 1979; Little 1991). Wie diese Arbeitsdefinition nahelegt, hängt Lernerautonomie wesentlich von Reflexion und Selbstbeurteilung ab. Wir machen Lernende nicht auf einen Schlag autonom, indem wir ihnen mitteilen, dass sie für ihr Lernen verantwortlich sind; sie werden allmählich autonom, indem sie die reflexiven Fähigkeiten zur Planung, Überwachung und Bewertung ihres Lernens entwickeln und ausüben. Dies nämlich ist der Kern reflexiven Lehrens und Lernens.

Eine der besten Darstellungen reflexiven Sprachunterrichts/-lernens findet sich bei Leni Dam (1995). Sie beschreibt, wie sie dänische Englischlernende der unteren Sekundarstufe in zunehmend anspruchsvolle Reflexion einband, indem sie wiederholt die folgenden fünf Fragen stellte: Was lernen wir? Warum lernen wir es? Wie lernen wir? Wie erfolgreich ist unser Lernen? Was werden wir als nächstes lernen? Diese fünf Fragen treiben einen Lernkreislauf an, in dem Planung von Durchführung gefolgt ist, nach welcher Beurteilung zu weiterer Planung führt. Bemerkenswert ist, dass Reflexion in jeder dieser Phasen eine Rolle spielt; ebenfalls bemerkenswert ist, dass formative Selbstbeurteilung ein wesentlicher Bestandteil sowohl der Planungs- und Durchführungsphase als auch der Beurteilungsphase ist. Denn um effizient zu sein, muss die Planung der Lernenden realistisch sein, was bedeutet, dass sie ihre Absichten ständig gegen entwickelte Kompetenzen messen müssen; erfolgreiche Durchführung wiederum ist abhängig von effizienter Überwachung, welche als „on-line“ Selbstbeurteilung beschrieben werden könnte.

Der Schwerpunkt des ESP auf Selbstbeurteilung und den reflexiven Prozessen, die sich daran anschließen, fordert die Frage heraus: Wird das Lernen, wie man Sprachen lernt, dem (eigentlichen) Sprachlernen im Weg stehen? Schließlich ist die Zeit, die Lernende damit verbringen, über ihre Fertigkeiten in ihren Zielsprache(n) zu reflektieren, Zeit, die sie nicht darauf verwenden können, weitere Fertigkeiten zu erlernen. Doch dieser Einwand ist nur dann gültig, wenn die Prozesse der Reflexion und Selbstbeurteilung in der Muttersprache ausgeführt werden. Wenn sie andererseits in der Zielsprache ausgeführt werden (wie bei Dam 1995), werden sie zu einem wesentlichen und äußerst wichtigen Bestandteil des Sprachlernens; denn nur wenn Lernende allmählich die Fähigkeit entwickeln, sich mit Reflexion und Bewertung in ihren Zweit-/Fremdsprachen zu befassen, können wir von ihnen erwarten, dass sie zu einem der gehobenen Kompetenzniveaus fortschreiten, welche eine ebensolche Fähigkeit voraussetzen. Dieses Argument hat eine klare Implikation für die Gestaltung des ESP: Um den Gebrauch der Zielsprache für Reflexion und Selbstbeurteilung zu ermöglichen, sollten Selbstbeurteilungs-Checklisten in jeder Sprache, die ein ESP-Inhaber lernt, verfügbar sein.

An diesem Punkt ist es angemessen, zwei Beispiele zu geben. Das erste betrifft die Entwicklung der Fähigkeit zur Produktion gesprochener Sprache (ZUSAMMENHÄNGEND

SPRECHEN) auf dem ersten Niveau des GER. Ich habe den zusammenfassenden Deskriptor im Raster zur Selbstbeurteilung bereits zitiert: *Ich kann einfache Wendungen und Sätze gebrauchen, um Leute, die ich kenne, zu beschreiben und um zu beschreiben, wo ich wohne.* Im irischen ESP für die Sekundarstufe I und Sekundarstufe II (Authentik 2001) weitet die Checkliste für A1 ZUSAMMENHÄNGEND SPRECHEN diese Zusammenfassung auf drei Deskriptoren aus:

- *Ich kann einfache Angaben zu meiner Person machen (z. B. Alter, Adresse, Familie, Hobbies)*
- *Ich kann in einfachen Worten und Sätzen beschreiben, wo ich wohne*
- *Ich kann in einfachen Worten und Sätzen Personen beschreiben, die ich kenne*

Jeder dieser Deskriptoren definiert eine einfache kommunikative Aufgabe, und jeder kann verwendet werden, um eine Lernphase zu gestalten. So kann zum Beispiel die Aufgabe, persönliche Angaben zu machen, als Lernziel in der Sprachenbiographie identifiziert werden; Lernende können in ihrem Dossier Vokabular und einfache Satzstrukturen notieren, die sie benötigen, um diese Aufgabe auszuführen; sie können sich in der Aufgabe individuell und in Kleingruppen üben; und sie können ihre Fähigkeit, die Aufgabe durchzuführen, in der Checkliste festhalten, sobald der Lehrer oder einer ihrer Klassenkameraden bestätigen, dass sie dazu in der Lage sind – nach dieser Herangehensweise besteht Selbstbeurteilung darin, Behauptungen aufzustellen, deren Gültigkeit von anderen bezeugt werden kann. In der Beurteilungsphase können Lernende darüber hinaus in ihrer Sprachenbiographie alles das festhalten, was sie über das Sprachenlernen als solches gelernt haben.

Das zweite Beispiel bezieht sich auf Niveau B2 und ist dementsprechend komplexer. Auf diesem Niveau definiert das Raster zur Selbstbeurteilung die Schreibfähigkeit der Lernenden wie folgt: *Ich kann über eine Vielzahl von Themen, die mich interessieren, klare und detaillierte Texte schreiben. Ich kann in einem Aufsatz oder Bericht Informationen wiedergeben oder Argumente und Gegenargumente für oder gegen einen bestimmten Standpunkt darlegen. Ich kann Briefe schreiben und darin die persönliche Bedeutung von Ereignissen und Erfahrungen deutlich machen* (Europarat 2001: 36). Die korrespondierende Checkliste im irischen ESP für Sekundarstufe I und Sekundarstufe II (Authentik 2001) schließt die folgenden Deskriptoren mit ein:

- *Ich kann zu sehr vielen persönlichen, kulturellen, interkulturellen und sozialen Themen klare, detaillierte Beschreibungen abgeben*
- *Ich kann einen Standpunkt zu einem aktuellen Thema erläutern und die Vor- und Nachteile verschiedener Standpunkte erörtern*
- *Ich kann eine Argumentationskette klar aufbauen, die Ideen logisch verknüpfen und die verschiedenen Punkte mit entsprechenden Beispielen untermauern*

Hier haben wir es nicht, wie im ersten Beispiel, mit einer Liste klar abgegrenzter Aufgaben zu tun, sondern mit allgemeineren Funktionen des geschriebenen Textes: detaillierte Beschreibungen abgeben, Standpunkte erläutern und erörtern, Argumentationsketten klar aufbauen, Ideen logisch verknüpfen, Punkte mit Beispielen untermauern. Und die viel größere Komplexität der kommunikativen Fähigkeit der Lernenden auf diesem Niveau impliziert eine weitaus größere Komplexität der Lernaktivität. Im Rahmen einer Portfolio-Herangehensweise wird die Entwicklung dieser Fähigkeit mit großer Wahrscheinlichkeit durch eine Folge von projektbasierten Lernkreisläufen verfolgt. Der erste

Schritt in jedem Keislauf besteht darin, mithilfe der Checkliste Schwerpunkt und Reichweite des Projekts zu identifizieren. Wird das Thema persönlich, kulturell, interkulturell oder sozial sein? Welche Art von Texten wird verwendet werden, um thematischen und linguistischen Input zu liefern? Werden sie deskriptiv oder argumentativ sein, oder beides? Was werden die strukturellen Eigenschaften der Texte sein, die die Lernenden produzieren werden? Wie genau werden detaillierte Beschreibungen artikuliert werden? Welche Hilfsmittel hinsichtlich Lexik und Syntax benötigt man, um die Vor- und Nachteile verschiedener Standpunkte zu erörtern oder eine Argumentationskette klar aufzubauen? Diese und viele andere Fragen, die die Deskriptoren implizieren, werden zu Beginn jeden Projektkreislaufs gestellt und von Lehrer und Lernenden interaktiv ergründet, und zwar in der Zielsprache. Als Teil dieses Prozesses können spezifische Kriterien etabliert werden, um die Effizienz einer Beschreibung oder Argumentationskette zu beurteilen. Daraufhin werden die Projekte entwickelt, evtl. im Rahmen von (Klein-)Gruppenarbeit. Reflexivität – Selbstbeurteilung, aber auch Peer-Beurteilung, welche beide als mentale Gewohnheiten kultiviert werden – wird dadurch gewährleistet, dass die Lernenden zu keinem Zeitpunkt die aus der Planungsphase hervorgegangenen Kriterien und die Notwendigkeit, mit der das Projekt diesen Kriterien entsprechen muss, aus den Augen verlieren. Am Ende des Kreislaufs, wenn die Projekte dem Rest der Klasse präsentiert werden, kann es eine Phase der Peer- und Selbstbeurteilung geben, die sich auf die kommunikative Effizienz der produzierten Texte konzentriert, aber auch auf ihre formale Genauigkeit. Angenommen, dass der Ausgangspunkt der Lernenden der B1 Kenntnisstand im Schreiben war – *Ich kann über Themen, die mir vertraut sind oder mich persönlich interessieren, einfache, zusammenhängende Texte schreiben. Ich kann persönliche Briefe schreiben und darin von Erfahrungen und Eindrücken berichten* (Europarat 2001: 36) –, so werden sie eine Reihe von Projektzyklen durchlaufen müssen, ehe sie zuversichtlich und plausibel von sich behaupten können, dass sie die allgemeinen Kriterien textueller Kompetenz, wie sie die Checklisten-Deskriptoren definieren, erfüllen.

#### 4. Das Europäische Sprachenportfolio und eine neue Beurteilungskultur

Portfolio-Beurteilung wurde ursprünglich als Gegenmaßnahme zu standardisierten Tests entwickelt, deren Form und Inhalt Lehre und Lernen nur allzu leicht einschränken. Jedoch gab es stets nicht nur pädagogische, sondern auch politische Einwände gegen standardisierte Tests. Die kritische Bewegung im Sprachtesten, insbesondere in der englischsprachigen Welt (z. B. Shohamy 2001; McNamara und Roever 2006), fand ihre Inspiration teilweise in Foucaults Auffassung (1976) von Prüfungen als Instrumenten der Kontrolle und Bestrafung. Aus dieser Perspektive waren auch Schritte in Richtung kommunikativer Sprachtests nicht frei von Verdacht. Doch entgegen dieser negativen Tendenz bietet das ESP einen Weg, eine neue Komplementarität zwischen Lehren/Lernen und Beurteilung zu entwickeln und die Entstehung einer neuen Beurteilungskultur anzuregen.

Zu Beginn des GER ist zu lesen: „Der *Gemeinsame Referenzrahmen* stellt eine gemeinsame Basis dar für die Entwicklung von zielsprachlichen Lehrplänen, curricularen Richtlinien, Prüfungen, Lehrwerken usw. in ganz Europa“ (Europarat 2001: 14). Die Anord-

nung der Elemente im Untertitel – „lernen, lehren, beurteilen“ – ist kein Zufall und lenkt die Aufmerksamkeit auf den vielleicht innovativsten Charakterzug des GER. Sein handlungsorientierter Ansatz beschreibt Kommunikation im Hinblick darauf, was der Lernende/Benutzer mit Sprache tun kann; jeder „Ich kann“-Deskriptor kann drei sich gegenseitig beeinflussende Funktionen erfüllen. Er kann dazu verwendet werden, ein Lernziel zu spezifizieren, einen Schwerpunkt für die Lernaktivität zu identifizieren und Kriterien zur Beurteilung von Lernergebnissen zu generieren. Die Selbstbeurteilungsfunktion des ESP ist abhängig von der Tatsache, dass Lernende in der Lage sind, akkurate Urteile über ihre Verhaltensfähigkeiten zu fällen, insbesondere wenn diese Urteile von dem reflexivem Diskurs unterstützt werden, welcher das Portfolio-Lernen prägt. Wenn Struktur und Inhalt von Tests und Prüfungen das gleiche handlungsorientierte Verständnis des Sprachgebrauchs widerspiegeln, dann sollten sie das Portfolio-Lernen unterstützen und die Selbst- und Peer-Beurteilung, die für den ESP-Gebrauch von so zentraler Bedeutung ist, ergänzen.

Verwenden Lernende das ESP, um ihr Sprachlernen zu planen, zu überwachen und zu überprüfen, so bringen sie ihre Subjektivität in explizite Interaktion mit objektiven (empirisch gewonnenen) Beschreibungen von Sprachkompetenz. Lernende konstruieren nach und nach ein Bild ihrer sich erweiternden sprachlichen und kulturellen Identität, das Zeugnis ihrer Errungenschaften wie auch Ansporn zu weiterführender Reflexion ist. Sie entwickeln ein gesteigertes Verständnis ihrer selbst, insbesondere was ihre Fähigkeiten in Zweit- und Fremdsprachen betrifft, und werden zunehmend sensibler mit Blick auf die sprachlichen und kommunikativen Komplexitäten, die scheinbar einfachen Deskriptoren zugrunde liegen. Doch Lernende verstehen ebenfalls den Sinn von Sprachtests, die auf der Grundlage des handlungsorientierten Ansatzes des GER erdacht und somit für ihre eigenen Lernziele relevant sind.

## 5. Schluss: Zukunftsaussichten

Die Art der Beurteilungskultur, die ich soeben skizziert habe, in welcher Selbst- und Peer-Beurteilung einerseits und formale Tests und Prüfungen andererseits einander ergänzen, liegt in der Zukunft. Die Wirkung des GER war bisher eher partiell als holistisch. Seine Kompetenzniveaus wurden von Institutionen, die Sprachtests durchführen, europaweit (und zum Teil darüber hinaus) eifrig übernommen, hauptsächlich, weil sie den Vergleich zwischen Tests ermöglichen. Doch zu behaupten, dass ein Test einem bestimmten GER-Niveau entspricht, ist nicht das gleiche, wie zu behaupten, dass dieser Test unter Berücksichtigung des deskriptiven Apparates des GER entworfen wurde. Die Kompetenzniveaus sind außerdem in Curricula und Lehrbücher eingegangen, doch sagt dies wenig darüber aus, in welchem Ausmaß die betrachteten Curricula kommunikativen Inhalt vorgeben oder die Lehrbücher einen pädagogischen Ansatz verkörpern, für den Sprachlernen eine Variante der Sprachverwendung ist (Europarat 2001: 21). Mir ist lediglich ein Fall bekannt, in dem ein Zweitsprachen-Curriculum als kommunikatives Repertoire definiert ist, welches durch „Ich kann“-Deskriptoren zum Ausdruck gebracht, durch eine Version des ESP vermittelt, und unter Verwendung von handlungsorientierten und aufgabenbasierten Tests bewertet wird. Bei diesem Curriculum handelt es sich um das zum Unterricht von Englisch als Zweitsprache an irischen Grundschulen (IILT 2003)

verwendete. Das ESP (IILT 2004) wurde teilweise als Verbreitungsinstrument für das Curriculum konzipiert und wurde sehr weitreichend verwendet; und in Umfang und Inhalt spiegeln die Tests die „Ich kann“-Deskriptoren des Curriculums wider.

Unter den Herausforderungen, mit denen DaF/DaZ als Disziplin in den deutschsprachigen Ländern zu Beginn des 21. Jahrhunderts konfrontiert ist, stechen zwei besonders hervor: die Notwendigkeit, sicherzustellen, dass eine große Anzahl von Kindern und Jugendlichen mit Migrationshintergrund die Deutschkenntnisse entwickeln, die ihnen vollen Zugang zu Bildung ermöglichen; und die Übernahme von Bildungsstandards, die eine Bewegung weg von schulbasierter Beurteilung und hin zu einer stärker zentralisierten Kontrollinstanz mit sich bringen. Um diese Probleme erfolgreich anzusprechen, wird es unter anderem nötig sein, gegen die großflächige Entfremdung der Lernenden vom Bildungssystem und damit von der Hauptströmung der Gesellschaft anzugehen. In der Auseinandersetzung mit diesen Herausforderungen haben Portfolios und andere Versionen informeller Beurteilung eine wichtige Rolle zu spielen.

## 6. Literatur in Auswahl

Alexander, Robin

2006 *Towards Dialogic Learning. Rethinking Classroom Talk*. York: Dialogos.

Authentik

2001 *European Language Portfolio/Punann na dTeangacha Eorpacha*. Dublin: Authentik.

Black, Paul und Dylan Wiliam

1998 *Inside the Black Box*. London: King's College London.

Council for Cultural Cooperation

1992 *Transparency and Coherence in Language Learning in Europe: Objectives, Evaluation, Certification. Report on the Rüschlikon Symposium*. Strasbourg: Council of Europe.

Council for Cultural Cooperation

1997a *European Language Portfolio: Proposals for Development. With contributions by I. Christ, F. Debyser, A. Dobson, R. Schärer, G. Schneider/B. North & J. Trim*. Strasbourg: Council of Europe (CC-LANG 97. 1).

Council for Cultural Cooperation

1997b *Language Learning for European Citizenship. Final report (1989–96). Compiled and edited by John Trim, Project Director*. Strasbourg: Council of Europe.

Council of Europe

2006 *European Language Portfolio: Key Reference Documents*. Strasbourg: Council of Europe.

Dam, Leni

1995 *Learner Autonomy 3: From Theory to Classroom Practice*. Dublin: Authentik.

Europarat

2001 *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.

Foucault, Michel

1976 *Überwachen und Strafe. Die Geburt des Gefängnisses*. Frankfurt a. M.: Suhrkamp.

Holec, Henri

1979 *Autonomy and Foreign Language Learning*. Strasbourg: Council of Europe.

IILT

2003 *English Language Proficiency Benchmarks for Non-English-Speaking Pupils at Primary Level*. Dublin: Integrate Ireland Language and Training.

## IILT

- 2004 *European Language Portfolio (Primary): Learning the Language of the Host Community*. Dublin: Integrate Ireland Language and Training.

## Little, David

- 1991 *Learner Autonomy 1: Definitions, Issues and Problems*. Dublin: Authentik.

## McNamara, Tim und Carsten Roever

- 2006 *Language Testing: The Social Dimension*. Malden, MA/Oxford: Blackwell.

## Müller, Andreas

- 2005 *Erlebnisse durch Ergebnisse. Das Lernportfolio als multifunktionales Werkzeug im Unterricht. Grundschule 6: 8–18.*

## Ruf, Urs, Stefan Keller und Felix Winter (Hg.)

- 2008 *Besser Lernen im Dialog. Dialogisches Lernen in der Unterrichtspraxis*. Seelze-Velber: Klett/Kallmeyer.

## Schärer, Rolf

- 2000 *European Language Portfolio: Final Report on the Pilot Project*. Strasbourg: Council of Europe. ([http://www.coe.int/T/DG4/Portfolio/documents/DGIV\\_EDU\\_LANG\\_2000\\_31Erev.doc](http://www.coe.int/T/DG4/Portfolio/documents/DGIV_EDU_LANG_2000_31Erev.doc) – 16. 12. 2008)

## Shohamy, Elana

- 2001 *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. London: Pearson.

## Winter, Felix

- 2004 *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schulleistungen*. Baltmannsweiler: Schneider.

## Portfolios für Deutsch als Fremd- und Zweitsprache

Für den Unterricht DaF/DaZ werden in vielen Ländern Portfolios genutzt, die die Pilotversionen des Europäischen Portfolios der Sprachen jeweils altersgemäß adaptieren und in die Landessprache übertragen (vgl. die Liste der vom Europarat validierten Portfolios in den einzelnen Ländern: Europarat, Validated Portfolios:

[http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main\\_pages/portfolios.html](http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/portfolios.html)).

Zunehmend bieten Lehrbuchverlage begleitend bzw. integriert in das Lehrwerk eigene Versionen an, vgl. z. B. die Portfolios in zu den Lehrwerken „Optimal“ (Langenscheidt Verlag):

[http://www.langenscheidt-unterrichtsportal.de/sprachenportfolio\\_1457.html](http://www.langenscheidt-unterrichtsportal.de/sprachenportfolio_1457.html) und „Team Deutsch“ (Klett Verlag): [http://www.klett.de/sixcms/media.php/71/352641/A08089\\_67594001\\_PORTFOLIO\\_TD.pdf](http://www.klett.de/sixcms/media.php/71/352641/A08089_67594001_PORTFOLIO_TD.pdf).

Die Universitäten Bremen und Stuttgart bieten inzwischen auch ein elektronisches Portfolio (EPOS) an:

<http://epos21.demo.fremdsprachenzentrum-bremen.de/portfolio/>.

Eine spezifische, sprachliche und beruflichbezogene Dimensionen integrierende Form stellt das „Sprachen- & Qualifikationsportfolio für MigrantInnen und Flüchtlinge“ (Plutzer und Haslinger 2005: <http://www.integrationshaus.at/portfolio/>) dar.

*David Little, Dublin (Irland)*

*Übersetzung aus dem Englischen von Cordula Politts*