

Zuverlässig – Objektiv – Gültig?

DSH und TESTDAF – Sprachprüfungen auf dem Prüfstand¹

Bernd Wintermann

Fast drei Jahre nach Verabschiedung der DSH-Rahmenordnung durch die Hochschulrektorenkonferenz, fast zwei Jahre nach Inkrafttreten, hat sich die »Deutsche Sprachprüfung für den Hochschulzugang ausländischer Studienbewerber« als feste Größe in der deutschen Hochschullandschaft etabliert. Der größere Teil der Hochschulen verfügt inzwischen über eine eigene Prüfungsordnung auf der Basis der Rahmenordnung, und wo dies noch nicht der Fall ist, wird vielerorts doch entsprechend den Regularien der Rahmenordnung geprüft.

Kritische Anfragen und Mißverständnisse bleiben dennoch nicht aus. Ein besonders häufig anzutreffendes Mißverständnis ist in diesem Zusammenhang die Ansicht, mit der DSH könnten oder sollten die für ein Fachstudium geeigneten Bewerber *ausgewählt* und die *Zulassungszahlen gesteuert* werden. Demgegenüber muß klar hervorgehoben werden: Es ist mit der Zielsetzung der DSH – Feststellung ausreichender Sprachfertigkeiten – unvereinbar, in Zeiten hoher Bewerberzahlen die Anforderungen hochzuschrauben und in Zeiten zurückgehender Bewerber-

zahlen mit den Anforderungen wieder herunterzugehen oder gar ganz auf eine Sprachprüfung zu verzichten, wie der Präsident der TU München dies für richtig gehalten hat. Gegen zurückgehende Bewerberzahlen helfen vielmehr:

- eine ansprechende und informative Werbung für ein Studium in Deutschland,
 - attraktive Fachstudienangebote,
 - eine verbesserte Anerkennung von bereits im Heimatland erbrachten Studienleistungen
 - eine intensive, insbesondere auch sprachliche Betreuung und Begleitung.
- Sprachprüfungen gleich welcher Art dürfen im Zusammenhang mit dem Ausländerstudium nicht als Berechtigungsnachweis mißbraucht werden. Ihre Aufgabe ist diagnostischer Art. Diagnose bedeutet: Entscheidungshilfe für Beratung und Förderung: Wieviel und welche sprachliche Vorbereitung ist vor Aufnahme des Fachstudiums noch erforderlich, wieviel und welche sprachliche Begleitung ist innerhalb des Fachstudiums noch nötig? Freilich kann das Ergebnis der Diagnose für manche Bewerber auch lauten: Der

1 Für die kritische Durchsicht eines ersten Manuskriptentwurfs und für vielerlei Anregungen danke ich Dietrich Eggers, Rolf Ehnert, Frank Königs, Evelyn Müller-Küppers, Winfried Welter. Alle verbliebenen Unzulänglichkeiten gehen jedoch voll zu meinen Lasten.

Zeitaufwand für den Erwerb der erforderlichen Deutschkenntnisse wäre so hoch, daß der Nutzen eines Studienaufenthalts in Deutschland dazu in keinem Verhältnis stünde.

Es ist ein naheliegender und verlockender Gedanke, sich für Diagnoseverfahren aller erdenklicher *technischer Hilfsmittel* zu bedienen. Ein Mißverständnis ist es jedoch, zu meinen, die technischen Hilfsmittel selbst könnten die Diagnose leisten. Ihr Einsatz ist in Kfz-Werkstätten bereits weit fortgeschritten: Motor, Getriebe, Elektrik, Elektronik werden an ein Diagnosegerät angeschlossen, und der Mechaniker kann, ohne daß er viel von den Funktionsabläufen verstehen müßte, ablesen, welche Teile ausgetauscht werden müssen, um das Fahrzeug wieder voll funktionsfähig zu machen.

In der Medizin wurden in den 70er Jahren – nach dem gleichen Muster – mit großem Aufwand Diagnosekliniken errichtet, die mit immer aufwendigeren, immer differenzierteren Geräten und Verfahren die Funktionsfähigkeit bzw. etwaige Funktionsstörungen des menschlichen Organismus mit absoluter Zuverlässigkeit feststellen sollten. Es ist inzwischen sehr still um diese Diagnosezentren geworden. Statt dessen sind die Intuition und das Einfühlungsvermögen des erfahrenen Arztes, die Einbeziehung der gesamten Person bei der Ausarbeitung eines Therapiekonzepts wieder gefragt. Die großen Hoffnungen, die in die Diagnosekliniken gesetzt wurden, haben sich offensichtlich nur zu einem kleinen Teil erfüllt.

Ein gleichsam automatisches und hochzuverlässiges Verfahren auch für die Feststellung sprachlicher Fertigkeiten und Defizite einzusetzen – das ist eben-

falls ein verlockender Gedanke. Er ist nicht nur für die Administration verlockend, sondern auch für uns Sprachlehrerinnen und -lehrer. Denn wer prüft schon gern? Wer hat schon Spaß daran, -zig in der Regel nicht besonders originelle und höchst fehlerhafte Textproduktionen zu einem Thema zu lesen, das weder einen selbst noch die Kandidaten wirklich interessiert? Oder ein Dutzend Gespräche zu führen, in denen es gar nicht um das Thema geht, sondern nur um die Sprachfertigkeiten?

Klar, daß sich aus diesen und ähnlichen Gründen ein Testinstrument wie der amerikanische *TOEFL*¹ – eingeführt in der Hochblüte behavioristisch-strukturalistischer Euphorie – lange Zeit hoher Beliebtheit erfreute. Und die Testforschung hat ja, was die Zuverlässigkeit und Objektivität von Testverfahren angeht, auch ungeheure Fortschritte gemacht – Fortschritte, die hier in keiner Weise kleingeredet werden sollen. Es steht auch uns Sprachlehrerinnen und -lehrern gut an, diese Fortschritte sorgfältig und selbstkritisch zur Kenntnis zu nehmen.

Trotzdem: auch dem *TOEFL* scheint es so zu gehen wie den Diagnosekliniken: Er erfüllt die in ihn gesetzten Erwartungen nur zum Teil. Das zeigt sich einmal daran, daß schon jetzt – innerhalb des *ETS*² wie auch ergänzend dazu – Zusatztests herangezogen werden (*Test of Written English: TWE; Test of Spoken English: TSE*), deren Zuverlässigkeit und Objektivität nicht mehr so einfach zu optimieren sind wie im eigentlichen *TOEFL*, der nach wie vor aus *multiple-choice*-Aufgaben besteht.

Zum anderen zeigt es sich daran, daß für die nächsten Jahre eine gründliche *Revisi-*

1 *Test of English as a Foreign Language.*

2 *Educational Testing Service Princeton NJ*; die Testzentrale des *TOEFL*.

on des TOEFL (TOEFL 2000) avisiert ist, mit dem Ziel, das Testformat zu diversifizieren und damit neben der Zuverlässigkeit und Objektivität gerade auch die *Gültigkeit* des Tests zu erhöhen.¹

Diese Entwicklung bestätigt das, was Helmut J. Vollmer in seinem Überblick über Leistungsmessung² zum »Aktuellen Erkenntnisstand« schreibt:

»In der psychometrisch-strukturalistischen Phase werden besondere Anstrengungen unternommen, um Objektivität und Reliabilität als notwendige Vorbedingungen für die Gültigkeit von Meßinstrumenten zu erhöhen. Dennoch wird die Frage danach, was ein Test denn nun eigentlich mißt, kaum gestellt, ja sogar vermieden; statistisches Denken überwiegt. Alles, was quantifizierbar und in diesem strikten Sinne als meßbar erscheint, erfreut sich besonderer Vorliebe. Aus der Unzufriedenheit über die theoretische Begrenztheit dieses Ansatzes und damit verbundener Testverfahren wiederum entwickelte sich die psycholinguistisch-soziolinguistische Phase, die sich nun verstärkt den Fragen der Validität zuwendet, ohne die beiden anderen Gütekriterien zu vernachlässigen. Zum einen wird jetzt genauer untersucht, welche kognitiven Operationen der Lösung einzelner Sprachaufgaben zugrunde liegen und wie die zu messende(n) Dimension(en) der Sprachkompetenz theoretisch angemessen zu strukturieren sind. Zum anderen konzentriert sich das Interesse (wieder) auf produktive Fremdsprachenleistungen, insbesondere auf Tests zur Überprüfung funktionaler Sprachverwendung in unterschiedlich komplexen Handlungskontexten, wie sie in Realsituationen vorkommen könnten. Damit wird die Testforschung verstärkt

wieder an den Erkenntnis- und Reflexionsstand der Angewandten Linguistik sowie an die Diskussionen innerhalb dieser Disziplin herangeführt.«

Die *Gültigkeit*, dieses dritte – und in meinen Augen entscheidende – Gütekriterium von Tests wird – das läßt sich auch im Bereich der Sprachtests beobachten – oft schamhaft verschwiegen. Wie bei den Intelligenztests befindet sich der Tester dann in der Situation: »Ich weiß zwar nicht, *was* ich eigentlich messe, aber *das* messe ich jedenfalls sehr zuverlässig und objektiv.«

Die Gültigkeit eines Tests läßt sich nur anhand eines geeigneten *Außenkriteriums* bestimmen.

Für die DSH – die ja die Frage klären soll: Reichen die sprachlichen Fertigkeiten für das beabsichtigte Fachstudium aus? In welcher Hinsicht müssen die sprachlichen Fertigkeiten ggf. noch weiter gefördert werden? – könnte ein solches Außenkriterium z. B. das Abschneiden der Absolventen in akademischen Zwischen- und Abschlußprüfungen sein. Dies ist sicherlich ein ganz entscheidendes Kriterium, für sich allein genommen aber noch nicht sehr aussagekräftig. Denn erstens sind akademische Prüfungen zwar auch, aber doch nur zum Teil Sprachprüfungen und zweitens genügen auch sie in aller Regel nicht den Anforderungen an Zuverlässigkeit und Objektivität, die an ein Prüfungsverfahren gestellt werden müssen.

Als Kriterium bleibt dann die möglichst sorgfältige Erhebung derjenigen *sprachli-*

1 Vgl. dazu Michaela Perlmann-Balme: *TESTDAF – Vorstudie zu dem Projekt »Entwicklung eines Tests Deutsch als Fremdsprache für ausländische Studienbewerber«*. In dieser sehr informativen und übersichtlichen Studie – zu beziehen über den DAAD in Bonn – wird die Praxis der Überprüfung von Sprachkenntnissen ausländischer Studienbewerber in verschiedenen Industrieländern (USA, Kanada, Großbritannien, Australien, Japan, Spanien, Niederlande, Belgien) gegenübergestellt. Die Studie zeigt sehr deutlich, daß das TOEFL-Format nur *eine* Möglichkeit unter mehreren anderen darstellt und daß gerade die anderen anglophonen Länder (Großbritannien, Australien) das TOEFL-Format *nicht* übernommen haben.

2 Bausch/Christ/Krumm (Hrsg.): *Handbuch Fremdsprachenunterricht*. 3. Aufl. 1995, 274f.

chen Handlungen, die für ein akademisches Studium konstitutiv sind,¹ und deren möglichst repräsentative Abbildung in Prüfungsaufgaben. Dieser Ansatz, diese Intention lag – bei aller Unvollkommenheit in den Details – der PNdS zugrunde, und gleiches gilt für die DSH. Jedes alternative Verfahren zur Feststellung der für ein Studium benötigten Sprachfertigkeiten muß sich daran messen lassen, ob es diesen Anspruch mindestens genauso gut einlöst wie die DSH, möglichst aber besser.

Im Rahmen seines Aktionsprogramms zur Förderung des Ausländerstudiums setzt sich nun der DAAD für die Einführung eines weltweit angebotenen Sprachtests TESTDAF ein, der sich am Rahmen und an den Standards der DSH orientiert. Der DAAD erhofft sich von diesem Test Verbesserungen im Bereich des Ausländerstudiums und sieht insbesondere die folgenden Vorteile:

- Der Test kann grundsätzlich in jedem Land der Welt abgenommen werden.
- Durch eine einheitliche Aufgabenstellung, Prüfungskorrektur und Punktevergabe ist dem Teilnehmer eine Überprüfung möglich, wo er mit seinen Sprachkenntnissen steht.
- Die einzelne deutsche Hochschule kann die für die Zulassung erforderliche Punktzahl festlegen, so daß dem ausländischen Studienbewerber die Erfordernisse genau bekannt sind.
- Die einzelne Hochschule hat hierbei die Möglichkeit, die erforderliche Punktzahl zu differenzieren, um damit den unterschiedlichen Spracherfordernissen je nach Studienziel und Studienfach des Ausländers Rechnung zu tragen.

- Der Test gibt dem ausländischen Studienbewerber hinsichtlich der Spracherfordernisse eine Zulassungsgewißheit vor der Einreise.
- Der Test führt zu einer Verkürzung des Vorstudienaufenthalts für den Ausländer in Deutschland, Selbstlernmaterialien, Fernsprachkurse und der Einsatz neuer Medien verbreitern und erleichtern die Spracherwerbsmöglichkeiten.

Die Vorüberlegungen sind relativ weit gediehen, und zum Zeitpunkt des Erscheinens dieses Beitrags steht das Projekt zur Ausschreibung an.

Der Vorstand des FaDaF wurde vom DAAD zur Mitarbeit an diesem Projekt eingeladen. Der Vorstand begrüßt und unterstützt die Bemühungen des DAAD zur Förderung des Studiums von Ausländern und beteiligt sich in diesem Rahmen auch nach Kräften an der Einführung des TESTDAF. Die Mitarbeit des FaDaF steht dabei selbstverständlich unter der Prämisse, daß geeignete Regularien zur Überprüfung der Deutschkenntnisse ausländischer Studierender in der DSH-Rahmenordnung niedergelegt sind. Daher ist für uns die Aussage wichtig, daß der geplante TESTDAF sich an der DSH als Rahmen und Standard orientiert. Auch der TESTDAF muß also wie die DSH Hörverstehen, Leseverstehen, Schreibfertigkeit und mündliche Kommunikationsfähigkeit in hochschulnahen Kontexten überprüfen.

Von diesem Ansatz her ist zu fragen: Welche Erwartungen kann man an die Einführung eines solchen weltweit einsetzbaren deutschen Sprachtests knüpfen? Welche dieser Erwartungen sind realistisch, welche werden sich bei näherem Hinsehen als unberechtigt erweisen? Ich

1 Vgl. dazu die knappe Übersicht von Dietrich Eggers: »Erwartungen der Fächer und Fachbereiche an ausländische Studierende in sprachlicher Hinsicht«, *fadaf-aktuell* 4 (1996), 5–7.

versuche im folgenden, diese Fragen für einige wichtige Aspekte aus der Sicht des FaDaF zu beantworten.

- 1) Uneingeschränkt teilt der FaDaF-Vorstand die Auffassung, daß mit einem solchen Test die Möglichkeit der Selbsteinschätzung für Interessenten und der rechtzeitigen Planung von Fördermaßnahmen am künftigen Studienort verbessert werden.
- 2) Ebenso unstrittig ist die Forderung, daß ausländische Studienbewerber vor der Einreise die Gewißheit erhalten müssen, auch zum Studium zugelassen zu werden. Wie oben ausgeführt, sollte diese Zulassungsgewißheit jedoch nicht am Nachweis ausreichender Sprachkenntnisse festgemacht werden. An vielen Hochschulen werden inzwischen Modelle erprobt, wie diese Sprachkenntnisse zum Teil auch noch studienbegleitend, jedenfalls aber bei vollgültigem Studierendenstatus erworben und nachgewiesen werden können.
- 3) Für sinnvoll halten wir es auch, potentiell in allen Ländern der Welt Sprachtests anzubieten, mit denen die für ein Studium ausreichenden Deutschkenntnisse nachgewiesen werden können. Zu fragen ist jedoch, ob es nicht der Transparenz und Klarheit der Information dienlicher gewesen wäre, die bereits vorhandenen Prüfungen – DSH an Universitäten im Ausland; Zentrale Oberstufenprüfung des Goethe-Instituts, Kleines Deutsches Sprachdiplom, Sprachdiplom der Kultusministerkonferenz – auszubauen, statt mit erheblichem finanziellem und organisatorischem Aufwand diesen Prüfungen eine fünfte an die Seite zu stellen. Es wird erheblicher Aufklärungsarbeit bedürfen, den ausländischen Partnern die Notwendigkeit dieser fünften Prüfung plausibel zu machen.
- 4) Der Wunsch, die Vorstudienzeit in Deutschland zu verkürzen, ist verständlich. Zu bedenken ist dabei jedoch einerseits die unter 2) angedeutete Möglichkeit, Teile des Spracherwerbsprozesses stärker als bisher in das Fachstudium zu integrieren, und zum anderen darf nicht übersehen werden, daß der Erwerb einer Fremdsprache im Heimatland beträchtlich viel mehr Zeit kostet als im Zielsprachenland. Die Vorbereitungszeit im Heimatland eingerechnet, wird also durch die Verlagerung des Spracherwerbs ins Heimatland die Gesamtstudiendauer eher verlängert als verkürzt.
- 5) Aus den oben genannten Gründen – eine Sprachprüfung darf nicht zur Steuerung der Zulassungszahlen mißbraucht werden – halten wir es nicht für sachgerecht, daß jede Hochschule, womöglich noch nach Fachrichtungen differenziert, die für eine Zulassung erforderliche Punktzahl selbst festlegt. Der Wettbewerb unter den Hochschulen um die besten und qualifiziertesten ausländischen Studierenden darf nicht darauf hinauslaufen, daß sie sich bei den sprachlichen Anforderungen gegenseitig unterbieten.
- 6) Die Differenzierung der erforderlichen Mindestpunktzahl je nach Studienfach beruht auf der Überlegung: Wer die Aufgaben zu einem Text zu 90% gelöst hat, ist für ein Philosophiestudium geeignet, wer sie nur zu 60% gelöst hat, ist zu einem Informatikstudium geeignet. Die Annahme, ein einziger Test könnte die unterschiedlichen sprachlichen Anforderungen unterschiedlicher Studienrichtungen skalar abbilden, ist unrealistisch. Sinnvoll wäre es sicherlich, nach Fächergruppen differenziert, unterschiedliche Tests anzubie-

ten. Das würde aber den organisatorischen Aufwand beträchtlich erhöhen und den Rahmen des TESTDAF-Projekts mit Sicherheit sprengen.

- 7) Sicherlich hat man bei der Konzeption des TESTDAF-Projekts auch die Möglichkeit im Blick gehabt, Kosten zu sparen. Die Erwartung, daß sich das Unternehmen nach einer gewissen Anlaufzeit durch Erhebung von Prüfungsgebühren eines Tages selbst tragen wird, erscheint aber sehr optimistisch. Zunächst einmal werden über mehrere Jahre hin erhebliche zusätzliche Kosten entstehen. Auch die Einsparungsmöglichkeiten durch maschinelle Auswertung dürfen nicht überschätzt werden. Wie die Studie von Michaela Perlmann-Balme (s. Anm. oben) zeigt, läßt sich mit den allein maschinell auswertbaren geschlossenen Prüfungsformen auf der Ebene der für ein akademisches Studium erforderlichen Sprachkenntnisse das Gültigkeitskriterium nicht erreichen.
- 8) Damit sind wir bei der m. E. entscheidenden Frage der Sicherung der Qualitätsstandards. Aus dem bisher Gesagten sollte deutlich geworden sein, daß die Einhaltung der Gütekriterien keine eindimensionale Angelegenheit ist, sondern daß die Kriterien Zuverlässigkeit und Objektivität auf der einen Seite, Gültigkeit auf der anderen Seite in einem Spannungsverhältnis zueinander stehen. Maschinell auswertbare Aufgaben scheiden zur Überprüfung der für ein Studium erforderlichen sprachlichen Kompetenz weitgehend aus. Wenn aber die für ein Studium konstitutiven sprachlichen Handlungen in Prüfungsaufgaben abgebildet werden sollen, müssen erhebliche Anstrengungen unternommen werden, damit die Kriterien Zuverlässigkeit und Objektivität nicht zu

kurz kommen. Unzweifelhaft ist das Prüfungsverfahren, wie es gegenwärtig auf der Grundlage der DSH-Rahmenordnung praktiziert wird, in dieser Hinsicht weiterhin *verbesserungsfähig und auch -bedürftig*. Die dezentrale Durchführung und Bewertung der Prüfung, für die es gute Gründe gibt, bedingt, daß die *regionale und überregionale Zusammenarbeit der Lehrgebiete und Studienkollegs* in dieser Hinsicht verstärkt wird. Dies betrifft insbesondere

- die Einigung über Auswahlkriterien für geeignete Textvorlagen und Aufgabentypen,
- die Verabredung von Bewertungsmaßstäben und Korrekturverfahren,
- den Austausch von Prüfungsaufgaben einschließlich der Bewertungsschlüssel und einiger Musterkorrekturen,
- die gemeinsame Erarbeitung von Musterprüfungen,
- die gleichzeitige Durchführung der Prüfung an verschiedenen Orten mit den gleichen Aufgabenstellungen,
- den gegenseitigen Besuch bei schriftlichen und mündlichen Prüfungen,
- den Austausch von Prüfern und Korrektoren.

Der FaDaF hat hierzu bereits mehrere Tagungen selbst veranstaltet und die Regionen ermuntert, diese Thematik auf Regionaltagungen weiter zu verfolgen. Er wird seine Anstrengungen in dieser Hinsicht fortsetzen.

Die genannten Arbeiten müssen mutatis mutandis für den geplanten TESTDAF jedoch ebenfalls geleistet werden: Auswahlkriterien, Bewertungsmaßstäbe, Korrekturverfahren müssen erarbeitet und abgestimmt werden, Prüfer und Korrektoren müssen vorbereitet und

kontinuierlich geschult werden. Nur wenn die erforderlichen Ressourcen zur Verfügung gestellt werden, könnte sich hier mittelfristig ein Prä für eine zentrale TESTDAF-Institution ergeben.

Der große und unschätzbare Vorteil des derzeitigen dezentralen DSH-Prüfungsverfahrens liegt aber darin, daß die Überprüfung von erfahrenen und engagierten Sprachlehrerinnen und -lehrern vorgenommen wird, die

- die Studiensituation in den verschiedenen Fachbereichen überblicken,
- regelmäßige Rückmeldungen aus den Fachbereichen über die sprachlichen Fertigkeiten und Defizite ausländischer Studierender erhalten

- die Kandidatinnen und Kandidaten großenteils aus prüfungsvorbereitenden Kursen kennen
- in der Prüfung nicht nur fertige Endprodukte zur Korrektur vorgelegt bekommen, sondern den Entstehungsprozeß von Textproduktionen z. B. anhand der Notizen und der Reparaturhandlungen mit überblicken oder in der mündlichen Prüfung die gesamte Kommunikationssituation beurteilen und nicht nur eine Tonbandaufzeichnung.

Um im Bild zu bleiben: Die Diagnoseklinik TESTDAF ersetzt nicht den erfahrenen Arzt, der die Gesamtpersönlichkeit und die Gesamtsituation berücksichtigt, um dem Kandidaten eine optimale Förderung angedeihen zu lassen.