

TestDaF: Konzeption, Stand der Entwicklung, Perspektiven

Projektgruppe TestDaF*

The article describes the new Test of German as a Foreign Language TestDaF – a standardised language test for foreign students applying for entry to an institution of higher education in Germany. TestDaF will be administered world-wide and is in its function and scope comparable to the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL).

TestDaF measures the candidate's performance in each of the four skills separately and relates it to one of three levels of proficiency, ranging approximately from level three (independent user) to five (good user) on the ALTE scale or levels B2.1 (lower vantage level) to C1.2 (higher effective proficiency) of the Council's of Europe "Common European Framework of Reference".

In the TestDaF development project, ending July 31, 2000, three model test sets in a paper-and-pencil format were developed and piloted. TestDaF will be operational in winter 2000. Development, assessment, grading and evaluation of TestDaF will be centralized in an institution whose organizational structure is still to be determined.

1. Einleitung

TestDaF (Test Deutsch als Fremdsprache) ist eine Prüfung für ausländische Studienbewerber, die ihr Studium in Deutschland aufnehmen bzw. ein bereits im Heimatland begonnenes Studium in Deutschland weiterführen möchten. Er ist damit in seiner Funktion vergleichbar mit dem IELTS (*International English*

* Der Projektgruppe TestDaF gehören an: Hans-Georg Albers, K.-Richard Bausch, Dörthe Üstünsöz-Beurer, Sibylle Bolton, Rüdiger Grotjahn, Bärbel Gutzat, Dieter Gutzen, Margareta Hauschild, Beate Helbig, Rainer Hofmann, Karin Kleppin, Gabriele Kniffka, Hans-Werner Six, Josef Voss.

Die Projektgruppe TestDaF ist mit der konkreten Testentwicklungsarbeit befasst, die insgesamt von einem Konsortium getragen wird. Das Konsortium besteht aus der FernUniversität Hagen und dem Seminar für Sprachlehrforschung der Ruhr-Universität Bochum sowie dem Goethe-Institut München und den Carl Duisberg Centren Köln.

Der vorliegende Text wurde von den Projektmitgliedern Rüdiger Grotjahn und Karin Kleppin abgefasst; eine gekürzte Version war Gegenstand eines Vortrags mit Diskussion auf dem 18. Kongress der Deutschen Gesellschaft für Fremdsprachenforschung, der vom 4. bis 6. Oktober 1999 an der Universität Dortmund stattfand. Rüdiger Grotjahn und Karin Kleppin tragen die Verantwortung für die Formulierungen und für die Bezüge zur aktuellen testwissenschaftlichen Literatur.

Kontaktadresse:

Priv. Doz. Dr. Rüdiger Grotjahn, Seminar für Sprachlehrforschung, Ruhr-Universität Bochum, 44780 Bochum; E-mail: ruediger.grotjahn@ruhr-uni-bochum.de

Language Testing System) und mit dem TOEFL (*Test of English as a Foreign Language*) – den beiden bedeutendsten internationalen Sprachtests für den Zugang zu englischsprachigen Universitäten und Institutionen.

Die Idee zur Entwicklung von TestDaF ist im Rahmen der Diskussion um die Sicherung des Wissenschaftsstandorts Deutschland entstanden. Der Auftrag wurde 1998 vom DAAD (Deutscher Akademischer Austauschdienst) an ein Konsortium vergeben. Entsprechend spiegelt die folgende Darstellung die gemeinsame Entwicklungsarbeit des Konsortiums wider (Diskussionsstand 1.3.2000). Sie beruht zudem in wesentlichen Teilen auf internen Projektberichten sowie auf folgenden Publikationen: Bolton (2000), Das Konsortium für das TESTDAF-Projekt (2000).

Die erste Entwicklungsphase des TestDaF-Projekts dauert vom 1. August 1998 bis zum 31. Juli 2000. In dieser Zeit werden drei Modellprüfungssätze entwickelt und in großem Umfang weltweit erprobt. Testerstellung und Erprobung werden testmethodisch begleitet, d.h., die Lösungen der Erprobungskandidaten werden statistisch analysiert, Fragebögen für die Testerprober und für die Kandidaten erstellt und ausgewertet. Diese Daten sowie die eingeholten wissenschaftlichen Gutachten bilden die Basis für die Revision der Testsätze.

Es wurde großer Wert darauf gelegt, von Anfang an möglichst viele Institutionen im Bereich DaF in die Erprobung und Evaluation einzubinden, wie z.B. den Fachverband Deutsch als Fremdsprache, mehrere Studienkollegs und Lehrgebiete DaF an deutschen Hochschulen.

TestDaF wurde z.T. sehr kontrovers diskutiert (vgl. exemplarisch Bickes, 1998; Wintermann, 1998), vor allem auch aufseiten von Praktikern.¹ Zum Start des Projektes erschienen Pressenotizen, die sich nicht nur auf Fachzeitschriften beschränkten. Man sprach vom kleinen Bruder des TOEFL (Ruhr-Nachrichten vom 11.8.98), der nun endlich auch Deutschland wieder zu einem attraktiveren Standort machen könnte, weil für ausländische Studienbewerber der Zeit- und Kostenaufwand bis zum Bestehen eines Studieneingangstests gesenkt werden könnte (vgl. Süddeutsche Zeitung vom 21.7.1998 und Westfälische Post vom 10.7.1998). Erwähnt wurde gleichzeitig immer, dass es sich um ein mit 1,7 Millionen Mark gefördertes Projekt handelte; sicherlich fragten sich nicht wenige Leser, warum denn soviel Geld für einen Test bereit gestellt werden müsse.

¹ So wurde z.B. häufig die Befürchtung geäußert, dass die Einführung von TestDaF zu einem Stellenabbau bei Institutionen führen könnte, die auf die "Deutsche Sprachprüfung für den Hochschulzugang ausländischer Studienbewerber" (DSH) bzw. auf die "Prüfung zum Nachweis deutscher Sprachkenntnisse" (PNdS) vorbereiten.

2. Politische Zielvorstellungen

Mit dem TestDaF-Projekt sollen im Wesentlichen folgende vier – auch von Gegnern von TestDaF weitgehend akzeptierte – übergeordnete politische Ziele verfolgt werden:

- TestDaF soll Sprachfähigkeiten im akademischen Bereich messen. Die Mehrzahl der anvisierten Testkandidaten hat allerdings Deutsch in ihrem Heimatland ohne direkte Kontakte zu deutschen Universitäten und deren Umfeld gelernt. Außerdem handelt es sich um Studienbewerber aus allen Fachrichtungen. Um eine Benachteiligung einzelner Kandidaten weitgehend auszuschließen, sind deshalb Inhalte und Themen der Prüfung schwerpunktmäßig aus dem allgemeinen akademischen Umfeld zu wählen.
- Die Studieneingangstests an den Universitäten sollen überschaubarer und transparenter werden; denn trotz Verbesserungen in den letzten Jahren konnte die jetzt gültige DSH keine Vergleichbarkeit über die Universitäten hinweg gewährleisten, was u.a. zu einem Testtourismus innerhalb von Deutschland geführt hat: Nicht wenige Studenten, die den Test an einem Studienort nicht bestanden hatten, haben den Test erfolgreich an einem anderen Ort abgelegt.
- TestDaF soll eine sprachliche Zulassungssicherheit für Studienbewerber aller Fachrichtungen bereits im Heimatland gewährleisten, was insbesondere für Herkunftsländer mit Visumpflicht Vorteile bietet; denn bisher erhielten die Studienbewerber in vielen Ländern nur ein Visum für die Studieneingangsprüfung und mussten dann ein weiteres Visum für ein Studium beantragen. Dieses Verfahren konnte zu hohen Kosten für die jeweiligen Studienbewerber führen, da diese gegebenenfalls mehrfach nach Deutschland reisen mussten.
- Der Test soll die zu überprüfenden Fähigkeiten mit einem möglichst hohen Maß an Objektivität, Reliabilität, Validität und Fairness messen.

Diese Zielvorstellungen führten zur Entscheidung für eine standardisierte, zentral erstellte Prüfung mit zentraler Korrektur.

3. Testmethodische Prinzipien

In testmethodischer Hinsicht haben die Konsortialpartner u.a. folgende Entscheidungen getroffen:

- TestDaF erfasst sprachliche Kenntnisse und Fertigkeiten auf den Stufen drei bis fünf einer fünfstufigen Skala. Die Niveaustufen sind in aufsteigender Anordnung folgendermaßen benannt: TestDaF-Niveaustufe 3 (TDN 3),

TestDaF-Niveaustufe 4 (TDN 4) und TestDaF-Niveaustufe 5 (TDN 5). Unterhalb der Stufe TDN 3 sowie oberhalb der Stufe TDN 5 wird bisher nicht weiter differenziert.² Bei TDN 3 und TDN 4 sind die Zwischenstufen TDN 3 + und TDN 4 + möglich. TestDaF orientiert sich an den Skalen des *Common European Framework of Reference* des Europarats sowie an den Kompetenzstufen der *Association of Language Testers in Europe (ALTE)*³, die eng aufeinander bezogen und zudem in Teilen bereits validiert sind (vgl. *Association of Language Testers in Europe (ALTE)*, 1998, Kap. 2; Coste, North, Sheils & Trim, 1998, Kap. 8). Der TestDaF-Spezifika wird u.a. dadurch Rechnung getragen, dass der akademische Kontext bei den Stufenbeschreibungen eine stärkere Berücksichtigung findet.

Für jede Niveaustufe sind in Abhängigkeit von den jeweiligen Zielsetzungen und Adressaten mehrere parallele Beschreibungen mit einem unterschiedlichen Grad an Komplexität und Detailliertheit anzufertigen. So erfordert eine Beschreibung im Hinblick auf die Aufgabenkonstruktion oder auch im Hinblick auf die Beurteilung der Leistung der Kandidaten durch die Testkorrektoren ein weit höheres Maß an Komplexität und Detailliertheit als das z.B. bei der allgemeinen Beschreibung der TestDaF-Niveaustufen für das TestDaF-Zeugnis der Fall ist.

Die TestDaF-Stufen TDN 3, TDN 4 und TDN 5 umfassen in etwa folgenden Kompetenzbereich: Bezogen auf die Stufen des Europarats decken sie das Spektrum von B2.1 (*Lower Vantage Level*) bis C1.2 (*Higher Effective Proficiency*) ab; bezogen auf die ALTE-Stufen entsprechen sie in etwa dem Bereich von Stufe 3 (*Independent User*) bis Stufe 5 (*Good User*). In Bezug auf deutsche Sprachprüfungen im Bereich DaF gelten laut ALTE Handbuch (*Association of Language Testers in Europe (ALTE)*, 1998) folgende Zuordnungen: Stufe 2 entspricht dem Zertifikat DaF, der Stufe 3 ist bisher keine deutsche Sprachprüfung explizit zugeordnet. Stufe 4 entspricht der Zentralen Mittelstufenprüfung (ZMP) und Stufe 5 der Zentralen Oberstufenprüfung (ZOP) sowie dem Kleinen Deutschen Sprachdiplom (KDS). Damit setzt TestDaF deutlich oberhalb des Niveaus des Zertifikats DaF an. Da sowohl die ZOP als auch das KDS von der DSH freistellen, bedeutet dies außerdem, dass auf einer formalen Ebene die DSH höchstens der ALTE-Stufe 5 zuzu-

² Die derzeitige Beschränkung auf die Stufen TDN 3 bis TDN 5 schließt nicht aus, dass zu einem späteren Zeitpunkt der Kompetenzbereich von TestDaF in Richtung auf eine Differenzierung unterhalb von TDN 3 und oberhalb von TDN 5 ausgeweitet wird.

³ In der *Association of Language Testers in Europe* haben sich Prüfungsinstitutionen zusammengeschlossen, die Prüfungen für die jeweilige Landessprache als Fremdsprache anbieten.

ordnen ist. Aufgrund des dezentralen und wenig standardisierten Charakters der DSH und den hieraus resultierenden Unterschieden in der Schwierigkeit einzelner Prüfungen dürften in der Praxis allerdings nicht wenige DSH-Versionen unterhalb der ALTE-Stufe 5 anzusiedeln sein. In Anbetracht dieses Sachverhalts sowie aufgrund mangelnder teststatistischer Untersuchungen zur DSH ist eine begründete Aussage über den Schwierigkeitsgrad und die Messgenauigkeit **der** DSH nicht möglich und deshalb auch die vielfach geforderte Ausrichtung des Schwierigkeitsgrads von TestDaF an den Anforderungen der DSH höchst problematisch (vgl. auch Bickes, 1998; Wintermann, 1998).⁴

- TestDaF misst die vier Fertigkeiten Hörverstehen, Leseverstehen, Schriftlicher Ausdruck und Mündlicher Ausdruck getrennt in Form von vier Teiltests. Dies hat vor allem Konsequenzen für das Testformat der Prüfungsteile Leseverstehen und Hörverstehen. Im Gegensatz zur DSH (vgl. z.B. Lee, 1998) werden im TestDaF keine offenen Aufgabentypen verwendet, die freies Schreiben erfordern, da diese keine eindeutigen Aussagen zulassen, ob die Defizite des Kandidaten mehr im Lesen/Hören oder mehr im Schreiben liegen. Bei den Testteilen Schriftlicher Ausdruck sowie Mündlicher Ausdruck handelt es sich eindeutig um Performanztests, bei den Testteilen Leseverstehen sowie Hörverstehen eher um Kompetenztests (vgl. zu dieser Unterscheidung Grotjahn, 2000b sowie McNamara, 1997).
- Die Leistungen werden – getrennt für die vier Fertigkeiten – nicht in Form von Punktwerten, sondern mittels verbaler Deskriptoren ausgewiesen, die dem Kandidaten verständliche Informationen über spezifische Schwächen und Stärken vermitteln. Damit wird dem Testkandidaten u.a. die Möglichkeit eröffnet, festgestellte Defizite auszugleichen. Die Bewertung erfolgt zudem nicht norm-, sondern kriterienorientiert, d.h. nicht in Bezug auf die Leistung der anderen Testteilnehmer, sondern anhand vorher festgelegter Leistungskriterien.
- Aus Gründen der Testökonomie soll längerfristig auch eine computerbasierte oder sogar eine computeradaptive Testversion erstellt werden.

3. Das Format von TestDaF

Die Prüfungszeit für den gesamten Test beträgt drei Stunden. Hinzu kommen 10 Minuten für das Übertragen der Lösungen im Testteil “Hörverstehen” auf das

⁴ Die im Rahmen des TestDaF-Projekts von D. Üstünsöz-Beurer, FernUniversität Hagen, durchgeführten empirischen Analysen zur Vergleichbarkeit von TestDaF und DSH stützen diese Einschätzung.

Antwortblatt durch die Kandidaten. Die Themen werden den Zielen des Tests entsprechend z.B. aus folgenden Bereichen entnommen: Studium und Umfeld der Studierenden, Umwelt und Umweltprobleme, Forschung und Wissenschaft (keine einzelwissenschaftlichen Themen), Neuere Technologien, Massenmedien, gesellschaftliche und menschliche Beziehungen, wirtschaftliche Verhältnisse, Sprache und Kommunikation. Es wurde darauf geachtet, dass keine potentiellen Tabuthemen angesprochen werden (z.B. Religion, Sexualität).

Es wird häufig gefordert, dass Testaufgaben möglichst authentisch gestaltet werden sollten. U.E. sollte das Kriterium der Authentizität bei der Entwicklung von TestDaF nicht zu sehr in den Vordergrund gestellt werden (zu den unterschiedlichen Bedeutungen des Begriffs Authentizität vgl. Grotjahn, 2000b; Lewkowicz, 2000). Eine hohe Authentizität dürfte zwar zur Erhöhung der Akzeptanz des TestDaF bei Testadministratoren und Lehrenden beitragen, sie führt allerdings nicht zwangsläufig auch zu valideren Testergebnissen. Im Fall eines für die Teilnehmer persönlich wichtigen Tests wie dem TestDaF ist vielmehr zu erwarten, dass die Probanden versuchen, auch Aufgaben, die von ihnen als wenig authentisch wahrgenommen werden, möglichst optimal zu lösen. Wie beim TOEFL 2000 (vgl. Chapelle, Grabe & Berns, 1997) steht deshalb zu Recht beim TestDaF die Sicherung der Konstruktvalidität im Vordergrund (vgl. auch Bachman, 2000; Douglas, 2000).

3.1. Leseverstehen

Ziel des Testteils "Leseverstehen" (LV) ist es zu überprüfen, wie weit der Kandidat in der Lage ist, in Bezug auf den akademischen Kontext relevante geschriebene Texte angemessen zu rezipieren. Der Testteil "Leseverstehen" besteht aus drei Aufgaben mit insgesamt 30 Items und dauert 60 Minuten.

Da die Schwierigkeit einer Leseverstehensaufgabe sich nicht allein durch die Textvorlage bestimmt, sondern auch durch die Schwierigkeit der Aufgabenstellung sowie der jeweils zum Text gehörenden Items (vgl. Grotjahn, 2000a), erfolgt die Steuerung der Aufgabenschwierigkeit sowohl über die Texte als auch über die Aufgabenstellung und Einzelitems.

Die Auswahl der Lesetexte orientiert sich an den zwei großen Anwendungsbereichen der Fremdsprache für die Adressaten: Hochschulalltag (soziales Umfeld) und allgemeiner wissenschaftlicher Lehr- und Lernbetrieb. Bei der Messung des Konstrukts "Leseverstehen" werden u.a. folgende Verarbeitungsebenen unterschieden: selektive Informationsentnahme, Verstehen von Hauptaussagen und Einzelinformationen, Verstehen von expliziten Standpunkten und Meinungen sowie komplexe Informationsverarbeitung unter Einschluss des Entnehmens impliziter Bedeutungen und Standpunkte. Die Messung erfolgt anhand von fol-

genden drei Aufgabentypen: In der ersten Aufgabe werden mehrere sprachlich und inhaltlich einfach strukturierte Kurztex te präsentiert; der Aufgabentyp besteht aus einer Zuordnung. Die zweite Aufgabe hat als Textgrundlage einen sprachlich und inhaltlich komplexen Text aus einer Zeitschrift, einer Fachzeitschrift, einem Sachbuch, einem Fachbuch u.Ä.; bei den Items handelt es sich um dreigliedrige Multiple-Choice-Aufgaben (MC-Aufgaben). Im dritten Teil ist jeweils über die Richtigkeit von Aussagen zu entscheiden, die sich auf ein und denselben komplexen Text aus einer wissenschaftlichen Fachzeitschrift oder einem Fachbuch beziehen. Dabei hat der Kandidat die Wahl zwischen “Ja”, “Nein”, “Text sagt dazu nichts”. Wir bezeichnen diese Form im Weiteren als trichotome Auswahlform.

Der Schwierigkeitsgrad der drei Aufgaben verteilt sich aufsteigend über die drei TestDaF-Niveaustufen.

Zur Bewertung der Leseverstehensleistung wird zunächst einmal die Zahl der korrekt gelösten Items ermittelt und mit Hilfe des einparametrischen Rasch-Modells dem Kandidaten ein numerischer Fähigkeitswert zugewiesen. Um den Informationswert und die Transparenz für den Kandidaten und andere Testbenutzer zu erhöhen, wird sodann der ermittelte Wert einer der drei TestDaF-Niveaustufen zugeordnet. Für das TestDaF-Zeugnis sind die Niveaustufen als kurze *can-do statements* mit einem geringen Komplexitäts- und Detailliertheitsgrad formuliert. Diese für den Kandidaten und andere Testbenutzer bestimmten sehr allgemeinen Niveaustufenbeschreibungen sollen im Folgenden als “generische Niveaustufenbeschreibungen” bezeichnet werden. Für das Leseverstehen lauten diese folgendermaßen:⁵

TestDaF-Niveaustufe 3 (TDN 3):

Kann sprachlich und inhaltlich einfach strukturierte schriftliche Texte zu hochschulbezogenen Themen in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen und diesen Texten explizite Bedeutungen und Standpunkte entnehmen.

TestDaF-Niveaustufe 4 (TDN 4):

Kann sprachlich und inhaltlich komplex strukturierte schriftliche Texte zu hochschulbezogenen oder allgemein-wissenschaftlichen Themen in ihrem Gesamtzu-

⁵ Wir beschränken uns im Weiteren auf die Angabe der generischen Niveaustufenbeschreibungen für den jeweiligen Testteil.

Die Projektgruppe TestDaF ist sich des vorläufigen Charakters der vorgelegten Niveaustufenbeschreibungen bewusst. Wir stellen diese hiermit zur Diskussion und bitten um Rückmeldung an die in Anmerkung 1 angegebene Kontaktadresse.

sammenhang und in ihren Einzelheiten verstehen und diesen Texten explizite Bedeutungen und Standpunkte entnehmen.

TestDaF-Niveaustufe 5 (TDN 5):

Kann sprachlich und inhaltlich komplex strukturierte schriftliche Texte zu allgemein-wissenschaftlichen Themen in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen und diesen Texten sowohl explizite als auch implizite Bedeutungen und Standpunkte entnehmen.

Tabelle 1 gibt einen Überblick über das Format des Testteils “Leseverstehen”.

Tabelle 1: Format des Testteils “Leseverstehen”*

	LV 1	LV 2	LV 3
Prüfungsziel	Verstehen von Gesamtzusammenhängen und Einzelinformationen	Verstehen von Gesamtzusammenhängen und Einzelinformationen	Komplexe Informationsverarbeitung (u.a. Entnehmen impliziter Bedeutungen und Standpunkte)
Themenauswahl	Hochschulbezogene Themen	Hochschulbezogene oder allgemein-wissenschaftliche Themen	Allgemein-wissenschaftliche Themen
Texttyp/ Textquelle	Sprachlich und inhaltlich einfach strukturierte Texte (Programm, Katalog, Vorlesungsverzeichnis o.Ä.; insg. 8 Kurztexte)	Sprachlich und inhaltlich komplexe Texte (Zeitschrift, Sachbuch, Fachbuch o.Ä.)	Sprachlich und inhaltlich komplexe Texte (Fachzeitschrift, Fachbuch o.Ä.)
Autor	Diverse	Journalist, Hochschullehrer, Wissenschaftler	Experten, Wissenschaftler, Hochschullehrer
Itemtyp	Zuordnung	Multiple Choice	Trichotome Auswahlform (ja, nein, Text sagt dazu nichts)
Zahl der Items	10	10	10
Stufe	TDN 3	TDN 4	TDN 5

*leicht modifizierte Fassung einer Vorlage von S. Bolton und R. Hofmann, Goethe-Institut München

In den ersten beiden Testsätzen wurde auch ein *summary cloze test* aufgrund der zumeist positiven Bewertung in der Fachliteratur als Aufgabentyp eingesetzt. In der Erprobung stellte sich jedoch heraus, dass ein Großteil der Lücken der vorgelegten *summary clozes* – zumindest durch Muttersprachler – ohne Kenntnis des Ausgangstextes gelöst werden konnten. Es wurde deshalb entschieden, auf dieses Testformat zu verzichten und anstelle dessen zur Überprüfung des Leseverstehens MC-Aufgaben zu verwenden. Dabei wurde darauf ge-

achtet, die für ein tiefer gehendes Textverständnis wichtigen Makrostrukturen zu erfassen.

Insbesondere von Praktikern wird zwar immer wieder bezweifelt, dass mit MC-Aufgaben das Verstehen von Makrostrukturen in valider Weise gemessen werden kann. In der testtheoretischen Literatur existieren allerdings eine Reihe von Untersuchungen, in denen u.a. anhand eines Vergleichs unterschiedlicher Aufgabenformate gezeigt worden ist, dass es durchaus möglich ist, mittels MC-Aufgaben wichtige Aspekte makrostrukturellen Verstehens zu erfassen (vgl. z.B. Bennett & Ward, 1993).

3.2. Hörverstehen

Ziel des Testteils “Hörverstehen” (HV) ist es zu überprüfen, wie weit der Kandidat in der Lage ist, in Bezug auf den akademischen Kontext relevante gesprochene Texte angemessen zu rezipieren. Der Testteil “Hörverstehen” besteht aus drei Aufgaben mit insgesamt 25 Items. Der eigentliche Hörverstehentest dauert 30 Minuten; hinzu kommen 10 Minuten für das Übertragen der Lösungen auf das Antwortblatt durch den Kandidaten.

Bei der Messung des Konstrukts “Hörverstehen” werden u.a. folgende Verarbeitungsebenen unterschieden: selektive Informationsentnahme, Verstehen von Hauptaussagen und Einzelinformationen, Verstehen von Standpunkten und Meinungen sowie komplexe Informationsverarbeitung unter Einschluss der Entnahme impliziter Bedeutungen. Die Messung erfolgt anhand von drei verschiedenen Textsorten und Aufgabentypen: Im ersten Teil wird ein Dialog aus dem studentischen Alltagsleben präsentiert; der Aufgabentyp besteht aus dem Notieren von Stichwörtern. Im zweiten Teil besteht die Textgrundlage aus einem Interview zu hochschulbezogenen Handlungsfeldern oder allgemein-wissenschaftlichen Themen; beim Aufgabentyp handelt sich um eine Alternativform vom Typ “richtig”/“falsch”. Dem dritten Teil liegt ein Text mit komplexen monologischen Passagen bzw. ein komplexer monologischer Text zu Grunde; gesteuerte Notizen bilden den Aufgabentyp. Die drei Hörverstehenteile steigen, wie beim Leseverstehen, im Schwierigkeitsgrad an.

Tabelle 2 gibt einen Überblick über das Format des Testteils “Hörverstehen”.

Tabelle 2: Format des Testteils “Hörverstehen”*

	HV 1	HV 2	HV3
Prüfungsziel	Verstehen von Gesamt-zusammenhängen und Einzelinformationen	Verstehen von Gesamt-zusammenhängen und Einzelinformationen	Komplexe Informationsverarbeitung (u.a. Entnehmen impliziter Bedeutungen und Standpunkte)
Themenauswahl	Hochschulbezogene Themen	Hochschulbezogene oder allgemein-wissenschaftliche Themen	Allgemein-wissenschaftliche Themen
Texttyp/ Textquelle	Sprachlich und inhaltlich einfach strukturiertes Gespräch	Interview, Gesprächsrunde (komplexe Darstellung von Sachverhalten, Erfahrungen, Meinungen o.Ä.)	Komplexer monologischer Text oder Interview mit komplexen monologischen Textpassagen
Sprecher	Personen aus dem Umfeld Hochschule	Studenten, Dozenten	Experten, Wissenschaftler, Hochschullehrer
Hörfrequenz	1-mal hören	1-mal hören	2-mal hören
Itemtyp	Durch Vorgaben gesteuertes Notieren von Stichwörtern	Alternativform (Richtig/Falsch)	Durch Vorgaben gesteuertes Notieren von Kurzantworten und Stichwörtern
Zahl der Items	8	10	7
Stufe	TDN 3	TDN 4	TDN 5

*leicht modifizierte Fassung einer Vorlage von S. Bolton und R. Hofmann, Goethe-Institut München

Analog zum Leseverstehen wird bei der Bewertung der Hörverstehensleistung zunächst einmal die Zahl der korrekt gelösten Items ermittelt und mit Hilfe des einparametrischen Rasch-Modells dem Kandidaten ein numerischer Fähigkeitswert zugewiesen. Anschließend wird der ermittelte Wert einer der drei TestDaF-Niveaustufen zugeordnet. Die generischen Beschreibungen der Niveaustufen lauten für das Hörverstehen folgendermaßen:

TestDaF-Niveaustufe 3 (TDN 3):

Kann sprachlich und inhaltlich einfach strukturierte gesprochene Texte zu hochschulbezogenen Themen in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen und diesen Texten explizite Bedeutungen und Standpunkte entnehmen, sofern klar und deutlich gesprochen wird und keine akustischen Störfaktoren vorhanden sind.

TestDaF-Niveaustufe 4 (TDN 4):

Kann sprachlich und inhaltlich komplex strukturierte gesprochene Texte zu hochschulbezogenen oder allgemein-wissenschaftlichen Themen in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen und diesen Texten explizite Bedeutungen und Standpunkte entnehmen.

TestDaF-Niveaustufe 5 (TDN 5):

Kann sprachlich und inhaltlich komplex strukturierte gesprochene Texte zu allgemein-wissenschaftlichen Themen in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen und diesen Texten sowohl explizite als auch implizite Bedeutungen und Standpunkte entnehmen.

3.3. Schriftlicher Ausdruck

Ziel des Testteils “Schriftlicher Ausdruck” (SA) ist es zu überprüfen, wie weit der Kandidat in der Lage ist, sich im akademischen Kontext schriftlich angemessen zu äußern. Der Testteil “Schriftlicher Ausdruck” besteht aus einer zweiteiligen Aufgabe und dauert 60 Minuten.

Dem Kandidaten wird die Aufgabe gestellt, auf der Basis von Vorgaben die studienrelevanten Schreibhandlungen “Beschreibung” und “Argumentation” zu realisieren. Die Vorgaben beziehen sich jeweils auf die gleiche übergeordnete Thematik. Als Vorgabe für die Beschreibung dienen Graphiken, Tabellen oder Diagramme. In der Aufgabenstellung wird angegeben, wie der Kandidat bei der Beschreibung vorgehen soll. Bei der Vorgabe für die Argumentation ist der Kandidat gehalten, zu einem Aspekt eines bestimmten Themas verschiedene Positionen abzuwägen und seinen eigenen Standpunkt zu begründen. Die Vorgabe für die Argumentation besteht aus Kurztexten, gegensätzlichen Thesen oder Zitaten.

Den Aufgabenteilen “Beschreibung” und “Argumentation” wird – im Gegensatz zum Vorgehen bei den übrigen Testteilen – kein unterschiedlicher Schwierigkeitsgrad zugewiesen. In Bezug auf den vom Kandidaten realisierten Text wird allerdings zwischen unterschiedlichen Leistungsmerkmalen (z.B. Korrektheit der verwendeten Strukturen, Breite des Wortschatzes, Grad der Textualität, Situationsangemessenheit) differenziert und auf dieser Basis eine Zuordnung zu den Leistungsstufen vorgenommen.

Der Schriftliche Ausdruck wird nach folgenden drei Kriterien bewertet:⁶

⁶ Die Bewertungsanleitungen für den Schriftlichen Ausdruck wurden von B. Gutzat, FernUniversität Hagen, erstellt.

a) Gesamtwirkung

Dieses Kriterium erfasst den Text als Ganzes in seiner Wirkung auf den Rezipienten. Die Bewertung erfolgt danach, inwieweit ein muttersprachiger Rezipient den Text flüssig lesen und dem Gedankengang folgen kann.

b) Behandlung der Aufgabe

Dieses Kriterium erfasst die gedankliche Organisation des Textes sowie die Ausführlichkeit und Komplexität, mit der die Aufgabenstellung behandelt wird. Die Bewertung erfolgt danach, inwieweit der Text logisch aufgebaut sowie inhaltlich und sprachlich angemessen ist in Bezug auf Adressat, Situation und Textsorte.

c) Sprachliche Realisierung

Dieses Kriterium erfasst die sprachlichen Mittel des Textes. Dazu gehören das Maß an Kohärenz und Ausdrucksfähigkeit sowie die Richtigkeit der verwendeten Sprachmittel. Die Bewertung erfolgt nach der Breite und Korrektheit der eingesetzten sprachlichen Mittel.

Eine detaillierte Bewertungsanleitung definiert für jedes dieser drei Kriterien die sprachliche Leistung, die auf den drei TestDaF-Niveaustufen erwartet wird.

Verglichen mit den Bewertungsanleitungen für die Korrektoren haben die für den Kandidaten und andere Testabnehmer bestimmten generischen Niveaustufenbeschreibungen einen weit geringeren Detaillierungsgrad. Sie lauten für den Schriftlichen Ausdruck:

TestDaF-Niveaustufe 3 (TDN 3):

Kann im universitären Alltag sowie in allgemein-wissenschaftlichen Kontexten schriftlich so handeln, dass eine der Schreibintention entsprechende Rezeption des Textes nicht behindert wird.

TestDaF-Niveaustufe 4 (TDN 4):

Kann im universitären Alltag sowie in allgemein-wissenschaftlichen Kontexten intentions- und situationsangemessen schriftlich handeln.

TestDaF-Niveaustufe 5 (TDN 5):

Kann im universitären Alltag sowie in allgemein-wissenschaftlichen Kontexten intentions- und situationsangemessen sowie klar strukturiert schriftlich handeln.

Tabelle 3 gibt einen Überblick über das Format des Testteils "Schriftlicher Ausdruck".

Tabelle 3: Format des Testteils “Schriftlicher Ausdruck”*

	Prüfungsziel	Vorgabe	Aufgabentyp
Beschreibung (ca. 20 Minuten)	1. Beschreiben von Diagrammen und Tabellen: <ul style="list-style-type: none"> • Verbalisierung der Informationen • Vergleich von Diagrammen und Tabellen 2. Beschreiben von Abläufen/Vorgängen	<ul style="list-style-type: none"> • Diagramme (Kurven-, Säulen-, Tortendiagramme) und Tabellen • Schematische Darstellung von Experimenten u.Ä. in Form von Flussdiagrammen oder Sequenzen von Bildern 	Durch Aufgabenstellung gesteuertes freies Schreiben
Argumentation (ca. 40 Minuten)	Stellungnahme zu kontroversen Fragen mit Darstellung und Begründung verschiedener Standpunkte	Kurztexte, Thesen, Zitate, Statements	Durch Aufgabenstellung gesteuertes freies Schreiben

*leicht modifizierte Fassung einer Vorlage von B. Gutzat, FernUniversität Hagen

3.4. Mündlicher Ausdruck

Ziel des Testteils “Mündlicher Ausdruck” (MA) ist es zu überprüfen, wie weit der Kandidat in der Lage ist, sich im akademischen Kontext mündlich angemessen zu äußern. Der Testteil “Mündlicher Ausdruck” besteht aus insgesamt 10 Aufgaben und dauert 30 Minuten.

Für den Prüfungsteil “Mündlicher Ausdruck” hatte die Forderung nach zentraler Korrektur und Standardisierung die weitgehendsten Auswirkungen auf das Testformat: Es wird kein face-to-face-Interview – wie z.B. das *Oral Proficiency Interview* (OPI) des *American Council on the Teaching of Foreign Languages* – verwendet, sondern ein dem *Simulated Oral Proficiency Interview* (SOPI) des *Center for Applied Linguistics* in Washington angelehnter kassettengesteuerter mündlicher Test. Das SOPI wird seit ca. 10 Jahren eingesetzt und ist empirisch vielfach überprüft worden (vgl. z.B. Kenyon, 2000 sowie auch Konsortialpartner TestDaF, 2000).

Der Testteil “Mündlicher Ausdruck” kann insgesamt gesehen wie folgt beschrieben werden:

- Die Kandidaten hören die Testaufgaben von einer Kassette und lesen sie zugleich in ihrem Aufgabenheft mit. Im ersten Teil der Aufgabe wird jeweils die Situation beschrieben, in die sich der Testkandidat hineinversetzen soll,

und es wird gesagt, wie dieser in der Situation sprachlich handeln soll. Dabei wird in einer Reihe von Aufgaben Bezug auf visuelle Stimuli (z.B. Graphiken) genommen. Nach einer Denkzeit spricht ein deutschsprachiger Gesprächspartner (Stimulus). Diese Äußerungen sind **nicht** im Aufgabenheft abgedruckt. Danach spricht der Kandidat. Zu jeder Aufgabe gibt es Angaben zu der dem Kandidaten zur Verfügung stehenden Denk- und Sprechzeit.

- Die Äußerungen der Kandidaten werden auf einer zweiten Kassette aufgenommen, die zur zentralen Korrektur eingeschickt wird. Die Korrektur erfolgt durch zwei Korrektoren; bei einer unterschiedlichen Beurteilung wird ein dritter Korrektor hinzugezogen.
- Der Testteil ist in vier inhaltliche Aufgabenbereiche (Einleitungssequenz als warm-up, situationsbezogene Sprechhandlungen, Beschreibung, Argumentation) mit Unteraufgaben aufgeteilt.
- Die Unteraufgaben sind jeweils in aufsteigender Schwierigkeit angeordnet.

Dem simulierten Interview fehlen sicherlich wichtige Merkmale genuiner mündlicher Kommunikation. Durch die Standardisierung des Inputs und durch die zentrale Korrektur der Tonträger ist jedoch die Objektivität und Reliabilität des Verfahrens und damit – zumindest potentiell – auch dessen Validität höher als die eines traditionellen mündlichen Interviews (vgl. auch Grotjahn, 2000b). So müsste im Fall eines face-to-face-Interviews u.a. auch gewährleistet sein, dass weltweit eine genügende Zahl an kompetenten Interviewern zur Verfügung steht. Außerdem ist das SOPI, vor allem wenn es als Gruppenprüfung z.B. im Sprachlabor durchgeführt werden kann, ökonomischer als ein face-to-face-Interview.

Allerdings hat sich bei der Testerprobung gezeigt, dass in einigen Ländern an immer weniger Institutionen Sprachlabore zur Verfügung stehen. Sofern nur wenige Kandidaten zu prüfen sind, bietet sich als Ausweg der gleichzeitige Einsatz mehrerer Sätze von hochwertigen Kassettenrekordern an. In Zukunft ist auch eine computerunterstützte Administration des Mündlichen Ausdrucks vorgesehen.

Der Mündliche Ausdruck wird analog zum Schriftlichen Ausdruck nach folgenden drei Kriterien bewertet:⁷

a) Gesamtwirkung

Dieses Kriterium erfasst die Äußerung als Ganzes in ihrer Wirkung auf den Hörer. Dazu gehören Flüssigkeit, Aussprache, Intonation und die gedankliche Or-

⁷ Die Bewertungsanleitungen für den Mündlichen Ausdruck wurden von G. Kniffka, Carl Duisberg Centren Köln, erstellt.

ganisation. Die Bewertung erfolgt danach, inwieweit ein muttersprachiger Hörer die Äußerung aufgrund ihrer phonetischen Realisierung verstehen und inwieweit er dem Gedankengang folgen kann.

b) Umsetzung der Aufgabenstellung

Dieses Kriterium erfasst die Realisierung der durch die Aufgabenstellung geforderten Sprechhandlung. Die Bewertung erfolgt danach, inwieweit die Äußerung inhaltlich und sprachlich in Bezug auf Adressat, Situation und Textsorte angemessen ist. Der Gebrauch/Einsatz von Kompensationsstrategien wird positiv berücksichtigt, wenn sie eingesetzt werden, um sprachliche oder inhaltliche Defizite zu umgehen/auszugleichen. Werden keine Kompensationsstrategien verwendet, so wird dies nicht negativ bewertet.

c) Sprachliche Realisierung

Dieses Kriterium erfasst die sprachlichen Mittel der Äußerung. Dazu gehören das Maß an Ausdrucksfähigkeit und die Richtigkeit der verwendeten Sprachmittel. Die Bewertung erfolgt nach der Breite und Korrektheit der eingesetzten sprachlichen Mittel.

Interessant ist beim Testteil “Mündlicher Ausdruck” insbesondere, dass in der Bewertung davon ausgegangen wird, dass die Leistung der Kandidaten bei den einzelnen Aufgaben maximal der jeweiligen Stufe (d.h. TDN 3, TDN 4, TDN 5), die der Aufgabe zugeordnet worden ist, entsprechen kann. Bei einer leichten Aufgabe, d.h. einer TDN 3-Aufgabe, oder einer mittelschweren, d.h. einer TDN 4-Aufgabe, kann also immer nur festgestellt werden, dass der Kandidat diese Aufgabe adäquat gelöst hat. Er kann deshalb in dem vorgestellten Fall maximal die Niveaustufe TDN 3 bzw. TDN 4 erreichen, nicht jedoch TDN 5.

Bei der Auswertung der Aufnahmen wird aus Ökonomiegründen folgendermaßen vorgegangen: Zunächst wird überprüft, ob der Kandidat die TDN 5-Aufgaben adäquat gelöst hat. Ist dies der Fall, dann bleiben die weiteren Aufgabenteile unberücksichtigt. Anderenfalls wird überprüft, ob der Kandidat die TDN 4-Aufgaben adäquat gelöst hat. Trifft dies zu, bleiben die TDN 3-Aufgaben unberücksichtigt. Trifft dies nicht zu, wird überprüft, ob der Kandidat die TDN 3-Aufgaben adäquat gelöst hat. Ist dies der Fall, erhält er die Einstufung ‘TDN 3’, anderenfalls die Einstufung ‘unter TDN 3’.

Die generischen Niveaustufenbeschreibungen lauten für den Mündlichen Ausdruck wie folgt:

TestDaF-Niveaustufe 3 (TDN 3):

Kann im universitären Alltag sowie in allgemein-wissenschaftlichen Kontexten sprachlich handeln, ohne dass die Kommunikation behindert wird.

TestDaF-Niveaustufe 4 (TDN 4):

Kann im universitären Alltag sowie in allgemein-wissenschaftlichen Kontexten intentions- und situationsangemessen sprachlich handeln.

TestDaF-Niveaustufe 5 (TDN 5):

Kann im universitären Alltag sowie in allgemein-wissenschaftlichen Kontexten intentions- und situationsangemessen sowie flüssig und spontan sprachlich handeln.

Tabelle 4 gibt einen Überblick über das Format des Testteils “Mündlicher Ausdruck”.

4. Erprobung der Modellprüfungssätze

Bevor die Testsätze in die eigentliche Erprobung gingen, wurden sie durch die Konsortialpartner mit ausländischen Kursteilnehmern bzw. ausländischen Studenten sowie mit muttersprachigen Studenten vorerprobt, um bereits in diesem Stadium problematische Items identifizieren, herausnehmen bzw. verbessern zu können.

Die Erprobung der ersten drei Testsätze erfolgte mit jeweils ca. 200 Probanden für die (relativ) geschlossenen Aufgabenteile LV und HV, für die eher freien Aufgabenteile SA und MA mit ca. 120 Probanden im In- und Ausland. Partner für die Erprobung waren z.B. im Inland: RWTH Aachen, Lehrgebiet DaF; TU Dresden, Lehrbereich DaF; Universität Göttingen, Lehrgebiet DaF; Universität Kaiserslautern, Akademisches Auslandsamt; Universität Mainz, Sprachlehranlage; Universität Münster, Lehrgebiet DaF; Studienkolleg bei den wissenschaftlichen Hochschulen des Freistaates Bayern; Studienkolleg der Universität Potsdam. Im Ausland erfolgte die Erprobung u.a. an folgenden Orten: Azhar Universität, Kairo (Ägypten); TU Sofia (Bulgarien); Zweite Fremdsprachenschule Beijing (China); University of Oxford (England); Nehru Universität, Neu Delhi (Indien); Universitas Indonesia, Jakarta (Indonesien); Pusan National University (Korea); Universidad de Guadalajara (Mexiko); Universität Minsk (Weißrussland); Studienzentrum der FernUniversität Hagen, Budapest (Ungarn).

Zur Erhöhung der Reliabilität wurden für die Testteile Schriftlicher Ausdruck und Mündlicher Ausdruck Ratertrainings (Training der Korrektoren) durchgeführt; die Resultate der Ratings wurden jeweils einer kritischen Analyse unterzogen.

Tabelle 4: Format des Testteils “Mündlicher Ausdruck”*

	Prüfungsziel	Vorgabe	Aufgabentyp
Teil 1: Einleitungssequenz (“Warm-up”) (1 Aufgabe)	Ein Anliegen vortragen	Situativ eingebettete Fragen	Durch Aufgabenstellung gesteuertes freies Sprechen
Teil 2: Situationsbezogene Sprechhandlungen (4 Aufgaben)	<ol style="list-style-type: none"> In (teilweise standardisierten) Situationen des universitären Alltags Informationen einholen/geben In nicht-konventionalisierten Situationen des universitären Alltags angemessen agieren und reagieren (z.B. jdn um etwas bitten, jdn überreden o.Ä.) 	<ul style="list-style-type: none"> Anzeigen Auszüge aus Vorlesungsverzeichnissen Aushänge u.Ä. kurze Darstellung einer Situation Lagepläne u.Ä. 	Durch Aufgabenstellung gesteuertes freies Sprechen
Teil 3: Beschreibung (2 Aufgaben)	<ol style="list-style-type: none"> Verbalisierung der in einer Graphik/einem Diagramm enthaltenen Informationen Verbalisierung von schematisch dargestellten Abläufen und Vorgängen 	<ul style="list-style-type: none"> Statistische Graphiken: Säulen-, Kurven-, Tortendiagramme Sequenz von Bildern, Flussdiagramme 	Durch Aufgabenstellung gesteuertes freies Sprechen
Teil 4: Argumentation (3 Aufgaben)	<ol style="list-style-type: none"> Meinung äußern und begründen Aus Alternativen eine wählen und diese Wahl begründen Begründete Stellungnahme zu (komplexem) Sachverhalt geben 	<ul style="list-style-type: none"> Kurze Darstellung eines Sachverhaltes Kurze Darstellung eines Sachverhaltes, zu dem es mehrere Handlungsalternativen gibt Kurze Darstellung eines Sachverhaltes 	Durch Aufgabenstellung gesteuertes freies Sprechen

*leicht modifizierte Fassung einer Vorlage von G. Kniffka, Carl Duisberg Centren Köln

5. Testmethodische Begleitung

Bei allen Erprobungsphasen wurde den Teilnehmern ein Fragebogen für die wissenschaftliche Begleitforschung vorgelegt. Mit diesem wurden Informationen zur qualitativen Ergänzung der statistischen Itemanalyse sowie Daten zur Beschreibung der Stichprobe erhoben. Die Prüfer erhielten einen Protokollbogen zur Erfassung der Durchführungsmodalitäten. Die erhobenen Daten wurden statistisch ausgewertet.

Ergänzend wurden die im Rahmen der wissenschaftlichen Begleitforschung erhobenen Kommentare der Teilnehmer zu den einzelnen Aufgaben und Items ausgewertet.

Für das Lese- und Hörverstehen wurden in Zusammenarbeit mit der *University of Cambridge Local Examinations Syndicate* (UCLES)⁸ Rasch-Analysen sowie klassische Itemanalysen durchgeführt.

Für die produktiven Testteile Mündlicher und Schriftlicher Ausdruck wurde die Auswertungsobjektivität anhand von Auswerter-Übereinstimmungen überprüft.

Weiterhin wurden jeweils mehrere externe Gutachten eingeholt. Bei den Gutachtern handelte es sich um internationale Testexperten und um mit der DSH befasste Wissenschaftler. Es wurde dabei explizit darauf geachtet, dass auch gegenüber dem TestDaF-Projekt kritisch eingestellte Wissenschaftler in die Begutachtung einbezogen wurden. Insgesamt gesehen wurde TestDaF überwiegend positiv beurteilt. Dies gilt insbesondere für die internationalen Testexperten.

Auf der Grundlage der qualitativen und quantitativen Informationen wurde bei der Revision z.B. entschieden, welche Items jeweils beibehalten werden konnten und welche zu revidieren waren.

6. Informatik

Zu den Aufgaben der Informatik im TestDaF-Projekt gehört die Realisierung einer informationstechnologischen Infrastruktur für die Testerprobung und den anschließenden Routinebetrieb. Dabei ist ein zweigleisiges Vorgehen gefordert, das zunächst die Durchführung des Tests auf Papier und mit Audiokassetten unterstützt, mittelfristig aber auch den Einsatz computerbasierter oder auch computeradaptiver Tests vorsieht. Die informationstechnologische Infrastruktur besteht im Wesentlichen aus drei Komponenten (vgl. auch Six, 2000):

⁸ UCLES hat eine Vielzahl (internationaler) Sprachtests entwickelt – darunter auch den IELTS.

- **Itembank.** In der Itembank werden Testsätze, Tests und Fragebögen zu den Tests gespeichert. Eine Aufgabe oder ein Item kann durch Attribute charakterisiert werden, die z.B. zur Aufgabensuche und Testzusammenstellung herangezogen werden können.
- **Teilnehmerdatenbank.** Hier werden die Stammdaten der Teilnehmer wie Name, Anschrift, Testdatum usw. gesammelt.
- **Testdatenbank, auch Testdurchführungsdatenbank.** Hier befinden sich die Testantworten einschließlich der Bewertung, die Fragebogenantworten und die Ergebnisse der Itemanalysen.

Für alle Komponenten wird eine einheitliche, leicht zu bedienende web-basierte Benutzungsoberfläche entwickelt, womit nach aktuellem Stand der Technik auch größtmögliche Flexibilität und Unabhängigkeit von bestimmten Software-Produkten gewährleistet ist.

7. Perspektiven

Es ist ein zentrales Testinstitut als organisatorischer Rahmen für alle im Zusammenhang mit TestDaF künftig anfallenden Arbeiten aufzubauen. Das Testinstitut hat Aufgaben zu erfüllen wie z.B. die kontinuierliche Erstellung neuer Testsätze und deren Erprobung, Erstellung von Begleit- und Vorbereitungsmaterialien, kontinuierliche Begleitforschung zur Sicherung der Qualität, Fortbildung von Testautoren und Testkorrektoren, Weiterentwicklung der Informatikinfrastruktur sowie Entwicklung einer computerbasierten/-adaptiven Testform.

Da zukünftige Testkandidaten schon in ihrem Heimatland TestDaF ablegen sollen, müssen ferner weltweit Testzentren aufgebaut werden, die durch das zentrale Testinstitut zu akkreditieren sind.

Der erste offizielle Einsatz von TestDaF ist im Winter 2000 vorgesehen.

Eingang des revidierten Manuskripts: 14.3.2000

Literaturverzeichnis

- Association of Language Testers in Europe (ALTE). (1998). *ALTE Handbuch europäischer Sprachprüfungen und Prüfungsverfahren*. Cambridge: The University of Cambridge Local Examinations Syndicate.
- Bachman, Lyle. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Bennett, Randy E. & Ward, William C. (Hrsg.). (1993). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*. Hillsdale, NJ: Erlbaum.

- Bickes, Gerhard. (1998). DSH – TESTDAF – TOEFL: Thesen zu einer Umgestaltung der Prüfungspraxis in Deutsch als Fremdsprache für ausländische Studienbewerber. *Info DaF*, 25(1), 97-103.
- Bolton, Sibylle. (2000). TESTDAF. In Armin Wolff & Harald Tanzer (Hrsg.), *Sprache – Kultur – Politik*. Regensburg: Fachverband Deutsch als Fremdsprache. (Materialien Deutsch als Fremdsprache, Bd. 53). (im Druck)
- Chapelle, Carol A., Grabe, William & Berns, Margie. (1997). *Communicative language proficiency: Definition and implications for TOEFL 2000* (TOEFL Monograph Series MS-10). Princeton, NJ: Educational Testing Service.
- Coste, Daniel, North, Brian, Sheils, Joseph & Trim, John L.M. (1998). *Modern languages: Learning, teaching, assessment. A common European framework of reference*. Strasbourg: Council of Europe.
- Das Konsortium für das TESTDAF-Projekt. (2000). TESTDAF: eine Zwischenbilanz (Stand: November 1999). *Info DaF*, 27(1), 68-71. auch in: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [on-line], 4(3) [<http://www.ualberta.ca/~german/ejournal/projekte.htm>]
- Douglas, Dan. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Grotjahn, Rüdiger. (2000a). Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TestDaF. In Sibylle Bolton (Hrsg.), *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar* (S. 7-55). Köln: VUB-Gilde.
- Grotjahn, Rüdiger. (2000b). Testtheorie: Grundzüge und Anwendungen in der Praxis. In Armin Wolff & Harald Tanzer (Hrsg.), *Sprache – Kultur – Politik*. Regensburg: Fachverband Deutsch als Fremdsprache. (Materialien Deutsch als Fremdsprache, Bd. 53). (im Druck)
- Kenyon, Dorry M. (2000). Tape-mediated oral proficiency testing: Considerations in developing Simulated Oral Proficiency Interviews (SOPIs). In Sibylle Bolton (Hrsg.), *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*. Köln (S. 87-106). VUB-Gilde.
- Konsortialpartner TestDaF. (Redaktion). (2000). *TestDaF: Hinweise für Testautoren*. München: Goethe-Institut. (in Vorbereitung)
- Lee, Wonkyung. (1998). *Prüfungen zum Nachweis deutscher Sprachkenntnisse bei ausländischen Studienbewerbern/Studienbewerberinnen (PNdS/DSH): Ihre Praxis und ihr Prüfprofil*. Regensburg: Fachverband Deutsch als Fremdsprache.
- Lewkowicz, Jo A. (2000). Authenticity in language testing: Some outstanding questions. *Language Testing*, 17(1), 43-64.
- McNamara, Tim F. (1997). Performance testing. In Caroline Clapham & David Corson (Hrsg.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment* (S. 131-139). Dordrecht: Kluwer.
- Six, Hans-Werner. (2000). Informatikunterstützung für TESTDAF. In Sibylle Bolton (Hrsg.), *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar* (S. 131-140). Köln: VUB-Gilde.
- Wintermann, Bernd. (1998). DSH und TESTDAF – Sprachprüfungen auf dem Prüfstand. *Info DaF*, 25(1), 104-110.