

# **TestDaF: Stand der Entwicklung und einige Perspektiven für Forschung und Praxis\***

Rüdiger Grotjahn (Bochum) & Karin Kleppin (Leipzig)

## **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 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 in wesentlichen Teilen die gemeinsame Entwicklungsarbeit des Konsortiums wider.

Die erste Entwicklungsphase des TestDaF-Projekts dauerte vom 1. August 1998 bis zum 31. Juli 2000. In dieser Zeit wurden drei Modellprüfungssätze entwickelt und in großem Umfang weltweit erprobt. Testerstellung und Erprobung wurden testmethodisch begleitet, d.h., die Lösungen der Erprobungskandidaten wurden statistisch analysiert, Fragebögen für die Testerprober und für die Kandidaten erstellt und ausgewertet. Diese Daten sowie die eingeholten wissenschaftlichen Gutachten bildeten die Basis für die Revision der Testsätze. Außerdem wurden eine Reihe von Informationsmaterialien und Handreichungen verfasst (vgl. Projektgruppe TestDaF 2000a, b, c, d). Der erste offizielle Einsatz von TestDaF ist für 2000/2001 vorgesehen.

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.

---

\* Bei den Kapiteln 1, 2, 3, 4 und 6 des vorliegenden Textes handelt es sich im Wesentlichen um eine gekürzte und geringfügig aktualisierte Fassung von Projektgruppe TestDaF (2000e). Der Projektgruppe gehören an: Hans-Georg Albers, Karl-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 Gesamthochschule in Hagen, dem Seminar für Sprachlehrforschung der Ruhr-Universität Bochum, dem Goethe-Institut München und den Carl Duisberg Centren Köln.

TestDaF wurde z.T. sehr kontrovers diskutiert (vgl. exemplarisch Bickes 1998; Wintermann 1998), vor allem auch aufseiten von Praktikern.

## **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 im jeweiligen 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 im Bereich Deutsch als Fremdsprache an den Universitäten sollen überschaubarer und transparenter werden; denn trotz Verbesserungen in den letzten Jahren konnte die jetzt gültige DSH (Deutsche Sprachprüfung für den Hochschulzugang ausländischer Studienbewerber) 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 die 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 Bewerber 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-Niveau-

stufe 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. 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), die eng aufeinander bezogen und zudem in Teilen bereits validiert sind (vgl. ALTE 1998: Kap. 2; Coste, North, Sheils & Trim 1998: Kap. 8). Der TestDaF-Spezifik wird u.a. dadurch Rechnung getragen, dass der akademische Kontext bei den Stufenbeschreibungen eine stärkere Berücksichtigung findet.

Für jede Niveaustufe werden in Abhängigkeit von den jeweiligen Zielsetzungen und Adressaten parallele Beschreibungen mit einem unterschiedlichen Grad an Komplexität und Detailliertheit angefertigt.

- Die TestDaF-Stufen TDN 3, TDN 4 und TDN 5 umfassen in etwa folgenden Kompetenzbereich: Bezogen auf die Europaratsstufen 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 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 zuzuordnen ist. Aufgrund des dezentralen und wenig standardisierten Charakters der DSH und den hieraus resultierenden Unterschieden in der Schwierigkeit einzelner DSH-Prüfungen dürften in der Praxis allerdings nicht wenige DSH-Prüfungen 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 Schwierigkeitsgrades von TestDaF an den Anforderungen der DSH höchst problematisch (vgl. auch Bickes 1998; Wintermann 1998).
- 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 dann keine eindeutigen Aussagen möglich sind, 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 bzw. eine computeradaptive Testversion erstellt werden.

#### **4. 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 Antwortblatt durch die Kandidaten. Die Themen werden den Zielen des Tests entsprechend z.B. aus folgenden Bereichen entnommen: Studium und Umfeld, 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 potenziellen Tabuthemen angesprochen werden (z.B. Religion, Sexualität).

Es wird häufig gefordert, dass Testaufgaben möglichst authentisch gestaltet werden sollten. Unseres Erachtens 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 muss allerdings nicht gleichzeitig auch zu valideren Testergebnissen führen. 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).

##### **4.1 Leseverstehen**

Ziel des Testteils „Leseverstehen“ (LV) ist es zu überprüfen, inwieweit der Kandidat in der Lage ist, im akademischen Kontext relevante geschriebene Texte angemessen zu rezipieren. Der Testteil „Leseverstehen“ besteht aus drei Aufgaben mit insgesamt 30 Items und dauert 60 Minuten. Der Schwierigkeitsgrad der drei Aufgaben verteilt sich aufsteigend über die drei TestDaF-Niveaustufen.

Da die Schwierigkeit einer Leseverstehensaufgabe sich nicht allein durch die Textvorlage, sondern auch durch die Schwierigkeit der Aufgabenstellung sowie der jeweils zum Text gehörenden Items bestimmt (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 Informationen. Die Messung erfolgt anhand von drei verschiedenen Aufgaben: 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“.

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 MC-Aufgaben durchaus in der Lage sind, wichtige Aspekte makrostrukturellen Verstehens zu erfassen (vgl. z.B. Bennett & Ward 1993).

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. In ihrer aktuellen Form – als *can-do statements* mit einem geringen Komplexitäts- und Detailliertheitsgrad formuliert – lauten diese für das Leseverstehen folgendermaßen:

**TestDaF-Niveaustufe 5 (TDN 5):**

Kann geschriebene Texte aus dem studienbezogenen Alltag (z.B. Material zur Studienorganisation) sowie Texte zu fächerübergreifenden wissenschaftlichen Themen (z.B. allgemeine Umweltprobleme, gesellschaftspolitische Fragestellungen), die sprachlich und inhaltlich komplex strukturiert sind, in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen und diesen Texten auch implizite Informationen entnehmen.

**TestDaF-Niveaustufe 4 (TDN 4):**

Kann geschriebene Texte aus dem studienbezogenen Alltag (z.B. Material zur Studienorganisation) sowie Texte zu fächerübergreifenden wissenschaftlichen Themen (z.B. allgemeine Umweltprobleme, gesellschaftspolitische Fragestellungen), deren

Struktur sich an der Allgemeinsprache orientiert, in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen.

**TestDaF-Niveaustufe 3 (TDN 3):**

Kann geschriebene Texte aus dem studienbezogenen Alltag (z.B. Material zur Studienorganisation) in ihrem Gesamtzusammenhang und in wesentlichen Einzelheiten verstehen; kann jedoch Texte zu fächerübergreifenden wissenschaftlichen Themen (z.B. allgemeine Umweltprobleme, gesellschaftliche Fragestellungen) nicht hinreichend verstehen.

## **4.2 Hörverstehen**

Ziel des Testteils „Hörverstehen“ (HV) ist es zu überprüfen, inwieweit der Kandidat in der Lage ist, im 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 die Kandidaten.

Die Messung des Konstrukts „Hörverstehen“ erfolgt analog zum Leseverstehen auf unterschiedlichen Verarbeitungsebenen. Gemessen wird 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 Hörverstehensteile steigen, wie beim Leseverstehen, im Schwierigkeitsgrad an.

Beim Hörverstehen wird wie beim Leseverstehen für jeden Kandidaten ein numerischer Fähigkeitswert ermittelt und einer der drei TestDaF-Niveaustufen zugeordnet.

## **4.3 Schriftlicher Ausdruck**

Ziel des Testteils „Schriftlicher Ausdruck“ (SA) ist es zu überprüfen, inwieweit 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 Grafiken, Tabellen oder Diagramme. 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. Der vom Kandidaten realisierte Text wird allerdings anhand von

unterschiedlichen Leistungskriterien bewertet wie Gesamtwirkung auf den Rezipienten, Behandlung der Aufgabe (gedankliche Organisation des Textes, Ausführlichkeit, Komplexität, Angemessenheit in Bezug auf Adressat, Situation und Textsorte) und sprachliche Realisierung (z.B. Korrektheit der verwendeten Strukturen, Breite des Wortschatzes, Grad der Textualität, Situationsangemessenheit). Anschließend wird auf dieser Basis eine Zuordnung zu den TestDaF-Niveaustufen vorgenommen.

#### 4.4 Mündlicher Ausdruck

Ziel des Testteils „Mündlicher Ausdruck“ (MA) ist es zu überprüfen, inwieweit 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. Kenyon 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. Grafiken) genommen. Nach einer Denkzeit spricht ein deutscher 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 potenziell – auch dessen Validität höher als die eines traditionel-

len 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 Kassettenrecordern an. In Zukunft ist auch eine computerunterstützte Administration des „Mündlichen Ausdrucks“ vorgesehen.

Der „Mündliche Ausdruck“ wird in enger Analogie zum „Schriftlichen Ausdruck“ bewertet.

## **5. Erprobung der Modellprüfungssätze und testmethodische Begleitung**

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

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 dem *University of Cambridge Local Examinations Syndicate* (UCLES) Rasch-Analysen sowie klassische Itemanalysen durchgeführt.

Für die produktiven Testteile „Mündlicher Ausdruck“ 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 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 Information wurde bei der Revision z.B. entschieden, welche Items jeweils verbleiben konnten und welche zu revidieren waren.

## 6. Anerkennung der Prüfungsleistung und Zulassung zum Studium

Wie bereits erwähnt, können ausländische Studienbewerber die sprachliche Zulassungssicherheit für ein Studium an einer deutschen Hochschule mit TestDaF bereits im Heimatland erhalten. Laut der Informationsbroschüre für die Hochschulen (vgl. Projektgruppe TestDaF 2000d) gelten folgende Modalitäten: Wird TDN 5 in allen vier Subtests erreicht, dann bedeutet dies – aufgrund eines Plenarbeschlusses der HRK vom Februar 2000 – die uneingeschränkte Zulassung zu einem Fachstudium. Hat der Kandidat TDN 5 in zwei Subtests und TDN 4 in den beiden übrigen Subtests erreicht, dann wird den Hochschulen die Zulassung empfohlen. Bei einem Ergebnis von TDN 4 in zwei und mindestens TDN 3 in den beiden anderen Subtests kann die Hochschule unter der Auflage studienbegleitender Sprachkurse eine Zulassung zum Fachstudium aussprechen. Ein Ergebnis von TDN 3 in drei oder vier Subtests bedeutet, dass die sprachliche Leistung des Kandidaten noch nicht für die Zulassung zu einem Fachstudium ausreicht. Die Hochschulen können unter bestimmten Voraussetzungen – wie z.B. im Fall bilingualer Studiengänge – Ausnahmen zulassen.

## 7. 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):

- **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.

Zu den Aufgaben der Informatik gehört weiterhin die Pflege der Homepage (<http://www.testdaf.de>).

## 8. Perspektiven

### 8.1 Organisatorischer Rahmen für die Weiterarbeit

Als organisatorischer Rahmen für alle im Zusammenhang mit TestDaF künftig anfallenden Arbeiten ist ein An-Institut an der FernUniversität Gesamthochschule in Hagen und der Ruhr-Universität Bochum mit Sitz in Hagen in Planung. Träger des An-Instituts ist ein eingetragener Verein mit folgenden Trägerinstitutionen: Deutscher Akademischer Austauschdienst (DAAD), Fachverband Deutsch als Fremdsprache e.V. (FaDaF), FernUniversität Gesamthochschule in Hagen, Goethe-Institut, Hochschulrektorenkonferenz (HRK), Ruhr-Universität Bochum, Universität Leipzig. Das Testinstitut hat Aufgaben zu erfüllen wie z.B. kontinuierliche Erstellung neuer Testsätze und deren Erprobung, Erstellung von Begleit- und Vorbereitungs-materialien, kontinuierliche Begleitforschung zur Sicherung der Qualität, Fortbildung von Testautoren und Testkorrektoren, Weiterentwicklung der Informatikinfrastruktur und Entwicklung einer computerbasierten/-adaptiven Testform.

Ferner müssen weltweit Testzentren eingerichtet werden, die durch das zentrale Testinstitut zu lizenzieren sind.

### 8.2 Begleitforschung

Für große, international anerkannte Tests, wie z.B. TOEFL oder IELTS, ist kennzeichnend, dass zur Qualitätssicherung eine kontinuierliche und umfassende Begleitforschung durchgeführt wird. Auch für TestDaF ist eine entsprechende Begleitforschung von zentraler Bedeutung. Erste Vorstellungen sind im Rahmen der konzeptuellen Planungen für das TestDaF-Institut bereits mehrfach diskutiert worden. Im Folgenden sollen einige weiterführende Überlegungen zur Diskussion gestellt werden.<sup>1</sup> Zu nennen sind zunächst einmal Untersuchungen zur Konstruktvalidität wie z.B.

- empirische Untersuchung und weitere Spezifizierung der für ein Studium in Deutschland notwendigen sprachlichen Handlungskompetenzen
- Auswirkung von Testmethoden auf das Messergebnis; dabei ist z.B. zu denken an eine gleichzeitige Messung der mündlichen Kommunikationsfähigkeit anhand des TestDaF-Subtests „Mündlicher Ausdruck“ und anhand von anderen Testformaten wie z.B. das *Oral Proficiency Interview (OPI)*
- Wechselwirkung zwischen Testmethoden (z.B. Multiple-Choice-Items, freies Schreiben, computerbasierte und computeradaptive Testformen) und Kandida-

---

<sup>1</sup> Bei den in den Kapiteln 8.2, 8.3 und 8.4 vorgestellten Überlegungen handelt es sich um Vorschläge und Ideen, die nicht mit dem Konsortium und der Projektgruppe TestDaF abgestimmt sind.

tenmerkmalen (z.B. sozio-kultureller Hintergrund, Lerntraditionen, kognitive Stile, Argumentationsstrukturen, Testangst, Sprechhemmungen)

- Untersuchung der mentalen Prozesse, die beim Bearbeiten von Testaufgaben ablaufen, z.B. anhand nachträglichen Lauten Denkens und retrospektiver Interviews
- Validierung der TestDaF-Niveaustufenbeschreibungen z.B. anhand der von ausgewählten Kandidaten im Studium gezeigten sprachlichen Leistungen
- Überprüfung der Äquivalenz von Papier-Bleistift-Version und der noch zu erstellenden computerbasierten bzw. computeradaptiven Version.

Im engen Zusammenhang mit der Überprüfung der Konstruktvalidität stehen Untersuchungen zum sog. Backwash-Effekt, d.h. Forschung zur Rückwirkung des Tests auf Unterricht, Lehrplan, gesellschaftliche Leistungsanforderungen oder Art der Vorbereitung auf den Test. Hier ist u.a. zu denken an Untersuchungen zur

- Auswirkung von Testmethoden und Testinhalten auf institutionelle und gesellschaftliche Strukturen, wie z.B. auf die Unterrichtsmethodik, auf die Lehrmaterialien und Lehrpläne
- Auswirkung von Testmethoden und Testinhalten, wie z.B. von gewählten Aufgabenformaten oder dem Grad der Wissenschaftsorientierung, auf das Kandidatenverhalten außerhalb der Testsituation (z.B. auf die Vorbereitung, die Kurswahl und das Ausmaß an selbst gesteuertem Lernen).

Die notwendige Forschung könnte durch das TestDaF-Institut selbst erfolgen, als externe Auftragsarbeit vergeben oder auch in Form von Magisterarbeiten, Dissertationen oder anderen Qualifikationsarbeiten durchgeführt werden. Insbesondere solche Forschung, die nur mit einem sehr hohen finanziellen bzw. personellen Aufwand zu realisieren ist, könnte auch im Rahmen von Drittmittelprojekten erfolgen.

### **8.3 Vorbereitung auf TestDaF im Heimatland der Studierenden**

Das Konsortium TestDaF und die Projektgruppe TestDaF haben damit begonnen, sowohl die Frage der Vorbereitung auf TestDaF im Heimatland als auch das Problem der sprachlichen Begleitung der für ein Studium an einer deutschen Hochschule zugelassenen ausländischen Studierenden zu diskutieren. Im Folgenden wollen wir einige weiterführende Gedanken zur inhaltlichen Ausgestaltung dieser beiden unterschiedlichen Situationen vorstellen.

Die lizenzierten ausländischen Testinstitute müssen in Zukunft sowohl auf TestDaF vorbereiten als auch remediale Angebote für solche Studierende machen, die aufgrund ihrer TestDaF-Ergebnisse nur eingeschränkt oder überhaupt nicht für ein Studium in Deutschland zugelassen sind (s. Kap. 5). Hierfür sind einerseits Kursmaterialien zu entwickeln, die sowohl auf die in TestDaF gestellten Aufgaben vorbereiten als auch eine Sprachstandsdiagnostik erlauben. Andererseits sind Selbstlernmaterialien (z.B. in Form von Fernstudienmaterialien) zu erstellen. Diese sollten sowohl für ein Selbststudium mit Unterstützung durch professionelle Berater und ergänzt durch Kontaktphasen als auch zur Selbstevaluation im Hinblick auf TestDaF einsetzbar sein. In beiden Fällen sollten zunächst einmal die TestDaF-spezifischen Fertigkeiten und die notwendigen Strategien (z.B. Strategien für das globale oder

auch selektive Lesen und Hören) und Themenbereiche im Vordergrund stehen. Darüber hinaus sollte auch schon auf die authentische Situation „Studium in Deutschland“ vorbereitet werden, z.B. durch Berücksichtigung von effizienten Strategien spontaner mündlicher Interaktion (z.B. Gesprächseröffnung, Gesprächsbeendigung, Aufrechterhalten der Kommunikation).

Die Materialien sollten eine hinreichende Zahl von Mustertestsätzen enthalten, damit sich die Kandidaten auf möglicherweise wenig vertraute Aufgabenformate gründlich vorbereiten können. Hiermit soll sichergestellt werden, dass bestimmte Teilnehmergruppen nicht aufgrund einer mangelnden Vertrautheit mit dem jeweiligen Format benachteiligt werden.

Für nicht wenige ausländische Studienbewerber dürfte insbesondere das kassettengesteuerte mündliche Testformat ein besonderes Problem darstellen. Hierauf könnte in Einzelgesprächen mit dafür speziell ausgebildeten Beratern vorbereitet werden, die sich gemeinsam mit dem Kandidaten auf nächste Zieletappen und Vorgehensweisen beim Lernen und Vorbereiten auf den Test verständigen (zu Vorgehensweisen in einer solchen Lernberatung vgl. allgemein Brammerts, Calvert & Kleppin 2000).

Ergänzt werden kann eine Vorbereitung auf TestDaF durch authentische Kontakte über das Internet mit deutschen Studenten, z.B. durch Internet-Tandemarbeit – möglicherweise sogar mit Studenten, die ihrerseits in dem entsprechenden Land ein Studium aufnehmen möchten. Die Ziele der beiden Partner können sich dabei durchaus unterscheiden (zu Informationen über Internet-Tandemlernen vgl. u.a. Brammerts & Kleppin 1998 sowie <http://www.slfruhr-uni-bochum.de>).

#### **8.4 Sprachliches Begleitangebot für ausländische Studierende in Deutschland**

Es ist ein Begleitkursangebot zu entwickeln, das den Studierenden im Fall einer Zulassung zum Studium mit Auflagen durch die jeweilige Universität erlaubt, vorhandene sprachliche Defizite gezielt auszugleichen. Dazu wäre es wünschenswert, wenn zumindest langfristig die Leistungen der Kandidaten nicht nur in Form von globalen Niveaustufenzuweisungen, sondern zusätzlich als detailliertes Profil ausgewiesen werden könnten. Doch selbst bei einer uneingeschränkten Zulassung ist eine weitere Förderung der für ein Studium relevanten sprachlichen Fähigkeiten und Fertigkeiten in vielen Fällen sinnvoll. Bisher schon vorhandene universitäre Infrastrukturen, die in wesentlichen Teilen der Vorbereitung auf die PNdS oder die DSH dienen, könnten dafür genutzt werden. Eine analoge Argumentation gilt im Hinblick auf die Studienkollegs. Deshalb muss es auch nicht – wie zuweilen befürchtet – aufgrund der Einführung von TestDaF zu einem Stellenabbau an den betreffenden Institutionen kommen.

Das Begleitangebot könnte Begleitmodule u.a. mit folgenden Schwerpunkten umfassen:

- Verfolgen von Vorlesungen (z.B. Training von Seh- und Hörstrategien anhand von entsprechenden Aufnahmen; ökonomisches Verfassen von Notizen; Ermutigung zu sozialen Strategien wie kurze Nachfragen bei Kommilitonen zu schwer verständlichen Passagen)

- Mitarbeit in Seminaren (z.B. Training von Seh- und Hörstrategien für Seminar-gespräche; Interpretation nonverbalen Verhaltens von Dozenten und Kommilitonen; Ermutigung zum Nachfragen; Training von kurzen zusammenhängenden Diskussionsbeiträgen)
- Referate (z.B. Training von freien zusammenhängenden längeren mündlichen Beiträgen unter Zuhilfenahme von Clusters oder Notizen; Reagieren auf Einwürfe und Nachfragen)
- Hausarbeiten (wissenschaftliches Schreibtraining wie z.B. zum Exzerpieren, Gliedern, Zitieren, Strukturieren, Beschreiben, Argumentieren)
- Mündliche Prüfungen (z.B. Training von spontanen Reaktionen auf Fragen; Nachfragen bei nicht verstandenen Äußerungen; Simulationen von Prüfungen unter Peers).

In allen Modulen sollte ein permanenter Bezug zur Praxis gewährleistet sein. Die Studierenden werden z.B. – wenn sie ein mündliches Referat halten – dazu angeregt, einen Partner aus dem betreffenden Kurs oder auch einen Face-to-Face-Tandempartner aus der jeweiligen Universität zu bitten, das wissenschaftssprachliche Verhalten zu beobachten und Feedback zu geben. Professionelle Berater können dabei helfen, dass der Studierende sich zunächst Einblick in seine eigenen Lernbedingungen und -voraussetzungen verschafft. Der Berater erreicht dies durch gezielte, entscheidungsrelevante Fragen und macht durch Einbringen von Fachkenntnissen und Erfahrungen auf Zusammenhänge aufmerksam, die für das Lernen wichtig sein könnten. Er drängt auf Evaluation der bisherigen Praxis und stellt sich als Gesprächspartner zur Verfügung, dem der Lerner eigene Entscheidungen vorträgt und dem gegenüber er sich auf Einhaltung verpflichtet. In diesem Zusammenhang kann der Berater auch eine individuelle (gesprächsbasierte) Diagnostik sowohl der sprachlichen Fertigkeiten als auch der individuellen kognitiven Stile der Kandidaten anbieten. Letztere hat insbesondere zum Ziel, sowohl die Lernenden als auch die Lehrenden für interindividuelle Unterschiede beim Gebrauch und Erlernen von Sprache zu sensibilisieren (vgl. Grotjahn 1998).

## Literatur

- Association of Language Testers in Europe (ALTE) (1998) *ALTE Handbuch europäischer Sprachprüfungen und Prüfungsverfahren*. Cambridge: The University of Cambridge Local Examinations Syndicate.
- Bachman, L. (2000) "Modern language testing at the turn of the century: Assuring that what we count counts". *Language Testing* 17/1:1–42.
- Bennett, R.E. & Ward, W.C. (eds.) (1993) *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*. Hillsdale, N.J.: Erlbaum.
- Bickes, G. (1998) „DSH – TESTDAF – TOEFL: Thesen zu einer Umgestaltung der Prüfungspraxis in Deutsch als Fremdsprache für ausländische Studienbewerber“. *Info DaF* 25/1:97–103.
- Bolton, S. (2000) „TESTDAF“. In: Wolff, A. & Tanzer, H. (eds.) *Sprache – Kultur – Politik*. Regensburg: Fachverband Deutsch als Fremdsprache (Materialien Deutsch als Fremdsprache 53), 342–353.

- Brammerts, H.; Calvert, M. & Kleppin, K. (2000) „Ziele und Wege bei der individuellen Lernberatung“. In: Brammerts, H. & Kleppin, K. (eds.) *Selbstgesteuertes Lernen im Tandem: Ein Handbuch*. Tübingen: Stauffenburg (erscheint).
- Brammerts, H. & Kleppin, K. (1998) „Autonom lernen mit einem fernen deutschsprachigen Tandempartner“. In: Donath, R. (ed.) *Deutsch als Fremdsprache. Projekte im Internet*. Stuttgart: Klett, 42–46.
- Chapelle, C.A.; Grabe, W. & Berns, M. (1997) *Communicative language proficiency: Definition and implications for TOEFL 2000*. Princeton, N.J.: Educational Testing Service (TOEFL Monograph Series MS-10).
- Coste, D.; North, B.; Sheils, J. & Trim, J.L.M. (1998) *Modern languages: Learning, teaching, assessment. A common European framework of reference*. Strasbourg: Council of Europe.
- Douglas, D. (2000) *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Grotjahn, R. (1998) „Lernstile und Lernstrategien: Definition, Identifikation, unterrichtliche Relevanz“. *Der fremdsprachliche Unterricht Französisch* 34/4:11–15.
- Grotjahn, R. (2000a) „Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TestDaF“. In: Bolton, S. (ed.) *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*. Köln: VUB-Gilde, 7–55.
- Grotjahn, R. (2000b) „Testtheorie: Grundzüge und Anwendungen in der Praxis“. In: Wolff, A. & Tanzer, H. (eds.) *Sprache – Kultur – Politik*. Regensburg: Fachverband Deutsch als Fremdsprache (Materialien Deutsch als Fremdsprache 53), 304–341.
- Kenyon, D.M. (2000) “Tape-mediated oral proficiency testing: Considerations in developing Simulated Oral Proficiency Interviews (SOPIs)”. In: Bolton, S. (ed.) *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*. Köln: VUB-Gilde, 87–106.
- Lee, W. (1998) *Prüfungen zum Nachweis deutscher Sprachkenntnisse bei ausländischen Studienbewerbern/Studienbewerberinnen (PNdS/DSH): Ihre Praxis und ihr Prüfprofil*. Regensburg: Fachverband Deutsch als Fremdsprache.
- Lewkowicz, J.A. (2000) “Authenticity in language testing: Some outstanding questions”. *Language Testing* 17/1:43–64.
- McNamara, T.F. (1997) “Performance testing”. In: Clapham, C. & Corson, D. (eds.) *Encyclopedia of language and education*. Vol. 7: *Language testing and assessment*. Dordrecht: Kluwer, 131–139.
- Projektgruppe TestDaF (2000a) *Informationen für Prüfungsteilnehmer*. (Ms.)
- Projektgruppe TestDaF (2000b) *Handreichungen für Prüfungsbeauftragte*. (Ms.)
- Projektgruppe TestDaF (2000c) *Handreichungen für Testautoren*. (Ms.)
- Projektgruppe TestDaF (2000d) *Informationsbroschüre für die Hochschulen*. (Ms.)
- Projektgruppe TestDaF (2000e) „TestDaF: Konzeption, Stand der Entwicklung, Perspektiven“. *Zeitschrift für Fremdsprachenforschung* 11/1:63–82.
- Six, H.-W. (2000) „Informatikunterstützung für TESTDAF“. In: Bolton, S. (ed.) *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*. Köln: VUB-Gilde, 131–140.
- Wintermann, B. (1998) „DSH und TESTDAF – Sprachprüfungen auf dem Prüfstand“. *Info DaF* 25/1:104–110.