

DER TestDAF. KONZEPT UND PRINZIPIEN DES STANDARDISIERTEN TESTS DEUTSCH ALS FREMDSPRACHE

Ulrike Arras
TestDaF-Institut Hagen (Deutschland)

Zusammenfassung

Der TestDaF (Test Deutsch als Fremdsprache) ist ein seit 2001 vom TestDaF-Institut administriertes Testsystem für ausländische StudienbewerberInnen. Die Prüfung testet Deutschkenntnisse auf fortgeschrittenem Niveau, die für ein Hochschulstudium in Deutschland relevant sind. Der TestDaF fungiert damit als Nachweis ausreichender Sprachkenntnisse, um ein Studium aufzunehmen. Die vier Fertigkeiten Lesen, Hören, Schreiben und Sprechen werden getrennt in je eigenen Subtests gemessen und die Leistungen entsprechend getrennt auf einem Zeugnis ausgewiesen. Die Prüfung kann weltweit mehrmals im Jahr an lizenzierten Testzentren abgelegt werden.

Die Erfüllung zentraler Testgütekriterien wie Validität, Reliabilität und Objektivität unter Einschluss von Fairness sind bei solchen High-Stakes-Tests besonders wichtig. Die Qualitätssicherung erfolgt über verschiedene Maßnahmen, insbesondere durch festgelegte Prüfungsformate, festgeschriebene Durchführungsmodalitäten, standardisierte Beurteilungsverfahren und testmethodische Verfahren. Damit ist der TestDaF eine standardisierte Prüfung.

Schlüsselwörter: Test, Prüfung, Standardisierung, Validität, Objektivität, Reliabilität, Fairness, Gemeinsamer europäischer Referenzrahmen für Sprachen

Resumen

El TestDaF (Test Deutsch als Fremdsprache / Examen de Alemán como Lengua Extranjera) es un examen coordinado desde 2001 por el Instituto TestDaF en Alemania. El TestDaF es un examen de idioma de nivel avanzado dirigido a estudiantes internacionales que desean acceder a la universidad en Alemania. Por tanto los contenidos del examen se basan en el uso de la lengua alemana dentro del contexto universitario. Como consecuencia, el certificado del TestDaF sirve como prueba de acceso lingüística para realizar estudios superiores en Alemania. El TestDaF evalúa

por separado las cuatro destrezas lingüísticas en cuatro pruebas y los resultados aparecen por separado en el certificado. El examen tiene lugar varias veces al año en varios centros colaboradores autorizados en todo el mundo.

Cumplir los criterios de calidad de un examen como validez, fiabilidad y objetividad incluido el criterio del fairness es especialmente importante para exámenes tipo high stakes. Se han tomado varias medidas para controlar la calidad del examen, entre otras la estandarización de las tareas, de los modos de realización del examen y los métodos estadísticos de análisis.

Descriptor: examen, prueba, estandarización, validez, fiabilidad, objetividad, fairness, Marco de Referencia Europeo.

1. Allgemeines

Ausländische StudienbewerberInnen, die an einer Hochschule in Deutschland studieren wollen, können den Nachweis über ausreichende Sprachkenntnisse mit dem „Test Deutsch als Fremdsprache“ (TestDaF) erbringen¹. Bedingt durch das Prüfungsziel ist der Test thematisch und hinsichtlich der sprachlichen Anforderungen im akademischen Kontext angesiedelt. Dank seiner Standardisierung und Orientierung an wichtigen Referenzsystemen wie dem Gemeinsamen europäischen Referenzrahmen für Sprachen des Europarats und der Skala der *Association of Language Testers in Europe* (ALTE)² ist ein TestDaF-Zeugnis jedoch auch für berufliche Zwecke, insbesondere für akademische Berufsfelder, relevant. Die Prüfung wird weltweit durchgeführt, deshalb kann sie bereits im Heimatland abgelegt werden und vereinfacht dadurch den Zugang zu einem Hochschulstudium in Deutschland. (Arras & Grotjahn, 2002).

Seit seiner Einführung im Frühjahr 2001 haben knapp 28.200 Personen (Stand Juni 2005) den TestDaF weltweit an ca. 370 Testzentren in mehr als 70 Ländern abgelegt. Im Jahr 2005 wurde der TestDaF an 7 Terminen (inklusive 2 Termine in der Volksrepublik China) durchgeführt. 2006 wird ein weiterer Prüfungstermin angeboten werden.

1. Damit übernimmt der TestDaF die gleiche Funktion wie die „Deutsche Sprachprüfung zum Hochschulzugang“ (DSH), die bislang der wichtigste Prüfungstyp zum Nachweis ausreichender Sprachkenntnisse für ein Studium in Deutschland war. Bei der DSH handelt es sich jedoch um keine standardisierte Prüfung, auch werden zentrale Testgütekriterien nicht hinreichend erfüllt. Zur Problematik des Vergleichs der TestDaF-Prüfung mit der DSH s. v. a. die Beiträge in Casper-Hehne & Koreik (eds.) (2004).

2. Die Niveaustufen-Beschreibungen des Gemeinsamen europäischen Referenzrahmens siehe <http://www.sprachenportfolio.ch> sowie Europarat 2001. Die Niveaustufen-Beschreibung der ALTE s. <http://www.alte.org>.

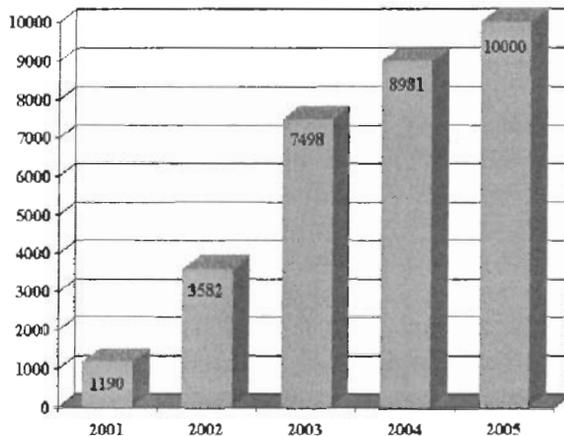


Abbildung 1: Prüfungsteilnehmende seit 2001

Die Erstellung der Prüfungsaufgaben, die Auswertung der Ergebnisse sowie die testmethodische Absicherung und die wissenschaftlich gestützte Weiterentwicklung des Testformats erfolgen **zentral im TestDaF-Institut in Hagen**³.

Beim TestDaF handelt es sich um einen standardisierten **Sprachstandtest**, der Sprachkenntnisse im akademischen Kontext im oberen Leistungsspektrum, d.h. auf **fortgeschrittenem Niveau**, misst. Die vier Fertigkeiten werden getrennt voneinander **gemessen und einer der drei TestDaF-Niveaustufen (TDN) 3, 4 und 5 zugeordnet**. **Unterhalb von TDN 3 differenziert der TestDaF nicht**; eine Einstufung auf „unter TDN 3“ besagt lediglich, dass das **Eingangsniveau** der Prüfung TestDaF nicht erreicht wurde. Das **TestDaF-Leistungsspektrum** liegt nach den Kompetenzskalen des Europarats ungefähr im Bereich B 2.1 bis C 1.2. Allen TestDaF-Niveaustufen sind in jedem der vier Fertigungsbereiche so genannte **Can-Do-Statements**⁴ zugeordnet, die den Teilnehmenden ebenso wie den abnehmenden Instanzen, z.B. eine Hochschule in Deutschland, Auskunft über den erreichten Sprachstand geben.

3. Das TestDaF-Institut ist ein An-Institut an der FernUniversität in Hagen und an der Ruhr-Universität Bochum. Es wurde im Februar 2001 gegründet und ist eine Einrichtung der Gesellschaft für Akademische Testentwicklung e.V. Gegründet wurde das TestDaF-Institut von Einrichtungen, die die deutschen Hochschulen vertreten und mit auswärtiger Kulturpolitik befasst sind (die Hochschulrektorenkonferenz, der Deutsche Akademische Austauschdienst e.V. sowie das Goethe-Institut e.V.), und von Institutionen, die das TestDaF-Institut wissenschaftlich beraten (die FernUniversität in Hagen, die Ruhr-Universität Bochum, die Universität Leipzig und der Fachverband Deutsch als Fremdsprache).

4. Zur Beschreibung der TestDaF-Niveaustufen siehe <http://www.testdaf.de>. Die *Can-Do-Statements* befinden sich auch auf der Rückseite des TestDaF-Zeugnisses.

Die Teilnehmenden erhalten ein Zeugnis, auf dem die Ergebnisse in den vier Teilbereichen Leseverstehen, Hörverstehen, Schriftlicher und Mündlicher Ausdruck getrennt voneinander ausgewiesen werden. Das bedeutet es gibt keine Gesamtnote, alle Prüfungsteile sind gleich gewichtet. Auch wird kein Gesamturteil „bestanden“ oder „nicht bestanden“. Damit wird dem individuell unterschiedlichen Leistungsspektrum Rechnung getragen. Es ist durchaus möglich, dass eine individuelle Leistung heterogen ausfällt, dass beispielsweise im Leseverstehen die TestDaF-Niveaustufe 5 erreicht wird, im Schriftlichen Ausdruck hingegen nur die Niveaustufe 3. Die Prüfungsteilnehmenden ebenso wie die Hochschulen und Fachbereiche erhalten damit also ein differenziertes Bild über die Stärken und evtl. Schwächen in den verschiedenen Fertigungsbereichen. Zum einen soll dieses Verfahren den Hochschulen die Möglichkeit einräumen, die sprachlichen Anforderungen vor der Zulassung zum Studium nach Fächern beziehungsweise Studienziel und Studiendauer zu differenzieren. Das bedeutet, dass je nach Anforderung in sprachlicher Hinsicht eine differenzierte Zulassung möglich wird, also beispielsweise in einem philologischen Studienfach andere sprachliche Voraussetzungen festgeschrieben werden können als in einem englischsprachigen Studiengang, bei dem Deutsch in erster Linie im Alltag, weniger jedoch für das Studium selbst von Bedeutung ist. Zum anderen wird die Zulassungssicherheit für die StudienbewerberInnen dadurch erhöht, dass die Fachbereiche die erforderlichen Sprachkenntnisse festlegen.

Im Folgenden sollen Prinzipien sowie Konzeption des TestDaF erläutert und Aufbau, Ziele und Inhalte der Prüfung skizziert werden. Im Anschluss sollen einige Informationen zum TestDaF in Spanien bzw. zu den Ergebnissen spanischsprachiger PrüfungsteilnehmerInnen aufgeführt und sodann Möglichkeiten der Prüfungsvorbereitung diskutiert werden.

2. Prinzipien des TestDaF

Bei der Prüfung TestDaF handelt es sich um einen so genannten High-Stakes-Test. Das bedeutet, vom individuellen Ergebnis hängt für die einzelnen Prüfungsteilnehmenden ab, ob sie zu einem Studium an einer Hochschule in Deutschland zugelassen werden und damit, wie sich ihr weiterer Lebensweg gestaltet. Aus diesem Grunde müssen bei der Konzeption der Prüfung, bei der Entwicklung der Prüfungsaufgaben sowie bei der Auswertung der Prüfungsergebnisse bestimmte Standards⁵ eingehalten werden. Zu den klassischen Testgütekriterien zählen Validität, Reliabilität und Objektivität. Von

5. Zu den Testgütekriterien im Kontext von Sprachtests s. besonders den Ansatz der Testnützlichkeit (test usefulness) bei Bachman & Palmer 1996.

zentraler Bedeutung besonders bei weltweit zum Einsatz kommenden Tests sind zudem (zeit-)ökonomische und Praktikabilitätsabwägungen, d.h. z.B. dass die Testergebnisse innerhalb eines bestimmten Zeitrahmens an die Prüflinge zurückgemeldet werden müssen, damit sie sich fristgerecht an einer deutschen Hochschule bewerben können. Nicht minder wichtig - gerade bei einem High-Stakes-Test - ist die Fairness bei der Durchführung und bei den Prüfungsanforderungen sowie die Authentizität der Prüfungsaufgaben bzw. die Relevanz der geforderten Leistung. Um die Qualität der Prüfung zu sichern, basiert die Prüfung auf den folgenden Säulen:

- zentrale Testerstellung durch eigens geschulte TestautorInnen anhand eines festgelegten Prüfungsformats,
- dezentrale Durchführung der Prüfung weltweit an lizenzierten Testzentren zu zentral festgelegten Prüfungsterminen,
- zentrale Beurteilung der Prüfungsergebnisse durch eigens geschulte BeurteilerInnen anhand festgelegter Bewertungsverfahren und -kriterien,
- testmethodische Absicherung der Testqualität mit Hilfe statistischer Verfahren.

2.1 Durchführung der Prüfung

Um eine faire und objektive Durchführung der Prüfung zu gewährleisten, wird der TestDaF an vom TestDaF-Institut zentral festgesetzten Terminen⁶ weltweit dezentral an so genannten Testzentren abgelegt. Diese werden vom TestDaF-Institut lizenziert. Es handelt sich um Institutionen (i.d.R. Hochschulen, aber auch Goethe-Institute u.ä. Einrichtungen), die die vorgeschriebenen Kriterien erfüllen⁷. Die Durchführung der Prüfung selbst erfolgt anhand festgelegter Regeln unter Aufsicht von so genannten Prüfungsbeauftragten, die für den korrekten Ablauf verantwortlich sind. Der Ablauf der Prüfung ist immer gleich, d.h. die einzelnen Prüfungsteile (s. u.) erfolgen in einer bestimmten Reihenfolge. Auch die Zeit für Pausen ist genau vorgegeben. Alle Prüfungsunterlagen werden aus Gründen der Testsicherheit vom TestDaF-Institut geliefert und müssen komplett wieder an das Institut zurück geschickt werden. Für die Teilnahme an der Prüfung wird eine Prüfungsgebühr erhoben, die derzeit je nach Region zwischen € 60 und € 130 beträgt.

6. Derzeit werden pro Jahr insgesamt sieben Prüfungstermine angeboten, wobei zwei davon den Testzentren in der Volksrepublik China vorbehalten sind. Eine Steigerung der Prüfungsfrequenz ist geplant.

7. Zu den Lizenzierungskriterien s. die Internetseite des TestDaF-Instituts: www.testdaf.de.

2.2 Aufgabenerstellung und Erprobung

Die Erstellung der Prüfungsaufgaben erfolgt zentral am TestDaF-Institut zusammen mit TestautorInnen, die sich mit dem Testformat in Form von Schulungen vertraut gemacht haben. Die Standardisierung der Prüfung erfordert, dass bei der Entwicklung der Testaufgaben bestimmte Vorgaben bzw. Merkmale (*task specifications*) wie Textsorte, Aufgabentyp, Schwierigkeitsniveau etc. zu berücksichtigen bzw. einzuhalten sind. Alle Prüfungsaufgaben werden vor ihrem Einsatz in einer TestDaF-Prüfung in einem mehrschrittigen Zyklus erprobt, und zwar zunächst in zwei kleinen Stichproben in Deutschland mit Hilfe von MuttersprachlerInnen und Deutschlernenden, sodann weltweit in einer hinreichend großen Versuchsgruppe mit Deutschlernenden, die der Zielgruppe des TestDaF entsprechen, also über die zu messenden Sprachkenntnisse verfügen. Die Ergebnisse der Erprobungsprüfungen werden testmethodisch ausgewertet. Sollten sich Aufgaben als ungeeignet erweisen (z.B. wegen geringer Trennschärfe oder weil sie nicht geeignet sind, die gewünschte Leistung zu elizitieren), werden sie verworfen oder überarbeitet und in einer revidierten Fassung erneut einer Erprobung unterzogen. Die Erprobung der Aufgaben dient auch der Festlegung der Schwierigkeitsniveaus. Um die Schwierigkeit der Lese- und Hörverstehensaufgaben zu kalibrieren, werden während der Erprobung neuer TestDaF-Aufgaben so genannte Ankeritems, d.h. Items mit feststehendem Schwierigkeitsgrad, sowie das Rasch-Modell der probabilistischen Testtheorie eingesetzt (Arras, Eckes, & Grotjahn 2002, Eckes 2003). Die produktiven Fertigkeiten Schriftlicher Ausdruck und Mündlicher Ausdruck werden von speziell geschulten BeurteilerInnen bewertet. Zur Sicherung der Itemqualität bei den offenen Aufgabentypen in diesen Subtests werden daher zudem qualitative Daten erhoben, indem BeurteilerInnen von schriftlichen bzw. mündlichen Leistungen aus der Erprobungsphase um kritische Rückmeldungen zur Eignung der Aufgabenstellungen gebeten werden. Denn oftmals zeigt sich erst an den Texten, die auf der Grundlage der Aufgaben produziert werden, ob diese Aufgaben das elizitieren können, was seitens der Testerstellung angestrebt wird bzw. was der Test messen tatsächlich will (Konstruktvalidität).

2.3 Aufbau und Inhalte der Prüfung

Beim TestDaF werden die vier Fertigkeiten, Leseverstehen, Hörverstehen, Schriftlicher Ausdruck und Mündlicher Ausdruck in vier so genannten Subtests gemessen, wobei die beiden rezeptiven Fertigkeiten in Kompetenztests, die beiden produktiven Fertigkeiten in Performanztests überprüft werden. Nach den Prüfungsteilen Lese- und Hörverstehen ist eine längere Pause einzuhalten. Danach werden die produktiven Fertigkeiten in den Prüfungsteilen

Schriftlicher und Mündlicher Ausdruck getestet. Insgesamt dauert die Prüfung ungefähr 3 Stunden und 10 Minuten. Alle Prüfungsteile werden am selben Tag i.d.R. in der gleichen Reihenfolge durchgeführt.

Der TestDaF prüft die vier Fertigkeiten getrennt, um die Validität der Prüfung sicherzustellen. Das bedeutet: Um beispielsweise Aussagen über die Lesekompetenz einer Kandidatin machen zu können, darf nur die Lesekompetenz, nicht aber z.B. die Schreibkompetenz gemessen werden. Bestünde die Aufgabe zur Messung von Lesekompetenz beispielsweise darin, einen Text zu lesen und diesen sodann schriftlich zusammenzufassen, so ließe das Produkt - die zu bewertende schriftliche Leistung - keine eindeutige Aussage zur Lesekompetenz zu. Denn die (z.B. schwache) schriftliche Leistung könnte sowohl ein Hinweis auf die (schwache) Lesekompetenz oder aber auf die (schwache) Schreibkompetenz sein. Integriertes Testen ist also für eine eindeutige Erfassung der zu messenden Fähigkeit ungeeignet. Geeignet sind vielmehr Aufgaben, die die Fähigkeit isoliert erfassen. Dies hat zur Folge, dass geeignete Itemtypen gewählt werden müssen, um die zu messende Fähigkeit genau zu erfassen. Prinzipiell werden folgende Itemtypen unterschieden:

- geschlossene Items (z.B. Multiple-Choice-Aufgaben),
- halboffene Items (z.B. Lückentexte oder Fragen, die stichwortartig zu beantworten sind),
- offene Items (z.B. Aufgaben, die einen eigenständigen längeren Redebeitrag oder schriftlichen Text elizitieren)

Da es sich beim TestDaF um eine Prüfung handelt, mit der man den Nachweis über ausreichende Sprachkenntnisse für ein Studium in Deutschland erbringen kann, haben alle Aufgaben bzw. Texte und geforderten Sprachhandlungen einen Bezug zur Hochschule.

Für die Prüfungsteile zur Erfassung der rezeptiven Fähigkeiten (hören und lesen) hat dies zur Folge, dass sowohl die Textsorten als auch die Sprachhandlungen hochschulrelevant sind. So werden im Subtest „Leseverstehen“ (LV) z.B. Kurztex te wie sie im studentischen Alltag an der Hochschule vorkommen oder wissenschaftssprachliche längere Texte präsentiert. Im Subtest „Hörverstehen“ (HV) werden Dialoge im studentischen Kontext sowie Wissenschaftsbeiträge mit Monologcharakter präsentiert. Die Aufgabe besteht z.B. im Falle der dritten Hörverstehensaufgabe darin, ähnlich einer Situation Vorlesung einem thematisch anspruchsvollen längeren Redebeitrag zu folgen und dabei Notizen in Form von Stichworten zu machen. Prüfungsziele sind je nach Schwierigkeitsniveau das Verstehen von Gesamtzusammenhängen und Einzelinformationen sowie ggf. das Verstehen impliziter Informationen. Entsprechend erforderlich sind geeignete

Verstehensstrategien. Der Prüfungsteilteil „Leseverstehen“ besteht aus drei Leseverstehensaufgaben unterschiedlichen Schwierigkeitsgrades mit insgesamt 30 geschlossenen Items. Je nach Aufgabe handelt es sich bei den Lesetexten um unterschiedliche Textsorten und bei den Items um Zuordnungs-, Mehrfachwahl- oder Auswahlitems. Der Prüfungsteil „Hörverstehen“ besteht aus drei Hörtexten, ebenfalls unterschiedlichen Schwierigkeitsgrades, die unterschiedliche Textsorten präsentieren. Insgesamt sind 25 Items (gesteuerte Notizen bzw. Alternativformen) zu lösen.

Die Überprüfung der schriftlichen Ausdrucksfähigkeit erfolgt anhand lediglich einer Texterstellungsaufgabe. Es handelt sich um eine direkte Erfassung der Fähigkeit auf der Basis eines offenen Itemformats, wobei jedoch schriftlich bzw. visuell präsentierte Vorgaben die Aufgabe steuern. Gefordert werden im Wesentlichen zwei Schreibhandlungen: Das Beschreiben und Zusammenfassen von statistischen Daten, die in Form einer Grafik oder tabellarisch präsentiert werden sowie das Argumentieren, indem beispielsweise zu einer Frage oder einem Problem begründet Stellung genommen werden soll und dabei unterschiedliche Meinungen zu paraphrasieren und zu berücksichtigen sind. Die Prüfungsteilnehmenden sollen zeigen, ob sie in der Lage sind, zu einem bestimmten Thema einen zusammenhängenden und klar aufgebauten, diskursiven Text zu schreiben. Zur Erfassung der mündlichen Ausdrucksfähigkeit wird ein kassetten- oder CD-gesteuertes Format verwendet. Die Prüfungsteilnehmenden hören (i.d.R. über Kopfhörer) die Aufgaben und Stimuli vom Band, ihre Antworten werden auf Band aufgenommen. Es handelt sich somit um ein semidirektes Testformat⁸. Der Prüfungsteil umfasst sieben⁹ situativ eingebettete Aufgaben, in denen die Prüfungsteilnehmenden sprachlich reagieren müssen und die unterschiedlich schwierig sind. Das zunächst ungewohnte Format hat den für einen standardisierten und weltweit eingesetzten Test wichtigen Vorteil, dass alle Prüfungsteilnehmenden unter den gleichen Bedingungen ihre Leistung zeigen (Durchführungsobjektivität) und diese Leistung ungeachtet der Person zentral von eigens geschulten BeurteilerInnen bewertet werden (Beurteilungsobjektivität). Die Prüfungsteilnehmenden sollen in unterschiedlichen Situationen, die für den Alltag an einer deutschen Hochschule typisch sind, verbal reagieren, z.B. in einem Gespräch mit KommilitonInnen oder im Kontext eines Fachseminars. Die geforderten

8. Zur Entwicklung dieses Testformats, dem das Konzept des SOPI (Simulated Oral Proficiency Interview) zugrunde liegt, s. Kniffka & Üstünsöz-Beurer 2001.

9. Der Prüfungsteil Mündlicher Ausdruck wurde nach einer mehrjährigen Überarbeitungs- und Erprobungsphase revidiert, so dass der Subtest seit Sommer 2005 statt der ursprünglich 10 nunmehr 7 Aufgaben umfasst. Die Veränderungen betrafen auch die Anordnung der Schwierigkeit der einzelnen Aufgaben. Zudem wurden die Bewertungskriterien zum Mündlichen Ausdruck strukturell revidiert und dem neuen Aufgabenformat angepasst. Darüber hinaus wurde die Datenerhebung präzisiert.

Sprechhandlungen umfassen beispielsweise die Beschreibung statistischer Daten, begründete Stellungnahmen, das Abwägen von Vor- und Nachteilen etc.

2.4 Beurteilung der Prüfungsleistungen

Die Beurteilung der Prüfungsleistungen erfolgt zentral durch das TestDaF-Institut. Was die Auswertung der Ergebnisse aus den rezeptiven Prüfungsleistungen anbelangt, so wird jeweils die Anzahl der korrekt gelösten Antworten (Testscore) ermittelt. Die Zuweisung des Testscores zu den TDN-Stufen erfolgt sodann anhand der auf der Datenbasis der Erprobung berechneten so genannten Cut-Offs (Eckes 2003).¹⁰

Ein standardisierter Test, der weltweit abgenommen wird, muss sich eines kriterienorientierten Beurteilungsverfahrens bedienen. Das bedeutet: Ausschlaggebend bei der Bewertung von Leistungen ist nicht die (durchschnittliche) Leistung der Gesamtgruppe, i. e. alle Teilnehmenden an einem Prüfungsereignis, sondern die durch die TestDaF-Niveaustufen ausgewiesene Leistung selbst. Aus diesem Grunde werden die Leistungen aus den Prüfungsteilen zur Erfassung der produktiven Fähigkeiten, d.h. die schriftlichen und mündlichen Texte, von BeurteilerInnen bewertet, die eigens geschult werden und regelmäßig Monitoring-Maßnahmen durchlaufen. Das zentrale Instrument der Beurteilung sind die Bewertungskriterien in Form von skalierten Deskriptoren. Dabei sieht das standardisierte Bewertungsverfahren vor, dass die individuelle Prüfungsleistung hinsichtlich vorgegebener, das Testkonstrukt widerspiegelnder, Aspekte mit den Deskriptoren des Bewertungsrasters abgeglichen wird. Diese Kriterien erfassen zum einen die Gesamtwirkung bei der Rezeption eines Textes. Es handelt sich um eine eher holistische Erfassung der Leistung. Zum anderen werden die sprachliche und die inhaltliche Umsetzung der jeweiligen Aufgaben beurteilt, was ein eher analytisches Vorgehen bei der Bewertung erfordert. Um die Schwierigkeit der verschiedenen Aufgaben über verschiedene Testsätze bzw. Testereignisse hinweg konstant zu halten bedarf es darüber hinaus einer testsatzspezifischen Kalibrierung. Die entsprechenden Instrumente bestehen aus zwei Teilen: Zum einen werden die Anforderungen aufgabenspezifisch festgehalten. Das bedeutet, zu jeder Aufgabe wird anhand von Texten aus den Erprobungen

10. Die Scores aus den geschlossenen Itemtypen sind unmittelbar maschinell erfassbar, d.h. eine Auswertung „per Hand“ ist nicht notwendig. Bei den halboffenen Items jedoch wie sie z. T. im Prüfungsteil Hörverstehen verwendet werden, muss zuvor eine Auswertung durch geschulte BeurteilerInnen erfolgen. Sie entscheiden anhand von festgelegten Beurteilungsrichtlinien, ob eine Kurzantwort als korrekte Lösung zu akzeptieren ist. Zur Bewertung halboffener Items im Kontext der Prüfung TestDaFs. Bärenfänger & Kecker 2004.

eruiert, welche Leistungen bzw. Anforderungen die Aufgabe eliziert und welche Maßstäbe bei der Umsetzung der jeweiligen Aufgabe anzulegen sind. Hiermit wird also versucht, die Schwierigkeit insofern über verschiedene Testereignisse und Aufgaben konstant zu halten, indem festgelegt wird, inwieweit bei eher komplexen Aufgaben geringere Anforderungen an die Umsetzung zu stellen sind als bei Aufgaben, die z.B. eher einfach zu erfassende Darstellungen statistischer Daten aufweisen. Zum anderen werden ebenfalls anhand von Leistungen aus der Erprobungsphase der jeweiligen Aufgabe Texte auf unterschiedlichen Leistungsniveaus ausgewählt und durch ein ExpertInnen-Gremium, bestehend aus geschulten BeurteilerInnen und TestentwicklerInnen, bewertet. Die Urteile und entsprechenden begründeten Einstufungen werden schriftlich fixiert und stellen eine weitere Orientierungshilfe für alle BeurteilerInnen dar, die an der Bewertung schriftlicher bzw. mündlicher Leistungen aus einem Testereignis eingesetzt werden. Diese Kalibrierungsmaßnahmen haben zum Ziel, die Beurteilungsmaßstäbe zu bestimmen. Die konstante Interpretation dieser Maßstäbe ist zum einen wichtig, um die Reliabilität der Beurteilungen zu erhöhen und zum anderen, um die Schwierigkeit der Aufgabe zu justieren. Die Konstanzhaltung der Schwierigkeit schließlich ist erforderlich, um das Gütekriterium der Validität zu erfüllen.

Da trotz Schulung, Kalibrierung und weiteren Monitoring-Maßnahmen Menschen unterschiedlich strenge Beurteilungsmaßstäbe anlegen, wird ein weiteres Instrument eingesetzt, um zu zuverlässigen und damit fairen Leistungsbeurteilungen zu gelangen, nämlich die Erfassung der individuellen Strenge der einzelnen BeurteilerIn mit Hilfe des Multi-Facetten-Rasch-Modells. Hierbei wird bei der Ermittlung der tatsächlich erreichten Leistungsstufe u. a. auch der Strengekoeffizient der individuellen Beurteilerin bzw. des individuellen Beurteilers einbezogen (Eckes 2004). Die Erfassung verschiedener Determinanten - also die Leistungseinstufungen hinsichtlich der verschiedenen Aspekte wie sie im Kriterienraster vorgegeben sind, die Aufgabenschwierigkeit sowie die individuelle Strenge und Konsistenz der BeurteilerIn - ermöglicht schließlich eine faire endgültige Stufenzuweisung.

3. Ergebnisse Spanischer PrüfungsteilnehmerInnen

Bislang haben ca. 300 SpanierInnen den TestDaF abgelegt, die Mehrzahl jedoch in Deutschland. In Spanien selbst existieren derzeit sechs Testzentren (genaue Informationen hierzu s. www.testdaf.de). Was die Ergebnisse anbelangt, so zeigt sich, dass 70% der KandidatInnen Resultate im hohen und mittleren Bereich (d.h. TDN 5 und TDN 4) erzielen. Interessant hierbei ist eine genauere Analyse der Ergebnisse hinsichtlich der einzelnen Subtests. Es wird

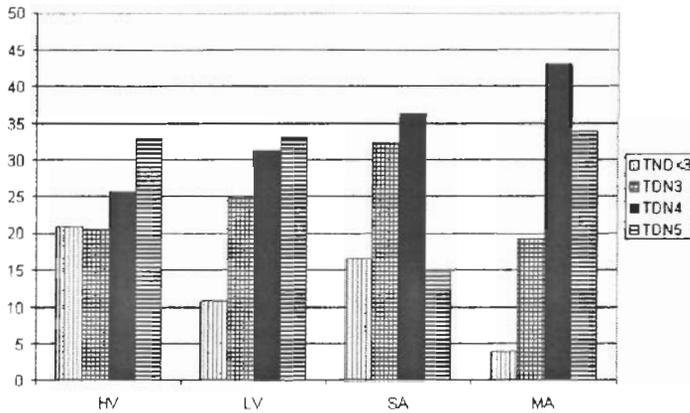


Abbildung 2: Ergebnisse nach Subtest in Prozent, Herkunftsland Spanien

deutlich, dass die besten Ergebnisse im Subtest Mündlicher Ausdruck erzielt werden. Die Resultate bei den Prüfungsteilen Lese- und Hörverstehen liegen tendenziell im mittleren Leistungsbereich. Die größten Probleme bereitet den spanischen Prüfungsteilnehmenden offensichtlich der Schriftliche Ausdruck. Die Gründe hierfür sollten eruiert und geeignete Maßnahmen zur Verbesserung der schriftlichen Ausdrucksfähigkeit entwickelt werden.

4. Vorbereitung auf den TestDaF

Entwicklung bzw. Implementierung und Etablierung eines neuen Tests werfen Fragen hinsichtlich der adäquaten Testvorbereitung auf. Zum einen geht es hierbei um Fragen zur Konzeption von entsprechendem Unterricht (z.B. Intensiv- vs. Extensivkurse), zum anderen um Unterrichtsinhalte und passende Lern- bzw. Lehrmaterialien. Was Vorbereitungsmaterialien anbelangt, so sind inzwischen einige Lehrwerke auf dem Markt, die gezielt auf den TestDaF vorbereiten wollen¹¹. Ebenso etablieren sich vielerorts, sowohl in Deutschland als auch im Ausland, spezielle TestDaF-Vorbereitungskurse. Eine gebündelte Erfassung und Evaluierung dieser Kurse sowie ihrer Konzeption und inhaltliche Gestaltung steht noch aus.

Es zeichnet sich ab, dass Vorbereitung auf den TestDaF oftmals dahingehend verstanden wird, lediglich dem TestDaF-Format entsprechende

11. Eine aktuelle Aufstellung von Materialien, die speziell für die Vorbereitung auf den TestDaF entwickelt worden sind findet sich auf der Internetseite des TestDaF-Instituts, www.testdaf.de. Neben Print-Materialien, die im Buchhandel erhältlich sind, werden mittlerweile auch online zugängliche Materialien und spezielle Kurse angeboten. Auch hierzu finden sich näheren Informationen bzw. Links auf den Seiten des TestDaF-Instituts.

Aufgaben aus den verschiedenen Subtests zu üben. Zweifelsohne ist es für die Prüfungsteilnehmenden unabdingbar, sich mit dem Testformat vertraut zu machen. Beispielsweise ist i.d.R. den Lernenden das kassettengesteuerte Format des Subtests Mündlicher Ausdruck nicht unbedingt vertraut, so dass es dringend notwendig ist, vor der Teilnahme am TestDaF das Sprechen bzw. Reagieren auf Kassette zu üben¹². Gleiches gilt für den Teil Hörverstehen. Hier erfordert das Format, dass in der ersten und dritten Hörverstehensaufgabe die Antworten in Form schriftlicher Notizen und Stichwörter zu geben sind. Auch das wird den meisten Deutschlernenden zunächst ungewohnt erscheinen, da diese Übungsform - vor allem auch in Verbindung mit der Fertigkeit Hörverstehen - möglicherweise nicht zum festen Bestandteil des DaF-Unterrichts gehört. Allerdings ist davor zu warnen, TestDaF-Vorbereitung darauf zu beschränken, möglichst viele TestDaF-ähnliche Übungen vorzunehmen. Die Gefahr besteht darin, lediglich *Testwiseness* zu trainieren und die undifferenzierte Verwendung von Redemitteln bzw. das Auswendiglernen von Mustertexten für die produktiven Prüfungsteile zu unterstützen. *Testwiseness*-Strategien bestehen z.B. darin, im Subtest Leseverstehen bei der Lösung der geschlossenen Items nach plausiblen, sich gegenseitig erklärenden Antworten zu suchen, ohne die entsprechende Textaussage zu verstehen oder gar ohne den Text überhaupt zu lesen. Die Konzentration auf ein *teaching to the test* wäre ein negativer Effekt des Tests auf den Unterricht, aber auch auf die Wahrnehmung seitens der Lernenden (*Impact*), denn nicht der Ausbau der Fähigkeiten stünde im Vordergrund, sondern lediglich die Testresultate wären von Interesse.

Nicht unproblematisch ist es zudem, mangels Aufgaben im TestDaF-Format, diese selbst zu Übungszwecken zu erstellen. Der TestDaF folgt bei der Aufgabenerstellung festen Kriterien z.B. was die Textsorte anbelangt. Die Item-Erstellung selbst ist komplex, die Qualität der Items kann nur sichergestellt werden, indem die Aufgaben erprobt und geeignete Itemanalysen, z.B. zur Ermittlung von Trennschärfen, erfolgen. Denn oftmals wird erst anhand einer solchen Erprobung deutlich, ob Items plausibel sind und deshalb eine hohe Lösungsrate erzielt wurde, ob also **mit** anderen Worten nicht Lese- oder Hörverstehenskompetenz, sondern vielmehr *Testwiseness* zur korrekten Lösung des Items geführt haben.

Anstatt möglichst intensiv anhand von dem TestDaF-Format entsprechenden Aufgaben zu trainieren sollte es bei der Prüfungsvorbereitung vielmehr darum gehen, die Kursteilnehmenden sprachlich auf ein Hochschulstudium in Deutschland vorzubereiten, denn das Konstrukt, das dem TestDaF zugrunde liegt lautet sprachliche Kompetenzen, die im

¹² Zum Zusammenhang zwischen Leistung und Vertrautheit mit dem kassettengesteuerten Format beim Mündlichen Ausdruck des TestDaF s. Kniffka & Üstünsöz-Beurer 2001.

Hochschulkontext relevant sind. Zur Ausbildung dieser Kompetenzen bedarf es keiner TestDaF-adäquater Items und Aufgaben. Vielmehr bieten kommunikativ konzipierte Mittelstufen-Lehrwerke i.d.R. mannigfaltige Materialien und Anregungen zum Training von Verstehensstrategien, von Lese- und Hörstilen sowie von kommunikativen Strategien, die für ein angemessenes sprachliches Handeln an der Hochschule erforderlich sind.

Verschiedene Verlage in Deutschland, aber auch in anderen Ländern bieten Materialien an, um gezielt auf das Prüfungsformat des TestDaF vorzubereiten.¹³ Es handelt sich hierbei um Materialien, die in erster Linie direkt auf die Prüfung TestDaF vorbereiten wollen, also eine Progression wie sie in extensiv konzipierten Sprachkursen vorgesehen ist, nicht berücksichtigen. Das TestDaF-Institut bietet zudem auf seinen Internetseiten nicht nur kostenlose Modellprüfungen an, mit Hilfe derer sich die KandidatInnen gezielt vorbereiten können, sei es im Selbststudium, sei es im Rahmen einer Lerngruppe oder im Kurs, sondern auch prüfungsrelevante Hinweise, welche die Prüfungsziele beschreiben, Tipps für die angemessene Umsetzung der Aufgaben aus den verschiedenen Subtests liefern und die Beurteilungsmaßstäbe erläutert.

Gerade die Standardisierung und die damit verbundene Transparenz hinsichtlich der Prüfungsformate und -inhalte ermöglicht eine gezielte Vorbereitung auf die Prüfung TestDaF. Die damit verbundene Fairness den Prüfungsteilnehmenden gegenüber ist ein nicht unwesentliches Gütekriterium der Prüfung TestDaF.

literaturverzeichnis

- Association of Language Testers in Europe (ALTE). [<http://www.alte.org>].
- ARRAS, U. & Grotjahn, R., 2002. "TestDaF: Aktuelle Entwicklungen". *Fremdsprachen und Hochschule*, 66: 65-88.
- ARRAS, U., Eckes, T. & Grotjahn, R., 2002. "*C-Tests im Rahmen des 'Test Deutsch als Fremdsprache' (TestDaF): Erste Forschungsergebnisse*". In: R. Grotjahn (ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (Bd. 4). Bochum: AKS-Verlag: 175-209.
- BACHMAN, L.F. & Palmer, A.S., 1996. *Language testing in practice. Designing and developing useful language tests*. Oxford University Press.
- BÄRENFÄNGER, O. & Kecker, G., 2004. "*Beurteilerkonsistenz und Beurteilerübereinstimmung bei der Bewertung von Kurzaufgaben: Eine empirische Studie zum Bewertungsverhalten von Beurteilern im*

13. Zu einigen in Deutschland erschienenen Materialien s. die Sammelrezension von Lehker 2003.

- Subtest Hörverstehen des Tests Deutsch als Fremdsprache (TestDaF)* “. Zeitschrift für Fremdsprachenforschung, 15: 237-268.
- BOLTON, S. (ed.), 2000. TESTDAF: *Grundlagen für die Entwicklung eines neuen Sprachtests*. Beiträge aus einem Expertenseminar. München: Goethe-Institut.
- CASPER-HEHNE, H. & KOREIK, U. (eds.). (2004). *DSH und TestDaF als hochschulbezogene Prüfungssysteme für Deutsch als Fremdsprache: Standortbestimmungen und Entwicklungslinien*. Hohengehren: Schneider.
- ECKES, T., 2003. “*Qualitätssicherung beim TestDaF: Konzepte, Methoden, Ergebnisse*“. Fremdsprachen und Hochschule, 69: 43-68.
- ECKES, T., 2004. “*Facetten des Sprachtestens: Strenge und Konsistenz in der Beurteilung sprachlicher Leistungen*“. In: A. Wolff, T. Ostermann & C. Chlosta (eds.), *Integration durch Sprache (Materialien Deutsch als Fremdsprache, Bd. 73)*. Regensburg: FaDaF: 485-518.
- EUROPARAT, 2001. *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin. [<http://www.sprachenportfolio.ch/>].
- KNIFFKA, G. & ÜSTÜNSÖZ-BEURER, D., 2001. “*TestDaF: Mündlicher Ausdruck. Zur Entwicklung eines kassettengesteuerten Testformats*“. Fremdsprachen Lehren und Lernen, 30: 127-149.
- KOREIK, U. (ed.), 2005. *DSH und TestDaF – eine Vergleichsstudie*. Hohengehren: Schneider.
- KOREIK, U. & SCHIMMEL, D., 2002. “*Hörverstehenstests bei der DSH, der Feststellungsprüfung und TestDaF: Eine Vergleichsstudie mit weiterführenden Überlegungen zu TestDaF und DSH*“. InfoDaF, 29,5: 409-440.
- LEHKER, M., 2003. “*TestDaF-Sammelrezension*“. Info DaF 30,2/3: 289-292. TestDaF-Institut [<http://www.testdaf.de>]