

ELiS_e

[e'li:zə]

<Essener Linguistische Skripte – elektronisch>

Christian Krekeler

**TestDaF und DSH –
ungleiche Sprachtests im Vergleich**

elise@uni-essen.de

<http://www.elise.uni-essen.de>

Christian Krekeler

TestDaF und DSH – ungleiche Sprachtests im Vergleich

Zusammenfassung

Anlass für die Studie waren Unsicherheiten von Zulassungsstellen im Umgang mit dem TestDaF-Zeugnis. Mit einer quantitativen Studie (Durchführung der DSH und des TestDaF mit C-Test), welche durch eine qualitative Studie ergänzt wurde (Befragung einzelner Teilnehmer), wurde die Frage verfolgt, welches Ergebnis im TestDaF einer bestandenen DSH entspricht bzw. bei welchem Ergebnis die sprachlichen Voraussetzungen für ein Fachstudium erfüllt sind.

Ein Vergleich der Testergebnisse wird durch die unterschiedlichen Formate und die fehlende Standardisierung der DSH erschwert. Da die Tests der Studie unter unterschiedlichen Voraussetzungen (mit/ohne Konsequenzen für die Kandidaten) und in einem zeitlichen Abstand von drei Monaten stattfanden, waren weitere Einschränkungen zu beachten. Die Studie kommt zu dem Ergebnis, dass ein Recht auf Befreiung von der DSH ausschließlich bei einem TestDaF-Ergebnis „TestDaF-Niveaustufe (TDN) 5 in allen vier Subtests“ unangemessen ist. Studienbewerber sollten mit einem TestDaF-Ergebnis von „durchschnittlich TDN 4“ von der DSH befreit werden. Ob das Ergebnis „DSH bestanden“ auch einem niedrigeren TestDaF-Ergebnis entspricht, kann mit dieser Studie weder bestätigt noch ausgeschlossen werden. Möglicherweise sollte die Frage nach der Äquivalenz jedoch umgekehrt gestellt werden: Sollte sich die DSH nicht vielmehr am standardisierten TestDaF ausrichten? Voraussetzung ist eine Festlegung von angemessenen TestDaF-Ergebnissen für die Studienzulassung.

1. Vergleichbarkeit der Tests

Mit dem „Test Deutsch als Fremdsprache für Studienbewerber“ (TestDaF) existiert neben der „Deutschen Sprachprüfung für den Hochschulzugang“ (DSH) seit dem Jahr 2000 ein weiterer Test, der sich auf die Prüfung der Deutschkenntnisse ausländischer Studienbewerber konzentriert. Der TestDaF, ein zentral gestellter und zentral ausgewerteter Sprachtest für den Hochschulzugang, bietet eine Alternative zur DSH. Studienbewerber oder Studieninteressierte können den TestDaF schon im Heimatland ablegen, sich vor der Einreise nach Deutschland ein Bild über ihre Deutschkenntnisse machen

und den erforderlichen Nachweis der Deutschkenntnisse erbringen. Die Verbreitung des TestDaF wird in den nächsten Jahren zunehmen, denn die Möglichkeit, den Test bereits im Heimatland abzulegen, ist für viele Studienbewerber trotz der Prüfungsgebühren in Höhe von € 90 bis 110 sicherlich attraktiv. Damit wird die Zahl derjenigen zunehmen, die sich mit einem TestDaF-Zeugnis um einen Studienplatz in Deutschland bewerben.

Aus dem Nebeneinander zweier Sprachtests mit unterschiedlichen Vorgehensweisen ergibt sich die Frage nach der Vergleichbarkeit der Ergebnisse. Bislang wird die Zulassung von Studienbewerbern mit TestDaF-Zeugnis unterschiedlich gehandhabt. Die Studie soll Hinweise geben, mit welchem TestDaF-Zeugnis ausländische Studienbewerber zum Studium zugelassen bzw. mit welchem TestDaF-Ergebnis sie von der DSH befreit werden sollten. Grundlage des Beitrags ist eine qualitative und eine quantitative Studie mit Studierenden des Studienkollegs der Fachhochschule Konstanz¹. Zunächst möchte ich einige Unterschiede zwischen beiden Tests vorstellen, die im Rahmen der Studie von Belang sind.

TestDaF: differenzierte Beschreibung der sprachlichen Fertigkeiten

DSH: „bestanden – nicht bestanden“

Ein wichtiger Unterschied zwischen beiden Tests liegt in der unterschiedlichen Rückmeldung über die Ergebnisse. Zunächst zur DSH: Das DSH-Zeugnis enthält in der Regel folgenden Satz:

„Frau/Herr ... hat die ‘Deutsche Sprachprüfung für den Hochschulzugang’ (DSH) bestanden und damit die für ein Studium an einer deutschen wissenschaftlichen Hochschule erforderlichen Kenntnisse der deutschen Sprache nachgewiesen.“
(FaDaF 2001: 10/3)

Wer das Zeugnis vorlegen kann, erfüllt die sprachlichen Voraussetzungen für ein Studium und kann zugelassen werden. Wer ein derartiges Zeugnis nicht vorlegen kann, kann nicht zugelassen werden. Damit erübrigt sich eine Interpretation. In Ausnahmefällen kann ein „Zeugnis über die bestandene DSH mit Auflagen“ ausgestellt werden (vgl. FaDaF 2001: 3/11), differenzierte Aussagen über die Leistungen in den Subtests sind im Unterschied zum TestDaF nicht vorgesehen. Wenn die DSH Teil der Feststellungsprüfung für Studierende von Studienkollegs ist, erhalten die Kandidaten eine Zensur, welche in die Gesamtnote der Feststellungsprüfung eingerechnet wird. Aber auch hier gibt es den Schwellenwert, so dass das wichtigste Resultat „bestanden“ oder „nicht bestanden“ bleibt. Im Studienkolleg der Fachhochschule Konstanz müssen die Kandidaten in der DSH als Teil der Feststellungsprüfung mindestens die Gesamtnote 4,0 auf einer Notenskala von 1,0 bis 5,0 erreichen.

¹ Ich spreche bei den Teilnehmerinnen und Teilnehmern des Studienkollegs von „Studierenden“. Sie haben zwar ihr Fachstudium noch nicht aufgenommen, sind aber im Besitz eines „Studierendenausweises“.

Der TestDaF stellt die Ergebnisse nach sprachlichen Fertigkeiten getrennt dar. In den Subtests Leseverstehen (LV), Hörverstehen (HV), Schriftlicher Ausdruck (SA) und Mündlicher Ausdruck (MA) erfolgt jeweils eine Rückmeldung in TestDaF-Niveaustufen (TDN): TDN 3, TDN 4, TDN 5 (vgl. Tabelle 1). Liegt das Ergebnis unter der TDN 3, wird nicht weiter differenziert, sondern das Ergebnis „unter TDN 3“ vermerkt.

Tabelle 1: TestDaF-Niveaustufen am Beispiel des Subtests Schriftlicher Ausdruck

TDN 5	Kann sich in studienbezogenen Alltagssituationen (z. B. Bericht für Stipendiengeber) sowie im fächerübergreifenden wissenschaftlichen Kontext (z. B. Protokoll, Thesenpapier) zusammenhängend und strukturiert sowie der jeweiligen Textsorte angemessen äußern.
TDN 4	Kann sich in studienbezogenen Alltagssituationen (z. B. Bericht für Stipendiengeber) sowie im fächerübergreifenden wissenschaftlichen Kontext (z. B. Protokoll, Thesenpapier) weitgehend zusammenhängend und strukturiert sowie der jeweiligen Textsorte weitgehend angemessen äußern, allerdings mit sprachlichen Mängeln, die jedoch das Textverständnis nicht behindern.
TDN 3	Kann sich in studienbezogenen Alltagssituationen (z. B. Bericht für Stipendiengeber) nur eingeschränkt zusammenhängend und strukturiert schriftlich äußern; kann sich im fächerübergreifenden wissenschaftlichen Kontext (z. B. Protokoll, Thesenpapier) schriftlich nicht so äußern, dass das Textverständnis hinreichend gesichert ist.

(TestDaF-Institut 2001)

Ein derartig differenziertes Ergebnis ermöglicht eine differenzierte Zulassungspraxis. Es stellt gleichzeitig die Abnehmer der Zeugnisse, ausländische Studienbewerber und Zulassungsstellen, vor die Aufgabe, das Zeugnis zu interpretieren. Es muss festgelegt werden, mit welchem TestDaF-Ergebnis eine Zulassung ausländischer Studierender sinnvoll ist. Bislang gelten folgende Regelungen:

„Wird TDN 5 in allen vier Subtests erreicht, dann bedeutet dies – aufgrund eines Plenarbeschlusses der HRK [Hochschulrektorenkonferenz] 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.“ (Grotjahn/Kleppin 2001: 427)

Das TestDaF-Institut hat eine Handreichung mit Empfehlungen ausgegeben, aus der das Bestreben deutlich wird, Hochschulen in Bezug auf die Sprachkenntnisse ausländischer Studienbewerber differenzierte Aufnahmeregelungen zu ermöglichen.

Tabelle 2: Vorschlag des TestDaF-Instituts zur Zulassung ausländischer Studierender

Subtest Fächergruppe	Leseverstehen	Hörverstehen	Schriftlicher Ausdruck	Mündlicher Ausdruck
Geistes-, Sprach-, Kulturwissenschaften	TDN 5	TDN 4	TDN 4	TDN 4
Jura, Sozial-, Wirtschaftswissenschaften	TDN 4	TDN 4	TDN 4	TDN 4
Informatik, Ingenieurwissenschaften, Kunst und Musik, Mathematik, Naturwissenschaften	TDN 3	TDN 3	TDN 3	TDN 3
	unter TDN 3 keine Zulassung in deutschsprachigen Studiengängen			

(TestDaF-Institut 2002)

Der differenzierte Ergebnisausweis des TestDaF hat zu Diskussionen geführt. Zwei Aspekte sind hervorzuheben: Erstens die Möglichkeit einer nach Studienfach differenzierter Zulassung und zweitens die Zulassung mit Auflagen.

Die Frage, mit welchem TestDaF-Ergebnis Studienbewerber zu einem bestimmten Studiengang aufgenommen werden können, scheint aktueller als die Äquivalenz mit einer bestandenen DSH. Die folgenden Gründe haben mich bewogen, die Vergleichbarkeit der Prüfungen dennoch in den Mittelpunkt der vorliegenden Untersuchung zu stellen: Erstens sind die sprachlichen Anforderungen einzelner Studiengänge meiner Kenntnis nach nicht systematisch erhoben und vor allem nicht verglichen worden. Ob Maschinenbau-Studierende weniger Deutschkenntnisse benötigen als Geisteswissenschaftler, ob die sprachlichen Anforderungen für Kunststudierende deutlich unter denen für Studierende der Kulturwissenschaften liegen, sind offene Fragen, die einer eigenen Untersuchung bedürfen. Zweitens ist die Frage nach der Äquivalenz für die Zulassungsstellen von großer Dringlichkeit. Der Beschluss der Hochschulrektorenkonferenz wird von einigen Zulassungsstellen so ausgelegt, dass nur Studienbewerber von der DSH befreit werden, welche ein TestDaF-Zeugnis mit der TDN 5 in allen vier Subtests vorlegen. Schließlich sollte die Frage nach der Äquivalenz beider Prüfungen – bei allen Schwierigkeiten auf Grund der Uneinheitlichkeit der DSH – zufriedenstellend beantwortet sein, bevor gesonderte Regelungen für einzelne Studienfächer aufgestellt werden. Verzichtet man auf diesen Schritt, bedeutet das auch, auf die langjährigen Erfahrungen mit der DSH zu verzichten.

Zum zweiten Aspekt: Durch den differenzierten Ergebnisausweis des TestDaF ist häufig von einer Zulassung mit Auflagen die Rede. Bei einem mittleren TestDaF-Ergebnis könnte beispielsweise eine bedingte Zulassung ausgesprochen werden mit der Auflage, am studienbegleitenden Sprachunterricht teilzunehmen. Der Sinn studienbegleitender Sprachförderung steht außer Zweifel. Ich habe aber Zweifel, ob man im Sinne ausländischer Studienbewerber handeln würde. Im Normalfall ist es sicherlich günstiger,

wenn ausländische Studierende bereits zu Beginn des Studiums über fortgeschrittene Deutschkenntnisse verfügen. Im Rahmen dieser Studie werden Überlegungen zu einer Zulassung mit Auflagen daher nicht weiter berücksichtigt. Im Mittelpunkt stehen die Fragen, bei welchem TestDaF-Ergebnis ausländische Studienbewerber von der DSH befreit werden sollten bzw. bei welchem TestDaF-Ergebnis ausreichende Sprachkenntnisse für die Aufnahme eines Fachstudiums in Deutschland vorliegen.

TestDaF: standardisiert

DSH: nicht standardisiert

Der TestDaF ist testmethodisch „auf der Höhe der Zeit“. Er ist ein standardisierter Test, d. h. er wird an einer repräsentativen Stichprobe geeicht. Die Entwicklung der Testsätze verläuft in mehreren Stufen, die Aufgaben werden vorerprobt, erprobt, gegebenenfalls revidiert und geeicht. Das Ergebnis ist ein Sprachtest, der die Gütekriterien von Sprachtests in hohem Maße erfüllt. Man kann davon ausgehen, dass die Ergebnisse von zwei Kandidaten unabhängig von der Testversion vergleichbar sind. Ein Kandidat, welcher an unterschiedlichen Testversionen teilnimmt, würde bei gleicher Leistung gleiche Ergebnisse erzielen.

Die DSH ist ein Beispiel für eine Prüfungstradition, welche vom Unterricht, von den Unterrichtenden und von der Institution ausgeht. Die Rahmenordnung für die DSH wurde von der Hochschulrektorenkonferenz 1995 in Kraft gesetzt und inzwischen wohl von allen Hochschulen in Deutschland umgesetzt. Im Auftrag des „Fachverbandes Deutsch als Fremdsprache“ (FaDaF) wurde 2001 ein Handbuch für Prüferinnen und Prüfer herausgegeben, das die Rahmenordnung durch Anwendungsbeispiele und Hinweise zur Durchführung konkretisieren soll. Auch über eine „Zertifizierung“ von Prüfungen soll eine Vergleichbarkeit gewährleistet werden. Dennoch: Die DSH ist keine standardisierte Prüfung, die Gütekriterien von Sprachtests dürften kaum zufriedenstellend erfüllt werden. Angesichts der fehlenden Standardisierung der DSH bezweifelt Alderson (2002), dass die DSH überhaupt als Test anzusehen ist. Seiner Meinung nach ist es angemessener, von einem Gerüst („framework“) zu sprechen. Vergleicht man DSH-Prüfungen miteinander, wird deutlich, dass die DSH-Rahmenordnung ein breites Spektrum an Tests ermöglicht bzw. dass sie unterschiedlich ausgelegt wird. Deutliche Abweichungen im Subtest Hörverstehen offenbarte beispielsweise die Studie von Koreik/Schimmel (2002). Kandidaten, welche am DSH-Subtest Hörverstehen verschiedener Institute teilnahmen, erzielten sehr unterschiedliche Ergebnisse. Die Prüfungsordnung erlaubt außerdem die Kombination von Testteilen, so dass eine DSH aus zwei bis vier Subtests bestehen kann. Für Untersuchungen über die DSH hat diese Problematik Folgen: Aussagen über den Schwierigkeitsgrad einer DSH oder einer ihrer Subtests sind immer unter der Einschränkung zu betrachten, dass die Tests möglicherweise stark voneinander abweichen. Auch für Kandidaten hat die eingeschränkte Vergleichbarkeit Folgen: Sie können nicht wissen, ob sie bei einer anderen DSH das gleiche Ergebnis erzielt hätten.

*TestDaF: integriertes Testen der sprachlichen Fertigkeiten**DSH: zusätzlicher Grammatiktest*

Der TestDaF besteht aus vier Subtests, in denen die sprachlichen Fertigkeiten integriert geprüft werden (vgl. Tabelle 3). Ein getrennter Test sprachlicher Elemente (Grammatik, Phonetik, Lexik) findet nicht statt. Bei den Subtests Schriftlicher Ausdruck und Mündlicher Ausdruck handelt es sich um direkte Performanztests. Sie stellen die Frage, ob die Kandidaten ihre Sprachfähigkeit in einer spezifischen Situation anwenden können. Auch die Subtests Leseverstehen und Hörverstehen sind als Performanztest anzusehen, auch wenn sie eher indirekt vorgehen müssen, denn das Verstehen eines Textes wird anhand von Fragen über den Text geprüft.

Tabelle 3: Subtests des TestDaF und der DSH im Vergleich

TestDaF	DSH
Leseverstehen Der Subtest LV besteht aus drei Texten mit unterschiedlichem Schwierigkeitsgrad. Insgesamt sind 30 Items zu bearbeiten (Zuordnung, Mehrfachauswahlaufgaben, Auswahlitems). Die TestDaF-Niveaustufen werden aus den Rohwerten ermittelt (max. 30 Punkte). Dauer: 60 Minuten.	Verstehen und Bearbeiten eines Lesetextes (DSH-LV) Aufgaben zu einem Lesetext, der authentisch, studienbezogen und wissenschaftsorientiert ist.
Hörverstehen Der Subtest HV besteht aus drei Hörtexten mit unterschiedlichem Schwierigkeitsgrad. Insgesamt sind 25 Items zu bearbeiten (gesteuerte Notizen, Auswahlitems). Die TestDaF-Niveaustufen werden aus den Rohwerten ermittelt (max. 25 Punkte). Dauer: 40 Minuten.	Verstehen und Verarbeiten eines Hörtextes (DSH-HV) Aufgaben zu einem Hörtext, der max. zweimal präsentiert wird. Der Text soll „der Kommunikationssituation Vorlesung/Übung angemessen Rechnung tragen“.
Schriftlicher Ausdruck Beim Subtest SA soll ein Text zu einer Aufgabenstellung, welche eine Grafik umfasst und eine begründete Stellungnahme verlangt, verfasst werden. Die Bewertung (TestDaF-Niveaustufen) erfolgt durch zwei, im Zweifel durch drei Prüfer. Dauer: 60 Minuten.	Vorgabenorientierte Textproduktion (DSH-TP) Texterstellungsaufgabe. Sie kann „erklärender, vergleichender oder kommentierender Art sein, sie kann auch die sprachliche Umsetzung von Grafiken, Schaubildern, Diagrammen zum Gegenstand haben.“
	Verstehen und Bearbeiten wissenschafts-sprachlicher Strukturen (DSH-Grammatiktest) Textgebundene Aufgabenstellung. Die Leistung wird nach sprachlicher Richtigkeit bewertet.
Mündlicher Ausdruck Der MA wird als „Simulated Oral Proficiency Interview“ (SOPI) durchgeführt. In diesem kassettengesteuerten Subtest müssen die Kandidaten auf zehn kurze Sprechanlässe reagieren. Bewertung erfolgt in TestDaF-Niveaustufen. Dauer: 30 Minuten.	Mündliche Prüfung Prüfungsgespräch, Dauer: max. 20 Minuten. Sie kann entfallen, wenn „für die Beurteilung der mündlichen Kommunikationsfähigkeit hinreichende Erkenntnisse vorliegen.“

(vgl. TestDaF-Institut 2001, FaDaF 2001. Prüfungsordnung der DSH, zit. nach FaDaF 2001)

Eine Besonderheit der Subtests MA, LV und HV sind Aufgabengruppen auf unterschiedlichen Anforderungsniveaus. Beim Leseverstehen müssen beispielsweise drei unterschiedliche Textsorten mit unterschiedlichen Aufgabentypen bearbeitet werden (vgl. Grotjahn 2000). Der Subtest Mündlicher Ausdruck wird als kassettengesteuertes Interview durchgeführt. Die Kandidaten sprechen nicht mit einem Prüfer, sondern erhalten einen Sprechanlass vom Band. Dieses Format bewährte sich als zuverlässig und valide (vgl. Kenyon 2000, Kniffka/Üstünsöz-Beurer 2001).

Die Subtests der DSH sind zwar in der Prüfungsordnung beschrieben und im „DSH-Handbuch für Prüferinnen und Prüfer“ erläutert, doch ob sie von den durchführenden Institutionen in dieser Weise umgesetzt werden, wird nicht geprüft. Einige Eckpunkte sind hervorzuheben: Die DSH-Subtests Textproduktion und Mündliche Prüfung sind ebenfalls den direkten Performanztests zuzurechnen. Allein beim DSH-Grammatiktest handelt es sich um einen Kompetenztest. Sprachliche Kompetenztests haben die allgemeine Beherrschung von Sprache unabhängig von einer konkreten Anwendungssituation zum Gegenstand. Sie bestehen eher aus dem Abfragen von Einzelfertigkeiten. Die übrigen Subtests sind durchaus mit denen des TestDaF vergleichbar, allerdings verzichten sie auf abgestufte Aufgabenstellungen. Verallgemeinernde Aussagen zum Format der einzelnen Subtests sind jedoch kaum zu treffen. Eine Besonderheit der DSH, die im Rahmen dieser Untersuchung durchgeführt wurde, ist ein Fachbezug der Subtests Textproduktion und Leseverstehen. Die Kandidaten konnten aus mehreren Aufgaben/Texten wählen. Bei der Textproduktion gab es beispielsweise Aufgaben aus den Bereichen Naturwissenschaften, Technik, Wirtschaft, Informatik, Design und Bauingenieurwesen.

Ich fasse zusammen: Ein Vergleich der Testergebnisse aus der DSH und dem TestDaF ist wünschenswert. In dieser Studie geht es allein um die Äquivalenz der Ergebnisse. Die Vergleichbarkeit beider Tests wird durch die mangelnde Standardisierung der DSH und die unterschiedlichen Ergebniskategorien erschwert. Die Stärke der DSH lässt sich nicht an den klassischen Kriterien der Testgüte festmachen. Sie liegt in der Nähe der Prüfer zu den Kandidaten, zum Unterricht und zum Studium. Beim TestDaF sind geringe Abweichungen zwischen einzelnen Testjahrgängen zwar unvermeidbar (vgl. Grotjahn 2000), doch die aufwändige Eichung gewährleistet eine hohe Vergleichbarkeit der Tests. Aussagen zum Anforderungsniveau des TestDaF können daher verallgemeinert werden.

2. Beschreibung der Studie

Grundlage dieses Beitrags ist eine quantitative Studie mit 67 Studierenden des Studienganges an der Fachhochschule Konstanz. Sie wurde durch eine qualitative Studie mit ausgewählten Teilnehmern der Testreihe erweitert (vgl. Tabelle 4).

Tabelle 4: Ablauf der Untersuchung

Quantitative Studie	
Phase 1 TestDaF-Erprobung, C-Test, DSH-Grammatiktest 4 Lerngruppen des Studienkollegs (58-67 Probanden)	Phase 2 (3 Monate später) DSH 2 Lerngruppen des Studienkollegs (39) und Externe (192 Probanden)
Qualitative Studie	
Befragung und Beobachtung ausgewählter Kandidaten	

Quantitative Studie

Im Rahmen der quantitativen Studie wurden mehrere Sprachtests mit einer Stichprobe von 67 ausländischen Studienbewerbern durchgeführt, die zum Zeitpunkt der Erhebung als Studierende am Studienkolleg der Fachhochschule Konstanz eingeschrieben waren. Zunächst nahmen die Studentinnen und Studenten an einer Erprobungsfassung des TestDaF teil, welche auch einen C-Test umfasste. Zeitnah (zwei Wochen später) machten sie den ersten DSH-Grammatiktest. Die DSH wurde etwa drei Monate später durchgeführt (vgl. Tabelle 4). Auch 192 externe Kandidaten nahmen an der DSH teil. Der TestDaF, welcher von den Studierenden bearbeitet wurde, war eine Erprobungsfassung in der zweiten Stufe. In dieser Phase werden die vorerprobten Aufgaben nochmals auf ihre Güte überprüft und die Ergebnisse der Subtests Leseverstehen sowie Hörverstehen werden den TestDaF-Niveaustufen zugeordnet (vgl. Arras/Eckes/Grotjahn 2002, TestDaF-Institut 2001). Die eingesetzte Erprobungsfassung unterschied sich nur geringfügig von der Endversion. Der erste DSH-Grammatiktest wurde dem DSH-Handbuch entnommen, wo er beispielhaft angeführt wird. Die DSH wurde von einem Team erstellt und korrigiert. Eine Besonderheit der DSH an der Fachhochschule Konstanz ist der Fachbezug der Testteile Textproduktion und Leseverstehen. Bei beiden Subtests werden den Kandidaten mehrere Texte bzw. Aufgaben zur Auswahl angeboten, welche sich auf unterschiedliche Studienfächer beziehen. Da bei den Kandidaten aus dem Studienkolleg insgesamt nur vier Mündliche Prüfungen durchgeführt wurden, wurde dieser Prüfungsteil nicht in die Untersuchung einbezogen.

Mit Ausnahme der TestDaF-Subtests Mündlicher Ausdruck und Schriftlicher Ausdruck lagen intervallskalierte Daten vor, das heißt die Abstände zwischen den einzelnen Skalenwerten sind konstant. Bei den TestDaF-Niveaustufen könnte das Vorliegen einer Intervallskala bestritten werden, was an dem Ergebnis „unter 3“ liegt. Beim TestDaF werden alle Ergebnisse unter der TDN 3 dem Ergebnis „unter 3“ zugeordnet. Die Erfahrung zeigt aber, dass sich auch das Ergebnis „unter 3“ wie eine weitere Ergebnisklasse auf einer quantifizierenden Skala verhält. Die bivariaten Korrelationen wurden mit Hilfe von Rangkorrelationskoeffizienten bestimmt, welche von der „niedrigeren“ Rang- oder Ordinalskala ausgehen.

Bei der Interpretation sind folgende Einschränkungen zu beachten:

- Auch in der zweiten Phase der TestDaF-Erprobung kann es zu Änderungen in der Testfassung kommen. Die Ergebnisse im Subtest LV könnten beispielsweise zu einer Überarbeitung der Items geführt haben. Das Schwierigkeitsniveau der Endfassung kann über oder unter der Erprobungsfassung liegen.
- Der TestDaF und der erste DSH-Grammatiktest traf alle Kandidaten völlig unvorbereitet. Die Tests wurden ohne Ankündigung am Beginn eines neuen Semesters durchgeführt. Man kann davon ausgehen, dass Teilnehmer bei einer Sprachprüfung besser abschneiden, wenn sie sich vorher mit dem Format vertraut gemacht haben. Wie groß die Auswirkungen einer Vorbereitung auf die Ergebnisse der einzelnen Testteile sind, ist jedoch unklar.
- Die Ergebnisse aus dem TestDaF und dem ersten DSH-Grammatiktest hatten für die Teilnehmer keine Auswirkungen. Es ist möglich, dass sich einige Kandidaten nicht in gleichem Maße angestrengt haben, bei anspruchsvollen Aufgaben beispielsweise eher aufgegeben haben oder erst gar keinen Antwortversuch unternommen haben als bei einer Prüfung mit gewichtigen Konsequenzen. Die ersten beiden Aspekte sprechen dafür, dass die Ergebnisse aus dem TestDaF und dem ersten Grammatiktests möglicherweise unter den Ergebnissen liegen, die sich in einer Testsituation ohne Probecharakter ergeben hätten.
- Zwischen der ersten (TestDaF, C-Test und DSH-Grammatiktest 1) und der zweiten Phase (DSH mit DSH-Grammatiktest 2) lag ein Semester Unterricht im Studienkolleg, welcher zehn Wochenstunden Deutschunterricht einschloss. In diesem Unterricht wurden die Studierenden auf die spezifischen Anforderungen der DSH vorbereitet und mit dem Prüfungsformat vertraut gemacht. Im Vergleich zu den externen Kandidaten haben die Studierenden aus dem Studienkolleg signifikant bessere Ergebnisse erzielt.
- Die DSH am Ende des Semesters hatte gewichtige Konsequenzen für die Kandidaten. Nur mit einer bestandenen DSH erhalten die Studierenden aus dem Studienkolleg das Feststellungszeugnis, welches sie für ein Fachstudium in Deutschland benötigen. Man muss davon ausgehen, dass sie ihr ganzes sprachliches Können aufgeboren haben, um ein gutes Ergebnis zu erzielen.
- Die mangelnde Standardisierung stellt eine Übertragbarkeit von Studien mit nur einer DSH grundsätzlich in Frage.

Für die Interpretation der Daten haben diese Umstände Folgen: Die Stichprobenergebnisse im TestDaF sind im Vergleich zur Population eher unterbewertet, die Ergebnisse in der DSH eher überbewertet.

Qualitative Studie

Bei der Analyse der Ergebnisse aus der quantitativen Studie traten zwei Schwierigkeiten auf: Zum einen war die Größe der Stichprobe gering. Lagen in der ersten Phase der quantitativen Studie mit 57 bis 65 Probanden noch ausreichend große Stichproben vor,

so waren statistisch signifikante Aussagen auf Basis der Ergebnisse von 29 bis 32 Probanden in der zweiten Phase nicht immer möglich. Zum anderen gaben die statistischen Daten wenig Hinweise auf die Ursachen auffälliger Erscheinungen. Die Interpretation der Ergebnisse stieß auf Hindernisse, weil Informationen über die individuellen Voraussetzungen der Kandidatinnen und Kandidaten fehlten. Vor allem die letztgenannte Schwierigkeit veranlasste mich dazu, im Anschluss an die statistische Auswertung der Tests aus der ersten Phase eine qualitative Studie durchzuführen. Die ausgesprochene Heterogenität als durchgehendes Merkmal der Population „ausländische Studienbewerber“ war ein weiterer Grund für die Berücksichtigung individueller Lernwege und Erfahrungen bei der Interpretation der Testergebnisse. Mit einer Kombination von quantitativen und qualitativen Methoden sollte ein umfassenderes Bild entstehen als dies allein mit statistischen Untersuchungen an der Stichprobe möglich war.

Leitgedanke der Auswahl war nicht, wie repräsentativ die Kandidaten waren; die Auswahl war vielmehr ein Ergebnis der Datenanalyse. Die Gespräche, welche in Form eines Leitfaden-Interviews geführt wurden, waren unterschiedlich formell und von unterschiedlicher Dauer: Manche Gespräche fanden nach der Lehrveranstaltung im Unterrichtsraum oder auf dem Flur statt, andere etwas formaler im Büro. Manchmal waren auch andere Studierende anwesend. In der Regel fertigte ich während der Gespräche Notizen an, in Ausnahmefällen danach. Die Kandidaten wurden nach biografischen Angaben und vor allem nach ihrer Sprachlernbiografie befragt. Außerdem wurden sie um Kommentare zu den Tests und den Testergebnissen gebeten.

Beschreibung der Stichprobe

Die Stichprobe entspricht in wesentlichen Merkmalen der Population „ausländische Studienbewerber“, etwa was die Verteilung der Herkunftsländer und Ausgangssprachen betrifft. Hinsichtlich der Herkunft und der Muttersprache(n) waren die Gruppen sehr heterogen. Die größten Gruppen stellten Studierende aus China, Marokko und dem Libanon. Die Stichprobe ist mit einer Einschränkung als spezifisch repräsentativ anzusehen. Die Einschränkung bezieht sich auf den Besuch des Studienkollegs. Nicht alle ausländischen Studienbewerber haben die Möglichkeit (bzw. die Pflicht), an einem einjährigen studienvorbereitenden Kurs teilzunehmen. Die Studierenden werden in zwei Semestern auf die DSH, die sie als Teil der Feststellungsprüfung ablegen, vorbereitet. An den Ergebnissen der DSH, an der 39 Studierende aus dem zweiten Semester des Studienkollegs und 192 externen Kandidaten teilnahmen, wird deutlich, dass sich die Studierenden aus dem Studienkolleg von externen Bewerbern unterscheiden, die sich – häufig im Heimatland – auf die sprachlichen Anforderungen des Studiums vorbereiten (vgl. Tabelle 5).

Die DSH gilt als bestanden, wenn sowohl die mündlichen als auch die schriftlichen Prüfungsleistungen mindestens mit 4,0 bewertet werden. Da sich nicht alle Kandidaten einer Mündlichen Prüfung unterziehen müssen, werden nur die Ergebnisse der schriftlichen Prüfung untersucht. 38 Studierende aus dem Studienkolleg haben den schriftlichen

Teil der DSH mit der Note 4,0 oder besser bestanden. (Dieses Resultat veränderte sich durch die Mündlichen Prüfungen nicht.) Nur eine Studentin aus dem Studienkolleg bestand nicht (Ergebnis der Schriftlichen Prüfung: 4,2). Zum Vergleich: Bei den 192 externen Kandidaten lautete das Ergebnis der DSH-Gesamprüfung (Mündliche und Schriftliche Prüfung) 56 Mal „nicht bestanden“, was etwa 30 Prozent entspricht. Im Mittel erzielten die internen Kandidaten in der Schriftlichen Prüfung ein deutlich besseres Ergebnis als die externen Kandidaten (Interne: $M = 2,67$; Externe: $M = 3,42$). Mit einem t -Test für unabhängige Stichproben konnte gezeigt werden, dass der Leistungsunterschied signifikant ist ($t(229) = 4,672$, $p < 0,01$ 2-seitig). In der Textproduktion war der Unterschied zwischen den externen und internen Kandidaten am größten.

Tabelle5: Statistische Kennwerte, DSH (Schriftliche Prüfung)

	Gruppe	Mittelwert (M)	Standardabweich. (s)	Min	Max	Anzahl (n)
DSH-Ergebnis der Schriftlichen Prüfung (Note 1,0-5,0)	Studienkolleg	2,67	0,83	1,1	4,2	39
	Externe	3,42	0,93	1,1	5,0	192
DSH-Leseverstehen (Note 1,0-5,0)	Studienkolleg	2,95	0,96	1,2	4,5	39
	Externe	3,53	0,96	1,2	5,0	192
DSH-Hörverstehen (Note 1,0-5,0)	Studienkolleg	2,26	1,04	1,0	4,7	39
	Externe	3,15	1,30	1,0	5,0	192
DSH-Textproduktion (Note 1,0-5,0)	Studienkolleg	2,35	0,92	1,0	4,0	39
	Externe	3,52	1,22	1,0	5,0	192
DSH-Grammatik 2 (Note 1,0-5,0)	Studienkolleg	3,14	1,08	1,1	5,0	39
	Externe	3,50	1,38	1,1	5,0	192

Welche Ursachen gibt es für die besseren Ergebnisse der internen Kandidaten? Zunächst dürfte ein nahe liegender Grund zutreffen, eine bessere Sprachkompetenz der Studierenden aus dem Studienkolleg. Sie halten sich seit mindestens einem Jahr in Deutschland auf, wurden anhand eines Aufnahmetests für das Studienkolleg aus einer großen Gruppe von Bewerbern ausgewählt und nahmen anschließend zwei Semester an einem intensiven Deutschunterricht teil. Sie sind außerdem mit dem Format der DSH gut vertraut, es werden aber keine Texte oder Aufgaben in die DSH aufgenommen, die bereits im Unterricht behandelt wurden. An den Leistungen im Subtest Textproduktion zeigt sich die Wirkung der wissenschaftspropädeutischen Arbeit im Studienkolleg. Nach der zweisemestrigen Ausbildung im Studienkolleg sind die Studierenden in der Lage, den Leseerwartungen deutscher Prüfer eher zu entsprechen als die externen Kandidaten. Außerdem sind einige Kandidaten nicht zur Prüfung angetreten, wenn sich aus den Leistungen im Semester abzeichnete, dass die Chancen auf Bestehen der DSH/Feststellungsprüfung gering sein würden. Bei den externen Kandidaten sind die Voraussetzungen sehr unterschiedlich, was sich auch in einer größeren Streuung der Ergebnisse zeigt (Externe: $s = 0,93$, Interne: $s = 0,83$; vgl. Tabelle 5). Gemeinsames Merkmal der externen Kandi-

daten ist ein Sprachzeugnis der Mittelstufe, das als Voraussetzung für die Teilnahme an der DSH verlangt wird.

Tabelle 6: Teilnehmer der quantitativen Studie

	Sem. 1 W-Kurs	Sem. 1 T-Kurs	Sem. 2 W-Kurs	Sem. 2 T-Kurs	externe Kandi- daten
Phase 1: TestDaF Erprobungsfassung mit C-Test, DSH-Grammatik 1	✓	✓	✓	✓	
Phase 2: DSH Gesamprüfung (einschl. DSH-Grammatik 2)			✓	✓	✓

Anmerkung: Der „W-Kurs“ (Wirtschaftskurs) bereitet auf wirtschaftswissenschaftliche Studiengänge vor. Der „T-Kurs“ (Technikkurs) bereitet auf technische Studiengänge vor (auch Informatik, Architektur).

Die Studierenden werden im Studienkolleg in vier Gruppen mit 18 bis 24 Teilnehmern unterrichtet. Es gibt zwei Lerngruppen für Teilnehmer, die ein technisches Studienfach anstreben (auch Informatik oder Architektur), und zwei Lerngruppen für Teilnehmer, die wirtschaftswissenschaftliche Studiengänge belegen möchten. Dass sich die sprachlichen Leistungen der beiden Gruppen unterscheiden, ging aus den Ergebnissen des vom TestDaF-Institut entwickelten C-Tests hervor, welcher zur Erprobungsfassung des TestDaF gehört. Der C-Test gilt als valider und ökonomischer Test der „allgemeinen Sprachkompetenz“ (vgl. Grotjahn 1995, 1997). Die Leistungen der Studierenden aus den W-Kursen lagen deutlich über denjenigen aus T-Kursen (W-Kurs: $M = 60\%$, T-Kurs: $M = 51\%$). Wenn man davon ausgeht, dass der C-Test ein gutes Maß für die allgemeine Sprachkompetenz darstellt, liegt die Sprachkompetenz der Studierenden aus dem W-Zweig signifikant über derjenigen von Studenten aus dem T-Zweig (t -Test für unabhängige Stichproben: $t(62) = 2,487$, $p < 0,05$ 2-seitig). Die Deutschkenntnisse der Studienbewerber mit den Studienzielen Informatik, Maschinenbau, Elektrotechnik usw. liegen demnach deutlich unter derjenigen von Studienbewerbern, welche wirtschaftswissenschaftliche Studiengänge anstreben. Dies ist mit Blick auf eine differenzierte Zulassungspraxis, welche der TestDaF ermöglicht, zwar interessant, aber kein Hinweis darauf, ob Studierende technischer Studienfächer weniger Deutschkenntnisse benötigen.

An der Studie nahmen Studierende aus dem ersten und dem zweiten Semester des Studienkollegs teil (vgl. Tabelle 6). Es war zu erwarten, dass die Ergebnisse der Studierenden aus dem zweiten Semester im C-Test im Mittel über denen aus dem ersten Semester lagen (1. Sem.: $M = 51,4\%$, 2. Sem.: $M = 59,6\%$). Auch dieser Unterschied ist signifikant ($t(62) = 2,17$, $p < 0,05$ 2-seitig). Die Studierenden aus dem ersten Semester hatten einige Wochen vorher an einem von uns konzipierten Aufnahmetest teilgenommen, der auch einen Deutschteil enthält. Die andere Gruppe der Teilnehmer befand sich am Beginn des zweiten Semesters. Sie absolvierten bereits ein Semester des Studienkollegs und wurden in das zweite Semester versetzt. Voraussetzung dafür sind u. a. mindestens ausreichende Leistungen im Fach Deutsch.

3. Ergebnisse

3.1 Ergebnisse der quantitativen Studie

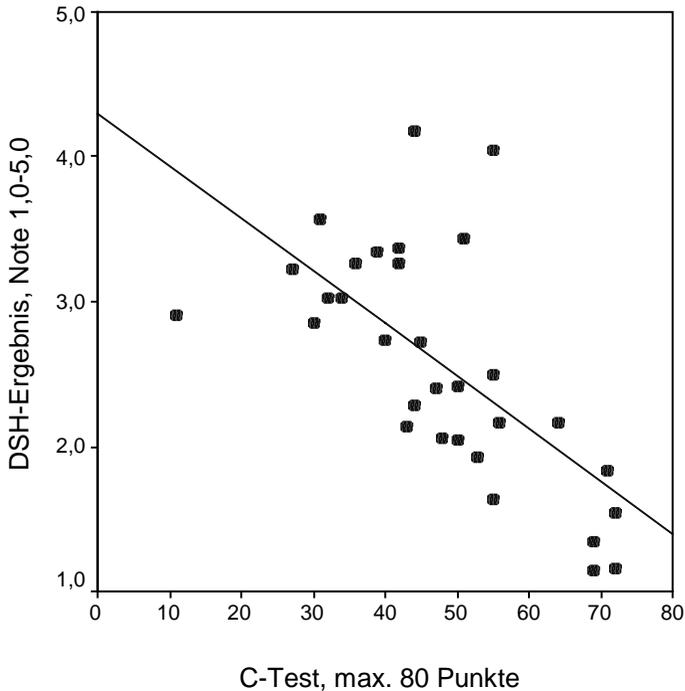
C-Test als „Screening Test“ für den TestDaF geeignet

Vor dem Vergleich der Ergebnisse aus dem TestDaF und der DSH möchte ich kurz auf den C-Test eingehen, der mit jeder TestDaF-Erprobungsfassung zur Verankerung der Subtests durchgeführt wird. In einer Studie von Arras, Eckes und Grotjahn (2002) zeigte der C-Test hohe Korrelationen mit allen vier Subtests des TestDaF. Auf Grund der hohen Korrelationen wird außerdem ein Einsatz des C-Tests als „Screening-Test“ für den TestDaF angeregt. Kandidaten könnten von den Ergebnissen eines frei zugänglichen C-Tests ihre Teilnahme am TestDaF abhängig machen. In der vorliegenden Untersuchung liegen die Rangkorrelationskoeffizienten (Spearman) der TestDaF-Subtests mit dem C-Test im Mittelbereich zwischen 0,526 und 0,589 (vgl. Tabelle 7). Sie sind etwas niedriger als die von Arras, Eckes und Grotjahn beobachteten, was eine Folge unterschiedlicher Auswertungsmethoden des C-Tests sein könnte. Sie weisen auf einen mittleren bis guten Zusammenhang zwischen dem C-Test und den TestDaF-Subtests hin, was den Einsatz des C-Tests als Prädiktor für den TestDaF als sinnvoll bestätigt.

C-Test möglicherweise auch als „Screening Test“ für die DSH geeignet

Wäre ein Einsatz des C-Tests als „Screening-Test“ auch für die DSH sinnvoll? Trotz des zeitlichen Abstandes findet man beträchtliche Korrelationen zwischen dem C-Test und den Subtests der DSH und nicht zuletzt mit dem DSH-Gesamtergebnis ($r_s = -0,676$; vgl. Tabelle 7). Das Gesamtergebnis der DSH wird als ungewichteter additiver Index aus den Ergebnissen der vier Subtests gebildet (Schriftliche Prüfung). Die enge Korrelation wird durch das Streudiagramm bestätigt (vgl. Abbildung 1). Mit Ausnahme der Ergebnisse von drei oder vier Kandidaten liegen die Ergebnisse dicht an der linearen Regressionsgeraden. Vor allem im oberen Leistungsspektrum besteht ein enger Zusammenhang. Kandidaten, welche zuvor im C-Test ein hohes Ergebnis erzielt haben, können bei der DSH ebenfalls mit einem guten Ergebnis rechnen. Einschränkend muss darauf hingewiesen werden, dass drei Monate zwischen den beiden Tests lagen und dass die Studierenden des Studienkollegs besonders gut auf die DSH vorbereitet waren. Außerdem haben 30 der 31 Kandidaten, welche sowohl an der DSH als auch am C-Test teilnahmen, die DSH bestanden, so dass die Frage nicht beantwortet werden kann, bei welchem C-Test-Ergebnis das Bestehen der DSH gefährdet wäre.

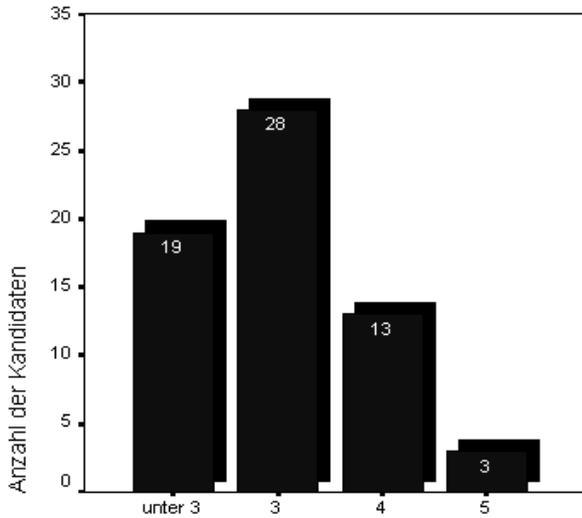
Abbildung 1: Ergebnisse aus der DSH und dem C-Tests im einem Streudiagramm mit linearer Regressionsgeraden (n=31)



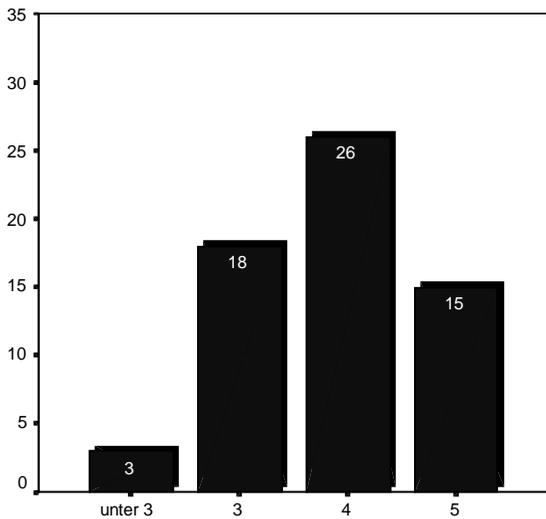
TestDaF: TDN 5 selten

Wie zu erwarten war, sind die mittleren Niveaustufen häufiger vertreten als die Extremwerte (vgl. Abbildung 2). Die TDN 5 ist unterschiedlich oft erzielt worden: Beim Mündlichen Ausdruck und beim Hörverstehen erzielte fast jeder vierte Kandidat die TDN 5 (24 %). Anders beim Leseverstehen. Dort erreichten nur 5 Prozent die höchste Niveaustufe. Das häufigste Ergebnis (Modus) war die TDN 3. Nur beim Hörverstehen wurde die TDN 4 am häufigsten erreicht. Da sich viele Kandidaten zum Testzeitpunkt bereits seit mehreren Monaten in Deutschland aufhielten, verwundern die relativ guten Leistungen in den Prüfungsteilen HV und MA nicht. Überraschend sind aber die Schwierigkeiten, welche die Kandidaten mit dem LV hatten.

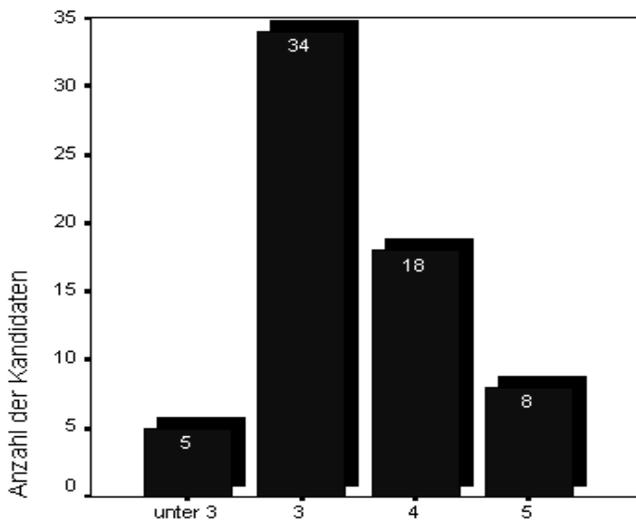
Abbildung 2: Ergebnisse der einzelnen TestDaF-Subtests (n = Summe der Säulenwerte)



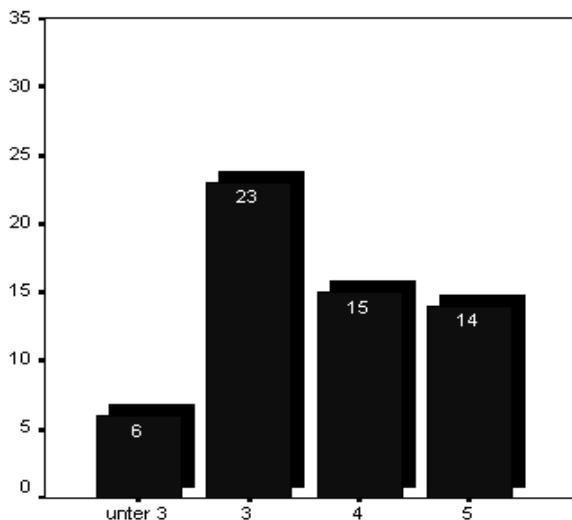
Leseverstehen, TDN



TestDaF Hörverstehen TDN



Schriftlicher Ausdruck, TDN



Mündlicher Ausdruck, TDN

Tabelle 7 zeigt die Korrelationen zwischen den Subtests der DSH und des TestDaF. Die Korrelationen wurden mit den Rangkorrelationskoeffizienten von Spearman (untere

Dreiecksmatrix) bzw. Kendalls-Tau-b (obere Dreiecksmatrix) berechnet. Korrelationen sind ein Maß für die Übereinstimmung von Messwerten. Der Kendalls-Tau-Koeffizient gibt Auskunft über die Wahrscheinlichkeit, mit der die Reihenfolge zweier zufällig ausgewählter Kandidaten wie beobachtet eintritt (vgl. Bortz/Döring 2002: 508-509). Die Werte liegen jeweils unter den Korrelationen von Spearman, sie folgen aber dem gleichen Muster.

Tabelle 7: Korrelationen zwischen den Tests der Studie (gesamte Stichprobe)

	DSH-LV	DSH-HV	DSH-TP	DSH-Grammatik 2	DSH-Grammatik 1	C-Test	Test-DaF LV	Test-DaF HV	Test-DaF SA	Test-DaF MA
Kendall-Tau-b (τ)										
DSH LV		,456** n=231	,394** n=231	,395** n=231	,490** n=34	,433** n=31	,408** n=31	,492** n=31	,430** n=31	,339* n=29
HV	,627** n=231		,371** n=231	,267** n=231	,420** n=34	,317** n=31	,537** n=31	,439** n=31	,390** n=31	,568** n=29
TP	,539** n=231	,518** n=231		,397** n=231	-,411** n=34	-,571** n=31	-,561** n=31	-,378** n=31	-,360* n=31	-,320* n=30
Grammatik 2	,554** n=231	,384** n=231	,552** n=231		,534** n=34	,398** n=31	,283* n=31	,227 n=31	,356* n=31	,164 n=29
Grammatik 1	,644** n=34	,549** n=34	-,564** n=34	,697** n=34		,340** n=59	,295** n=58	,324** n=57	,449** n=60	,164 n=55
C-Test	,613** n=31	,437* n=31	-,713** n=31	,555** n=31	,477** n=59		,395** n=62	,439** n=61	,431** n=63	,415** n=57
Test DaF-LV	,582** n=31	,733** n=31	-,714** n=31	,403* n=31	,411** n=58	,543** n=62		,386** n=61	,357** n=63	,400** n=57
HV	,610** n=31	,609** n=31	-,513** n=31	,316 n=31	,459** n=57	,589** n=61	,523** n=61		,470** n=61	,594** n=56
SA	,525** n=31	,480** n=31	-,472** n=31	,436* n=31	,537** n=60	,533** n=63	,440** n=63	,571** n=61		,437** n=57
MA	,423* n=29	,683** n=29	-,414* n=29	,213 n=29	,228 n=55	,526** n=57	,494** n=57	,718** n=56	,500** n=57	
Spearman's rho (r_s)										

** Korrelation ist auf dem Niveau von 0,01 signifikant (2-seitig).

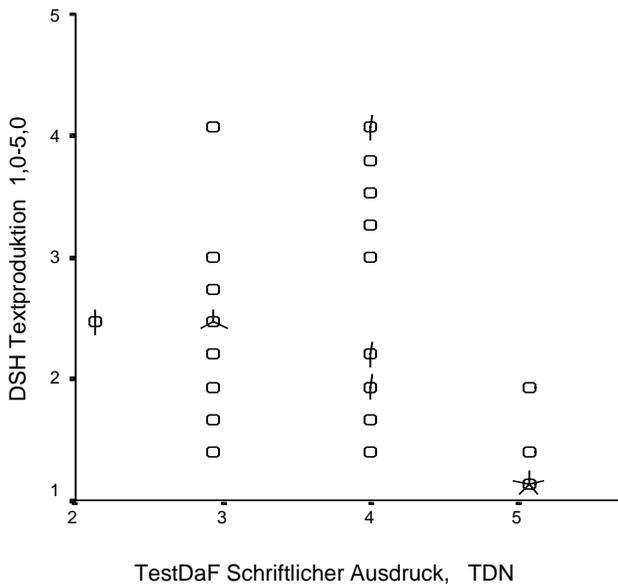
* Korrelation ist auf dem Niveau von 0,05 signifikant (2-seitig).

Subtests zum Schreiben messen ähnliche Fertigkeiten

In beiden Subtests zum Schreiben erzielten die Kandidaten gute Ergebnisse: Im TestDaF blieben nur fünf Kandidaten unter der TDN 3 und sieben der 31 Kandidaten, die auch an der DSH teilnahmen, erreichten die TDN 5. Das schlechteste Ergebnis der DSH war die Note 4,0, im Mittel wurde von den Kandidaten aus dem Studienkolleg 2,67 erzielt. Wie eng war der Zusammenhang zwischen den Ergebnissen in der DSH-TP und im TestDaF-SA? Die Korrelation zwischen den Subtests zum Schreiben liegt auf einem mittleren Niveau

($r_s = -0,472$; vgl. Tabelle 7). Eine hohe TDN im TestDaF-SA war in vielen Fällen auch ein Hinweis auf ein gutes Ergebnis in der DSH-Textproduktion. Besonders bei den Ergebnissen der Spitzengruppe ist ein enger Zusammenhang zu erkennen (vgl. Abbildung 3). Die besten fünf Kandidatinnen erzielten im TestDaF-SA die TDN 5 und in der DSH-Textproduktion eine 1,0. Mehrere Kandidaten haben in der DSH-TP ein Ergebnis erzielt, das über demjenigen im TestDaF liegen dürfte: Es gibt sieben Studierende, die am Anfang des Semesters im TestDaF-SA ein niedriges Ergebnis (TDN 3), in der DSH-Textproduktion am Ende des Semesters aber ein durchschnittliches oder gutes Ergebnis erzielten (2,5 oder besser). Allerdings konnten sich nicht alle Kandidaten verbessern: Zwei Studentinnen erreichten am Beginn des Semesters im Schriftlichen Ausdruck die TDN 4, in der DSH aber nur eine 4,0.

Abbildung 3: Ergebnisse aus der DSH Textproduktion und dem TestDaF Schriftlicher Ausdruck in einem Sonnenblumen-Streudiagramm (n=31)



(Anm.: In einem Sonnenblumen-Streudiagramm werden Wertepaare zu Gruppen zusammengefasst. Ein Kreis steht für ein Ergebnis, zusätzliche Striche jeweils für Kandidaten, die das gleiche Ergebnis erzielten.)

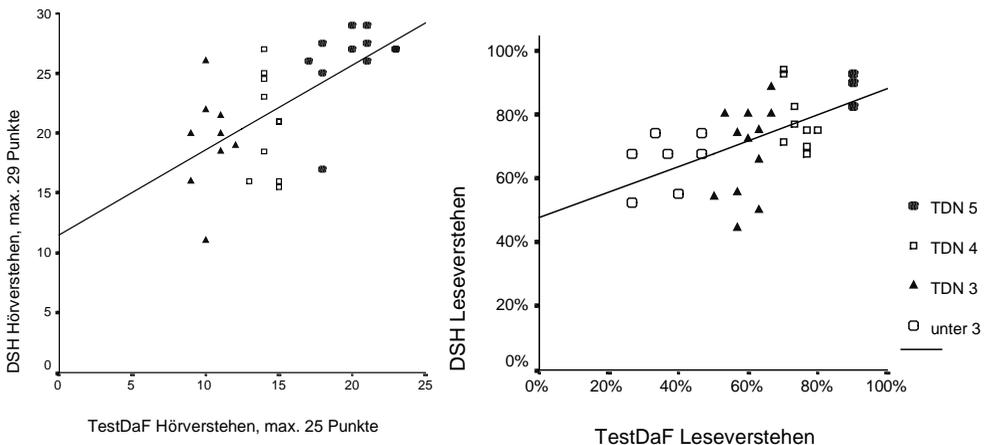
Die Aufgabenstellungen beider Subtests ähnelten sich, daher war ein Zusammenhang zwischen den Ergebnissen zu erwarten. Die Verbesserungen sind sicherlich mit dem Unterricht zu erklären. Wie konnte es zum Einbruch der zwei Studentinnen kommen? Möglicherweise ist der Fachbezug eine Ursache dafür. Im TestDaF wird ein Fachbezug vermieden, für die Textproduktion werden Themen gewählt, bei denen man erwartet,

dass alle Kandidaten sie unabhängig vom Studienziel und fachlichen Vorkenntnissen bearbeiten können. Thema des Subtests SA waren die Kosten für ein Studium, welche vor dem Hintergrund eines Schaubildes zu kommentieren waren. Im Subtest Textproduktion der vorliegenden DSH wird ein Fachbezug jedoch hergestellt. Die Kandidaten konnten aus mehreren Themen auswählen, z. B. ein Schaubild über Unternehmensfusionen kommentieren. Denkbar ist, dass die fachgebundene Aufgabe die beiden Kandidatinnen vor besondere Schwierigkeiten stellte.

Leseverstehen und Hörverstehen: Hohe Korrelationen zwischen TestDaF und DSH

Der Subtest Leseverstehen des TestDaF besteht aus drei Texten mit insgesamt 30 Items (Zuordnungs-, Mehrfachauswahl- und Auswahlaufgaben). Der TestDaF-Subtest LV ist als schwierig empfunden worden, nur drei Kandidatinnen erzielten die TDN 5. Beim DSH-Leseverstehen konnten die Kandidaten einen aus drei Texten auswählen. Die Texte wurden von unterschiedlichen Prüfern erstellt und korrigiert. Die Ergebnisse liegen im Mittel knapp über dem Mittel des Gesamtergebnisses, wurden also nicht als besonders schwierig empfunden (LV: M = 2,84, DSH-Gesamtergebnis M = 2,57). Wie aus dem Streudiagramm (Abbildung 4) hervorgeht, differenziert der Subtest DSH-LV vor allem im unteren Bereich nur wenig, der TestDaF differenziert deutlicher. Es gibt dennoch einen engen Zusammenhang zwischen den Ergebnissen der Subtests TestDaF-LV und DSH-LV ($r_s = 0,582$, vgl. Tabelle 7). Die Kandidaten konnten im DSH-LV mehr Items lösen als im TestDaF-LV, was z. B. an der geringen Steigung der Regressionsgeraden zu erkennen ist. Sie schneidet die y-Achse bei 45 % und nicht im Nullpunkt.

Abbildung 4: Ergebnisse DSH-HV und TestDaF-HV sowie DSH-LV und TestDaF-LV in einem Streudiagramm mit linearer Regressionsgeraden (n = 31)



Auf einen ähnlichen Sachverhalt trifft man beim Vergleich des DSH-Hörverstehens mit dem TestDaF HV: Die Kandidaten konnten in dem leichtesten Subtest der DSH drei Monate nach dem TestDaF mehr Items beantworten als im TestDaF (vgl. Tabelle 4). Die Korrelationskoeffizienten deuten auf einen engen Zusammenhang ($r_s = 0,609$; vgl. Tabelle 7).

TestDaF Mündlicher Ausdruck und Hörverstehen korrelieren hoch.

Auffällig sind hohe Korrelationen zwischen dem TestDaF-Subtest Mündlicher Ausdruck und den Subtests Hörverstehen der DSH und des TestDaF (DSH-HV: $r_s = 0,683$; TestDaF-HV: $r_s = 0,718$). Möglicherweise ist dies ein Hinweis auf Ähnlichkeiten bei der Verarbeitung von gesprochener Sprache. Erwähnenswert sind außerdem die hohen Interkorrelationen zwischen den jeweiligen Subtests Hörverstehen und Leseverstehen, welche methodenspezifische Ursachen haben könnte (z. B. TestDaF-LV und DSH-HV: $r_s = 0,733$, vgl. Tabelle 7).

DSH-Grammatiktest korreliert mit Subtests zum Schreiben und zum Lesen

Die DSH-Grammatiktests weisen hohe Korrelationen mit Subtests zum Schreiben auf (TestDaF-SA: $r_s = 0,537$; DSH-TP: $r_s = 0,552$; vgl. Tabelle 7). Bei der Aufgabe, Sätze bzw. einen Text zu produzieren, werden vergleichbare Fertigkeiten geprüft wie bei der Aufgabe, kontextualisierte Sätze zu transformieren. Hier spiegeln sich auch die Leitlinien für die Bewertung des Schriftlichen Ausdrucks, bei der die sprachliche Korrektheit ein wichtiges Kriterium darstellt. Ein hoher Zusammenhang zwischen Grammatiktests und Tests zum Leseverstehen ist in mehreren Studien beobachtet worden (vgl. Alderson 1993, Bernhardt 1999, Clapham 1996). In der vorliegenden Untersuchung konnten die Beobachtungen nur teilweise bestätigt werden. Der DSH-Grammatiktest 2 korrelierte hoch mit dem DSH-Leseverstehen ($r_s = 0,554$), die Korrelation des DSH-Grammatiktests 1 mit dem TestDaF-Leseverstehen war jedoch geringer ($r_s = 0,411$, vgl. Tabelle 7). Dieser Unterschied könnte seine Ursache auch in unterschiedlichen Formaten der Subtests zum Leseverstehen haben: Im TestDaF besteht er aus 30 Zuordnungs-, Mehrfachwahl- und Auswahlitems, in der DSH aus Fragen, die mit einem Stichwort oder in einem Satz beantwortet werden mussten. Es ist nicht auszuschließen, dass im Falle der DSH nicht nur das Textverständnis, sondern auch die Sprachproduktion in die Bewertung eingeflossen sind. Eine signifikante Korrelation des DSH-Grammatiktests mit dem TestDaF-Subtest Mündlicher Ausdruck konnte nicht beobachtet werden. Niedrig ist der Zusammenhang mit dem Subtest Hörverstehen ($r_s = 0,384$; vgl. Tabelle 7). Das Konstrukt des Grammatiktests hat demnach mehr mit geschriebener als mit gesprochener Sprache zu tun. Dies war auch eine mögliche Interpretation von zwei Faktoranalysen, von denen eine mit den Subtests der DSH und eine mit dem TestDaF und dem DSH-Grammatiktest durchgeführt wurde. Bei einer Extraktion von zwei Faktoren erwies sich der Grammatiktest jeweils als bestimmende Größe für den zweiten, durch den Umgang mit Schriftsprache geprägten Faktor. Der erste Faktor, welcher über einen deutlich höheren Eigenwert verfügt, wird eher durch den Umgang mit gesprochener Sprache bestimmt.

Zusammenfassung

Der C-Test korreliert mittel bis hoch mit den Subtests des TestDaF und hoch mit dem Gesamtergebnis der DSH, er dürfte sich daher als „Screening-Test“ sowohl für den TestDaF als auch für die DSH eignen. Er ist vor allem im oberen Leistungsspektrum sehr aussagekräftig. Obwohl TestDaF und DSH auf unterschiedlichen Skalen messen, wurde deutlich, dass den vergleichbaren Subtests (z. B. Textproduktion und Schriftlicher Ausdruck) vergleichbare Konstrukte zugrunde liegen. Dabei wurden der TestDaF-Subtest Mündliche Ausdruck und die Mündliche Prüfung der DSH nicht berücksichtigt. Deutliche Zusammenhänge konnten auch zwischen Subtests zum Hören und zum Lesen beobachtet werden, was methodenspezifische Ursachen haben könnte. Der DSH-Grammatiktest korreliert mit den Subtests zum Lesen und zum Schreiben. Gering bzw. nicht nachzuweisen waren Zusammenhänge des DSH-Grammatiktests mit Subtests zum Hören und zum Sprechen. Weitere Untersuchungen zur Austauschbarkeit vergleichbarer Subtests wurden nicht unternommen, da nicht die Verschiedenheit der Subtests, sondern die Verschiedenheit der Auswertungsmethoden die Vergleichbarkeit der Testergebnisse erschweren.

3.2. Ergebnisse der qualitativen Studie

TestDaF-Format erfreut sich hoher Akzeptanz.

Die Kandidatinnen und Kandidaten haben sich insgesamt positiv über den TestDaF geäußert. Der Test und seine Teile schienen den Kandidaten angemessen und fair zu sein. Zwei Aspekte wurden häufiger angesprochen: die Bearbeitungszeit sowie die Auswahlaufgaben des Subtests Leseverstehen. Die Bearbeitungszeit für die Testteile LV und HV wurde als zu knapp empfunden. Dieses Problem kann durch eine Vorbereitung auf das Testformat möglicherweise verringert werden. Die Auswahlaufgaben beim LV, welche nicht eindeutig genug gewesen seien, waren ein weiterer Kritikpunkt. Der kassettengesteuerte Testteil MA wurde zwar thematisiert und als ungewöhnlich empfunden, aber nicht negativ bewertet. Ein Grammatiktest wurde nicht vermisst.

Kandidaten empfanden DSH-Grammatiktest als schwierig.

In den Gesprächen über die DSH wurde der Grammatiktest häufig angesprochen. Er wurde als verwirrend und als zu schwierig empfunden. Die Aufgaben in der Prüfung waren nach Meinung einiger Studierender komplexer als diejenigen im Unterricht. Einige Kandidaten bemerkten zu spät, dass sich auch auf der Rückseite des Aufgabenblattes Aufgaben befanden und gerieten daher in Zeitnot. Deutlich wurde: Der Grammatiktest ist ein Testteil, über den die Kandidaten sprechen, denn es geht um richtig oder falsch; man möchte die Ergebnisse vergleichen. Er spielt für das Ansehen der DSH bei den Kandidaten eine wichtige Rolle; der Grammatiktest lässt die DSH „schwierig“, aber im Gegensatz zum Leseverstehen oder Hörverstehen als berechenbar erscheinen. Das bedeutet meiner Ansicht nach allerdings nicht, dass ein Verzicht notwendigerweise zu einem Verlust an Augenscheinvalidität führen muss.

Auswirkungen unterschiedlicher Zulassungskriterien.

Welche Auswirkungen das TestDaF-Ergebnis auf die Zulassung zum Studium bzw. auf die Befreiung von der DSH hat, hängt von den Maßstäben ab, die Anwendung finden. Die Auswirkungen der Ergebnisse bei unterschiedlichen Zulassungshürden sind in der Tabelle 8 dargestellt. Dabei sind auch Kandidaten berücksichtigt worden, bei denen ein Prüfungsteil fehlt. Sie wurden anhand der Tendenz der Ergebnisse aus den anderen Subtests eingeordnet.

Nur eine Kandidatin, welche sowohl am TestDaF als auch an der DSH teilnahm, bestand die DSH nicht. Die Libanesin erzielte in der DSH nur in der Textproduktion ein gutes Ergebnis (2,5), in den anderen Subtests fiel sie durch. Ihre Ergebnisse im TestDaF drei Monate vorher waren unterdurchschnittlich: Im Leseverstehen erzielte sie 17 von 30 Punkten (TDN 3), im Hörverstehen 10 von 25 (TDN 3), im Schriftlichen Ausdruck die TDN „unter 3“, im Mündlichen Ausdruck die TDN 3. Auffällig ist das Ergebnis im ersten DSH-Grammatiktest (0 Punkte), den sie wohl abgebrochen hat. Weitere Informationen über ihre Lernbiografie liegen mir nicht vor. Aus dem Bestehen der übrigen 30 Kandidaten zu folgern, ihre Ergebnisse im TestDaF seien mindestens äquivalent zum Bestehen der DSH, wäre unzutreffend. In Kapitel 2 wurde bereits auf Besonderheiten bei der Durchführung der Tests hingewiesen, welche zur Folge haben, dass die Ergebnisse der Stichprobe (im Vergleich zur Population) im TestDaF eher niedriger und in der DSH eher höher ausgefallen sein dürften.

Zulassung bei TDN 5-5-5-5?

Wenn Studienbewerber mit einem TestDaF-Ergebnis von TDN 5 in allen Subtests von der DSH befreit werden, würde das auf drei der 66 Kandidatinnen und Kandidaten zutreffen, deren Sprachlernbiografie ich kurz vorstellen möchte:

- Die Studentin aus Bulgarien wurde wegen ihrer außergewöhnlich guten Leistungen im Aufnahmetest zum Studienkolleg direkt in das zweite Semester des Studienkollegs aufgenommen. Sie lernte in Bulgarien vier Jahre Deutsch, spricht auch Englisch und verfügt bereits über Arbeitserfahrungen. In Deutschland besuchte sie einen Intensivkurs, in dem sie ein Zeugnis über die Mittelstufe II erhielt. Ihre guten Deutschkenntnisse schreibt sie einer engagierten Lehrerin in Bulgarien zu, welche sie zum Lernen motiviert habe. Dies allein erklärt allerdings nicht, warum sie auch in allen anderen Fächern sehr gute Noten erzielte.
- Die zweite Studentin, welche das Ergebnis TDN 5 in allen Testteilen erzielte, kommt aus der Slowakei. Sie war zum Testzeitpunkt 21 Jahre alt, lernte Deutsch in der Schule, arbeitete vor dem Besuch des Studienkollegs ein Jahr als Aupairmädchen in einer deutschen Familie und nahm gleichzeitig an einem Oberstufenkurs an einer Sprachschule am Ort teil. Sie besuchte das zweite Semester des Studienkollegs, das sie mit sehr guten Noten abschloss.
- Auch eine Studentin aus Ungarn erfüllt das Kriterium. Sie besuchte in Ungarn eine zweisprachige Schule, ist Fremdenführerin und legte bereits die Zentrale Oberstu-

fenprüfung des Goethe-Instituts ab, könnte also bereits auf dieser Grundlage von der DSH befreit werden.

Tabelle 8: TestDaF – Auswirkungen unterschiedlicher Zulassungskriterien

TDN 5 in allen Subtests (TDN 5-5-5-5)	
<ul style="list-style-type: none"> Hochschulrektorenkonferenz: „müssen zugelassen werden“ (Befreiung von DSH) 	
<i>3 Kandidatinnen erfüllen das Kriterium:</i>	
5555, W-Kurs, 2. Sem., Slowakei, DSH 1,3	5555, W-Kurs, 2. Sem., Ungarn, DSH 1,5
5555, W-Kurs, 2. Sem., Bulgarien, DSH 1,1	
TDN 5 im LV und TDN 4 in allen anderen Subtests (TDN 5 LV 4-4-4)	
<ul style="list-style-type: none"> TestDaF-Institut: Zulassung für Fächergruppe Geistes-, Sprach-, Kulturwissenschaften (Befreiung von DSH) 	
<i>Zusätzlich keine weiteren Kandidaten, insgesamt 3.</i>	
TDN 4 in allen Subtests (TDN 4-4-4-4)	
<ul style="list-style-type: none"> TestDaF-Institut: Zulassung für Fächergruppe Jura, Sozial-, Wirtschaftswissenschaften (Befreiung von DSH) Hochschulrektorenkonferenz: „sollen zugelassen werden“ 	
<i>Zusätzlich 7 Kandidaten, insgesamt 10:</i>	
4555, W-Kurs, 2. Sem., Slowakei, DSH 1,1	4555, W-Kurs, 2. Sem., Tschechische Rep., DSH 1,6
4555, W-Kurs, 2. Sem., Ungarn, DSH 1,8	4545, W-Kurs, 2. Sem., Slowakei, DSH 1,8
4555, W-Kurs, 2. Sem., Slowakei, DSH 2,4	4545, W-Kurs, 1. Sem., Russland
	4545, W-Kurs, 1. Sem., Peru
Durchschnitt TDN 4 (TDN 4-4-4-4, TDN 3-3-5-5, TDN 3-4-4-5, nicht „unter 3“)	
<ul style="list-style-type: none"> Vorschlag, der im Text diskutiert wird (Befreiung von der DSH) 	
<i>Zusätzlich 3 Kandidaten, insgesamt 13:</i>	
3535, T-Kurs, 2. Sem., Irak, DSH 1,9	4543, W-Kurs, 2. Sem., Indonesien, DSH 2,0
	3545, W-Kurs, 1. Sem., Slowakei
Durchschnitt 3,5 (TDN 3-3-4-4; TDN 3-3-3-5, nicht „unter 3“)	
<ul style="list-style-type: none"> Vorschlag, der im Text diskutiert wird. 	
<i>Zusätzlich 12 Kandidaten, insgesamt 25:</i>	
4434, T-Kurs, 2. Sem., Iran, DSH 3,2	3434, T-Kurs, 1. Sem., Kenia
3443, T-Kurs, 2. Sem., Mauretanien, DSH 2,5	3434, W-Kurs, 1. Sem., Kenia
3443, T-Kurs, 2. Sem., Korea, DSH 2,1	3434, W-Kurs, 1. Sem., Mexiko
4433, T-Kurs, 2. Sem., Vietnam, DSH 2,4	3444, T-Kurs, 1. Sem., Kenia
3344, T-Kurs, 2. Sem., Indonesien, DSH 2,1	3543, W-Kurs, 1. Sem., VR China
3435, W-Kurs, 1. Sem., Russland	4434, W-Kurs, 1. Sem., Südafrika
TDN 3 in allen Subtests (TDN 3-3-3-3)	
<ul style="list-style-type: none"> TestDaF-Institut: Zulassung für Fächergruppe Informatik, Ingenieurwissenschaften, Mathematik, Kunst und Musik, Naturwissenschaften (Befreiung von DSH) Hochschulrektorenkonferenz: „können zugelassen werden“ 	
<i>Zusätzlich 12 Kandidaten, insgesamt 37:</i>	
3433, T-Kurs, 2. Sem., Indonesien, DSH 2,3	343-, T-Kurs, 1. Sem., Ghana
344-, W-Kurs, 2. Sem., Peru, DSH 2,7	3433, W-Kurs, 1. Sem., Kenia
344-, W-Kurs, 2. Sem., Litauen, DSH 3,4	3433, T-Kurs, 1. Sem., Marokko
4333, T-Kurs, 2. Sem., VR China, DSH 2,0	3334, T-Kurs, 1. Sem., Marokko
3334, T-Kurs, 2. Sem., Syrien, DSH 2,9	3333, T-Kurs, 1. Sem., VR China
3333, T-Kurs, 2. Sem., Marokko, DSH 3,3	435-, T-Kurs, 1. Sem., Iran

nicht TDN 3 in allen Subtests

- TestDaF-Institut: keine Zulassung zum Studium

29 Kandidaten:	4332, W-Kurs, 1. Sem., VR China
3323, T-Kurs, 2. Sem., Libanon, DSH 4,2	333-, T-Kurs, 1. Sem., VR China
2443, T-Kurs, 2. Sem., Mauretanien, DSH 2,9	2434, W-Kurs, 1. Sem., Litauen
3424, T-Kurs, 2. Sem., Libanon, DSH 3,2	2423, T-Kurs, 1. Sem., Libanon
2545, W-Kurs, 2. Sem., Slowakei, DSH 4,0	2-33, T-Kurs, 1. Sem., VR China
2444, W-Kurs, 2. Sem., Senegal, DSH 3,3	2233, W-Kurs, 1. Sem., Rep. Kongo
2342, W-Kurs, 2. Sem., VR China, DSH 3,0	3232, T-Kurs, 1. Sem., Vietnam
3332, W-Kurs, 2. Sem., VR China, DSH 3,3	-3-4, T-Kurs, 1. Sem., Mongolei
2343, W-Kurs, 2. Sem., Vietnam, DSH 3,6	2-34, T-Kurs, 1. Sem., Marokko
2333, T-Kurs, 2. Sem., VR China, DSH 2,7	2444, T-Kurs, 1. Sem., Indonesien
2433, W-Kurs, 2. Sem., VR China, DSH 3,0	2433, T-Kurs, 1. Sem., Mauretanien
3232, W-Kurs, 1. Sem., VR China	2332, W-Kurs, 1. Sem., VR China
233-, T-Kurs, 1. Sem., VR China	--3-, T-Kurs, 1. Sem., Libanon
2323, T-Kurs, 1. Sem., Libanon	2433, T-Kurs, 1. Sem., Indonesien
2433, W-Kurs, 1. Sem., Libanon	--2-, W-Kurs, 1. Sem., Marokko

Erläuterungen:

vierstellige Nummer: TestDaF-Ergebnis als TDN (LV, HV, SA, MA), „unter 3“ = „2“, „-“ = Prüfungsteil fehlt;

T-Kurs = Vorbereitungskurs für technische Studiengänge; W-Kurs = Vorbereitungskurs für wirtschaftliche Studiengänge.

Zulassung bei TDN 5 (LV) 4-4-4?

Die differenzierte Rückmeldung des TestDaF ermöglicht auch unterschiedliche Interpretationen des Ergebnisses für unterschiedliche Zielgruppen. Das TestDaF-Institut regt an, für unterschiedliche Studienfächer unterschiedliche TestDaF-Ergebnisse vorauszusetzen (vgl. Tabellen 2 und 8). Für die Studiengänge Geistes-, Sprach-, und Kulturwissenschaften sollen Kandidaten von der DSH befreit werden, wenn sie im Leseverstehen die TDN 5 und in allen anderen Testteilen mindestens die TDN 4 erzielen. Für die vorliegende Stichprobe ergibt sich keine Veränderung zum Kriterium „TDN 5 in allen Subtests“, denn im Leseverstehen haben nur drei Kandidatinnen die TDN 5 erzielt. Diese erfüllen gleichzeitig aber auch das höhere Kriterium (vgl. Tabelle 8).

Zulassung bei TDN 4-4-4-4?

Kandidaten, die ein Studium der Wirtschaftswissenschaften oder der Rechtswissenschaften anstreben, müssten nach dem Vorschlag des TestDaF-Instituts in allen Testteilen mindestens die TDN 4 erzielen. Zehn der 66 Teilnehmer aus dem Studienkolleg erfüllen dieses Kriterium. Die Stichprobengröße der Untersuchung reicht nicht aus, um signifikante Aussagen über die Verteilung der Herkunftsländer machen zu können. Die Verteilung der Herkunftsländer ist jedoch so auffällig, dass sie erwähnt werden soll: Mit Bulgarien, der Slowakei (4), Russland, der Tschechischen Republik, Ungarn (2) und Peru sind fast nur Kandidaten aus Mittel- und Osteuropa vertreten. Nur der Student aus

der Slowakei besucht den Technikkurs, die acht Studentinnen besuchen den Wirtschaftskurs. Bei diesen Kandidaten stand das Bestehen der DSH drei Monate nach dem TestDaF nicht in Frage, sie gehörten zur sprachlichen Spitzengruppe des Studienkollegs, erzielten gute bis sehr gute Ergebnisse in der DSH. Auch zwei Studentinnen aus dem ersten Semester erfüllten das Kriterium:

- Die Studentin aus Peru erzielte zwei Mal die TDN 4 und zwei Mal die TDN 5. Nur im DSH-Grammatiktest ist ihr Ergebnis unterdurchschnittlich. In ihrem Heimatland besuchte sie in den zwei Jahren vor ihrer Ausreise nach Deutschland eine zweisprachige Schule. Ebenfalls in Peru erhielt sie das Kleine Deutsche Sprachdiplom (KDS). Sie absolvierte den TestDaF zu Beginn des ersten Semesters des Studienkollegs. Zu dem Zeitpunkt konnte sie bereits flüssig und verständlich sprechen. Auch in den anderen TestDaF-Subtests und im C-Test stellte sie ihre fortgeschrittenen Deutschkenntnisse unter Beweis. Mit Ausnahme des DSH-Grammatiktests lagen ihre Ergebnisse in den übrigen Testteilen durchweg deutlich über dem Durchschnitt. Im Gespräch gab sie an, dass sie das Testformat des DSH-Grammatiktests verwirrt habe und sie nicht gewusst habe, was von ihr verlangt wird. Die grammatischen Phänomene, die für die geforderten Transformationen notwendig sind, seien ihr eigentlich nicht fremd gewesen, was sie bei der Besprechung des Tests bemerkt habe. In ihrem Fall liefert ein unvorbereiteter DSH-Grammatiktest die zusätzliche Information, dass sie Schwierigkeiten mit dem Testformat hat. Schlussfolgerungen für die Studierfähigkeit lassen sich daraus nicht ziehen, denn ihrer Interpretation des Testergebnisses ist wahrscheinlich zuzustimmen. Sie müsste sich mit dem Testformat vertraut machen, dann würde sie auch in diesem Testteil ein besseres Ergebnis erzielen.
- Die Russin absolvierte in ihrem Heimatland bereits eine Ausbildung zur Deutschlehrerin und arbeitete vor dem Besuch des Studienkollegs ein Jahr als Aupair bei einer deutschen Familie. Das Studienkolleg besucht sie, um in Deutschland wirtschaftswissenschaftliche Fächer studieren zu können.

Zulassung bei TDN 3-3-3-3?

Kandidaten, die Informatik, Ingenieurwissenschaften, Mathematik, Kunst, Musik oder Naturwissenschaften studieren möchten, sollen nach dem Vorschlag des TestDaF-Instituts zugelassen werden, wenn sie in allen Testteilen mindestens die TDN 3 erzielten. Dieses Kriterium wird von insgesamt 37 der 66 Kandidaten erfüllt, auch viele Studierende aus dem ersten Semester gehören dazu. Bei dieser Gruppe ist eine regionale Häufung nicht mehr zu beobachten.

Einige Lernbiografien sollen kurz skizziert werden:

- Der 21-jährige Chinese bringt sehr gute Leistungen aus dem Heimatland mit. Nach der Schule studierte er in China zwei Semester Chemie – was ihn allerdings nicht interessierte. In der Schule und während des Studiums lernte er Englisch, aber noch kein Deutsch. Seinen Wunsch, ein Studium in Großbritannien oder in den USA,

konnte er aus finanziellen Gründen nicht verwirklichen. Ein Jahr vor der Teilnahme an der Prüfung kam er ohne Deutschkenntnisse nach Deutschland und nahm an einem einjährigen Intensivkurs an einer privaten Sprachschule teil. Der Intensivkurs schloss mit einer Mittelstufenprüfung ab. Wegen seiner sehr guten Leistungen im (schriftlichen) Aufnahmetest zum Studienkolleg wurde er gleich in das zweite Semester des Studienkollegs aufgenommen. Der intelligente und sprachbegabte junge Mann kann auf einen strukturierten Deutscherwerb in Deutschland zurückblicken. Durch seine Englischkenntnisse ist er sprachlernerfahren. Dies scheint ihm allerdings beim Mündlichen Ausdruck und beim Hörverstehen nicht zu helfen (TDN 4-3-3-3). Beim DSH-Grammatiktest gehörte er zur Spitzengruppe (90 % richtige Lösungen, $M = 64\%$).

- Die Kandidatin aus Kenia belegte nach ihrem Schulabschluss Sprachkurse in Deutsch und Französisch in ihrem Heimatland. Ein halbes Jahr vor dem TestDaF hat sie am Goethe-Institut in Nairobi das Zertifikat Deutsch erworben. Vom TestDaF fühlte sie sich überrascht, sie hätte gerne mehr Zeit für die Bearbeitung gehabt. Mit dem TestDaF-Ergebnis von TDN 3-4-3-3 könnte sie nach den Vorstellungen des TestDaF-Instituts nur zu einem technischen Fachstudium zugelassen werden.
- Der Kandidat aus Indonesien hätte das Kriterium TDN 3-4-3-3 ebenfalls erfüllt. Die DSH am Semesterende bestand er mit einer guten Note (Schriftliche Prüfung: 2,3). Er lernte Deutsch erst in Deutschland, wo er das Zertifikat Deutsch und ein Zeugnis der Mittelstufe I erwarb.
- Der Marokkaner aus dem ersten Semester des Technikkurses erfüllt das Kriterium ebenfalls mit dem Ergebnis TDN 3-4-3-3. Erst nach dem Besuch des technischen Gymnasiums lernte er Deutsch im Heimatland.

Dieses Kriterium wird nicht nur von Kandidaten erfüllt, die sich schon seit längerer Zeit in Deutschland aufhalten oder die im Heimatland viele Jahre Deutsch lernten. Auch ausländische Studienbewerber, die Deutsch erst als Erwachsene nach ihrem Schulabschluss als Vorbereitung auf das Studium in Deutschland erworben haben, gehören zu der Gruppe. In der DSH drei Monate nach dem TestDaF erzielten sie schlechtestenfalls befriedigende Noten.

Zulassung bei durchschnittlich TDN 4?

Bevor eine abschließende Wertung vorgenommen wird, sollen noch zwei weitere Varianten zur Diskussion gestellt werden. Bei diesen Varianten ist der Durchschnitt der Niveaustufen das Ausschlag gebende Kriterium. Dann besteht die Möglichkeit, ein niedriges Ergebnis in einem Testteil durch ein hohes Ergebnis in einem anderen auszugleichen. Bei der Berechnung des Durchschnitts sollen allerdings nur die TDN 3, 4 und 5 berücksichtigt werden. Wenn in einem Subtest „unter 3“ erzielt wurde, wird das Kriterium nicht erfüllt.

Zur ersten Variante: Wenn Kandidaten von der DSH befreit würden, welche im Durchschnitt eine TDN 4 erzielten, würde sich die Zahl im Vergleich mit dem Kriterium

„mindestens TDN 4 in allen Subtests“ um drei Kandidaten auf insgesamt dreizehn (von 66) erweitern.

- Der Iraker war zum Testzeitpunkt 20 Jahre alt. Nach seinem Schulabschluss im Heimatland ist er im Alter von 18 Jahren nach Deutschland geflohen, wo ihm Asyl gewährt wurde. Zu Zeitpunkt der Ausreise beschränkten sich seine Fremdsprachenkenntnisse auf geringe Englischkenntnisse. Achtzehn Monate nahm er an berufsvorbereitenden Maßnahmen teil, die einen Deutschunterricht mit EDV-Elementen beinhalteten. Anschließend besuchte er eine Sprachschule in Mannheim. Dort erhielt er nach drei Monaten ein Zeugnis über die Grundstufe DaF und weitere drei Monate später ein Mittelstufenzeugnis. Im Studienkolleg bestand der Student das erste Semester mit einer sehr guten Deutschnote. Schwierigkeiten hat er mit dem DSH-Grammatiktest. Nach zwei Jahren in Deutschland und nach zwei Jahren der Teilnahme an Deutschunterricht verschiedener Ausprägung fällt ihm die mündliche Kommunikation offensichtlich leichter als der Umgang mit anspruchsvollen Strukturen. Er bestand die DSH drei Monate im Anschluss an den TestDaF mit der Gesamtnote 1,9.
- Die Studentin aus der Slowakei hält sich bereits seit mehreren Jahren in Deutschland auf. Ihrer eigenen Einschätzung nach ist vor allem die inhaltliche Vorbereitung auf das Studium im Rahmen des Studienkollegs wichtig für sie. Ihre Deutschkenntnisse muss sie lediglich in Bezug auf fachbezogene Sprachverwendungssituationen erweitern.
- Der Student aus Indonesien kann bereits auf eine lange Deutschlernbiografie in seinem Heimatland zurückblicken. Mit dem Studienziel Wirtschaftsinformatik interessiert er sich nach eigener Auskunft allerdings mehr für Computer als für Sprachen. Der Fremdspracherwerb war nie sein Hauptanliegen, sondern lediglich Mittel zum Zweck. Da er nach eigener Auskunft für andere Prüfungsteile in der Feststellungsprüfung nicht viel üben musste, konnte er sich auf die DSH konzentrieren und erzielte ein gutes Ergebnis (2,0).

Zulassung bei durchschnittlich TDN 3,5?

Wenn man den Durchschnitt auf 3,5 verringert, würden insgesamt 19 der 66 Kandidaten das Kriterium erfüllen, acht mehr als beim Kriterium „Durchschnitt 4“. Auch hier gilt wieder, dass in allen Subtests mindestens die TDN 3 erzielt werden muss. Wer erreicht das Kriterium „durchschnittlich TDN 3,5“?

- Insgesamt zehn von 35 Studierenden aus dem ersten Semester erfüllen dieses Kriterium bereits.
- Auffällig ist auch eine große Gruppe aus dem zweiten Semester des Technikkurses. Während die Spitzengruppe des Wirtschaftskurses die TDN 4 in allen Subtests oder besser erzielt, erfüllen die guten Teilnehmer des Technikkurses lediglich diese Hürde.

- Außerdem beachtenswert ist die Verteilung der Herkunftsländer bzw. -sprachen: Es gibt einzelne Studierende aus China, Korea und Vietnam, welche das Kriterium erfüllen. Übrigens erzielte keiner der siebzehn Studierenden aus Ostasien im TestDaF-Subtest Mündlicher Ausdruck die TDN 4 oder TDN 5.

Diese Zulassungshürde würde den Personenkreis stark erweitern. Studierende mit sehr unterschiedlichen Lernbiografien und Ausgangssprachen sind in der Lage, dieses Kriterium zu erfüllen.

4. Empfehlungen

Externen Kandidaten ist der TestDaF zu empfehlen.

Der TestDaF hat sich in dieser Studie als fairer Test der Sprachkenntnisse erwiesen, der sich einer hohen Akzeptanz von Seiten der Prüfungsteilnehmerinnen und -teilnehmer erfreut. Vor allem Kandidaten, die nicht an institutseigenen Vorbereitungskursen teilnehmen oder die sich im Heimatland auf das Studium in Deutschland vorbereiten, ist die Teilnahme am TestDaF zu empfehlen. Der Vorteil der DSH kommt vor allem zum Tragen, wenn die Kandidaten an einem bestimmten Institut auf ein bestimmtes Studium vorbereitet werden. Die DSH wird jedoch auch von Kandidaten absolviert, die nicht an der hochschuleigenen Vorbereitung teilgenommen haben und die auch kein Studium an der betreffenden Hochschule anstreben. Die internen Kandidaten erzielten in unserer Studie bessere Ergebnisse in der DSH als die externen Kandidaten. Die internen Kandidaten waren mit dem Format der Prüfung vertraut und kannten die Leseerwartungen der Prüfer. Die externen Kandidaten verfügten (laut Prüfungsergebnis) über Deutschkenntnisse auf niedrigerem Niveau, möglicherweise waren sie auf die Prüfung nicht so gut vorbereitet wie die internen Kandidaten. Vor allem im Prüfungsteil Textproduktion waren die Leistungen deutlich niedriger, was an dem Fachbezug dieses Subtests liegen kann. Die Vermutung liegt nahe, dass sie bei einer dezentralen Prüfung, welche von Prüfern in Anlehnung an den Unterricht gestaltet wird, benachteiligt sind.

Keine Zulassung bei TDN „unter 3“ in einem Subtest.

Es scheint ein Konsens vorzuliegen, dass Studienbewerber nicht von der DSH zu befreien sind, wenn sie in einem Subtest nicht die TDN 3 erzielt haben („unter 3“). Ich halte diesen Konsens für unstrittig, denn die Einschätzung „unter 3“ beschreibt ein breites Spektrum an Leistungen. Wegen der fehlenden Differenzierung auf dieser Stufe kann die Einstufung „unter 3“ nicht als Basis für eine Entscheidung gelten. Im Rahmen der Untersuchung haben freilich neun der zehn Kandidaten, welche im TestDaF in einem Prüfungsteil nicht die TDN 3 erzielten, die DSH mit Ergebnissen im Bereich befriedigend bis ausreichend bestanden. Allerdings hatte der zeitliche Abstand zwischen den Prüfungen wohl zur Folge, dass die Leistungen der Kandidaten sich verbesserten.

Allgemeine Zulassung zum Fachstudium bei durchschnittlich TDN 4 im TestDaF.

Bei welchem TestDaF-Ergebnis sollen Studienbewerber von der DSH befreit werden bzw. zum Fachstudium zugelassen werden? Das Kriterium „TDN 5 in allen Subtests“ erscheint vor dem Hintergrund dieser Studie als zu hoch. In der vorliegenden Untersuchung hätten lediglich drei äußerst sprachbegabte und auch in den anderen Fächern überdurchschnittlich leistungsstarke Studentinnen das Kriterium erfüllt. Das entspricht fünf Prozent einer Stichprobe, zu der zweifelsohne mehr Kandidaten gehören, welche zum Zeitpunkt der Durchführung als studierfähig zu bezeichnen waren. Eine Erweiterung des Kriteriums auf „TDN 5 im Leseverstehen und TDN 4 in den anderen Subtests“ hatte bei der vorliegenden Stichprobe keine Auswirkungen, was an der niedrigen Anzahl derjenigen liegt, welche im LV die TDN 5 erzielten (drei von 66 Kandidaten). Es steht fest: Diese Kriterien liegen deutlich über „DSH bestanden“.

Das Kriterium „TDN 4 in allen Subtests“ erzielten insgesamt zehn Kandidaten. Es wurde (mit Ausnahmen) von sprachgewandten Osteuropäerinnen erreicht, die sich bereits lange in Deutschland aufhalten und/oder über eine gründliche Ausbildung aus ihrem Heimatland verfügen. Selbst unter ausdrücklicher Berücksichtigung der in Kapitel 2 genannten Einschränkungen kann mit großer Sicherheit behauptet werden, dass dieses Kriterium über dem Ergebnis „DSH bestanden“ liegt.

Eine Anerkennung von TestDaF-Zeugnissen mit „durchschnittlich TDN 4“ würde die Zielgruppe nur leicht verändern, aber Kandidaten eine Zulassung ermöglichen, die eine TDN 3 in einem Subtest erzielten und diese durch eine TDN 5 in einem anderen Subtest ausgleichen. Dies dürfte in Ausnahmefällen zutreffen. Der Personenkreis erweitert sich um Kandidaten, bei denen es ebenfalls kaum Zweifel über die ausreichenden Deutschkenntnisse gibt. Auch Studienbewerber, bei denen der Fremdspracherwerb nicht Mittelpunkt ihrer bisherigen Ausbildung darstellte, gehören zu dieser Gruppe. Sie dürften in der Lage sein, ihr Studium zu absolvieren, benötigen aber die Chance, TDN 3 in einem Prüfungsteil auszugleichen. Dies trifft beispielsweise auf Studierende zu, welche aufgrund ihrer Herkunftssprache Schwierigkeiten mit der Aussprache des Deutschen haben.

Zulassung bei durchschnittlich TDN 3,5 kann sinnvoll sein.

Ist ein Drittel bis ein Viertel der Studierenden aus dem ersten Semester des Studienkollegs bereits studierfähig? Eine Gruppe dieser Größenordnung könnte zum Studium zugelassen werden, wenn das Kriterium „durchschnittlich TDN 3,5“ lauten würde. Ob ausländische Studienbewerber ein Studienkolleg besuchen müssen, hängt nicht von den Sprachkenntnissen, sondern von der Vorbildung im Heimatland ab. Ich halte es daher für möglich, dass viele Studierende des Studienkollegs bereits im ersten Semester die sprachlichen Zulassungshürden erfüllen können. Die DSH-Ergebnisse der Kandidaten aus dem zweiten Semester waren schlechtestenfalls befriedigend (3,2), das Bestehen der DSH stand bei keinem Kandidaten in Frage. Schließlich ist zu bedenken, dass selbst die sprachlich fortgeschrittensten Kandidaten aus China, Vietnam oder Korea ein höheres

Kriterium nicht erzielen. Vor dem Hintergrund dieser Untersuchung spricht einiges dafür, Studierende bereits mit einem TestDaF-Zeugnis von durchschnittlich TDN 3,5 zum Fachstudium zuzulassen. Unter Beachtung der oben erwähnten Einschränkungen (vgl. Kapitel 2), gibt diese Studie wenig gesicherte Hinweise für eine weitere Öffnung der Zulassungskriterien. Die Ergebnissen der vorliegenden Studie schließen ein derartiges Ergebnis nicht aus. Weitere Studien sind nötig, um festzustellen, ob und unter welchen Umständen Kandidaten mit dem Ergebnis „TDN 3 in allen Subtests“ von der DSH befreit werden sollten. Dabei sollte es sich nicht nur um Vergleichsstudien, sondern auch um qualitative Langzeitstudien handeln.

Zulassung am TestDaF orientieren, nicht an der DSH.

Die Frage nach der Äquivalenz der Prüfungsergebnisse ist für Kandidaten und Zulassungsstellen relevant. Das eindeutige Ergebnis der DSH kann problemlos in das Zulassungsverfahren Eingang finden. Der TestDaF ist testmethodisch ausgereift und mit der DSH in diesem Aspekt nicht zu vergleichen. Der Umgang mit dem Ergebnis ist aber weniger eindeutig, daher auch anspruchsvoller. Möglicherweise ist der TestDaF mit der starken Differenzierung im oberen Leistungsspektrum über das Ziel hinausgeschossen: Diese Differenzierung mag als Sprachstandsdiagnose in einigen Fällen interessant sein, als Hochschulzugangsprüfung ist sie überflüssig. Dennoch sollte sich eine Beschreibung der sprachlichen Zulassungsvoraussetzungen in Zukunft am standardisierten TestDaF ausrichten und nicht an dem „Prüfungsgerüst“ DSH. Das hätte zur Folge, dass der Nachweis der Sprachkenntnisse verlässlicher und damit auch gerechter wird. In Form von qualitativen Studien, die in dieser Untersuchung bereits in begrenztem Umfang durchgeführt wurden, könnten Erkenntnisse über einer Zulassungspraxis gewonnen werden, die auch die Differenzierungsmöglichkeiten des TestDaF nutzt.

Die Stärken beider Tests nutzen.

Welche Rollen können die Tests in Zukunft übernehmen? Die DSH ist in Bezug auf die Gütekriterien dem TestDaF unterlegen. Obwohl eine Standardisierung der DSH wünschenswert wäre, kann eine Erhöhung der Qualität allein die Zukunft der DSH nicht sichern. Zu den Stärken der DSH gehört die Durchführung durch Prüfer, welche die Kandidaten kennen und mit den „Abnehmern“ in engem Kontakt stehen. Der DSH-Grammatiktest, den man ohne gezielte Vorbereitung kaum bewältigen kann, ist ein Hinweis darauf, dass die DSH wie eine Kursabschlussprüfung eingesetzt wird – allerdings ohne dass alle Kandidaten an einem spezifischen Vorbereitungskurs teilnehmen. Dies ist für eine Sprachstandsprüfung zwar ungewöhnlich, bietet aber auch Chancen in Bezug auf eine besonders zielgruppengenaue Gestaltung. Die DSH ist nicht isoliert von der hochschulnahen Studienvorbereitung zu bewerten. Als flexibles Instrument zur Förderung der Eigenständigkeit, mit dem institutionelle Besonderheiten und besondere Zielgruppen berücksichtigt werden können, hat die DSH weiterhin eine Berechtigung. Der TestDaF

macht den Studienzugang für ausländische Studienbewerber offener, transparenter und fairer. Eine Einbindung in studienvorbereitende Maßnahmen steht noch aus.

5. Literatur

- Alderson, J. Charles 1993. The Relationship Between Grammar and Reading in an English for Academic Purposes Test Battery. In Douglas, Dan; Chapelle, Carol (Hrsg.). *A New Decade of Language Testing Research: Selected Papers From the 1990 Language Testing Research Colloquium*. Alexandria, Virginia: Teachers of English to Speakers of Other Languages (TESOL), 203-219.
- Alderson, J. Charles 2002. *Criteria for test comparability studies*. Vortrag auf der Fachtagung „DSH und TestDaF: Politische Implikationen und wissenschaftliche Erforschung“ vom 22.-23.6.2002 in Braunschweig.
- Arras, Ulrike; Eckes, Thomas; Grotjahn, Rüdiger 2002. C-Tests im Rahmen des ‘Test Deutsch als Fremdsprache’ (TestDaF): Erste Forschungsergebnisse. In Grotjahn, Rüdiger (Hrsg.). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 4). Bochum: AKS-Verlag, S. 175-209.
- Bernhardt, Elizabeth 1999. If reading is reader-based, can there be a computer adaptive test of reading? In Chalhoub-Deville, Micheline (Hrsg.). *Issues in Computer-adaptive Testing of Reading Proficiency*. Cambridge: Cambridge University Press, 1-10.
- Bortz, Jürgen; Döring, Nicola 2002³. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin et al.: Springer.
- Clapham, Caroline 1996. *The development of IELTS: a study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.
- Fachverband Deutsch als Fremdsprache (FaDaF) 2001. *DSH-Handbuch für Prüferinnen und Prüfer*. Münster: Fachverband Deutsch als Fremdsprache.
- Grotjahn, Rüdiger 1995. Der C-Test: State of the Art. *Zeitschrift für Fremdsprachenforschung (ZFF)*, 6 (2), 37-60.
- Grotjahn, Rüdiger 1997. Der C-Test: Neuere Entwicklungen. In Gardenghi, Monica, O’Connell, Mary (Hrsg.). *Prüfen, Testen, Bewerten im modernen Fremdsprachenunterricht*. (Bayreuther Beiträge zur Glottodidaktik 8.) Frankfurt am Main et al.: Lang, 117-128.
- Grotjahn, Rüdiger 2000. Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TestDaF. In Bolton, Sybille (Hrsg.). *TestDaF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*. (Standpunkte zur Sprach- und Kulturvermittlung 8, Werkstattberichte des Goethe-Instituts.) München: Goethe-Institut; Köln: Gilde-Verlag, 7-56.
- Grotjahn, Rüdiger; Kleppin, Karin 2001. TestDaF: Stand der Entwicklung und einige Perspektiven für Forschung und Praxis. In: Aguado, Karin; Riemer, Claudia (Hrsg.). *Wege und Ziele: Zur Theorie, Empirie und Praxis der Deutschen als Fremdsprache (und anderer Fremdsprachen)*. *Festschrift für Gert Henrici zum 60. Geburtstag*.

- (Perspektiven Deutsch als Fremdsprache 15.) Baltmannsweiler: Schneider Verlag Hohengehren, 419-434.
- Kenyon, Dorry M. 2000. Tape-mediated Oral Proficiency Testing: Considerations in Developing Simulated Oral Proficiency Interviews (SOPIs). In Bolton, Sybille (Hrsg.). *TestDaF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*. (Standpunkte zur Sprach- und Kulturvermittlung 8, Werkstattberichte des Goethe-Instituts.) München: Goethe-Institut; Köln: Gilde-Verlag, 87-106.
- Kniffka, Gabriele; Üstünsöz-Beurer, Dörthe 2001. TestDaF: Mündlicher Ausdruck. Zur Entwicklung eines kassettengesteuerten Testformats. *Fremdsprachen Lehren und Lernen* 30, 127-149.
- Koreik, Uwe; Schimmel, Dagmar 2002. Hörverstehenstests bei der DSH, der Feststellungsprüfung und TestDaF – eine Vergleichsstudie mit weiterführenden Überlegungen zu TestDaF und DSH. *InfoDaF*. (In Druck)
- TestDaF-Institut 2001. *Informationen für Hochschulen in Deutschland*. Hagen: TestDaF-Institut/Gesamthochschule Hagen.
- TestDaF-Institut 2002. *Handreichungen zur Zulassung ausländischer Studierender mit TestDaF-Zeugnis*. (Faltblatt.) Hagen: TestDaF-Institut/Gesamthochschule Hagen.

Anschrift des Autors

Christian Krekeler

Fachhochschule Konstanz, Brauneggerstr. 55, D-78462 Konstanz,

Tel.: 07531 206-395, Fax: 07531 206-87395, Email: krekeler@fh-konstanz.de